Frontal-posterior loop integrating captured attention into visual consciousness in the human brain

Xilei Zhang[a,b], Chao Zhang[c], Zhihao Guo[c], Xiqian Wu[a,b], Yi Jiang[b,d], Kai Zhang[c,*], Liang Wang[a,b,*]

[a] CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

[b] Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

[c] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China

[d] State Key Laboratory of Brain and Cognitive Sciences, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

*Corresponding authors: Liang Wang (lwang@psych.ac.cn) or Kai Zhang (zhangkai62035@163.com).

Abstract

The emergence of subjective consciousness in the brain remains a profound puzzle, with attention at its core. Despite enduring debates on whether consciousness requires attention, the mechanisms by which attention becomes integrated into consciousness have been insufficiently explored. Here we show it is the bottom-up captured attention rather than top-down supplied attention that contributes to conscious experience, and revealed landmark neural dynamics underlying this process. Combining psychophysics with intracranial electroencephalography (iEEG) data from 21 epilepsy patients, we examined consciousness-related neural activity (visible versus invisible) under conditions of both poor and sufficient integration of attention into consciousness. In scenarios where attention was poorly integrated, we observed sustained and homogeneous neural representations in the posterior brain, despite its functional disconnection from the frontal brain. This observation supports the proposal of a preconscious buffer state situated between unconscious and conscious states. Conversely, in scenarios where attention was sufficiently integrated into consciousness, even before frontal-posterior coupling was established (thus still in a preconscious state), supplied attention primed the entire brain for subsequent attentional integration. This priming involved selecting signals-to-be-perceived by the posterior brain and stabilizing attentional supply by the frontal brain. Subsequently, the functional coupling between the frontal and posterior brain formed a closed loop. Notably, within this loop, what continuously broadcasted were coarse-grained binary signals of consciousness emergence, opposing views attributing consciousness to the global broadcasting of fine-grained contents. These findings show what and how attention gets integrated into visual consciousness, and the resulting outcomes call for a significant revision of current theories of consciousness. In response, we propose an attentional integration model of consciousness (AIM) aimed at reconciling the sharp discrepancies in this field.

Main text

The emergence of subjective consciousness in the brain presents a multifaceted puzzle, with attention occupying a central position. Ongoing debate revolves around the question of whether consciousness can manifest in the absence of top-down attention [1-3]. However, even when attention is actively directed, the precise mechanisms by which this attention becomes integrated into consciousness remain elusive. Within this context, attention is often involuntarily captured by salient signals, which subsequently become consciously perceived. Here, we hypothesize that this captured attention may play a pivotal role in the integration process leading to conscious experience. To explore this hypothesis, we have designed a psychophysical experiment aimed at discerning between captured and supplied attention. Our first objective is to rigorously investigate whether and how captured attention contributes to the subjective experience of consciousness.

Then, combining psychophysics with high spatial-temporal resolution intracranial recording techniques (iEEG; [4]), we sought to elucidate how attention integrates into visual consciousness within the awake human brain. Our focus was on discerning the distinct contributions of the posterior and frontal brain regions, as well as their interplay, in supporting visual consciousness of images which appeared under either limited (poorly integrated) or ample (sufficiently integrated) attention.

In the scenario where available attention is scarce, even highly salient visual signals may go unnoticed due to inattentional blindness ([5, 6]). Traditionally, researchers have posited that stimulus-evoked recurrent processing (RP) in the posterior brain is constrained to a preconscious buffer state ([7]), situated between subliminal (unconscious) and conscious awareness. However, critics ([8-10]) contend that this buffer state is non-existent, as unattended RPs can still be consciously perceived and reportable at the moment of presentation, but rapidly vanished thereafter ([10, 11]). If true, this would imply an early, transient shift from unconscious to conscious neural activity patterns. To test such transient hypothesis, we analyzed iEEG data recorded extensively across the brain, emphasizing the temporal profiles of neural representations under unattended conditions. Additionally, we investigated the involvement of the frontal brain in these processes.

Conversely, in the scenario where ample attention is pre-allocated, the posterior brain may exhibit an elevated baseline [12], enhancing early visual responses within the first 100 ms post-stimulus onset [13, 14]. We were particularly interested in understanding whether and how this enhanced early signal captures additional attention and becomes explicitly perceived. Specifically, we aim to identify the neural markers that distinguish between these two stages, and examine the mutual interaction between the frontal and posterior brain regions before and after successful integration of attention into visual consciousness.

A prevalent theory posits that the frontal brain acts as a global neuronal workspace (GNWT; [15]), which not only represents the contents of consciousness based on feedforward inputs from the posterior brain, but also broadcasts these contents back to widely distributed posterior regions in real-time ([15, 16]). While widely accepted, this content-broadcasting mechanism has yet to be rigorously tested. Moreover, we considered potential disadvantages (e.g., high redundancy, bandwidth congestion, and poor interpretation, see the Supplementary) associated with broadcasting detailed consciousness contents to local processors across the entire brain simultaneously. This led us to speculate that the frontal-posterior exchanged signals may be simpler, indexing binary consciousness presence rather than detailed contents. Accordingly, we designed our study to differentiate between these two forms of neural representations (**Method** section) to better understand the role of the frontal-posterior loop in consciousness. Collectively, these efforts drive

us towards novel explorations that may necessitate a revision of current consciousness theories.

Experiment 1: captured attention gets integrated into visual consciousness

We designed a highly demanding counting task (**Fig.1A, 1B**, Method) of which the performance would change as a function of the amount of consumed attention. During counting, an image was suddenly presented in the peripheral vision. Since the amount of attentional resources is limited [17], if attention is somehow distracted by an image, less attention available for the counting task would lead to a deteriorated counting performance.

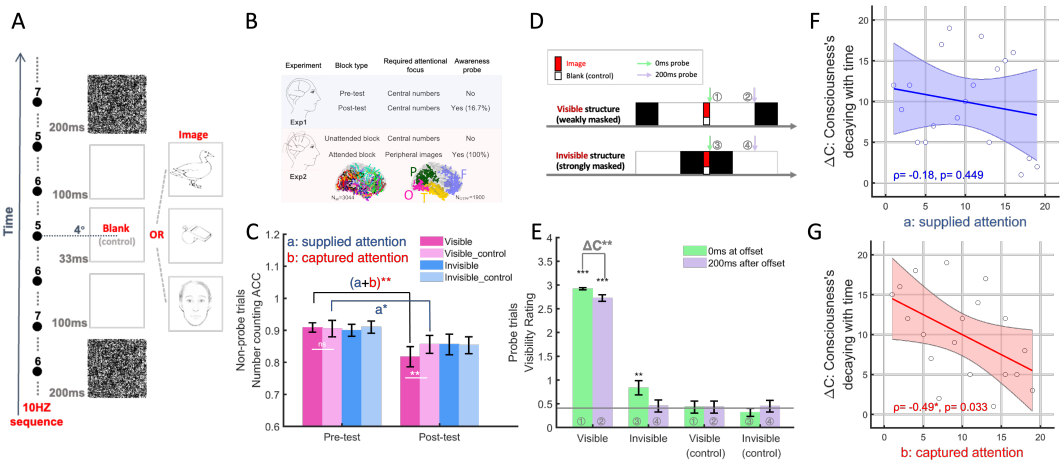


Figure 1. Captured attention gets integrated into conscious experience. (A) Schematic illustration of the timing structure of an exemplar visible trial consisting of a central number stream (10 Hz) and a peripheral image stream, displayed in parallel. (B) The supply of attention to the peripheral image was manipulated across two independent experiments. Intracranial local field potential signals were only recorded in experiment 2, of which regions of interest consisted of four neocortex lobes (including 1900 implanted contacts in sum). (C) Attention generally supplied to peripheral stream (component a) and attention selectively captured by visible images (component b) were indirectly measured by number counting performance following the logic that more attention split and less attention remaining for the counting task would lead to a deteriorated counting accuracy. (D) In 16% randomly selected trials (post-test of Exp.1), an awareness probe was introduced at either 0ms or 200ms post-stimulus offset, in order to spare partial attention for peripheral visual stream. Black and white square indicate visual mask and blank screen, respectively. (E) Visibility rating for these probes revealed how much subjective visibility decays over 200ms interval. Spearman correlation analysis revealed a negative correlation between this decay and attention for captured attention (G), but not for supplied attention (F), suggesting captured attention gets integrated into visual consciousness. See **Fig.S2** for replication study of this effect in an independent sample.

In the pretest session (i.e., without probe, **online video 1**), we required participants to focus on counting and neglect the peripheral image completely. The image was either invisible (heavily masked) or visible (weakly masked). As a result, counting performance did not vary with visibility (Fig. 1C, $F_{(3, 54)}= 0.16$, p=0.922), suggesting that in the pretest session peripheral salient images failed to capture attention.

In the posttest session (i.e., with probe**, online video 2**), we still required participants to set priority the counting task and perform their best. The only difference was that we designed to shift a bit of

attention to peripheral image, by occasionally probing its visibility in 16% randomly selected probe trials (**Fig.1D**, Method). In all trials (including the rest 84% non-probe trials), we anticipated participants would divide some attention away from the counting task, allocating divided attention to peripheral image, to cope with such randomly selected probes. If so, counting performance of non-probe trials (see following) would decrease because the attention resource available for the counting task was reduced (Method).

Indeed, there was a general decrease (**Fig.1C**, $t_{(18)}$=-3.57, p=0.002) on counting performance in the posttest session relative to the pretest baseline. Herein, critically, visible image trials relative to invisible image trials encountered an additional decrease (interaction: $F_{(1, 18)}$=6.82, p=0.018) which was irrespective of categorical contents ($F_{(2,36)}$=0.78, p=0.464). This finding showed that when attention became available, consciousness-related (visible versus invisible) signal automatically captured attention. By contrast, an invisible signal failed to capture attention because performance of the invisible image trials was equivalent with the other two control conditions where only masks were displayed in the absence of the target image. To avoid any confounding from masks, we subtracted the amount of attention generally divided to visible control trials (component **a:** supplied attention, $t_{(18)}$=2.65, p=0.016) from that of visible image trials (component **a+b:** supplied plus captured attention, $t_{(18)}$=3.62, p=0.002), obtaining a clean measure of the amount of captured attention (component **b**). Next, this measure was used to evaluate how the attention captured by the visible image contributed to its conscious experience.

Among 16% probe trials, a probe cue appeared either immediately at 0ms or shortly later at 200ms after image disappeared (**Fig.1D**). Whenever seeing the probe cue, participants orally reported whether the image displayed at that moment was 'seen' and if yes, they needed to report details about its category and identity (recorded online, **Fig.S1** and Methods). Above-chance category classification accuracy proved visible and invisible images were indeed consciously and unconsciously perceived, respectively (**Fig.S1B**), while the decrease of seen ratio (**Fig.1E,** $t_{(18)}$=-3.02, p=0.007) for 200ms versus 0ms probes indicated that the consciousness trace was decaying over a 200ms temporal gap. Calculated using probe and non-probe trials respectively, the magnitude of consciousness's decaying over time was negatively correlated with the amount of captured attention (spearman ρ= -0.49, p=0.033, **Fig. 1G**) but nonsignificant for supplied attention (**Fig.1F,** spearman ρ= -0.18, p=0.449). This finding had been well replicated (**Fig.S2**) in another experiment performed using the same paradigm and comparable sample size but recruiting different experimenter and audio encoders, thus providing specific and compelling evidence showing that the captured attention has been actively integrated into subjective consciousness.

Experiment 2: neural mechanisms of attentional integration into visual consciousness

Experiment 2 aimed to further unlock how attention gets integrated into visual consciousness in the human neocortex by contrasting an unattended condition (number counting task, **online video 3**) versus a fully attended (category classification task, **online video 4**) condition. We recruited 21 drug-resistant epilepsy patients (**Fig.1B** lower panel, **Table S1**), who were implanted a number of iEEG electrodes (3,044 contacts in total) solely based on clinical demands.

Unattended preconscious state beyond unconsciousness: large-scale activation despite functional isolation between the frontal and posterior brain

We first looked at neural responses at the unattended condition when peripheral images were in inattentional blindness [5, 6]. Replicating behavioral results of Exp1, visible signals failed to capture attention relative to the invisible signals in behavior (**Figure.S3**). In terms of neural response, taking a representative contact in the primary visual area (V1) as an example (**Fig. 2A**, responsive for contralateral image; **Fig.S4**, non-responsive for ipsilateral image), V1 response to invisible images was weak and transient (amplitude 40μV, duration 10ms, upper panel) while response for visible images (400μV, 460ms) was about ten times stronger and much more durable (**Fig.2A**, middle and bottom panels). This finding provides direct evidence favoring the existence of recurrent processing despite lack of attention [7].

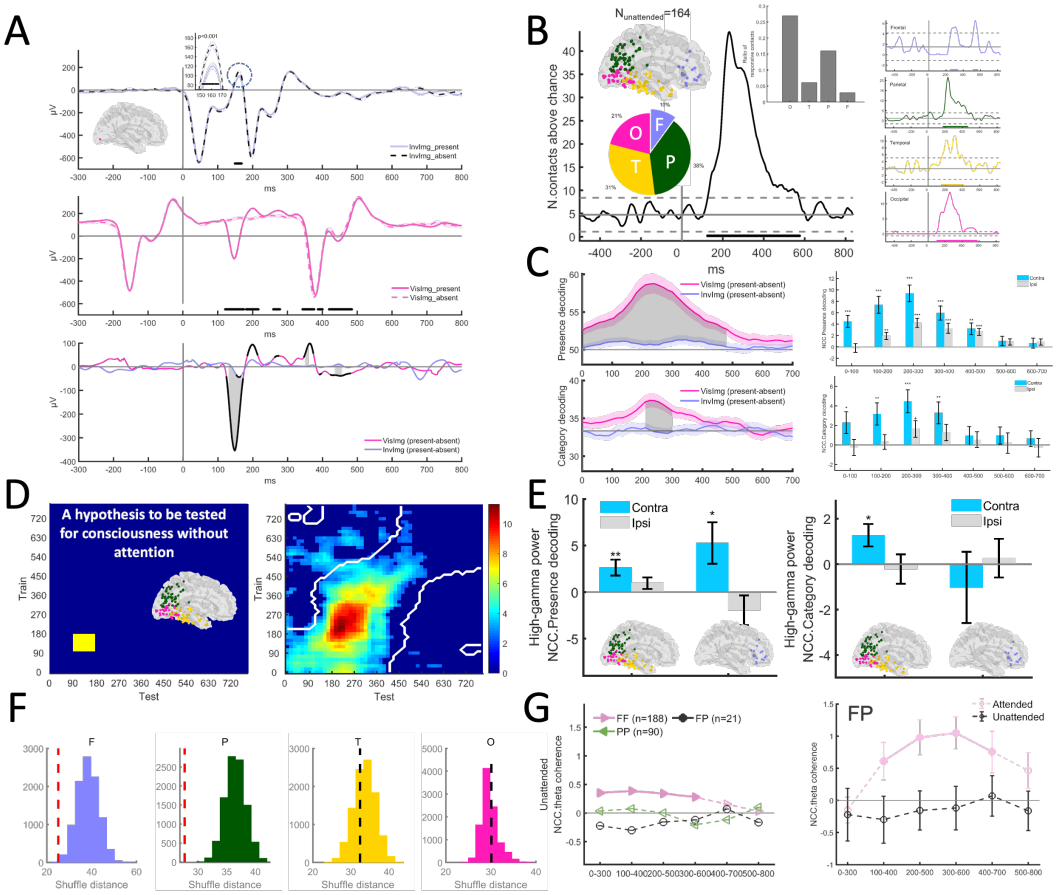


Figure 2. Preconscious visual processing in the unattended condition. (A) An exemplar responsive contact at the primary visual cortex (V1) showed a weak and transient response to an invisible image, whereas a visible image evoked a much stronger and more durable response even though attention was directed elsewhere. (B) Responsive contacts (N=164) extended to widespread regions of the posterior brain and a small subset of the frontal brain. Gray dotted lines indicate threshold of chance level (see Methods). For each lobe of interest, the pie chart and bar chart show percentage among all responsive and implanted contacts, respectively. (C) Posterior activity (pooling responsive contacts at O, T and P lobes) decoded both coarse-grained binary presence and fine-grained categorical contents of the consciousness-related signal (left panel), especially by the contralateral hemisphere (right panel). (D) Coarse-grained representations generalized homogeneously across time for approximately 450 ms (right panel), opposing the hypothesis (left panel) for transient existence of consciousness without attention and suggesting a preconscious state between

unconscious and conscious. (E) A small subset of the frontal brain was also responsive in the preconscious state, with its local high-gamma activity decoding coarse-grained representations only, unlike the posterior brain which decoded both coarse-grained and fine-grained representations. (F) In the frontal and parietal lobes, responsive contacts were closely clustered with a significantly smaller mutual distance than shuffled chance. (G) However, frontal-parietal (FP) coupling was not established in the unattended condition (left panel); on the contrary, FP coupling got established only when attention was engaged (right panel).

Beyond V1, RP activity in the posterior brain also expanded to broader regions of the occipital (O), temporal (T) and parietal (P) lobes (**Fig.2B, Fig.S5**). This posterior activity encoded both binary presence (**Fig.2C**, upper panel) and categorical contents (**Fig.2C**, lower panel) of consciousness-related (visible versus invisible) signal, especially by slow theta oscillations (**Fig.S6**). Importantly, opposing the hypothesis for transient existence of consciousness without attention (**Fig.2D**, left panel), neural representations of binary presence were temporally homogenous across a duration as long as up to 450ms (**Fig.2D**, right panel; **Fig.S7** for transient and discrete representation of fine-grained contents). This length far exceeds 'a conscious but transient and non-reportable state [7, 10])'. As experienced in the demo video (**online video 1**), when you are fully concentrating on counting from the fast-alternating number sequence you cannot explicitly 'see' a blurred peripheral image presented durably for 450ms (unless you pay attention to the image). Alternatively, findings above support the tripartite distinction viewpoint [7] (dating back to Freud [18]) that salient stimuli enter into a 'preconscious' buffer state inter-between unconscious and conscious if not accessed by attention, i.e., 'visible yet not seen' [7].

A small set of the frontal brain was also activated in the preconscious state (**Fig.2B, Fig.S5**), encoding the presence (200-400ms, **Fig.2E** left panel) rather than the contents (right panel) of the preconscious signal by the power of high-gamma oscillations, whose activity correlates with local neuronal spiking [19, 20]. Furthermore, these frontal responsive contacts clustered with a significantly smaller mutual distance than chance (**Fig.2F**, permutation test, $p=0.0042$). This spatial clustering effect was also observed for parietal lobe (**Fig.2F**, permutation test, $p<0.0001$, Table.S2) but not for occipital and temporal lobes. Simultaneously, the peripheral salient signal failed to establish functional coupling (in theta band peaking at 4HZ, **Fig.S8**) between the frontal and posterior brain (**Fig.2G**, right panel showing frontal-parietal disconnection in the unattended condition). Given FP coupling is usually assumed to be associated with conscious perception [7], their disconnection further implies the image was processed in a preconscious buffer state. In this state, the peripheral salient image might have produced an 'attend-to-me' signal, reaching functionally specific regions of the frontal and parietal lobes (particularly the frontal-parietal network and dorsal attention network, **Fig.S9**) to request attention [21], but this request had just been overridden to prevent actual visual distraction [21-23].

Early attended preconscious state: attention reorganizes early posterior activities despite their functional disconnection with the frontal brain.

In the attended condition, the number of responsive contacts sharply increased relative to the unattended condition (**Fig.3A**), with the frontal brain, a critical hub controlling attention [24-26], showing the greatest increase (**Fig.3B**). Pre-stimulus alpha band power of the posterior brain was lateralized to the hemisphere ipsilateral to the attended side (**Fig. S10**), indicating that attention had been pre-allocated to the peripheral image location as required [12]. Also, expectedly, this pre-allocated attention rendered the posterior brain into an elevated baseline state, boosting post-stimulus early visual activity even within the first 100ms (for review see [13, 14], **Fig. 3C** replicating the early effect at the posterior brain, **Fig.3D** showing similar effect at the frontal brain but of an

about 100ms time lag).

Beyond that, intra-between the posterior brain and at 4-HZ peaked theta band (**Fig.S8**), we found attention had induced both coupled and decoupled pairs simultaneously at an early bin (0-300ms post-stimulus onset), with each sensitively detecting the salient emerging signal but leading to an overall zero coherence since they cancelled out each other (**Fig.3E**, red and gray curves for the attended and unattended condition, respectively; **Fig.S11** rigorously testing this U-shape relationship). For OO pairs (**Fig.3F**), notably, coexistence of decoupling and coupling even persisted across time till to late stages. Prior studies revealed similar attention-induced decoupling at the neuronal spiking level ([27, 28]). Extending this mechanism to a macroscopic scale in the human brain, findings here suggest the pre-allocated attention starts to exert its impact at the posterior brain by disrupting old ensembles and creating new ones, especially starting from the early bin.

Despite these early altered posterior activities and connectivity, we found the posterior brain was, however, still functionally isolated from the frontal brain at the same early bin. Frontal-posterior coherence (including the FP pair) neither increased nor decreased in response to the emerging signal (**Fig.3G**), akin to their isolation in the unattended condition (**Fig.2G**), suggesting consciousness had not yet emerged and the image (albeit attended) were still early processed in a preconscious state. Attention supplied does not equal to attention captured/integrated. In Exp.1, we have shown that it is the captured attention that contributes to conscious experience (referring to Exp.1, **Fig.1C, 1F**, and **1G**). Findings here further suggest attention had not yet been captured by the early signal, neither integrated into visual consciousness, even though this signal had been modified by pre-allocated attention.

Late attended conscious state: frontal-posterior coupling evolves within hundreds of milliseconds

Until the medium bins (**Fig.3G**), there appeared significant FP coupling (also for FF and FT pairs, attended condition) in response to visible versus invisible stimuli. This frontal-posterior coupling persisted across medium bins (**Fig.3H**), indicating occurrence of attentional integration and emergence of consciousness during this period. Albeit that, no conscious experience can last indefinitely. In the late bins as conscious experience finally faded, the posterior lobes (O, T, and P) first disentangled from each other (i.e., negative coherence), followed by FF and FT, and finally terminated by FP pairs (**Fig. 3G** and **3H**). These findings illustrate the dynamic nature of attentional integration process, engaging temporally varying whole-brain functional connectivity within hundreds of milliseconds likely accompanying conscious experience.

Researchers holding frontal-centric views of consciousness may treat frontal-posterior coupling as means of globally broadcasting real-time consciousness contents [15], feedback predictions [29] or higher-order meta-representations [30] across the whole brain, a process generating subjective experience. However, as opposed in the following section, such possibilities of broadcasting fine-grained representations had been falsified. Alternatively, we treat the frontal-posterior coupling as a mechanism for attentional integration, integrating frontal attention with posterior contents. In this way, findings here highlight FP coupling as a critical switch indicating either an 'on' or 'off' state of attentional integration into consciousness.

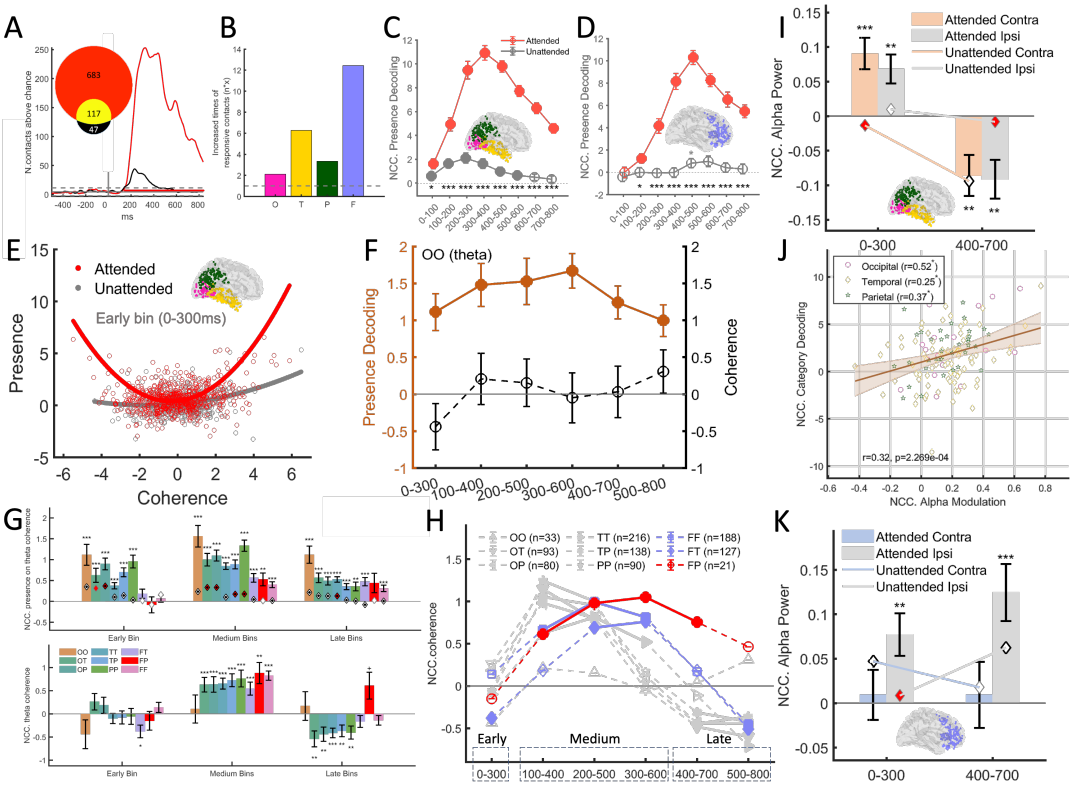


Figure 3. Neural dynamics switching early attentional modulation to late attentional integration. (A) Due to attentional engagement, the number of responsive contacts significantly increased relative to the unattended condition, with (B) the frontal brain showing the largest increase. Decoding performance (±SEM) of binary presence of consciousness was enhanced within the first 100 ms in the (C) posterior brain, followed by a similar enhancement in the (D) frontal brain about 100 ms later. (E) In terms of functional connectivity, theta-phase coupling (positive values on the x-axis) and decoupling (negative values) coexisted intra-lobally within the posterior brain in the early bin (0-300 ms), canceling out each other to zero coherence by average. (F) Such functional reorganization effect was particularly observed for OO pairs which persisted till to the late stage. Coherence values and coherence-based presence decoding performance were tied to the right and let axis, respectively. (G) Despite these early changes intra-between the posterior brain, neither coupling nor decoupling occurred between the posterior and frontal brain at the early bin (±SEM). Subsequently, all four lobes exhibited coupling at medium bins, followed by decoupling at the late bins. (H) FP pairs maintained positive coupling for the longest duration, persisting into the late bin when other pairs became decoupled. (I) Throughout the whole-time course, attention induced a two-stage reversed alpha power pattern in the posterior brain (±SEM), observed in both hemispheres. (J) The amplitude of this reversed modulation effect was positively correlated with the mean decoding performance of the posterior brain for categorical consciousness contents. (K) Conversely, in the frontal brain, alpha power (±SEM) was continuously lateralized to the hemisphere ipsilateral to the peripheral image side.

Posterior brain serves as a module of content while the frontal brain as of attention

To integrate attention with contents, the contents-to-be-perceived need to be precisely determined in the posterior brain. Our finding that attention induces functional reorganization at the posterior

brain agrees well with this idea. Furthermore, previous research suggests attention also engages the inhibitory alpha band oscillation to access stored knowledge [31]. Specifically, an early increase in alpha activity inhibits background noise first, allowing emerging signals to pop out, and then selects the popped-out signal for further amplification by a decrease in alpha activity [31]. Here, as measured by alpha band power, we found clear evidence of initial increase of alpha activity (0-300ms). Simultaneously, the early decrease in micro-saccade behavior (**Fig.S12**), which reflects attentional orienting [32-34], indicated that attention had been successfully captured by these emerging signals [34, 35]. After that, the decreased late (400-700ms) alpha-band power (**Fig.3I**), as well as the increased micro-saccade behavior and dilated pupil size (**Fig.S12** and **S13**; not confounded by horizontal or vertical eye gaze location, **Fig.S14** and **S15**), indicated that these signals were further amplified at a later stage (>300ms).

Notably, such a two-stage structure of alpha activity was only observed at the posterior brain and in the attended condition (**Fig. 3I, 3K** and **Fig.S16**). When we calculated the correlation between the two-stage modulation of alpha power and the mean capacity of content representation, we found significant correlations for the posterior brain (**Fig. 3J**) but nonsignificant for the frontal brain or at the unattended condition (**Fig.S17**). These findings provide direct support for Klimesch [31], implying that the two-stage alpha activity is engaged by attention to filter consciousness contents in the posterior brain.

To complete the process of attentional integration, we propose the supply of attention cannot be interrupted. It is well known that the two frontal brain hemispheres compete for control of attention through mutual inhibition, each striving to direct attention towards its contralateral hemifield [36, 37]. Therefore, disrupting their balance can bias the direction of attention. As found, in the frontal brain, inhibitory alpha power was consistently lateralized to the ipsilateral hemisphere (**Fig. 3K** and **Fig.S16**). This indicates that the ipsilateral frontal lobe was continuously suppressed, thus preventing attention from being distracted and stabilizing attention towards where the image was displayed.

Attentional integration: frontal-posterior loop exchanges coarse-grained signals of consciousness emergence rather than fine-grained consciousness stream

Almost all neuronal theories of consciousness agree that sensory analyses are primarily performed by the posterior brain. Indeed, the posterior brain encoded categorical contents using regional activity (**Fig. 4A**) and inter-regional connectivity (**Fig.4C**, lower left). Importantly, these content representations were continuously and durably maintained across time, aligning well with our smooth experience. Posterior-centric views of consciousness [9, 10] typically attribute conscious experience to this sustained posterior activity.

In contrast, frontal-centric views of consciousness consider this posterior activity as necessary but not sufficient for consciousness [15, 29, 30]. For example, the global neuronal workspace theory posits that consciousness arises from the frontal brain while broadcasting hybrid representations to widely distributed simple processors in the posterior brain [15]. This hypothesis was rigorously tested here and refuted as a result.

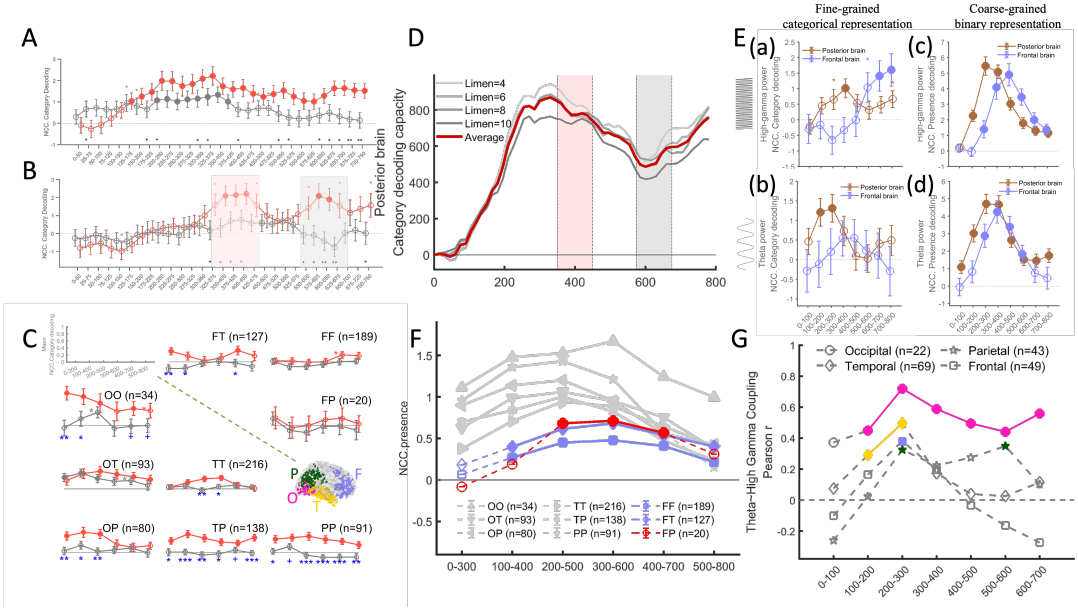


Figure 4. Persistently and globally exchanged signals are coarse-grained but not fine-grained representations of consciousness. (A-C) The posterior brain represented fine-grained categorical contents of consciousness in a persistent profile, as decoded by both regional activity and inter-lobe coherence measures, while the frontal brain represented contents in a phasic profile instead. (D) During the initial period when the frontal brain started to represent contents, the summed amount of content representations of the posterior brain decreased rather than increased. (E) The frontal brain (light blue lines) decoded fine-grained categorical contents using high gamma oscillations (not until the late stage, panel a) rather than slow theta oscillations (panel b), although both oscillations durably represented coarse-grained binary signal of consciousness (panel c and d). (F) This binary presence signal was also represented by measures of inter-lobe theta coherence, across the entire brain and persistently. (G) Conveyed by theta oscillations, these globally exchanged binary signals might coordinate local regional activity via power-based theta and high-gamma cross-frequency coupling, especially in the occipital and parietal lobes.

We first examined the temporal profile of content representations in the frontal brain. Running content decoding analysis on regional activity measures, we found a phasic profile (**Fig.4B**) characterized by an intermediate silence period (450~600ms) of which decoding performance decreased to non-significant ($t_{(56)}$=1.04, p=0.303) relative to an early (350~450ms, paired sample t-test, $t_{(56)}$=-2.80, p=0.007) and a late (>600ms, $t_{(56)}$=-2.24, p=0.029) surge of content representations in the frontal brain. The same analysis on inter-regional connectivity measures at the theta band [38] (**Fig.4C**, upper right) revealed a similar phasic profile for frontal-temporal (FT, corrected p<0.05) and frontal-frontal pairs (FF, uncorrected p<0.05), while the frontal-parietal pair (FP), proposed to gate attentional supply, never encoded contents. In sharp contrast, the posterior brain exhibited a persistent profile (**Fig.4C**, lower left). These discrete content representations of the frontal brain do not align with the temporal continuity of our conscious experience. Instead, they better align with post-perceptual processes [39, 40] (e.g., action), which may unfold either reflexively (possibly corresponding to the early surge) or be held under deliberate control (possibly corresponding to the late surge).

Another key proposal of the frontal view of consciousness suggests that the majority of posterior

brain are blind to consciousness contents before receiving global broadcasts from the frontal brain [2, 15, 29, 30]. If this is true, then, on the one hand, the total amount of content representations in the posterior brain should be significantly enriched immediately after the frontal brain starts broadcasting. However, as revealed, we found a general decrease instead () after the early surge of frontal content representations (**Fig.4D**, 350~450ms, same window outlined at **Fig.4B**). On the other hand, the frontal brain requires a messenger to broadcast contents to distant regions. Nonetheless, although frontal contents could be decoded (preferentially at a late stage) by high gamma oscillations (**Fig.4E**, panel a, light blue curve) correlating with local spiking activity [19, 20], they could never be encoded by slow theta oscillations (**Fig.4E** panel b) subserving long-range communications [26, 41, 42], thus making them less likely to reach distant regions.

Recently, a growing GNWT hypothesis [43] has suggested that the frontal brain can still generate a conscious experience even when neural activity there is silenced (e.g., silent working memory [44, 45]). However, in reality, the frontal brain never halts its response midway in the attended condition (**Fig.3D** and **Fig.S18**). Instead, it actively and persistently maintains another type of coarse-grained representations, which signal the emergence of consciousness in a binary (1/0) manner, in parallel with the posterior brain (**Fig.3C** and **Fig.S18**) and also in both high-gamma and 4-Hz peaked theta oscillations (**Fig.4E**, panel c and d). In addition to regional activity, these binary signals are also represented by inter-lobe connectivity across the whole brain (**Fig.4F**), also in a persistent profile. This profile aligns well with the temporal continuity of our conscious experience. Serving possibly as global coordinators, these binary signals (unaware of consciousness details) conveyed by theta oscillations also have the potential to coordinate distant local activity (particularly in the occipital and parietal lobes) through power-based theta and high gamma cross-frequency coupling (**Fig.4G** and **Fig.S19**).

General discussions

Bringing all elements together, we elucidate how attention becomes integrated into visual consciousness by delineating the distinct roles of supplied attention and captured attention. Initially, supplied attention primes both the frontal and posterior brain regions for subsequent integration before establishing frontal-posterior coupling. During this early phase, the posterior brain selectively filters signals-to-be-perceived by suppressing noise through early alpha inhibition and reshaping functional connectivity by forming new ensembles and disassembling old ones. Meanwhile, the frontal brain maintains attentional stability toward the emerging signal through early and sustained inhibition of the ipsilateral hemisphere. In the absence of supplied attention, as observed in the unattended condition, the lack of these inhibitory mechanisms leads to a failure in consciously experiencing objects.

While supplied attention is essential for consciousness, it's the portion seized by the emerging signal that significantly contributes to the conscious experience. Following the initial stage, the frontal and posterior brain regions become functionally intertwined within a closed loop, exchanging coarse-grained signals of consciousness emergence rather than broadcasting intricate details. Within this loop, the coupling between frontal and posterior regions serves as a gating switch. If this coupling fails to activate—for instance, when the stimulus intensity is too faint (resulting in unconsciousness), when attention is diverted elsewhere despite strong stimulus intensity (unattended preconsciousness), or when the late integration process hasn't commenced in time (attended preconsciousness)—functional isolation between the frontal and posterior brain regions occurs. This isolation prevents frontal attention from accessing posterior contents, effectively reducing integrated attention to zero and hindering the entry of contents into conscious experience.

By contrasting consciousness-related activity in the attended versus unattended conditions, these discoveries illuminate fresh insights into the neural mechanisms of visual consciousness in the human brain. In essence, consciousness emerges from an attentional integration process facilitated by the interplay between the posterior and frontal brain regions, yet neither is independently capable of sustaining consciousness alone.

The frontal brain does not primarily act as a global neuronal workspace for consciousness wherein it broadcasts highly detailed consciousness contents to broadly distributed posterior regions. Instead, it actively operates as an attentional module, detecting emerging signals early on and continuously sustaining attentional resources thereafter, spanning both the early preconscious and late conscious stages. Conversely, its intermittent representation of highly detailed consciousness contents is better suited for post-perceptual processes (e.g., introspective reporting and preparation for actions) in line with task requirements [39, 40]. As a hypothesis for future investigation, we speculate that its early and late content representations are inherently configured to support reflexive and deliberate actions, respectively.

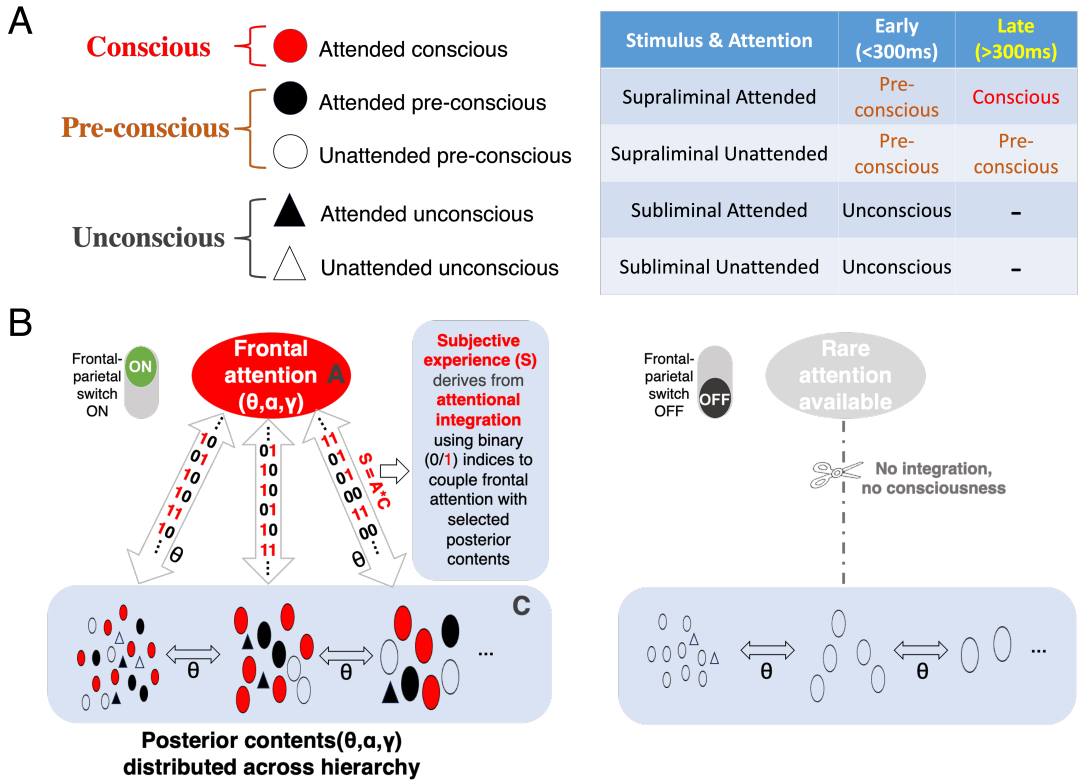
Consciousness cannot be solely attributed to the posterior brain. Previous studies employing no-report paradigms of binocular rivalry [46] and dreams [47] have predominantly identified neural correlates of consciousness (NCCs) in the posterior brain regions. However, our findings in the unattended condition significantly challenge this perspective. Here, despite persistent and uniform neural activity in the posterior brain, conscious experience failed to emerge due to its inability to integrate with frontal attention. Applying the same reasoning, the previous findings [46, 47], seemingly supporting the posterior-centric view of consciousness, might actually arise from posterior activities that are integrated with frontal attention to varying degrees (i.e., robustly versus weakly). In more extreme cases, such as during general anesthesia, a complete functional disconnection between the prefrontal cortex and other brain regions renders the brain entirely unresponsive and unconscious [48].

Combining our findings on the roles and mechanisms of attentional integration in consciousness helps resolve the longstanding debate [43, 49-52] regarding whether consciousness originates from the frontal or posterior brain regions. Our results indicate that it is indeed a collaborative process involving both areas. The structure of the attention-content loop, facilitating the transmission of simple binary signals between the frontal and posterior brain regions, offers fresh insights into the integration of attention into consciousness. This perspective circumvents the shortcomings (Supplementary) of perspectives advocating for a widespread dissemination of highly detailed conscious contents. Moreover, these binary signals could potentially function as a saliency map [53], wherein prominent signals in the posterior brain attract more attention. Experiment 1 illustrates that such attentional capture significantly enhances the conscious experience (**Fig.1C, 1G**). We suggest that this loop structure, operating through mechanisms that are still inadequately understood, might play a pivotal role by utilizing captured attention as a cost while translating corresponding contents into consciousness as a benefit.

Attentional Integration Model (AIM) of Consciousness

Based on these considerations, we originally propose an attentional integration model of consciousness (AIM). Conceptually, AIM holds the tripartite view (**Fig.5A**) proposing consciousness requires attention and can only emerge till to the late stage (>300ms) when preconscious signals get further integrated with captured attention. In terms of neural mechanisms

of consciousness (**Fig.5B**), the fundamental idea of AIM can be encapsulated by a formula: S = A * C. This formula suggests that subjective consciousness (S) arises from the integration of attention (A) with contents (C) via an attention-content loop structure. Within the context of visual consciousness in the human brain (**Fig.5B** left panel), this loop structure comprises the frontal brain, acting as a module for attentional control, and the posterior brain, serving as a module for representing the content stream. Crucially, their effective interaction, facilitated by attentional integration using binary signals, transforms this content stream into our conscious experience. Conversely, when attention is diverted elsewhere and contents (such as normal heartbeats) remain unattended, or when frontal-posterior connectivity is significantly compromised (e.g., during general anesthesia [49]), the disconnection of this loop (**Fig.5B**, right panel) results in contents existing in an unconscious or preconscious state rather than being consciously perceived.



Stimulus & Attention	Early (<300ms)	Late (>300ms)
Supraliminal Attended	Pre-conscious	Conscious
Supraliminal Unattended	Pre-conscious	Pre-conscious
Subliminal Attended	Unconscious	-
Subliminal Unattended	Unconscious	-

Figure 5. Attentional integration model (AIM) of consciousness. (A) AIM holds a tripartite view on the relationship between attention and consciousness. There is a preconscious buffer state inter-between unconscious and conscious (A, left panel). And, consciousness can only emerge till to the late stage (>300ms) when preconscious signals get further integrated with captured attention (A, right panel). (B) in terms of neural mechanism, AIM proposes that consciousness arises from an attention-content loop structure inter-between the frontal and posterior brain (B, left panel), where simple binary signals are exchanged to integrate frontal attention with posterior contents. In the context of visual consciousness in the human brain, contents are initially determined in the posterior brain, regardless of whether they are attended to or not. Subsequently, neural interactions (reflected by theta coherence measures) exchanging binary signals between the frontal and posterior brain enable these emerging contents (i.e., illustrated by red colored circles) to additionally capture more attention and thereby be consciously perceived. Conversely, failure in attentional integration (B, right panel) results in the absence of consciousness.

AIM aligns partially with the Global Neuronal Workspace Theory (GNWT) [15] concerning the

'ignition' aspect, proposing that stimulus-induced frontal-parietal (FP) coupling serves as a gating mechanism for the interaction between the frontal and posterior brain regions. When an appropriate combination of supplied attention and stimulus-evoked posterior activity occurs, the FP switch is activated, establishing a loop between the frontal and posterior brain regions. However, where AIM notably diverges from GNWT is in how this frontal-posterior loop functions post-'ignition'. The primary distinction lies in the type of signals exchanged between the two regions. In AIM, the loop operates by exchanging coarse-grained signals related to attention rather than fine-grained signals representing consciousness contents. This observation suggests that a frontal 'workspace' (although possibly required for post-perceptual processes) is not indispensable for consciousness per se. According to this model, sensory elements do not necessarily need to be routed to the workspace and consolidated there into a unified representation before attaining consciousness. Instead, distributed representations at various levels of the sensory hierarchy can be consciously perceived if effectively integrated with attention.

In the realm of consciousness research, various neuronal models have emerged over the past decades, akin to blind individuals exploring different aspects of an elephant. Alongside the notion of global broadcasting [15], these models encompass ideas such as local recurrent processing [10], higher-order representation [30], top-down prediction [29], causal-effect loop structure [9], and attentional control [54]. Many of these concepts can be discerned within the attention-content loop structure of AIM, albeit from distinct perspectives, yet none present a fully comprehensive view. For instance, the attention-content loop can also be construed as a causal-effect loop [9], where each module can causally influence itself through interaction with the other. The distinction lies in the spatial scales at which these loops operate-either encompassing the entire brain or focusing on the posterior brain. By amalgamating these disparate concepts, AIM stands poised to herald a new milestone in future consciousness research.

Limitations and future directions

This study is not without its limitations. One constraint is our inability to precisely pinpoint the moment at which consciousness emerges in the attended condition. While we interpret frontal-parietal (FP) coupling as a contemporary indicator delineating early preconscious from late conscious states, this aspect warrants more rigorous investigation in the future. Another limitation pertains to the absence of causality. Experimental interventions targeting the neural activity underlying the three components of AIM—accessible contents, supplied attention, and their coupling to capture attention—are necessary to empirically test this model in a causal manner. Lastly, it remains uncertain whether the current findings, derived from visual perception tasks, can be generalized to conscious experiences involving more abstract content such as goals and volition. In such cases, given that the frontal brain would act as a component of both attention and content modules [39, 40, 55], future studies are awaited to elucidate whether the AIM loop structure can persist within the frontal brain or if it necessitates alternative, yet undiscovered mechanisms.

In addition to the outlined limitations, future studies should explore the involvement of subcortical structures in consciousness. Subcortical regions, particularly thalamic nuclei such as the pulvinar, are known to play pivotal roles in regulating arousal [56-58] and directing attention [59-61], potentially through modulation of cortical connectivity [59, 60]. It remains to be investigated whether these subcortical activities and connectivity contribute indirectly to background factors that facilitate consciousness (e.g., arousal), or if they directly influence specific aspects [62] of consciousness. Clarifying this issue is essential for advancing our understanding of the neural mechanisms underlying consciousness.

Beyond the "easy problem" of consciousness lies the "hard problem," which delves into how matter, such as the brain, can give rise to subjective experience [63]. While AIM and the findings discussed here don't directly tackle this problem, they do suggest that addressing the hard problem necessitates a reexamination of the age-old question: what is attention? Essentially, attention can be viewed as a neurobiological form of energy. Understanding how the frontal-posterior loop efficiently utilizes this energy during visual consciousness could provide insights into tackling the hard problem of consciousness. Moreover, in the current era of exploring conscious artificial intelligence agents [64, 65], the attention-content loop proposed for attentional integration may also inspire novel engineering approaches.

Acknowledgement

This work was supported by grants from STI2030-Major Projects (2021ZD0204200, 2021ZD0203800, 2022ZD0205000), The National Natural Science Foundation of China (32100863, 32020103009, 31830037), and the Scientific Foundation of the Institute of Psychology, Chinese Academy of Sciences (E1CX1910).

Main references

1. Li, F.F., et al., *Rapid natural scene categorization in the near absence of attention.* Proceedings of the National Academy of Sciences, 2002. **99**(14): p. 9596-9601.
2. Seth, A.K. and T. Bayne, *Theories of consciousness.* Nature Reviews Neuroscience, 2022. **23**(7): p. 439-452.
3. Koch, C. and N. Tsuchiya, *Attention and consciousness: two distinct brain processes.* Trends in Cognitive Sciences, 2007. **11**(1): p. 16-22.
4. Nikolai, A., *Intracranial EEG: A Guide for Cognitive Neuroscientists*. 2023: Springer Nature 2023.
5. Wolfe, J.M., *Inattentional amnesia*, in *Fleeting memories: Cognition of brief visual stimuli.* 1999, The MIT Press: Cambridge, MA, US. p. 71-94.
6. Mack, A. and I. Rock, *Inattentional blindness*. Inattentional blindness. 1998, Cambridge, MA, US: The MIT Press. xiv, 273-xiv, 273.
7. Dehaene, S., et al., *Conscious, preconscious, and subliminal processing: a testable taxonomy.* Trends in Cognitive Sciences, 2006. **10**(5): p. 204-211.
8. Byrne, A., D.R. Hilbert, and S. Siegel, *Do we see more than we can access?* Behavioral and Brain Sciences, 2007. **30**(5-6): p. 501-+.
9. Tononi, G., et al., *Integrated information theory: from consciousness to its physical substrate.* Nature Reviews Neuroscience, 2016. **17**(7): p. 450-461.
10. Lamme, V.A.F., *Towards a true neural stance on consciousness.* Trends in Cognitive Sciences, 2006. **10**(11): p. 494-501.
11. Landman, R., H. Spekreijse, and V.A.F. Lamme, *Large capacity storage of integrated objects before change blindness.* Vision Research, 2003. **43**(2): p. 149-164.
12. Van der Lubbe, R.H.J., C. Bundt, and E.L. Abrahamse, *Internal and external spatial attention examined with lateralized EEG power spectra.* Brain Research, 2014. **1583**: p. 179-192.
13. Qin, N., et al., *Effects of selective attention on the C1 ERP component: A systematic review and meta-analysis.* Psychophysiology, 2022. **59**(12): p. e14123.
14. Slotnick, S.D., *Attentional modulation of early visual areas.* Cogn Neurosci, 2018. **9**(1-2): p. 1-3.
15. Dehaene, S. and L. Naccache, *Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework.* Cognition, 2001. **79**(1-2): p. 1-37.
16. Baars, B.J., *Global workspace theory of consciousness: toward a cognitive neuroscience of human experience*, in *Progress in Brain Research*, S. Laureys, Editor. 2005, Elsevier. p. 45-53.
17. Lavie, N., D.M. Beck, and N. Konstantinou, *Blinded by the load: attention, awareness and the role of perceptual load.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1641): p. 20130205.
18. Freud, S., *An outline of psychoanalysis*. An outline of psychoanalysis. 1949, Oxford, England: W. W. Norton. 127-127.
19. Ray, S., et al., *Neural Correlates of High-Gamma Oscillations (60-200 Hz) in Macaque Local Field Potentials and Their Potential Implications in Electrocorticography.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2008. **28**: p. 11526-36.
20. Nir, Y., et al., *Coupling between Neuronal Firing Rate, Gamma LFP, and BOLD fMRI Is Related to Interneuronal Correlations.* Current Biology, 2007. **17**(15): p. 1275-1285.
21. Sawaki, R. and S.J. Luck, *Capture versus suppression of attention by salient singletons: electrophysiological evidence for an automatic attend-to-me signal.* Atten Percept Psychophys, 2010. **72**(6): p. 1455-70.
22. Stilwell, B.T., et al., *The role of salience in the suppression of distracting stimuli.* Psychonomic Bulletin & Review, 2023.
23. Gaspelin, N. and S.J. Luck, *The Role of Inhibition in Avoiding Distraction by Salient Stimuli.*

Trends Cogn Sci, 2018. **22**(1): p. 79-92.
24. Gregoriou, G.G., et al., *Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4.* Nature Neuroscience, 2014. **17**(7): p. 1003-1011.
25. Fiebelkorn, I.C., M.A. Pinsk, and S. Kastner, *A Dynamic Interplay within the Frontoparietal Network Underlies Rhythmic Spatial Attention.* Neuron, 2018. **99**(4): p. 842-853.e8.
26. Helfrich, R.F., et al., *Neural Mechanisms of Sustained Attention Are Rhythmic.* Neuron, 2018. **99**(4): p. 854-865.e5.
27. Cohen, M.R. and J.H.R. Maunsell, *Attention improves performance primarily by reducing interneuronal correlations.* Nature Neuroscience, 2009. **12**(12): p. 1594-1600.
28. Mitchell, J.F., K.A. Sundberg, and J.H. Reynolds, *Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4.* Neuron, 2009. **63**(6): p. 879-888.
29. Clark, A., *Whatever next? Predictive brains, situated agents, and the future of cognitive science.* Behav Brain Sci, 2013. **36**(3): p. 181-204.
30. Brown, R., H. Lau, and J.E. LeDoux, *Understanding the Higher-Order Approach to Consciousness.* Trends Cogn Sci, 2019. **23**(9): p. 754-768.
31. Klimesch, W., *Alpha-band oscillations, attention, and controlled access to stored information.* Trends in Cognitive Sciences, 2012. **16**(12): p. 606-617.
32. Martinez-Conde, S., J. Otero-Millan, and S.L. Macknik, *The impact of microsaccades on vision: towards a unified theory of saccadic function.* Nature Reviews Neuroscience, 2013. **14**(2): p. 83-96.
33. Hafed, Z.M. and J.J. Clark, *Microsaccades as an overt measure of covert attention shifts.* Vision Research, 2002. **42**(22): p. 2533-2545.
34. Engbert, R. and R. Kliegl, *Microsaccades uncover the orientation of covert attention.* Vision Research, 2003. **43**(9): p. 1035-1045.
35. Wang, C.-A., et al., *Saliency and priority modulation in a pop-out paradigm: Pupil size and microsaccades.* Biological Psychology, 2020. **153**: p. 107901.
36. Duecker, F. and A.T. Sack, *The hybrid model of attentional control: New insights into hemispheric asymmetries inferred from TMS research.* Neuropsychologia, 2015. **74**: p. 21-29.
37. Kinsbourne, M., *Hemi-neglect and hemisphere rivalry.* Adv Neurol, 1977. **18**: p. 41-9.
38. Bosman, Conrado A., et al., *Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas.* Neuron, 2012. **75**(5): p. 875-888.
39. Fine, J.M. and B.Y. Hayden, *The whole prefrontal cortex is premotor cortex.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2022. **377**(1844): p. 20200524.
40. Haggard, P., *Human volition: towards a neuroscience of will.* Nature Reviews Neuroscience, 2008. **9**(12): p. 934-946.
41. Buzsáki, G. and M. Vöröslakos, *Brain rhythms have come of age.* Neuron, 2023. **111**(7): p. 922-926.
42. Kam, J.W.Y., et al., *Default network and frontoparietal control network theta connectivity supports internal attention.* Nature Human Behaviour, 2019. **3**(12): p. 1263-1270.
43. Consortium, C., et al., *An adversarial collaboration to critically evaluate theories of consciousness.* bioRxiv, 2023: p. 2023.06.23.546249.
44. Stokes, M.G., *'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework.* Trends Cogn Sci, 2015. **19**(7): p. 394-405.
45. Trübutschek, D., et al., *A theory of working memory without consciousness or sustained activity.* Elife, 2017. **6**.
46. Frässle, S., et al., *Binocular rivalry: frontal activity relates to introspection and action but not to perception.* J Neurosci, 2014. **34**(5): p. 1738-47.
47. Siclari, F., et al., *The neural correlates of dreaming.* Nature Neuroscience, 2017. **20**(6): p. 872-878.
48. Zelmann, R., et al., *Differential cortical network engagement during states of*

un/consciousness in humans. Neuron, 2023. **111**(21): p. 3479-3495 e6.
49. Melloni, L., et al., *Making the hard problem of consciousness easier.* Science, 2021. **372**(6545): p. 911-912.
50. Raccah, O., N. Block, and K.C.R. Fox, *Does the Prefrontal Cortex Play an Essential Role in Consciousness? Insights from Intracranial Electrical Stimulation of the Human Brain.* The Journal of Neuroscience, 2021. **41**(10): p. 2076-2087.
51. Koch, C., et al., *Posterior and anterior cortex — where is the difference that makes the difference?* Nature Reviews Neuroscience, 2016. **17**(10): p. 666-666.
52. Solms, M., *Dreaming and REM sleep are controlled by different brain mechanisms.* Behavioral and Brain Sciences, 2000. **23**(6): p. 843-850.
53. Li, Z., *A saliency map in primary visual cortex.* Trends in Cognitive Sciences, 2002. **6**(1): p. 9-16.
54. Graziano, M.S.A., *The Attention Schema Theory: A Foundation for Engineering Artificial Consciousness.* Frontiers in Robotics and AI, 2017. **4**.
55. Dolan, Ray J. and P. Dayan, *Goals and Habits in the Brain.* Neuron, 2013. **80**(2): p. 312-325.
56. Redinbaugh, M.J., et al., *Thalamus Modulates Consciousness via Layer-Specific Control of Cortex.* Neuron, 2020. **106**(1): p. 66-75.e12.
57. Mofakham, S., et al., *Electrocorticography reveals thalamic control of cortical dynamics following traumatic brain injury.* Communications Biology, 2021. **4**(1): p. 1210.
58. Mofakham, S., et al., *Injury to thalamocortical projections following traumatic brain injury results in attractor dynamics for cortical networks.* Prog Neurobiol, 2022. **210**: p. 102215.
59. Schmitt, L.I., et al., *Thalamic amplification of cortical connectivity sustains attentional control.* Nature, 2017. **545**(7653): p. 219-223.
60. Saalmann, Y.B., et al., *The pulvinar regulates information transmission between cortical areas based on attention demands.* Science, 2012. **337**(6095): p. 753-6.
61. Zhou, H., R.J. Schafer, and R. Desimone, *Pulvinar-Cortex Interactions in Vision and Attention.* Neuron, 2016. **89**(1): p. 209-20.
62. Koch, C., et al., *Neural correlates of consciousness: progress and problems.* Nature Reviews Neuroscience, 2016. **17**(5): p. 307-321.
63. Chalmers, D., *Facing up to the problem of consciousness.* Journal of Consciousness Studies, 1995. **2**(3): p. 200-19.
64. Butlin, P., et al., *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness.* ArXiv, 2023. **abs/2308.08708**.
65. Blum, L. and M. Blum, *A theory of consciousness from a theoretical computer science perspective: Insights from the Conscious Turing Machine.* Proceedings of the National Academy of Sciences, 2022. **119**(21): p. e2115934119.

Method

Participants

Twenty-one healthy participants (16 females, 24.19±2.54 years) took part in Experiment 1. For two participants, audio data recording oral reports for probe trials in the posttest session were not correctly recorded due to technical issue. These two participants were thus excluded from analysis of oral reports but were included in the analyses of counting performance. To replicate results of Exp 1, a follow-up experiment (more details see Fig.S2) was designed and performed using roughly the same paradigm and comparable sample size (20 healthy colleague students, 16 females, age=23.1 ± 1.88 years) but recruiting a different experimenter and two new audio encoders.

Twenty-one drug-resistant epilepsy patients (10 females, 24.48±6.91 years) participated in Experiment 2. Nineteen patients were from Tiantan Hospital, Capital Medical University, and two patients were from Yuquan Hospital, Tsinghua University. These patients underwent intracranial electrode implantation as part of their clinical evaluation for epilepsy surgery when noninvasive studies could not adequately localize the origin of their seizures. Among them, six patients did not have their pupil data recorded: the first four (S01-04) due to the eye tracker not being in place, and the other two (S16 and S17) due to poor eye signal quality, which prevented successful eye tracking. These six patients were excluded from pupil data analysis but were included in other analyses.

For both Experiment 1 and 2, all participants had normal or corrected-to-normal vision, had no history of head trauma or encephalitis, and provided written informed consent. The research protocol of Experiment 1 was approved by the review board at the Institute of Psychology, Chinese Academy of Sciences, and that of Experiment 2 was approved by the institutional review boards at the hospital sites.

Apparatus

The stimuli were presented using Psychophysics Toolbox [1, 2] for MATLAB (The Mathworks, Natick, MA, USA) on a 21-inch gamma-corrected CRT monitor with a resolution of 1920 × 1020 pixels and a refresh rate of 60 Hz. An infrared eye tracker system Eyelink Portable Duo (SR research, Kanata, Ontario, Canada) was used to monitor gaze location, (micro)saccade behavior and pupil size in both Exp1 and 2. Observers viewed the screen from about 60 cm away, using a chin rest to stabilize the head position in Exp1. The chin rest was not used in Exp 2, where, instead, epilepsy patients sitting in bed were told to minimize their head movement during the recording period (in remote recording mode).

Materials

Visual stimuli used consisted of central and peripheral streams. The central visual stream consisted of four numerical values (4, 5, 6, or 7) which were employed to create a rapidly alternating number sequence at a frequency of 10 Hz, presented within the central visual field.

The peripheral visual stream consisted of 48 visual stimuli, comprising 16 animal images, 16 object images, and 16 face images, and were presented at 4.0° visual eccentricity (Fig.1A). Additionally, eight visual masks originally constructed using line segments and then created variants by rotating and flipping this original mask were included in the peripheral visual stream. And a pair of nonmatching masks was randomly selected for each trial. The animal and object images, as well as the visual masks, were sourced from [3]. The face images which are all emotionally neutral and differ in gender half by half, were initially obtained from [4, 5] and were then converted into simple line drawings using an open-source tool (http://jianbihua.renrensousuo.com/).

Both central and peripheral streams were controlled using Psychophysics Toolbox version 3 for MATLAB. All stimuli were presented on a white background (1920 × 1080 px, 60Hz monitor; thinkpad T490). Participants were seated 60cm from the monitor, thus making the average size of the line drawings 4.8° in side length, and the masks 6°.

Paradigm design and procedure

In the experiment, participants were instructed to fixate on a central cross while a central fast-alternating number stream and a peripheral visual stream were presented simultaneously and unfolding across time in each own pace. Participants maintained fixation on a small black fixation dot (0.5° in diameter) that was continuously visible throughout the procedure. Four demo videos of

the stimulus presentation sequences are available on the Supplementary material.

Central number sequence

For each trial, the central visual stream was a pseudo-randomly ordered temporal sequence of single number characters altering at 10 Hz. For number-counting blocks, the target number to be counted was chosen randomly and informed of the participant at the beginning of each block.

Peripheral visual stream consisting of masks and image

For each block of 72 trials, forty-eight images were divided into four groups (A, B, C, and D), each containing four animals, four objects, and four faces, with images from only one group used in any given block. Within a block, 24 visible image trials (twice per image), 24 invisible image trials (twice per image), 12 visible control trials and 12 invisible control trials were conducted in random order. All trials were conducted in mixed and random order. There was a 1000ms blank intertrial interval that occurred after each trial and before the next.

Visibility manipulation closely followed the methodology outlined in [3, 6]. Visible trials in the peripheral stream comprised a 'mask-blank-image-blank-mask' sequence, with each mask and blank period lasting 200ms and 100ms, respectively, resulting in a total trial duration of 633ms. The image was surrounded by blanks, making it weakly masked and explicitly visible. In contrast, invisible trials followed a 'blank-mask-image-mask-blank' sequence, with the image closely surrounded by masks, rendering it heavily masked and thus invisible to observers. Both visible and invisible trials were followed by a 1000ms blank intertrial interval before the next trial.

For both visible and invisible trials, respectively, their control trials were set to mirror the mask and blank sequence, but replaced the critical image with a complementary blank of the same duration (33ms). We subtracted different measures of interest (see following) of control trials from that of invisible and visible trials (exampled at Fig.2A), obtaining a clean mask-free measure tied to visibility state when the image was either invisible or visible, respectively.

General procedure of Exp1 and Exp2

Both Experiment 1 and Experiment 2 were conducted using a block design. In all trials and blocks, the central number stream was consistently presented alongside the peripheral image and mask stream. We manipulated attentional focus across different conditions: directing attention primarily to the central number stream (pretest of Exp1 and unattended blocks of Exp2), focusing mostly on the central number stream but also partially on the peripheral stream (posttest of Exp1, achieved by adding random and occasional probes), or directing attention primarily to the peripheral stream (attended block of Exp2).

In the pretest session of Experiment 1 (i.e., without probes, consisting of 2 blocks), participants were instructed to focus solely on counting while disregarding the peripheral stream. To achieve so, participants were not informed of the posttest session, which included probes, until after the pretest session concluded. The peripheral stream was presented in either the left or right visual field, with the order determined randomly for each trial, but no more than three successive trials featured the same side.

In the posttest session of Exp1 (i.e., with probe, 10 blocks), we still required participants to set priority the counting task and perform their best. The only difference was that we designed to shift a bit of attention to peripheral image, by occasionally probing its visibility in 16% randomly selected probe trials (**Fig.1D**, Method). In these trials, a probe cue, i.e., a square outlining the image (side length 4.8°), interrupted the trial sequence at two timepoints (0ms or 200ms after image offset). We set 12 probes per block, with any two probes interleaved by 4~8 trials though participants were not informed of this rule. These probes appeared in 4 visible image trials, 4 invisible image trials, 2 visible control trials and another 2 invisible control trials. Half probes appeared at 0ms after image offset, half at 200ms after image offset.

When seeing this probe, participants first pressed keys to complete counting as usual, to indicate how many target numbers they saw before probe. This was designed to hold the majority of attention towards counting, performing the counting task as best as possible while leaving only a small portion distracted to peripheral images. Otherwise, if too much attention was distracted to image, wo proposed too much attention supplied to image would lead to a saturation ceiling effect hindering us to detect to what extent consciousness might decay within a 200ms temporal gap (200ms versus 0ms). After the first key-press, color of the probe cue turned red, indicating the start

of audio recording. Seeing this, participants were required to orally reported whether they saw the target image surrounded by masks. If yes, further reported its category and detailed identity. A typical report for probe within a visible image trial was like 'I am sure I saw an animal image with a tail but I am not sure whether it was a cat or a dog'.

In experiment 2 (block design), attention was directed to different streams randomly alternating block by block but with no more than two repetitions, to peripheral image in attended blocks but to central counting stream in unattended blocks. In attended blocks of Exp2, participants focused on an image categorization task, but neglecting the central number stream completely. To help stabilize attention to peripheral image, peripheral stream was displayed consistently in one visual field (either left or right) for all trials but randomly set across blocks. At the end of each trial, participant was provided a forced choice to class the peripheral critical image into either animal, object or face by pressing a corresponding key. In attended blocks, counting task was not performed any more. In unattended blocks of Exp2, same stimulus streams were presented on the screen and the only difference was that participants performed the central counting task instead. The image categorization task was not performed anymore and participants were required to totally ignore the peripheral stream.

Audio data collection and offline coding.

Audio data were collected for probe trials during the posttest session of Experiment 1. The laptop's built-in recording device was activated immediately after participants pressed the response key which indicated the times they saw the target number. Then, participants orally reported whether they perceived the critical image (excluding masks) at the onset of the probe cue, if so, they identified the category and identity of the image. Audio data were recorded in real-time and later analyzed offline by two researchers. Importantly, both coders were blinded to the experimental conditions to ensure unbiased coding.

The audio data were coded into three dimensions. In the first dimension of visibility (i.e., whether the critical image was seen), scores of 0-3 corresponded to surely not seen, seemingly not seen, seemingly seen, and surely seen, respectively. In the dimensions of categorical and item-level accuracy (i.e., whether the reported category/item was correct), scores of 0-4 corresponded to surely wrong, likely wrong, neutrally not sure, likely correct, and surely correct, respectively. Between two coders, interrater reliability was high in all three dimensions (correlation coefficient close to 1, Fig.S1A).

Pupil data collection

During Experiment 1, participants rested their heads on a chin-rest to limit head movement. In Experiment 2, the chin-rest constraint was removed to increase comfort for epilepsy patients who were sitting in bed. However, participants in both experiments were instructed to minimize head movements during the recording period. In both experiments, eye-tracking data were acquired using the EyeLink Portable Duo SR Research eye-tracker with 1 ms temporal resolution (at 1000 Hz frequency) in remote head-free-to-move mode, recording data from the monocular eye with a higher signal-to-noise ratio. The camera-to-eye distance roughly fluctuated around 60 cm.

Pupil data preprocessing

The raw sampling data was preprocessed using the Matlab interface to Eyelink EDF files (Version 1.0.1) provided by SR research to identify eye-tracking indices. Relevant eye-tracking events, such as fixations, blinks, and saccades (especially microsaccades), were detected for each participant. This was achieved using the functions edfExtractInterestingEvents.m and edfExtractMicrosaccades.m. The relevant variable values and timing information were then extracted using edfExtractVariables.m and edfExtractKeyEventsTiming.m, respectively.

For each participant, time series of eye-related variables (pupil size and microsaccade ratio, see following) were calculated independently for each combination of attention (attended versus unattended) and visibility (visible versus invisible, as well as corresponding control trials). This was based on epochs lasting from 500 ms before image onset to 1500 ms after image onset. Trials with missing values (typically caused by eye blinks) exceeding 25% of the timepoints were discarded from subsequent analysis.

Pupil data analysis.

Pupil Size. Following the protocol provided by SR Research (version 1.0), we utilized an artificial eye to convert arbitrary pupil size units to millimeters. This process involved calculating a scaling factor (0.106) based on a distance parameter set at 60 cm. To convert pupil data from arbitrary units to diameter, we multiplied the square root of the pupil area data by the scaling factor (0.106). For instance, for a typical pupil area data value of 1216, which represents the average mean of trials from our first recording block, the corresponding pupil diameter would be sqrt (1216) * 0.106 = 3.70 mm.

It's important to note that this scaling factor is tied to the distance parameter and would change if the distance parameter were altered. Additionally, pupil data in both experiments was recorded in a head-free-to-move mode (i.e., remote mode), which could lead to fluctuations in the distance parameter. Although we reminded participants to maintain the ideal distance setting (60 cm), especially in Experiment 2 where a chin rest was not used, these fluctuations may have been more pronounced. To this end, all subsequent analyses on pupil size were conducted using the pupil area data in arbitrary units.

Microsaccade detection and analysis

Microsaccade events were extracted from the raw eye position data using the algorithm described in Engbert & Kliegl (2003), implemented with default settings in edfExtractMicrosaccades.m. Velocity was computed using a time span of 20 ms around the current sample. The onset of a microsaccade event was defined as the time point at which velocity exceeded the trial's median velocity by 6 or more standard deviations. The minimal microsaccade duration was set to 12 ms, and the minimal time allowed between two microsaccades was also set to 12 ms; otherwise, they were merged into a single event spanning from the onset of the first microsaccade to the offset of the second one. This approach enabled the detection of microsaccade events for every trial. Timepoints corresponding to microsaccade events were coded as 1, while other timepoints were coded as 0.

The microsaccade ratio was then calculated as the percentage of trials containing a microsaccade event for each timepoint (1 ms), resulting in a percentage time course. This time course was smoothed using a 100 ms moving window (following [7]), replacing each center value by averaging values within its window. The smoothed data was then detrended to remove baseline drift or slow trends from signals using a polynomial detrending algorithm with the break point set to image onset.

iEEG: electrode/contact reconstruction and localization

Postoperative computed tomography (CT) images were coregistered to preoperative T1-weighted MR images using FreeSurfer (v6.0.0, http://surfer.nmr.mgh.harvard.edu/). The registration was visually verified and manually adjusted if necessary. The implanted electrodes and contacts were reconstructed using the stereotactic localization software [8].

Initially, the intracranial electroencephalography (iEEG) contacts were labeled with a cortical area based on Freesurfer's anatomical parcellation [9] in native space. Subsequently, all converted iEEG contacts were further labeled in standard space using the Desikan-Killiany (DK) labeling protocol [10], which is a widely used human cortical labeling protocol. During this process, the coordinates of all contacts were mapped onto a standard MNI space. Based on the MNI coordinates, contacts were further labeled to functional parcellations using atlases like the Human Connectome Atlas [11] and its extension [12], as well as the Yeo 7 network [13], to evaluate their functional roles when necessary.

To distinguish gray matter (GM) contacts from white matter (WM) ones, a spherical region of interest (ROI) with a radius of 3.5 mm was first created for each contact around the electrode coordinate in native space and then transformed to the standard space. The percentage of voxels labeled as GM within the ROIs (inclusion) was calculated and compared to the percentage of voxels outside the ROIs (exclusion). If the probability of inclusion exceeded that of exclusion, the ROI was classified as GM; otherwise, it was categorized as WM or an unknown area. For each included ROI, the probability values were compared between brain areas involved, and the ROI contact was

labeled as the structural area with the maximal probability value.

iEEG: data collection and preprocessing

For all 21 patients, intracranial electroencephalography (iEEG) data were recorded using a Nihon-Kohden recording system at a sampling rate of 2000 Hz. To ensure data quality and avoid potential confounds from interictal activity or seizures, we verified that no subclinical or clinical seizures occurred during or immediately before the experiment.

After acquisition, the iEEG data were analyzed using MATLAB combined with open-source toolboxes. The data were first bandpass-filtered from 0.5 to 200 Hz using a zero-phase delay finite impulse response filter and then downsampled to 500 Hz. Noise at 50 Hz and its harmonics were removed using a notch filter. Subsequently, iEEG traces from each contact were visually inspected to identify and exclude bad channels and time segments containing artifacts such as epileptic activity, large-amplitude slow-wave drifts, or high-frequency activity. Importantly, this visual inspection was performed in a blinded manner to the experimental conditions. For each patient, the time series of all remaining valid contacts were re-referenced by subtracting the mean signal of all white matter (WM) contacts. This step aimed to remove globally shared background signals. This unbiased approach avoids introducing inter-contact variance due to the re-referencing procedure and enhances the replicability of the findings.

iEEG: ERP analysis

For event-related potential (ERP) analysis, preprocessed traces were first bandpass-filtered between 0.5 and 30 Hz before segmenting epochs. Since band filtering usually introduces edge artifacts at the head and tail of traces, this filtering step was crucial to ensure that data inter-between the head and tail (including all data of interest) were not contaminated by these artifacts. After filtering, the traces were segmented into epochs spanning from -500 ms before to 1000 ms after image onset. Subsequently, outliers in the remaining data (time * epochs) of each participant were identified and replaced by NaNs if their voltage value surpassed 5 standard deviations of all epochs at any time point. After baseline correction (200ms pre-stimulus baseline), the epoch data were submitted to statistical analyses. Note that these processes were performed upon non-NaN elements, while treating NaNs as missing values.

iEEG: time-frequency decomposition

For time-frequency analysis (TFA), we first bandpass-filtered the preprocessed traces between 0.5 and 200 Hz before segmenting epochs. Prior to epoch segmentation, outliers within each trace were identified and replaced with NaNs, following the same procedure as the event-related potential (ERP) analysis. After that, NaN values within traces were interpolated using the inpaint_nans.m function (https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans), which uses a simple plate metaphor, before time-frequency decomposition.

The TFA decomposition was conducted using Morlet wavelets via the newtimef.m function of the EEGLAB toolbox. The length of the wavelet increased linearly from 3 cycles at 2 Hz to 54 cycles at 180 Hz, corresponding to a sliding window length of 1670 ms. This modified wavelet length was chosen to optimize the tradeoff between temporal resolution at lower frequencies and stability at higher frequencies. The TFA outputted complex values combining both amplitude and phase values for each time-frequency point, resulting in a 2-D plane with a log-spaced frequency dimension (60 bins from 2 to 180 Hz) and a linearly spaced time dimension sampled at 250 Hz (4ms resolution). In the dimension of time, this 2-D plane was segmented into epochs ranging from -600ms to 1000ms around image onset.

iEEG: neural oscillation power

For each complex value at every time-frequency point, we extracted the amplitude and calculated the power as the squared amplitude. The power values of each epoch were then z-normalized across time to reduce sensitivity to noisy trials and allow for comparison of power values across different frequency bins. Subsequently, every normalized power epoch was baseline corrected using a 200-ms pre-stimulus baseline. Finally, the power values were averaged across frequencies to assess the change in mean band power following image onset for any frequency bands of interest. It is

important to note that mask-evoked power change was specifically measured by control trials, which could have been removed via the contrast between image (masks plus image) versus control (masks only) trials, resulting in a clear measure tied to visible or invisible processing in either the attended or unattended condition.

To test whether inhibitory alpha band power during the pre-stimulus period was selectively allocated to the hemisphere ipsilateral to the peripheral visual stream, we contrasted contralateral versus ipsilateral trials on its oscillation power for each contact located in either the left or right hemisphere. Baseline removal and normalization across time, both of which suppress baseline variance between trials, were not conducted anymore. Instead, power values were normalized across epochs for each time point to scale the data while preserving the variance between trials. Since visual masks and the target image were arranged in different timing sequences between visible versus invisible trials (but in the same sequence within each), which might differently alter the temporal profile of alpha power activity, we performed this analysis in parallel for visible trials and invisible trials, respectively.

iEEG: mutual distance between contacts

To assess the spatial clustering of responsive contacts in each lobe, particularly for the unattended condition, we computed the mutual distance between any two responsive contacts based on their MNI coordinates. To reduce variability, we first flipped the MNI coordinates of right hemisphere contacts to symmetric locations at the left hemisphere. Then, Euclidean distance was calculated for any two contacts. The mutual distance was finally calculated as the mean of all pairwise Euclidean distance.

Next, we performed a surrogate procedure to create a chance distribution of mutual distances. For each electrode, including the one with the maximum number of responsive contacts, we randomly replaced it with another electrode that contained an equivalent or greater number of contacts within the same lobe. From these contacts, an equal number of contacts were randomly selected, even if this included non-responsive contacts. Each electrode and contact could only be selected once at most for each surrogate. By repeating this procedure 10,000 times, we generated a chance-level distribution of mutual distances. This distribution was used to rank the true distance value, providing an indication of the spatial clustering of responsive contacts relative to chance.

iEEG: functional connectivity between contact pairs

Between any two responsive contacts, functional connectivity was calculated across time using the inter-site phase coherence method (ISPC-time). In this study, an approach based on imaginary coherence [14] was employed to mitigate the impact of a common input on interregional neural synchronization [15, 16]. Using code from [17], imaginary coherence was calculated as:

$$ISPC_{xy} = \frac{\left|Im\left(n^{-1}\sum_{n=1}^{n=N} S_{xy}\right)\right|}{\sqrt{\left(n^{-1}\sum_{n=1}^{n=N} S_{xx}\right)\cdot\left(n^{-1}\sum_{n=1}^{n=N} S_{yy}\right)}}$$

where S_{xy} denotes the cross-spectral density between activities at contacts x and y, S_{xx} and S_{yy} are the autospectral densities for contacts x and y, N denotes the number of time bins, and ISPC represents the absolute imaginary coherence. Cross- and autospectral densities were computed using Morlet wavelets introduced above.

ISPC-time analysis was conducted on time series of complex values obtained from time-frequency decomposition via wavelet convolution. For each trial, ISPC-time was computed between any two contacts over 300 ms time segments in a sliding manner. The resulting ISPC-time values were then averaged across corresponding trials to obtain coherence measures for each combined condition across attention and visibility. In this study, six sliding time segments of a 300 ms window size were used: 0-300 ms, 100-400 ms, 200-500 ms, 300-600 ms, 400-700 ms, and 500-800 ms, to illustrate how coherence measures evolved over time.

The selection of time segment length involves a trade-off between signal-to-noise ratio and temporal precision. Longer time segments, encompassing more oscillation cycles, yield more robust and

higher signal-to-noise estimates of ISPC-time. However, this comes at the expense of decreased temporal precision for time-varying and task-related modulations of ISPC-time, which are the main focus of this study. To reduce susceptibility of ISPC-time to noise or nonrepresentative data, the resulting ISPC-time value for each condition was transformed to $ISPC_Z$ (Rayleigh's Z, using code from [17]) using the formula:

$$ISPC_Z = N * ISPC^2$$

where N is the number of timepoints. A p-value can be computed from $ISPC_Z$ to determine statistical significance [17].

In addition to ISPC-time, ISPC can also be computed across trials (ISPC-trials), offering the advantage of higher temporal resolution similar to the raw data. Using ISPC-time, the preferred phase angle difference can vary between trials as long as the degree of clustering is similar across trials. In contrast, ISPC-trials additionally require the preferred phase angle difference to be similar at each time-frequency point across trials. It has been demonstrated that ISPC-time can be high while ISPC-trials is low, even when calculated using the same trials ([17], figure 26.2). In this study, our aim was to assess the degree of phase clustering for each condition without necessarily requiring phase differences to cluster around the same angles across trials. Therefore, we utilized ISPC-time rather than ISPC-trials.

iEEG: decoding analysis

The decoding methodology typically involved three main components: a classifier, label data and feature data, with the aim of classifying feature data into different labels. In terms of classifiers, classifiers used in the present study were Support Vector Machine (SVM) classifiers with a linear kernel. Without otherwise delineated, classifiers were typically trained in a sliding bin manner (bin size = 25ms, step size = 15ms), with 10ms overlap between two neighboring bins. In this way, models were trained independently for each contact and each bin.

Without otherwise delineated, the feature data typically used for decoding were multi-variate, consisting of mean ERP response and the mean normalized power of six contiguous frequency bands (2-4, 4-8, 8-13, 13-30, 30-60, and 60-150 Hz), with all averaged within the bin. In other scenarios, feature data could also consist of some single variate measures (e.g., theta power or high gamma power), including phase coherence measures calculated for each trial, to better evaluate contribution of specific features.

Setting of the label data reflect our design of interest. Basically, two types of labels varying in granularity were trained: one type of coarse-grained representation of image presence and another one of fine-grained representation of categorical content. The former (coarse-grained) utilized all trials, including both image trials (labeled as 'image present') and control trials (labeled as 'image absent'). Models were trained to predict whether an image had presented (chance level=1/2) under each combined condition of attention and visibility. Since the number of image trials was double that of control trials, each training was split into two runs, with each run using non-overlapping half of the image trials to distinguish with the control trials, and their decoding performance was averaged across two runs. In each run, an 8-fold cross-validation approach was utilized, where the feature data were divided into eight equivalent datasets. In each fold, seven datasets were used for training the classifier, while the remaining dataset was used for testing. This process was repeated for all eight folds, allowing all samples to be tested once. The performance of the classifiers was evaluated based on decoding accuracy, calculated as the ratio of correctly classified samples to the total number of samples.

By contrast, fine-grained decoding of categorical content was performed by utilizing only image trials (control trials were not used since no images were presented) with the aim to decode pattern difference between three image categories (chance level=1/3). Since different labels were of equal trial number, model training was performed in one run by utilizing the same 8-fold cross-validation approach. Finally, for each of four combined conditions of attention and visibility, decoding accuracy for the fine-grained representations was also calculated as the ratio of correctly classified samples to the total number of samples.

NCC algorithm

Consciousness-related activity (or termed neural correlates of consciousness, NCC), including both behavioral and neural measures, was a central focus of our study. Thus, our primary interest was the consciousness-related contrast between visible versus invisible activity. To this end, at first, activity associated with either invisible or visible processing of images were calculated by subtracting control trials (masks only) from corresponding image trials (masks and image), obtaining a mask-free measure linked to each visibility (visible or invisible). Following the logic of [3, 6], we got to quantify consciousness-related activity using the following formula.

$$NCC_{att} = (Vis_{image} - Vis_{control})_{att} - (Inv_{image} - Inv_{control})_{att}$$

Here, "Vis" and "Inv" denote the mean value of measures of interest in the visible and invisible conditions, respectively, while "att" denotes the attended or unattended condition for each. 'NCC_{att}' means the resulting activity tied to each attention condition. Subsequently, we calculated the difference between them (attended minus unattended) to evaluate how attention contributes to consciousness.

Taking decoding performance as an example. We first trained SVM models of different granularity for each combined condition of attention and visibility, namely attended visible, attended invisible, unattended visible and unattended invisible. Importing the decoding performance to the NCC algorithm, we then obtained neural representation of both coarse-grained and fine-grained representation of consciousness-related signal for each contact and each time bin. This allowed us to evaluate how NCC representations of different granularity evolved in the dimension of space and time. Replicating this procedure for attended and unattended condition, independently, and further making a contrast between them (attended versus unattended), we finally got to evaluate the impacts of attention on consciousness-related representations of different granularity.

Responsive contacts

In the current discourse, a hotly debated topic regarding the mechanisms of consciousness [18, 19] revolves around the functional differences between the anterior brain, particularly the frontal lobe, and the posterior brain, encompassing the occipital, temporal, and parietal lobes. Until this issue is resolved, elucidating the roles of more specific brain regions remains challenging. Hence, in this study, responsive contacts were specifically defined as contacts that are responsive to consciousness-related signal in either attended or unattended condition. More specifically, they were selected from the gray matter of the four neocortical lobes (Occipital, Temporal, Parietal and Frontal lobes), while excluding those from white matter or unknown regions. In terms of neural activity, responsive contacts were identified based on either event-related potential (ERP) responses contributed mainly by slow neural oscillations (here filtered between 0.5-30 Hz) or fast gamma oscillation (70-150 Hz) power responses whose activity correlates with local neuronal spiking [20, 21].

Taking the ERP activity as an example first. For each attention condition and timepoint, a true NCC measure were calculated using the NCC algorithm introduced above. Subsequently, the visibility labels of epochs were randomly shuffled while keeping block membership and attention label unchanged, and NCC measures were recomputed using the same formula. This process was repeated 10,000 times to create a null distribution for each timepoint. The true NCC measure was then ranked within the 10,000 surrogate distribution, with ranks of >=9995 or <=5 considered significant (corresponding to a two-tailed P <= 0.001). Timepoints showing significant activity were clustered across time, with clusters lasting >33ms considered responsive. Timepoints inside the identified clusters were coded as 1 while other timepoints outside as 0.

The same procedure was conducted for high gamma power in parallel with ERP analysis, resulting in another timeline of binary indices. For tach timepoint, above-chance contacts were defined as the union of contacts that turned out above-chance for either ERP or high gamma power analysis.

In the second level analysis, for each timepoint, we summed the number of above-chance contacts from the four neocortex lobes of interest, resulting in a new timeline of summed N. To control for false alarms, we calculated the mean value (*n*) and standard deviation (*std*) across timepoints during the 500ms pre-stimulus baseline period. We then calculated a Z value for each timepoint using the formula:

$$Z_t = \frac{(N_t - n)}{std}$$

where N_t denotes the number of above-chance contacts for each timepoint. Responsive window

included timepoints of a Z value >3 and falling within the range of [0 600]ms post-stimulus onset. Responsive window for the attended condition turned out much exceeding that for the unattended condition. For each attended and unattended condition, respectively, responsive contacts were identified as the union of above-chance contacts derived from all timepoints within the responsive window.

Responsive units reducing data homogeneity

For each lobe, patients may have multiple electrodes implanted, with each electrode potentially containing multiple responsive contacts. It is well known that data of adjacent contacts coming from the same electrode and same lobe are highly correlated. On the contrary, data of contacts from the same electrode but different lobes, as well as those from different electrodes but the same lobe, are much more heterogeneous. We thus conducted second-level statistical analyses by considering each responsive unit as an independent sample. Specifically, data from n responsive contacts coming from the same lobe and same electrode were averaged to represent the activity of a single unit. To further mitigate contamination from false alarm noise, unless specified otherwise, we only included units that contained n>=2 responsive contacts into the second-level analysis of regional power and decoding performance.

Similarly, imaginary coherence values were also computed at the unit level, with unit-level coherence calculated as the mean of all included contact-level pairs. To control for false alarm, unless specified otherwise, we only included unit pairs that contained n>=4 contact pairs into the second-level analysis of inter-unit coherence.

Statistics

Considering each responsive unit as an independent sample for second-level analyses, parametric tests used included one-way and two-way repeated measures analysis of variance (ANOVA), paired-sample t-test, one-sample t-test, as well as Pearson and Spearman correlation analysis. All statistical tests were conducted in a two-tailed manner with a threshold of $p = 0.05$. The FDR correction or cluster-based permutation test was used to correct for multiple comparisons with a threshold of corrected $p = 0.05$.

To replicate findings, we also conducted linear mixed effects (LME) models with conditions of interest (e.g., visibility and/or attention) as fixed effects, and with patient and electrode (not individual contacts) as random effects. Using the restricted maximum likelihood method to estimate LME parameters, and using Bonferroni correction to correct for multiple comparisons (corrected P=0.05), the LME-based approach, which requires significantly more calculations, resulted in similar findings (not shown) as those reported here.

Method references

1. Pelli, D.G., *The VideoToolbox software for visual psychophysics: Transforming numbers into movies.* Spatial Vision, 1997. **10**(4): p. 437-442.
2. Brainard, D.H., *The Psychophysics Toolbox.* Spat Vis, 1997. **10**(4): p. 433-6.
3. Cohen, M.A., et al., *Distinguishing the Neural Correlates of Perceptual Awareness and Postperceptual Processing.* J Neurosci, 2020. **40**(25): p. 4925-4935.
4. Tottenham, N., et al., *The NimStim set of Facial Expressions: Judgments from Untrained Research Participants.* Psychiatry research, 2009. **168**: p. 242-9.
5. Ekman, P. and W.V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. Unmasking the face: A guide to recognizing emotions from facial clues. 1975, Oxford, England: Prentice-Hall. xii, 212-xii, 212.
6. Dehaene, S., et al., *Cerebral mechanisms of word masking and unconscious repetition priming.* Nature Neuroscience, 2001. **4**(7): p. 752-758.
7. Engbert, R. and R. Kliegl, *Microsaccades uncover the orientation of covert attention.* Vision Research, 2003. **43**(9): p. 1035-1045.
8. Qin, C., et al., *Automatic and precise localization and cortical labeling of subdural and depth intracranial electrodes.* Frontiers in neuroinformatics, 2017. **11**: p. 10.
9. Fischl, B., et al., *Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain.* Neuron, 2002. **33**(3): p. 341-355.
10. Desikan, R.S., et al., *An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.* Neuroimage, 2006. **31**(3): p. 968-980.
11. Glasser, M.F., et al., *A multi-modal parcellation of human cerebral cortex.* Nature, 2016. **536**(7615): p. 171-178.
12. Huang, C.-C., et al., *An extended Human Connectome Project multimodal parcellation atlas of the human cortex and subcortical areas.* Brain Structure and Function, 2022. **227**(3): p. 763-778.
13. Yeo, B.T., et al., *The organization of the human cerebral cortex estimated by intrinsic functional connectivity.* Journal of neurophysiology, 2011.
14. Nolte, G., et al., *Identifying true brain interaction from EEG data using the imaginary part of coherency.* Clinical neurophysiology, 2004. **115**(10): p. 2292-2307.
15. Chen, S., et al., *Theta oscillations synchronize human medial prefrontal cortex and amygdala during fear learning.* Science advances, 2021. **7**(34): p. eabf4198.
16. Karalis, N., et al., *4-Hz oscillations synchronize prefrontal–amygdala circuits during fear behavior.* Nature neuroscience, 2016. **19**(4): p. 605-612.
17. Cohen, M.X., *Analyzing neural time series data: theory and practice*. 2014: MIT press.
18. Koch, C., et al., *Posterior and anterior cortex — where is the difference that makes the difference?* Nature Reviews Neuroscience, 2016. **17**(10): p. 666-666.
19. Consortium, C., et al., *An adversarial collaboration to critically evaluate theories of consciousness.* bioRxiv, 2023: p. 2023.06.23.546249.
20. Ray, S., et al., *Neural Correlates of High-Gamma Oscillations (60-200 Hz) in Macaque Local Field Potentials and Their Potential Implications in Electrocorticography.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2008. **28**: p. 11526-36.
21. Nir, Y., et al., *Coupling between Neuronal Firing Rate, Gamma LFP, and BOLD fMRI Is Related to Interneuronal Correlations.* Current Biology, 2007. **17**(15): p. 1275-1285.