Superior colliculus coordinates whole-brain dynamics for sudden unconscious visual insight

Tsutomu Murata[1], Komaki Kunishige[2], Shigeto Seno[3], Izumi Ohzawa[1,2], Kunihiko Kaneko[1,4], Toshio Yanagida[1,2], Masahiko Haruno[1,2], Kazufumi Hosoda[1,2,5*]

[1]Center for Information and Neural Networks, Advanced ICT Research Institute, NICT, Japan
[2]Department of Frontier Biosciences, Graduate School of Frontier Biosciences, Osaka University, Japan
[3]Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Japan
[4]Niels Bohr Institute, University of Copenhagen, Denmark
[5]Department of Rehabilitation Science, Kobe University Graduate School of Health Sciences, Japan
***Corresponding author**: Kazufumi Hosoda (hosodak@nict.go.jp)
Keywords: superior colliculus, whole-brain dynamics, unconscious, visual insight, Mooney image, Eureka, Aha, empirical dynamic modeling

Abstract

Recognition of Mooney images, highly degraded black-and-white stimuli, provides a classic paradigm for studying unconscious, sudden generative insight often accompanied by a "Eureka!" experience. This spontaneous recognition is believed to rely on top-down modulation by widespread cortical networks, a conclusion primarily drawn from contrasts between recognition and non-recognition. However, how the brain transitions from non-recognition to recognition remains unclear, particularly regarding the role of non-cortical regions such as the midbrain superior colliculus. Here we show that this transition can be captured by three data-driven large-scale functional clusters, reflecting stimulus-driven activation, suppression, and recognition-associated dynamics. Within these whole-brain dynamics, the superior colliculus displayed a distinct activation peak just before the behavioral response indicating recognition, and was the only region exhibiting mutual positive interactions with all three clusters. Although the superior colliculus has been recognized as a key node in binary perceptual decision-making in non-human primates, its contribution to human insight has remained largely unexplored. Our findings suggest that the superior colliculus coordinates large-scale neural activity for unconscious generative insight, revealing that evolutionarily conserved non-cortical structures may contribute directly to higher-level cognitive processes.

Introduction

The human brain can generate novel perspectives during complex problem solving, often marked by an “Aha!” or “Eureka!” moment [1, 2, 3, 4, 5]. This insight-related process has been studied using Mooney images, highly degraded black-and-white stimuli that initially appear meaningless but may suddenly yield a recognizable object after continued viewing [6, 7, 8, 9, 10, 11, 12]. Because the timing of recognition is unpredictable, most studies have explored the underlying neural processes by presenting participants with undegraded images beforehand or offering specific choices such as “face or not” [7, 13, 14, 15, 16, 17, 18, 19, 20]. However, recognition under such guided conditions is thought to differ qualitatively from spontaneous generative recognition without prior cues or response options [4, 12, 21, 22, 23, 24, 25]. The neural dynamics underlying such unconscious transitions transition from non-recognition to recognition remain poorly understood when no prior information is available.

Neuroimaging studies have shown that resolving the ambiguity of Mooney images involves widespread cortical networks. Top-down influences, particularly under conditions with prior information, have been linked to activity in the parietal and frontal cortices, the precuneus, and early and higher-order visual areas, and more broadly the frontoparietal and default mode networks [7, 13, 14, 15, 16, 17, 18, 19, 20, 26, 27, 28]. Even without prior information, recognition has been associated with differences in connectivity between posterior visual and frontal regions [29, 30]. Subcortical contributions have mainly been reported in the context of memory and reward processing, such as the involvement of the amygdala and nucleus accumbens [31, 32]. In other insight-related tasks, the midbrain, thalamus, and cerebellum have also been implicated [33, 34]. However, most studies have focused on cortical activity and on differences between recognition and non-recognition. Thus, the whole-brain dynamics underlying the transition from non-recognition to recognition, particularly those involving non-cortical regions such as the superior colliculus (SC) whose involvement has been suggested [17, 35], remain largely unknown.

The SC is of particular interest due to its well-established roles in saccades, unconscious visual perception (e.g., blindsight), and motor preparation [36, 37, 38, 39], as well as face detection [40, 41]. Beyond these sensorimotor functions, growing evidence suggests that SC also contributes to higher-order cognitive processes typically associated with the cortex, including selective attention, working memory, and decision-making [39, 42, 43, 44, 45]. Notably, SC has been implicated in determining the amount of evidence information required for decisions, a function once thought to be exclusive to the cortex [46, 47, 48, 49]. While this role has been described primarily in gradual evidence accumulation tasks, it may extend to detecting sudden increases in information generated through internal processes, as seen in Mooney image recognition [21]. Most of these findings are based on nonhuman primate or rodent studies, nevertheless, the multifunctional nature of SC has made it difficult to isolate its role even in those species. Investigating SC function in humans is even more challenging. However, given that whole-brain dynamics are accessible through noninvasive imaging, it remains feasible and

important to explore the timing and directed dynamical interactions of the SC within whole-brain activity.

Here, we used functional magnetic resonance imaging (fMRI) to perform a data-driven analysis of whole-brain neural time series during a task involving open-ended Mooney image recognition without prior information. Although the visual stimulus remained constant throughout the task, widespread activation was observed around the moment of recognition. Approximately half of the analyzed brain regions were significantly engaged, and their temporal profiles were effectively captured by three functional clusters, reflecting stimulus-driven activation, stimulus-driven suppression, and a pattern most strongly associated with the moment of recognition. The latter included SC, which showed a prominent peak prior to the behavioral response. Data-driven modeling further revealed directed positive interactions between the SC and the representatives of all three functional clusters. These findings offer a comprehensive view of the whole-brain dynamics and highlight a potential contribution of the SC underlying Mooney image recognition as one of the insight-related cognitive processes.

Results

Task Design and Extraction of Time-Series Data

We conducted an experiment in which participants were instructed to press a button with their right hand when they recognized the content of a black-and-white image (Fig. 1A; see Methods for details). Each trial began with a 15-second resting period, followed by the presentation of a visual stimulus. The stimulus remained on the screen until 10 seconds after the participant's button response. If the participant did not respond within 30 seconds, the stimulus was removed and the next trial began.

Each of the 14 participants completed 90 consecutive trials using 90 distinct images, with measurements acquired every second. Although data were collected from the whole brain, our analysis focused on 122 atlas-defined regions provided by SPM12, excluding areas that do not contain neuronal tissue [50]. While voxels within the same atlas-defined region tend to exhibit similar temporal dynamics, in some cases they display markedly different time courses. Therefore, we applied a pre-functional parcellation step to group voxels with similar time series within each atlas region. Specifically, we concatenated the time series data from all participants into a single sequence totaling 41,446 seconds. Hierarchical clustering was then performed using complete linkage based on pairwise distances between voxel time courses. A common threshold was applied across regions to produce a total of 300 functional subregions (see Methods).

Among the subregions, we retained only those in which the representative voxel, defined as the voxel whose

time series was closest to the mean within the subregion, was located within the set of voxels significantly associated with the stimulus. The identification of stimulus-related voxels was performed independently of the pre-functional parcellation, using standard first-level (task-based GLM for individual data) and second-level (group-level) analyses (see Methods). One subregion contained fewer than ten voxels and was excluded from further analysis, resulting in 148 subregions that collectively encompassed approximately 64% of the total voxel volume within the 122 atlas-defined regions (see Methods). For each remaining subregion, a representative BOLD time series was computed by averaging the time courses of the ten voxels that were most similar to the subregion's mean time course, to avoid losing features in noise. The resulting 148 representative time series, each spanning 41,446 seconds, were used in subsequent analyses. A portion of the time series from one such subregion is shown as an example in Figure 1B.

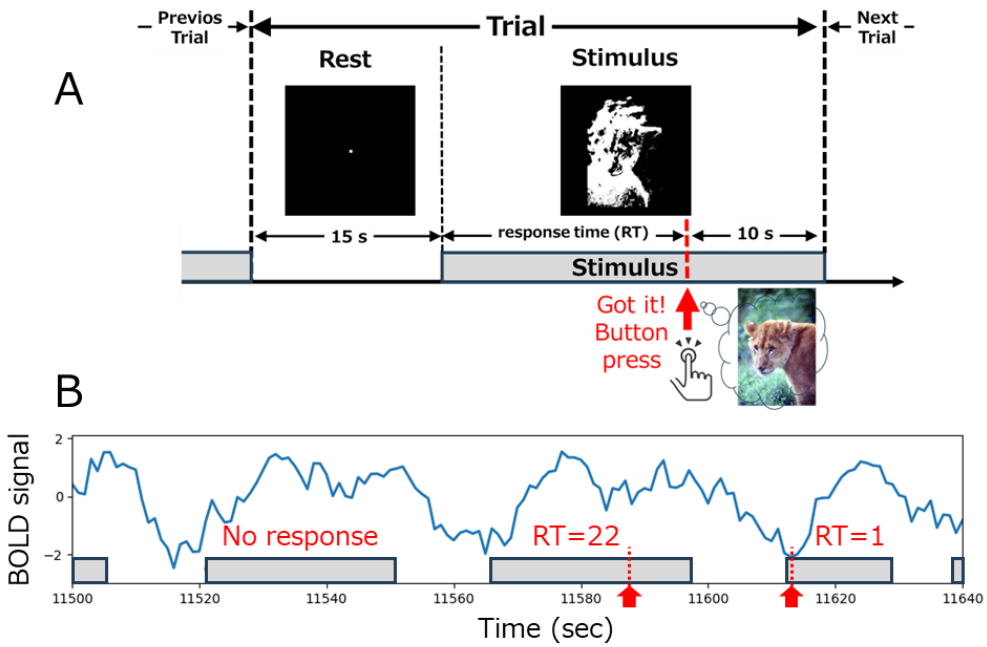

Figure 1. Experimental design and obtained time series data. **A**, Schematic of the experimental paradigm. Gray boxes indicate periods during which the stimuli were displayed, and the red arrow indicates the moment of the button press response. See main text for details. **B**, Example of the obtained time series data. The gray boxes and red arrow have the same meanings as in panel A. RT = 22 indicates that the Response Time, measured from stimulus onset, occurred between 22 and 23 seconds.

Overview of the temporal dynamics of brain activity during Mooney image recognition

To characterize whole-brain dynamics during the Mooney image recognition task, from non-recognition to recognition, we analyzed the overall profile of the time series data. For each of the 148 subregions, trial-averaged time series were computed to capture task-related activity patterns. Figure 2A (i) and (ii) show the averaged time series for trials in which the response occurred at 1 second (i.e., RT = 1, corresponds to Reaction Time between 1 and 2 seconds, N = 353) and at 8 seconds (RT = 8, N = 27), respectively. Each thin line in the plots represents the time course of a single subregion. For visualization, a moving average was applied with a 2-second window before and after each time point, and each time series was standardized individually.

Because the number of trials with RT = 8 were relatively few and thus noisier, we aggregated all trials in which the response occurred after 8 seconds, under the assumption that these trials share similar BOLD time series both before and after the response (designated as RT ≥ 8). The resulting averaged time series is shown in Figure 2A (iii) (N = 152). For example, in a trial of RT = 12, the measurement window from 8 to 11 seconds was trimmed and included in the average. Figure 2A (iv) shows the averaged time course for RT ≥ 18 trials (N = 43), and Figure 2A (v) shows trials in which no response occurred within the 30-second window (N = 202).

The time courses suggest that many subregions exhibit similar temporal profiles, implying that the activity can be effectively captured in a lower-dimensional representation. To evaluate the dimensional efficiency with which the temporal dynamics of the 148 subregions could be represented, we first performed principal component analysis (PCA) on the full time series data (148 subregions × 41,446 time points) to reduce the temporal dimensionality, which yielded 148 temporally compressed features (148 subregions × 148 temporal features). We then applied Gaussian mixture modeling using all 148 temporal features to perform clustering across a range of cluster numbers. Figure 2B shows the Bayesian Information Criterion (BIC) scores for each model, indicating that a three-cluster solution provides the most efficient representation. The log-likelihood of the three-cluster model reached 62% of the difference between the 1-cluster model and the full-dimensional model, indicating the efficiency of this solution. Although the clustering was performed in 148 dimensions, visualizing the data in the PCA space of the first two components suggests a reasonable separation into three groups (inset of Figure 2B).

The resulting three functional clusters are shown in Figure 2, with each color representing one cluster. Thick lines indicate the average time course within each functional cluster. These clusters reveal interpretable patterns in the temporal dynamics of brain activity. As an illustrative example, Figure 2A(iv) is used in the following explanation. The orange cluster shows sustained high activation throughout the stimulus presentation period, with additional transient activation surrounding the response. This cluster is referred to

as C_{SP} for its Stimulus-Positive pattern, indicating stimulus-driven activation. The blue cluster exhibits overall suppression during the stimulus period, yet demonstrates increases in activity at moments such as the response. This cluster is referred to as C_{SN} for its Stimulus-Negative pattern, indicating stimulus-driven suppression. The green cluster exhibits transient responses peaking approximately 5 seconds after key events, including stimulus onset, response, and stimulus offset, consistent with a canonical hemodynamic delay. Among the three, this cluster responds most specifically to the response, and is referred to as C_{RA}, denoting its Response-Associated profile that exhibits recognition-associated dynamics. Notably, all three clusters, C_{SP}, C_{SN}, and C_{RA}, share a common feature: a transient activity at the moment of the response, indicating widespread neural engagement at this cognitively critical event.

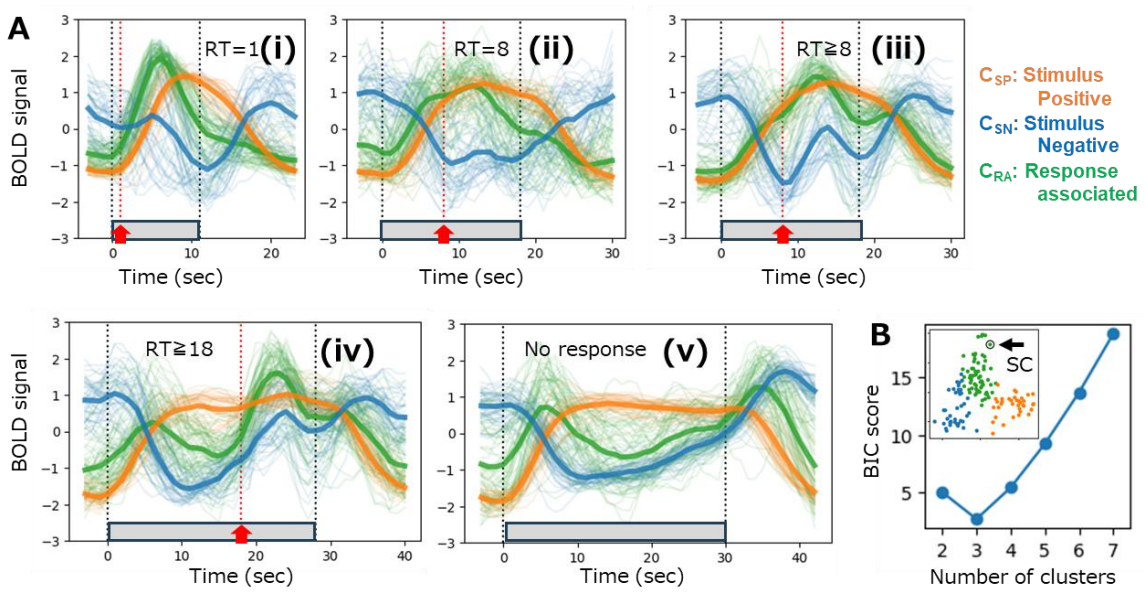

Figure 2. Overview of temporal dynamics. **A**, Time series of all 148 subregions across the brain. Thin lines represent individual subregions, color-coded by functional cluster identity (orange: C_{SP}, blue: C_{SN}, green: C_{RA}). Thick lines show the average time course within each functional cluster. The gray boxes and red arrow have the same meanings as in Figure 1. See main text for details on panels (i) through (v). **B**, Bayesian Information Criterion (BIC) scores for Gaussian mixture models with varying numbers of clusters. The inset shows the clustering result for three clusters, with each of the 148 subregions projected into the PC1–PC2 state space. The superior colliculus (SC) can be seen as an edge within the C_{RA} cluster.

Temporal Profile and Anatomical Distribution of the Three Functional Clusters

To assess how well the temporal profiles of the three functional clusters could be explained by discrete task events, we constructed a model comprising four event-related components: stimulus onset, the response (i.e., moment of recognition), stimulus offset, and sustained stimulus presentation. Each component was convolved with the canonical hemodynamic response function (HRF), and the model was fitted separately to the average time course of each functional cluster. Figure 3A(i) shows the result of this fitting, with 12 parameters in total (i.e., four components for each of the three clusters). Filled circles represent the empirical time courses (shown in Figure 2A(iv)), and the thick lines represent the model fits. The fitted parameters were then used to simulate time courses for the experimental conditions shown in Figure 2A(i) and (v), without any further fitting. These model-based reconstructions are shown in Figure 3A(ii) and (iii), respectively, and they qualitatively capture the key features of the observed dynamics. A conceptual summary of this modeling result is shown in Figure 3B.

Figure 3C shows the anatomical regions included in each of the three functional clusters, and Figure 3D presents the relationship between the three functional clusters and established large-scale brain networks, as defined by the CONN functional connectivity toolbox [51], as well as additional regions for the diencephalon and the brainstem.

Cluster C_{SP} is distributed across the occipital and parietal lobes, as well as the middle frontal gyrus. This cluster primarily includes the visual network (Vis) and the dorsal attention network (DAN), consistent with their sustained activation throughout the task. The division between cerebellar regions included in this cluster and those included in the following C_{SN} cluster also aligns well with previously reported findings [52].

Cluster C_{SN} is found mainly in the temporal and frontal lobes, the cingulate gyrus, and the precuneus. This cluster includes regions mainly associated with the default mode network (DMN) and the salience network (SN), both of which are typically suppressed during externally focused tasks. Interestingly, the frontoparietal network (FPN), typically associated with executive control and often considered functionally anti-correlated with the DMN, is also largely included in cluster C_{SN} alongside the DMN. Although the DMN is generally suppressed during task engagement, its activation increases at the response moment, suggesting a more integrated role in conjunction with the FPN. Given that recognition in this task likely requires creative processing [21], the co-involvement of DMN and FPN may reflect mechanisms previously associated with creative cognition [53].

Cluster C_{RA} included the diencephalon and the brainstem, with cortex such as the anterior insula, frontal operculum, opercular part of the inferior frontal gyrus, supplementary motor cortex, and the left primary motor

cortex, which was associated with the right-hand button press. Cluster C_{RA} encompasses nearly all core regions, including 44% of the brainstem, which corresponds to the entirety of the significantly prescreened brainstem mask. This cluster contains regions that show peak activation at the response moment, suggesting a potentially central role for these structures in the Mooney image recognition.

As described above, the whole-brain dynamics during the Mooney image recognition task, ranging from non-recognition to recognition, can be represented in a data-driven manner using characteristic three-dimensional functional clusters. Notably, when focusing on the moment of recognition, the cluster including the diencephalon and brainstem, and thus SC, became prominent.

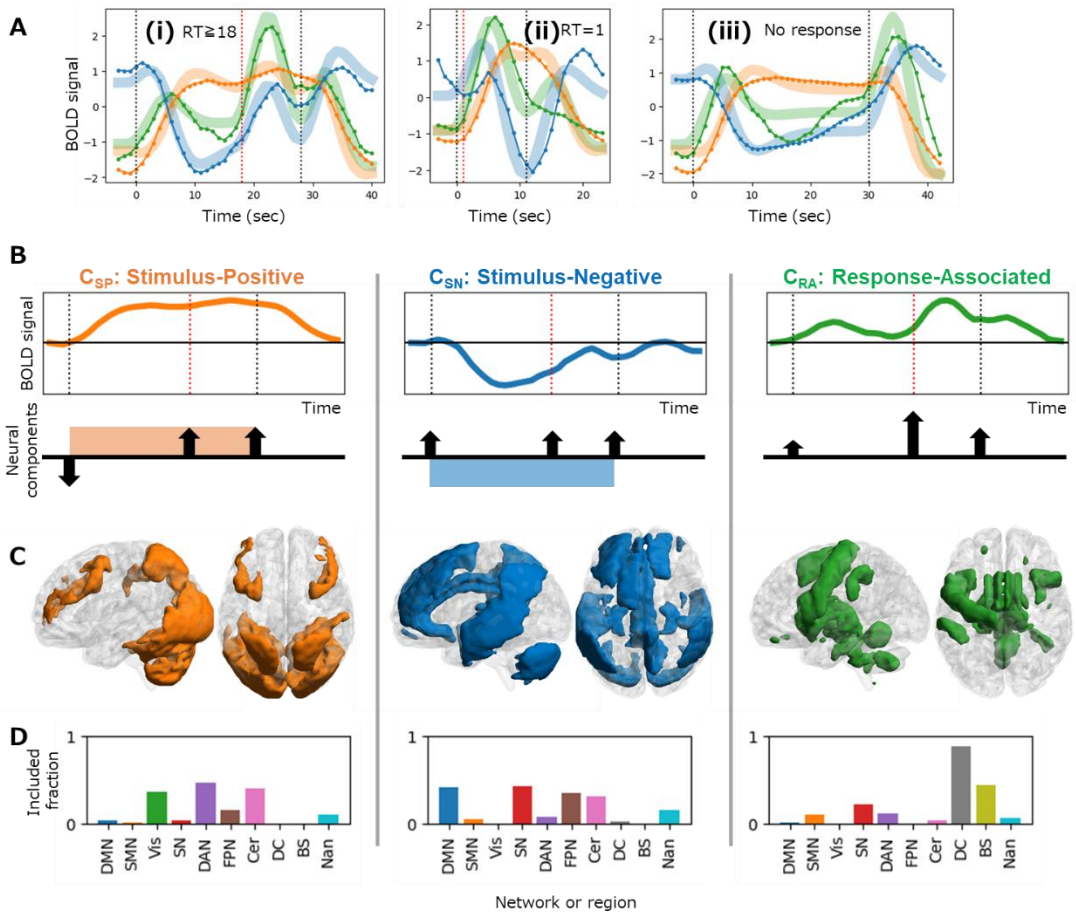

Figure 3. Temporal Profile and Anatomical Distribution of the Three Functional Clusters. **A**, Model-based explanation of time series dynamics. The model includes four event-related components, stimulus onset, response, stimulus offset, and sustained stimulus presentation, each convolved with the canonical hemodynamic response function (HRF). (i) Fit to the data from trials with RT ≥ 18 (corresponding to Figure 2A-iv). (ii, iii) Reconstructions for RT = 1

and no-response trials (corresponding to Figure 2A-i and 2A-v, respectively), using the same fitted parameters without refitting. B, Summarize of the Temporal Profile. The lines are the same as shonw in the panel A(i). Colored box show the stimulus presentation, and the arrows show the contribution of each clusters at the three transient events. **C**, Spatial distribution of the three functional clusters. **D**, Proportion of each known large-scale brain network, as well as the diencephalon and brainstem regions included in the three functional clusters. For example, the value of 44% for the brainstem in C_{RA} does not mean that 44% of C_{RA} consists of brainstem regions, but indicates that 44% of the brainstem voxels fall within cluster C_{RA}, which corresponds to the entirety of the significantly prescreened brainstem mask. DMN: Default Mode Network; SMN: Sensorimotor Network; Vis: Visual Network; SN: Salience Network; DAN: Dorsal Attention Network; FPN: Frontoparietal Network; Cer: Cerebellum; Thal: Thalamus; DC: Diencephalon; BS: Brainstem. The CONN functional connectivity toolbox includes the Language network in addition to the others, but since it is often considered part of the DMN, we treated it as part of the DMN in this study as well.

Subregions characterized by peak activation at the moment of response

As described above, the response was accompanied by transient activation across nearly all subregions. Therefore, we focused on identifying regions that exhibited especially early activation. Figure 4A shows subregions whose 95% confidence interval for the peak response occurred within 5 seconds after the response, using trials with RT ≥ 8. These subregions are sorted by the mean timing of their peaks. All of these subregions belonged to cluster C_{RA}, as expected.

This set included the left precentral gyrus (shown as a black open circle in Figure 4A), which corresponded to the left primary motor cortex (thus hereafter designated as M1), which governs right-hand responses. In contrast, the right precentral gyrus was properly excluded during prescreening (see Methods). Slightly earlier activation was observed in the bilateral supplementary motor area, consistent with its known role in the preparation of self-initiated movement [54].

All other regions that peaked earlier than the left M1 were located within the brainstem or diencephalon. Among these, we focused on two subregions that exhibited the earliest responses (highlighted in color in Figure 4A). The 95% confidence intervals for the peak responses of these two subregions did not overlap with that of the left M1. Anatomical localization identified the brainstem region as the SC (Figure 4B) by a detailed atlas of the brainstem [55]. The temporal dynamics of these two subregions are shown in Figure 4C. The left M1 peaked earlier than the average of its associated cluster, and indeed both of these subregions peaked even earlier

than the left M1.

Such early peaks were observed in trials with relatively long RT, as shown in Figure 4D, whereas the peak latency of the left M1 exhibited minimal dependence on RT. This difference between trials with RT = 1 or 2 and those with longer RTs was also apparent in the global dynamics, as revealed by the PCA-based analysis (Figure S3). Importantly, the lack of substantial differences at RT = 1 suggests that these early peaks are unlikely to be attributable solely to variability in the latency of the hemodynamic response.

To further determine whether the early activation observed in the two subregions was merely due to motor execution associated with the response, we conducted a simple reaction task. In this control experiment, because the signals in the two subregions were noisy and did not yield clear peaks, we calculated the mean BOLD signal within the 3 to 4 second window following the response for each subregion.

Figure 4E presents box plots of these average values for both the Mooney image recognition task and the simple reaction task, computed separately for each subregion and participant. In the Mooney image task, all subregions showed significantly greater-than-zero activation (one-sample Wilcoxon signed-rank test with Benjamini–Hochberg FDR correction [56], $q < 0.01$, black asterisks). In contrast, during the simple reaction task, only the left M1 showed a significant increase, while the other regions did not differ significantly from zero.

Moreover, when compared between the Mooney task and the simple reaction task, only the region corresponding to the SC exhibited significantly higher activation in the Mooney task (Mann-Whitney U test with Benjamini–Hochberg FDR correction [56], $q < 0.01$, orange asterisk), further supporting its involvement beyond simple motor output.

These results suggest that the early activation of the SC observed in the Mooney image recognition task cannot be explained solely by motor execution, and may reflect additional cognitive processes specific to the Mooney image recognition event. Indeed, the SC showed a distinct position even within the PCA space, among the subregions in the C_{RA} cluster (black arrow in Figure 2B inset).

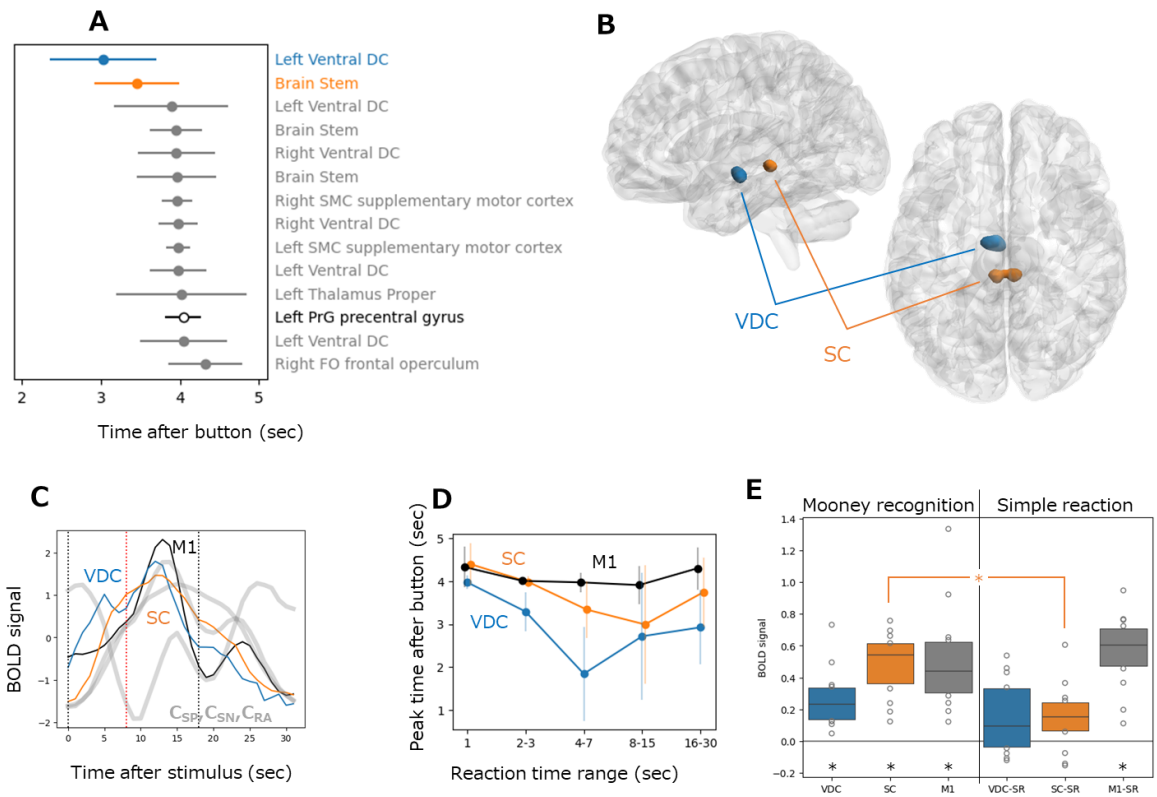

Figure 4. Subregions characterized by peak activation at the moment of response. **A**, Subregions showing early BOLD signal peaks following the response. The displayed names correspond to the labels defined in the Neuromorphometrics atlas provided in SPM12. Multiple subregions may originate from the same atlas region, resulting in repeated label names; however, the voxel territories of the individual subregions do not overlap. The color coding for the top two subregions and the left precentral gyrus (M1) is consistent across all panels below. **B**, Anatomical locations of the top two subregions identified in panel A. The brainstem region was identified as the SC. **C**, Time courses of the BOLD signal for the top two subregions and the left M1, averaged across trials with RT ≥ 8 seconds as shown in Figure 2A(iii). Gray bold lines indicate the three functional clusters for comparison. All time courses were standardized for display. **D**, Peak timing for the three subregions across different RT conditions. **E**, Comparison of BOLD signal amplitudes between the Mooney image recognition task and the simple reaction task for the three subregions. Data are shown as box plots across 14 participants. Black asterisks indicate values significantly different from zero, and orange asterisk indicate significant difference between the tasks. VDC: Ventral Diencephalon (top 1), SC: Superior Colliculus (top 2 brainstem), M1: Left Precentral Gyrus. Error bars in all panels indicate standard deviation.

Analyzing directed dynamical interactions using empirical dynamic modeling

The order of peak activations above does not necessarily reflect the directionality of interactions among regions. To capture such directional influences, it is necessary to adopt a dynamical systems approach. However, the complexity of brain activity makes it difficult to define explicit mechanistic models. Therefore, we employed empirical dynamic modeling (EDM), a data-driven method that reconstructs underlying dynamical relationships directly from observed time series [57, 58, 59]. Unlike approaches that rely on trial-averaged signals or simple correlations, EDM allows for the detection of directed interactions from whole time series directly. For instance, if $dx_1/dt = f(x_1, x_2)$ and $dx_2/dt = g(x_2)$, the dynamics of x_1 are influenced by x_2, but not vice versa. Such asymmetric relationships can be uncovered without the need for predefined model structures, making this approach particularly suitable for exploring dynamical causal interactions in complex neural systems.

We analyzed directed dynamical relationships among six representative time series: the three functional clusters (C_{SP}, C_{SN}, and C_{RA}), the SC subregion that peaked earlier than the response, the response-related subregion M1, and the stimulus on/off time series (Stim). It is important to note that although both the SC and M1 subregions belong to the C_{RA} cluster, the C_{RA} cluster analyzed here refers to the average time series representing the entire C_{RA} cluster, and is therefore not identical to either of the two subregions.

First, we assessed the presence of dynamical relationships using convergent cross mapping (CCM) [57]. To illustrate this method, for example, let us consider the relationship between cluster C_{SP} and the stimulus time series. It is reasonable to expect that activity in C_{SP}, which includes visual areas, would be influenced by the stimulus. To test this, we attempt to predict the stimulus time series using the C_{SP} time series. If the prediction accuracy using the full-length C_{SP} time series is significantly higher than that obtained using shorter segments, it can be interpreted as that the C_{SP} time series data contains embedded information about the stimulus. In dynamical systems terms, this corresponds to the presence of a stimulus-related term in the differential equation governing C_{SP} (i.e., $dC/dt = f(C, S, \ldots)$, where C and S are state of C_{SP} and stimulus, respectively), implying that the stimulus exerts a dynamical influence on C_{SP}. It is important to note that in this framework, the prediction target represents the dynamical cause, which is the inverse of the conventional terminology. Conversely, predicting C_{SP} from the stimulus may also yield above-chance accuracy, as C_{SP} is influenced by the stimulus; however, since C_{SP} does not influence the stimulus variation, the stimulus time series does not embed information about C_{SP}, and the prediction accuracy will not improve with increasing data length (library size). The pattern in which prediction accuracy increases and asymptotically converges as more data are used is referred to as convergence, and this is used as a criterion for detecting dynamical causality in CCM.

Specifically, we defined the presence of a dynamical relationship when the lower bound of the 99% confidence

interval (CI) for prediction accuracy using the maximum library size (i.e., the full time series of 41,446 data points) exceeded the upper bound of the 99% CI using the minimum library size (8 data points), each obtained from 1,000 predictions. Prediction accuracy was evaluated using Spearman's rank correlation coefficient between the predicted and observed signals [57, 58, 59].

Figure 5A shows the distribution of prediction accuracies for each combination of prediction library and prediction target. The blue histograms represent prediction accuracy distributions using the minimum library, and the orange histograms show the distributions using the maximum library. Subplots in which a significant dynamical relationship was detected are outlined with bold borders. The color of each border corresponds to the direction of influence, as described below. The matrix is not symmetric across the diagonal, because the relationships are directed, as illustrated by the gray arrows.

To assess the direction (positive or negative) of influence for each detected dynamical relationship, we applied the multivariate sequential locally weighted global linear maps (S-map) method [57, 59]. This approach estimates, e.g., the derivative $\partial C/\partial S$, representing the change in C_{SP} in response to a change in the stimulus, based on their respective time series. The sign of this derivative indicates whether the influence is excitatory (positive) or inhibitory (negative). Because the derivative is calculated at each time point, we summarized the sign of interaction (I_S) using a standardized index, defined as I_S = M/SD, where M and SD are the mean and standard deviation of the derivative values, respectively. For visualization in Figure 5A, the frame color of each subplot was assigned based on the I_S value. If I_S was greater than 2, the frame was rendered in red (RGB = 1, 0, 0); if less than –2, it was rendered in blue (RGB = 0, 0, 1). For intermediate values, the red or blue intensity was scaled proportionally to $I_S/2$, depending on whether I_S was positive or negative, respectively. Thus, frames are colored red when the influence is consistently positive, blue when consistently negative, and dark or near black when the influence fluctuates between positive and negative, centering around zero.

The results revealed several patterns that are both interpretable and consistent with known brain functions, and also highlight SC again. First, no neural signals were found to causally influence the stimulus time series, while causal influence from the stimulus was detected only in cluster C_{SP}, which includes visual regions. This pattern is reasonable given that sensory processing should not influence, but be driven by, external stimuli. A mutual negative causal relationship was detected between C_{SP} and C_{SN}, consistent with known antagonistic interactions between the dorsal attention network (DAN) and the default mode network (DMN). The cluster C_{RA} exhibited positive influences on all other components, consistent with its role in event-related activations around the response moment, which broadly impact large-scale brain systems. The observation that C_{SP} exerted a positive influence on C_{RA} and the SC also aligns with the notion that visual and attentional systems can drive those neural responses.

Notably, the SC was the only region to exhibit mutual, positive directional interactions with all three clusters. Despite the generally opposing dynamics of C_{SP} and C_{SN}, the SC showed positive causal influence on both, suggesting that within the C_{RA} cluster, the SC may play a critical role in the global activation occurring at the moment of response.

A schematic summary of these interaction directions is presented in Figure 5B. The stimulus exerted influence on C_{SP}, which in turn led to activation in C_{RA} and SC (which belongs to C_{RA}), as well as deactivation in C_{SN}. From there, the dynamics expanded into widespread bidirectional interactions across the entire brain. Among all regions, the SC exhibited bidirectional positive dynamical interactions with all three major clusters representing global brain activity, ultimately culminating in widespread activation associated with recognition. The M1 region, responsible for executing the button-press response, appeared to be influenced by both C_{RA} and C_{SN}. However, no influence from SC to M1 was detected, further suggesting that the SC is involved in functions beyond simple motor execution. These findings reinforce the notion that the SC holds a distinctive role in the Mooney image recognition.

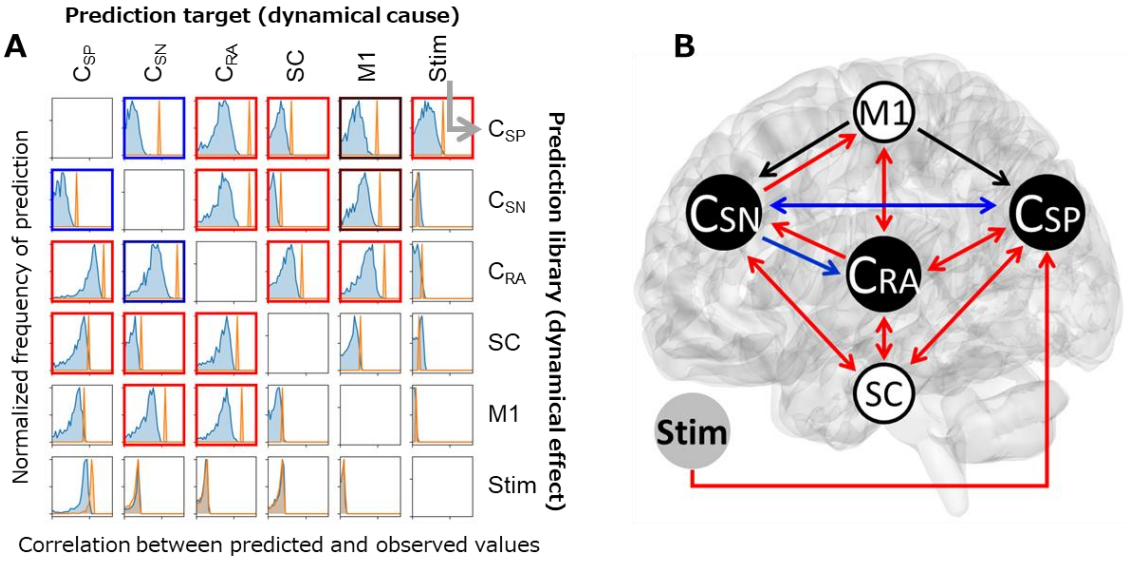

Figure 5. Directed dynamical interactions. **A**, Presence and sign of dynamical relationships as assessed using empirical dynamic modeling. Each subplot shows histograms of prediction accuracy for one prediction library to target pair. The orange histogram corresponds to predictions using the maximum library size (entire time series), while the blue histogram reflects predictions using the minimum library size (8 continuous time points). Subplots outlined with bold borders indicate statistically significant dynamical relationships. The color of the border represents the sign of the interaction: red indicates a consistently positive influence, blue indicates a consistently negative influence, and black indicates mixed polarity,

where the direction of influence varied over time. **B**, Schematic summary of the interactions shown in panel A. Arrows indicate the detected dynamical interactions. Blue, red, and black arrows represent positive, negative, and mixed influences, respectively, following the same color scheme as in panel A. SC corresponds to the superior colliculus as shown in Figure 4B, M1 refers to the left precentral gyrus, and Stim indicates stimulus onset and offset events. C_{SP}, C_{SN}, and C_{RA} refer to the three functional clusters shown in Figure 3.

Discussion

In this study, we investigated the whole-brain neural dynamics underlying the spontaneous recognition of Mooney images, a classic paradigm of unconscious and sudden visual insight. Using data-driven analysis of fMRI time series, we identified three large-scale functional clusters reflecting distinct stimulus-related and recognition-associated dynamics. Notably, we found that the SC, a midbrain structure traditionally associated with sensorimotor functions, exhibited a uniquely early activation peak and mutual positive interactions with all three clusters. These findings demonstrate, for the first time, that the SC may serve as a central coordinator of brain-wide activity during perceptual insight in humans. While previous studies have primarily focused on cortical mechanisms or post-recognition differences, our work reveals the moment-to-moment transition from non-recognition to recognition and uncovers a previously underappreciated non-cortical contribution. This suggests that even evolutionarily conserved midbrain structures may support high-level cognitive operations, including sudden generative insight. Our results bridge the gap between simple sensorimotor decision-making and complex problem-solving, expanding current models of cognitive processing beyond the cerebral cortex.

It is known that the process of recognizing Mooney images differs depending on whether predefined answer choices are available or not [21]. When choices are provided, recognition can be achieved by evaluating just a limited set of possibilities. In contrast, when no choices are given, the correct interpretation must be derived from an open-ended and potentially vast solution space. In the case of Mooney images, the identity of individual parts often cannot be determined until the entire image is recognized. This makes gradual information accumulation difficult, and recognition is thought to occur only when multiple missing elements are coincidentally generated [21, 22]. We here refer to this mechanism as the Coincidental Emergence Model (CEM). This view aligns with prior accounts of insight as an all-or-none generative transition from unawareness to solution, with no discernible intermediate stages [4, 23], and often yielding accurate outcomes [24, 25]. The generation itself is considered independent from the accompanying “Aha” experience and is thought to have a stronger impact on both long-term memory and emotional responses [12].

In contrast, when predefined choices are available, the recognition process is often explained by gradual evidence accumulation models such as the drift-diffusion model (DDM) [21, 50], whose neural substrates have been identified in both humans and non-human primates [49, 60, 61, 62]. The DDM posits that sensory evidence is gradually accumulated over time until it reaches a decision threshold, at which point a response is triggered. At the neural level, areas such as the lateral intraparietal area (LIP) in macaques [49], and the intraparietal sulcus (IPS) and dorsal attention network (DAN) regions in humans [63], have been implicated in this accumulation process. The SC, known to exhibit burst-type firing patterns, has been proposed to serve as the structure that evaluates whether the accumulated evidence has reached the threshold required for decision commitment [49].

A comparison between the two models, CEM and DDM, suggests that the exploratory phase of our task is better accounted for by the CEM, whereas the mechanism underlying the decision moment appears to be shared by both models. In our data, regions associated with the dorsal attention network (DAN), corresponding to the C_{SP} cluster, did not exhibit neural signatures of gradual evidence accumulation or threshold-dependent correlations with recognition. Instead, C_{SP} regions maintained high levels of activation throughout the task, prior to recognition (Figure 2A(iv) and (v)), suggesting continuous engagement in CEM rather than gradual accumulation in DDM.

Subsequently, the moment of recognition was accompanied by a peak activation of the SC (Figure 4C), consistent with its proposed role in threshold detection as described in DDM frameworks. These observations are not contradictory. Rather, the lack of an explicit decision termination mechanism in the CEM just reflects the absence of empirical data from human studies to date. Even within the CEM framework, the brain must evaluate whether sufficient information has been generated to allow recognition. Although the information level may remain low for prolonged periods, successful generation can abruptly yield a coherent percept. This rapid increase in information may be detected by the SC. Such functionality extends beyond the SC's classical sensorimotor role and reflects a more sophisticated form of information processing, as supported by findings in non-human primates. While such invasive verification is not possible in humans, the convergence of primate studies and our whole-brain, noninvasive observations strongly suggests that comparable high-level computations occur in the human SC.

The recognition of Mooney images can be discussed more broadly as a model for fundamental cognitive adaptability. From the perspective of theoretical biophysics, generative and sudden insight in response to difficult or novel problems reflects an ability to cope with unexpected situations, a function considered essential and deeply rooted in biological systems [64, 65]. Although such insight may appear to be a high-level cognitive capacity, it may in fact represent a more basic and evolutionarily conserved function, possibly implemented in phylogenetically older structures such as the SC. Such unexpected situations are analogous to extrapolation

in statistical or machine learning contexts [66], where the solution lies outside the range of training data. Similarly, in computational complexity theory, problems become dramatically harder when no predefined options are available, reflecting characteristics of decision problems in the class NP [67]. While extrapolation and NP-class problems remain difficult for artificial systems, it is suggested that evolutionarily conserved structures such as the SC might implement such computations through the dynamic integration of multiple neural signals, without requiring logically sophisticated architectures. Taken together, Mooney image recognition may serve not only as a neuroscience paradigm, but also as a window into adaptive intelligence, with implications that extend across theoretical biology, machine learning, and complexity science.

In this study, we found that trials with RTs longer than a few seconds exhibited distinct neural dynamics compared to faster trials. Specifically, when RT exceeded several seconds, the peak activation of the SC occurred earlier than that of the left M1, and this temporal difference remained relatively stable even as RT increased further (Figure 4D). A similar pattern was observed in the global state-space trajectories derived from principal component analysis (Figure S3). Previous studies have reported that recognition of typical visual stimuli can occur within approximately 500 milliseconds, whereas recognition of Mooney images typically takes longer than one second, even when the stimuli are preceded by exposure to undegraded versions [15]. This delay has been associated with the onset of top-down processing, which begins around 500 milliseconds after stimulus presentation [28]. The longer latencies observed in our study may thus reflect the time required for more extensive perceptual completion. Given that the interval between response and peak activation in left M1 remained constant, the earlier SC peak in long-RT trials likely indicates prolonged upstream processing. Interestingly, the distinction between responses occurring within a few seconds versus longer delays appears to be recognized in broader cognitive contexts [68]. Although the difference between analytical and intuitive decision-making has been well studied [69], the temporal dynamics underlying sudden insight remain poorly understood and warrant further investigation.

The characterization of large-scale neural dynamics, as demonstrated in this study, provides a predictability for a type of region-level statistical pattern. For example, when brain subregions were tested for BOLD signal changes that varied systematically with RT, those showing increased signal with longer RTs predominantly belonged to the C_{SN} and C_{RA} clusters, while those showing decreased signal were primarily classified within the C_{SP} cluster (Figure S4). This pattern is consistent with the time courses illustrated in Figure 2A(v). Both C_{SN} and C_{RA} exhibited signal trajectories that initially decreased and then gradually increased over time. Although the underlying reason for this gradual rise remains unclear, one possible explanation may involve mind wandering or related internal processes. In any case, a gradual increase in signal naturally results in higher values in trials with longer RTs. In contrast, regions within the C_{SP} cluster showed a slight but consistent decline in signal over time after the peak, leading to an inverse relationship between RT and BOLD amplitude. These observations suggest that capturing global dynamics provides a more predictive and integrated

understanding of neural activity than analyses based solely on statistical contrasts.

Future research will need to move beyond trial-averaged analyses and focus more deeply on the dynamics of individual trials. In our current study, the analysis of dynamical interactions using empirical dynamic modeling (EDM) was conducted without relying on trial averaging (Figure 5). Notably, the absence of averaging did not result in unstable outcomes; rather, the large volume of time-resolved data (41,446 seconds) available across trials enabled highly stable and interpretable results. These findings suggest that fMRI data, when analyzed with appropriate tools, can support the investigation of unaveraged neural dynamics. Although the brain is inherently complex, advances in measurement and analytical methods are making it increasingly feasible to examine trial-specific dynamics. This direction holds considerable promise for uncovering previously hidden mechanisms that cannot be captured through traditional averaging approaches.

Methods

Participants

Fourteen healthy native Japanese speakers (seven females; mean age: 23.2 ± 2.5 (SD) years) participated in an fMRI experiment investigating the Mooney image recognition. All participants had normal or corrected-to-normal vision and successfully passed the Visual Object and Space Perception Battery tests (Thames Valley Test Company, Bury St. Edmunds, Suffolk, England). All of these participants also took part in a separate fMRI experiment for Simple Reaction Time task, in which participants press a button as soon as they noticed a randomly displayed image of white fixation point at the center of a black background. This study was approved by the Ethics Committee for Human and Animal Research of the National Institute of Information and Communications Technology, Japan. All participants provided written informed consent in accordance with the committee's guidelines and were compensated for their participation.

Visual stimuli

Ninety digital color images were selected from clip-art collections (Sozaijiten, Image Navi Corp., Japan) and personal photographs. Each image featured a distinct object that could be easily recognized to describe verbally (e.g., "a chicken's head facing to the right"). To minimize categorical bias, the selected images represented diverse categories: human face (5 images), human with face and body (20), natural or artificial scenes (6), hand-held tools (3), animals such as dogs, fish, and frogs (43), and vehicles (13). These color images were degraded using monochromatic binarization. The binarization threshold for each image was manually adjusted using Adobe Photoshop (Adobe Inc., USA) to ensure that the subjective difficulty of recognizing the degraded images was evenly distributed within an identifiable range. The software

Presentation (ver. 23.1, Neurobehavioral Systems) was used for stimulus presentation and response collection. The LCD monitor was positioned 184 cm from the participant's eye level inside the MRI gantry. Stimuli were displayed at the center of the monitor, with each image measuring 640 pixels on its longer side, subtending a visual angle of 10° and featuring a white luminance of about 100 cd/㎡ against a black background.

Data acquisition

MRI data were acquired using a 3T MR scanner (Vida, Siemens) at the Center for Information and Neural Networks. For functional imaging, a simultaneous multi-slice (SMS) sequence with six bands and ascending T2*-weighted gradient-echo echo-planar imaging (EPI) was used. This procedure produced trans-axial slices covering the entire cerebrum and cerebellum, excluding the eyeballs, through oblique scanning. The imaging parameters were as follows: repetition time (TR) = 1000 ms, echo time (TE) = 30 ms, flip angle (FA) = 60°, matrix size = 100 ×100, number of slices = 72, and voxel size = 2.0 × 2.0 × 2.0 mm^3. To allow stabilization of functional imaging data, the scanner collected ten dummy volumes before EPI recording began. The acquired MRI data in raw DICOM format were converted to NIFTI format and preprocessed using Statistical Parametric Mapping 12 (SPM12 [70, 71]; University College London, United Kingdom; https://www.fil.ion.ucl.ac.uk/spm/), implemented in MATLAB 2020b (MathWorks). For each session, head motion-induced displacement in the scans was corrected through realignment relative to the mean image using affine transformation, with image headers adjusted to reflect the relative displacement. Slice timing correction was applied to account for differences in signal acquisition times across slices, adjusting each time series as if all slices were acquired simultaneously with the reference slice. For the SMS imaging with six bands, the reference slice was set at the middle of each scan band. Spatial normalization was performed using nonlinear three-dimensional transformations to align individual brain images with the ICBM template for East Asian brains, based on the MNI152 standard space. As a result of this normalization with resampling to 2.0 × 2.0 × 2.0 mm^3 isotropic resolution, the functional images were transformed into a common space with dimensions of 79 × 95 × 79 voxels (totaling 592,895 voxels). The normalized images were subsequently smoothed using a three-dimensional Gaussian kernel with a full width at half maximum (FWHM) of 6 mm. To eliminate low-frequency artifacts and task-unrelated trends, the data from each session were high-pass filtered with a cutoff frequency of 0.01 Hz. Subsequently, the data were standardized within each session to have zero mean and unit variance.

Experimental design

A challenge in fMRI experiments investigating the Mooney image recognition is determining whether participants arrive at the correct solution without requiring verbal responses during MRI scanning. Verbal responses can introduce signal changes unrelated to the Mooney image recognition itself in the MRI environment. To address this issue, we leveraged a key empirical property of the Mooney image recognition: once the solution is known, a previously ambiguous image becomes immediately recognizable. Using this

property, we developed a post-hoc verification method, in which each session consisted of the following two parts: The first part of a session, referred to as the scanning part, aimed to acquire brain activity data related to the Mooney image recognition and was conducted during MRI scanning. The second part, referred to as the verification part, was conducted immediately after the scanning part without MRI scanning, allowing participants to provide verbal reports to verify their answers post hoc. Details of the methods are described below.

Before data collection, participants were provided with instructions about the task, which required them to observe each Mooney image with the aim of recognizing meaningful objects hidden within it and to press a button as soon as they identified a stable object. The instructions emphasized that only the button press was required during the scanning part, while verbal reporting was additionally required in the verification part. To get familiar with the procedure, participants practiced with several example images beforehand.

The experiment consisted of six sessions, each featuring 15 different Mooney images, which were commonly used in both the scanning and verification parts of the session. A single trial of the scanning part of a session began with a 15-second rest period displaying a white fixation point at the center of a black background (Figure 1A). This was followed by a stimulus period, during which a Mooney image was presented. The order of the 15 images in each session was randomized for each participant. Participants were instructed to report their recognition moment by pressing an assigned button during the presentation of the Mooney image. The instruction allowed participants to press the button again if they identified another valid solution, in order to prevent them from hesitating to respond. After the final response, the Mooney image remained on display for an additional 10 seconds before proceeding to the next trial. If a participant did not press the button at all during the 30-second presentation, the trial was classified as a "no response," indicating the participant was unable to identify any solution. In such cases, the trial immediately transitioned to the next one.

In the verification part of each session, conducted immediately after the scanning part without MRI scanning, the same 15 Mooney images were presented in a different order. Participants were instructed to press a button as soon as they identified the object in the image. Upon pressing the button, the Mooney image was turned off, and participants were asked to verbally report their recognition to the experimenter. If a participant's answer was unclear, the experimenter asked follow-up questions to verify whether the object was correctly identified. A response in the scanning part was judged correct if the participant's response in the verification part was both rapid (response time within 3.5 seconds) and validly matched the correct solution. This verification method was empirically justified for its effectiveness, as detailed in the Results section. For each participant, the 90 trials (90 different images) from the scanning parts across the six sessions were categorized into correct responses (i.e., matches with the corresponding original color image; 930 trials), incorrect responses (128 trials), or no responses (202 trials). The overall accuracy was high, consistent with prior findings that, in insight-

based tasks, participants tend to either provide correct answers or omit responses, with incorrect responses being relatively rare [24]. However, because many Mooney images are substantially degraded and often lack sufficient information to define a single objectively correct answer, our analysis focused on participants' perceptual interpretations rather than strict correspondence with the original color images. Accordingly, all trials were included in subsequent analyses, regardless of response accuracy. By carefully setting the difficulty of the images, we were able to obtain 152 trials in which participants took more than 8 seconds to recognize the Mooney images. These trials included data from all 14 participants. The 8-second threshold corresponds to the time point at which the HRF returns to baseline, allowing us to dissociate recognition-related BOLD activity from stimulus-onset-related responses.

Pre-screening of task-associated regions

As a prescreening step, we excluded brain regions that did not show significant changes in BOLD activity during the recognition task, in order to focus subsequent analyses on regions potentially associated with Mooney image recognition. A general linear model (GLM) was constructed to identify brain regions exhibiting task-related activation. The model included three regressors corresponding to distinct temporal phases: pre-response (stimulus onset to response), response (a transient response event), and post-response (response to stimulus offset). Each regressor was designed to capture both activations and deactivations relative to baseline. This model was convolved with the canonical hemodynamic response function (HRF [72]; the parameters were the same as those in the previous study [50]) to generate the regressors. Although the canonical HRF was originally derived from activation-related BOLD responses, we applied the same HRF to model both activations and deactivations, as the GLM framework captures signal increases and decreases through the sign of the estimated beta weights. In the individual (first-level) analysis, preprocessed fMRI data for each participant, including realigned, slice-timing corrected, spatially normalized, smoothed, high-pass filtered, and standardized as described in the Data Acquisition section, were analyzed using the general linear model described above. Serial autocorrelation was modeled using a first-order autoregressive (AR(1)) approach. For the group-level (second-level) analysis, contrast estimates (beta values) obtained from each participant's first-level GLM were entered into voxel-wise one-sample t-tests across 14 participants [73, 74]. The resulting statistical map was thresholded using false discovery rate (FDR) correction for multiple comparisons (Benjamini–Hochberg procedure [56], $q < 0.01$), without cluster-level correction methods based on random field theory, as the analysis included small brainstem structures such as the superior colliculus (31 voxels in volume).

To restrict our analysis to anatomically meaningful regions, we excluded voxels that did not fall within any labeled area of the Neuromorphometrics atlas provided in SPM12. While the Neuromorphometrics atlas provides MNI-space volume data for 136 tissue-defined brain regions, we included 122 areas, excluding 14 structures lacking neuronal cell bodies (e.g., white matter) [50]. Among the 130,707 voxels corresponding to

these 122 atlas-defined regions, statistically significant effects were observed in 68,263 voxels, approximately half. Each of the three temporal regressors (pre-response, response, and post-response) was associated with significant activation in roughly 30,000 voxels, with substantial overlap observed among them. Furthermore, approximately half of the voxels that showed a positive beta value for one regressor overlapped with those showing a negative beta value for another regressor. These results and an overview of the anatomical distribution of these significantly positive and negative beta values is shown in Figure S1. The spatial distribution of these voxels was widespread and dispersed, suggesting that a broad network of brain regions was involved in this task. These activations tended to form contiguous clusters, rather than appearing as isolated points. Additionally, even when applying a more stringent multiple-comparisons correction (e.g., Benjamini–Yekutieli method), the overall conclusions remained unchanged, although the number of voxels included was reduced.

Pre-functional parcellation and extraction of representative subregion BOLD time courses

Given that functionally heterogeneous activity patterns can arise within a single atlas region, we applied a pre-functional parcellation procedure to further subdivide regions exhibiting internally divergent signal profiles. To maintain analytical flexibility, this parcellation step was performed independently of the GLM-based prescreening described above.

Specifically, BOLD time series from all participants were concatenated (totaling 41,446 seconds), and pairwise distances between voxel time courses were computed within each atlas region. Hierarchical clustering using complete linkage was then applied to define functional subregions. Given that voxel time series were standardized, the mean squared Euclidean distance between voxel pairs is mathematically proportional to $1-R$, where R is the Pearson correlation coefficient and is widely used in functional connectivity analyses. A uniform clustering threshold was applied across all atlas regions, yielding a total of 300 functional subregions. We confirmed that increasing the resolution of this parcellation (e.g., to approximately 1,000 subregions) did not qualitatively affect the conclusions of this study.

Following pre-functional parcellation, we defined the representative voxel for each subregion as the one whose time series exhibited the smallest Euclidean distance to the mean time series of all voxels within that subregion. This approach was adopted to reduce the influence of noisy or outlier voxels while preserving the central temporal characteristics of each subregion. We then retained only those subregions whose representative voxel fell within the GLM-based prescreened areas described above. Excluding a single subregion containing fewer than 10 voxels, this resulted in 148 subregions selected for further analysis. These 148 subregions collectively encompassed approximately 64% of the total voxel volume within the 122 atlas-defined regions (Figure S2). This spatial coverage was slightly larger than that of the original GLM-based prescreened regions because each subregion was selected based on the location of its representative voxel, which was within the

prescreened areas, even though the subregion itself could extend into regions outside the GLM-based mask. In these 148 subregions, the left precentral gyrus, which is expected to be involved in right-hand button presses, was preserved, whereas the corresponding region in the right hemisphere, which was not task-related, was excluded as expected.

For each of the 148 subregions, we computed a representative BOLD time series using a robust averaging approach. Rather than using a simple mean, which could be unduly influenced by noisy voxels distant from the central tendency, we selected the 10 voxels whose time series were closest (in Euclidean distance) to the subregion's mean time series and averaged them. The resulting 148 representative time series, each spanning 41,446 seconds, were used in subsequent analyses.

Large-scale functional clustering and model fitting

To identify the three large-scale functional clusters, we applied principal component analysis followed by Gaussian mixture modeling (Figure 2), implemented using scikit-learn [75]. We then fitted the time series of the three functional clusters using a model composed of four event-related components: stimulus onset, the response, stimulus offset, and sustained stimulus presentation (Figure 3A). This fitting was performed using Scipy [76]. Additional information is provided in the Results section.

Detection of peak time just after the response

To estimate the timing of peak activation following the response, we analyzed trials with RT of 8 seconds or longer (N = 152). For each trial, we extracted BOLD signal data from a 12-second window beginning 2 seconds before the response. Because peak detection at the single-trial level is highly susceptible to noise, moving average was applied with 2-second window before and after each time point, and we employed a non-parametric bootstrap approach to improve robustness. Specifically, 152 trials were resampled with replacement 1000 times to generate averaged time series, from which the peak time was computed for each iteration. This procedure yielded a bootstrap distribution of peak timings, from which we derived the mean, standard deviation, and 95% confidence interval. This resampling-based estimation is advantageous for capturing central tendencies and variability without relying on parametric assumptions.

To ensure that only subregions with clear and temporally localized peaks were included, we excluded those in which any of the 1000 bootstrap samples showed a peak at either edge of the 12-second window. We also excluded subregions in which the maximum signal within the window did not exceed both endpoints by at least 0.1. Among the remaining 30 subregions, those whose 95% confidence intervals were entirely within at most 5 seconds after the response are shown in Figure 4A. Figure 4D shows the same analysis performed across a range of RT bins. See the Results section for further information.

Empirical dynamic modeling

We performed empirical dynamic modeling using the pyEDM package [77], developed and maintained by the original creators of the method [57]. Analyses were conducted in accordance with the standard documentation [77], with minor adjustments to the embedding dimension parameter. While this parameter is typically optimized for each time series, its estimation is often unstable depending on the library (time window) used. We observed similar variability in our data; however, optimal values consistently fell within the range of 5 to 10, corresponding to 5 to 10 seconds given our 1 Hz sampling rate. Among the three functional clusters, which showed particularly stable dynamics, an embedding dimension of 6 was most commonly observed. Therefore, we used a fixed embedding dimension of 6 for all time series. Note that this choice was further justified by the fact that all time series were derived from BOLD signals and thus expected to share similar nonlinear temporal characteristics within the 5 to 10 second range. Additionally, the stimulus time series lacks a meaningful embedding structure, and a fixed embedding dimension ensured analytical consistency with the other BOLD-based signals. The Results section provides additional context and information.

References

1. Kaplan CA, Simon HA. In search of insight. *Cognitive psychology* **22**, 374-419 (1990).
2. Sternberg RJ, Davidson JE. *The nature of insight*. The MIT Press (1995).
3. Ahissar M, Hochstein S. Task difficulty and the specificity of perceptual learning. *Nature* **387**, 401-406 (1997).
4. Kounios J, Beeman M. The cognitive neuroscience of insight. *Annual review of psychology* **65**, 71-93 (2014).
5. Sprugnoli G*, et al.* Neural correlates of Eureka moment. *Intelligence* **62**, 99-118 (2017).
6. Mooney CM. Age in the development of closure ability in children. *Can J Psychol* **11**, 219-226 (1957).
7. Dolan RJ*, et al.* How the brain learns to see objects and faces in an impoverished context. *Nature* **389**, 596-599 (1997).
8. Tallon-Baudry C, Bertrand O. Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences* **3**, 151-162 (1999).
9. McKeeff TJ, Tong F. The timing of perceptual decisions for ambiguous face stimuli in the human ventral visual cortex. *Cerebral Cortex* **17**, 669-678 (2007).
10. Giovannelli F*, et al.* Involvement of the parietal cortex in perceptual learning (Eureka effect): An interference approach using rTMS. *Neuropsychologia* **48**, 1807-1812 (2010).
11. Castelhano J, Rebola J, Leitao B, Rodriguez E, Castelo-Branco M. To perceive or not perceive: the role of gamma-band activity in signaling object percepts. *PloS one* **8**, e66363 (2013).
12. Kizilirmak JM, Galvao Gomes da Silva J, Imamoglu F, Richardson-Klavehn A. Generation and the subjective feeling of “aha!” are independently related to learning from insight. *Psychological Research* **80**, 1059-1074

(2016).

13. Hsieh P-J, Vul E, Kanwisher N. Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *Journal of neurophysiology* **103**, 1501-1507 (2010).
14. Gorlin S, Meng M, Sharma J, Sugihara H, Sur M, Sinha P. Imaging prior information in the brain. *Proceedings of the National Academy of Sciences* **109**, 7935-7940 (2012).
15. Hegdé J, Kersten D. A link between visual disambiguation and visual memory. *Journal of Neuroscience* **30**, 15124-15133 (2010).
16. Chang R, Baria AT, Flounders MW, He BJ. Unconsciously elicited perceptual prior. *Neuroscience of Consciousness* **2016**, (2016).
17. Martinsen MM*, et al.* Facial ambiguity and perception: How face-likeness affects breaking time in continuous flash suppression. *J Vis* **24**, 18 (2024).
18. Jiang Y, He S. Cortical Responses to Invisible Faces: Dissociating Subsystems for Facial-Information Processing. *Current Biology* **16**, 2023-2029 (2006).
19. Andrews TJ, Schluppeck D. Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *NeuroImage* **21**, 91-98 (2004).
20. Lu Y, Singer W. Dynamic signatures of the Eureka effect: an EEG study. *Cerebral Cortex* **33**, 8679-8692 (2023).
21. Murata T, Hamada T, Shimokawa T, Tanifuji M, Yanagida T. Stochastic Process Underlying Emergent Recognition of Visual Objects Hidden in Degraded Images. *PLOS ONE* **9**, e115658 (2014).
22. Hosoda K, Seno S, Murata T. Simulating Reaction Time for Eureka Effect in Visual Object Recognition Using Artificial Neural Network. *IIAI Letters on Informatics and Interdisciplinary Research* **3**, (2023).
23. Smith RW, Kounios J. Sudden insight: All-or-none processing revealed by speed–accuracy decomposition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**, 1443 (1996).
24. Kounios J*, et al.* The origins of insight in resting-state brain activity. *Neuropsychologia* **46**, 281-291 (2008).
25. Salvi C, Bricolo E, Kounios J, Bowden E, Beeman M. Insight solutions are correct more often than analytic solutions. *Think Reason* **22**, 443-460 (2016).
26. Teufel C*, et al.* Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc Natl Acad Sci U S A* **112**, 13401-13406 (2015).
27. González-García C, Flounders MW, Chang R, Baria AT, He BJ. Content-specific activity in frontoparietal and default-mode networks during prior-guided visual perception. *eLife* **7**, e36068 (2018).
28. Flounders MW, González-García C, Hardstone R, He BJ. Neural dynamics of visual ambiguity resolution by perceptual prior. *Elife* **8**, (2019).
29. Imamoglu F, Kahnt T, Koch C, Haynes JD. Changes in functional connectivity support conscious object recognition. *Neuroimage* **63**, 1909-1917 (2012).
30. Imamoglu F, Koch C, Haynes J-D. MoonBase: Generating a database of two-tone Mooney images. *Journal of Vision* **13**, 50-50 (2013).

31. Ludmer R, Dudai Y, Rubin N. Uncovering camouflage: amygdala activation predicts long-term memory of induced perceptual insight. *Neuron* **69**, 1002-1014 (2011).
32. Becker M, Yu Y, Cabeza R. The influence of insight on risky decision making and nucleus accumbens activation. *Sci Rep* **13**, 17159 (2023).
33. Tik M*, et al.* Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment. *Hum Brain Mapp* **39**, 3241-3252 (2018).
34. Ogawa T, Aihara T, Shimokawa T, Yamashita O. Large-scale brain network associated with creative insight: combined voxel-based morphometry and resting-state functional connectivity analyses. *Scientific Reports* **8**, 6477 (2018).
35. Goold JE, Meng M. Visual Search of Mooney Faces. *Front Psychol* **7**, 155 (2016).
36. Horwitz GD, Newsome WT. Separate Signals for Target Selection and Movement Specification in the Superior Colliculus. *Science* **284**, 1158-1161 (1999).
37. Tamietto M, de Gelder B. Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience* **11**, 697-709 (2010).
38. Ajina S, Pollard M, Bridge H. The Superior Colliculus and Amygdala Support Evaluation of Face Trait in Blindsight. *Front Neurol* **11**, 769 (2020).
39. Benarroch E. What Are the Functions of the Superior Colliculus and Its Involvement in Neurologic Disorders? *Neurology* **100**, 784-790 (2023).
40. Le QV*, et al.* A Prototypical Template for Rapid Face Detection Is Embedded in the Monkey Superior Colliculus. *Front Syst Neurosci* **14**, 5 (2020).
41. Lim C, Inagaki M, Shinozaki T, Fujita I. Analysis of convolutional neural networks reveals the computational properties essential for subcortical processing of facial expression. *Scientific Reports* **13**, 10908 (2023).
42. Krauzlis RJ, Lovejoy LP, Zénon A. Superior colliculus and visual spatial attention. *Annu Rev Neurosci* **36**, 165-182 (2013).
43. Rahmati M, DeSimone K, Curtis CE, Sreenivasan KK. Spatially Specific Working Memory Activity in the Human Superior Colliculus. *J Neurosci* **40**, 9487-9495 (2020).
44. Basso MA, Bickford ME, Cang J. Unraveling circuits of visual perception and cognition through the superior colliculus. *Neuron* **109**, 918-937 (2021).
45. Thomas A*, et al.* Superior colliculus bidirectionally modulates choice activity in frontal cortex. *Nat Commun* **14**, 7358 (2023).
46. Lo C-C, Wang X-J. Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience* **9**, 956-963 (2006).
47. Crapse TB, Lau H, Basso MA. A Role for the Superior Colliculus in Decision Criteria. *Neuron* **97**, 181-194.e186 (2018).
48. Jun EJ*, et al.* Causal role for the primate superior colliculus in the computation of evidence for perceptual decisions. *Nat Neurosci* **24**, 1121-1131 (2021).

49. Stine GM, Trautmann EM, Jeurissen D, Shadlen MN. A neural mechanism for terminating decisions. *Neuron* **111**, 2601-2613.e2605 (2023).
50. Morito Y, Murata T. Accumulation System: Distributed Neural Substrates of Perceptual Decision Making Revealed by fMRI Deconvolution. *The Journal of Neuroscience* **42**, 4891-4912 (2022).
51. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity* **2**, 125-141 (2012).
52. Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BT. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 2322-2345 (2011).
53. Beaty RE, Benedek M, Barry Kaufman S, Silvia PJ. Default and Executive Network Coupling Supports Creative Idea Production. *Scientific Reports* **5**, 10964 (2015).
54. Cunnington R, Windischberger C, Deecke L, Moser E. The preparation and readiness for voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response. *NeuroImage* **20**, 404-412 (2003).
55. Bianciardi M. *Brainstem Navigator [Tool/Resource.]*, 1.0 edn. The Neuroimaging Tools and Resources Collaboratory (NITRC) (2021).
56. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
57. Sugihara G*, et al.* Detecting Causality in Complex Ecosystems. *Science* **338**, 496-500 (2012).
58. Wismüller A*, et al.* Nonlinear Functional Connectivity Network Recovery in the Human Brain with Mutual Connectivity Analysis (MCA): Convergent Cross-Mapping and Non-Metric Clustering. *Proc SPIE Int Soc Opt Eng* **9417**, (2015).
59. Ushio M*, et al.* Fluctuating interaction network and time-varying stability of a natural fish community. *Nature* **554**, 360-363 (2018).
60. Cook EP, Maunsell JHR. Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nature Neuroscience* **5**, 985-994 (2002).
61. Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci* **30**, 535-574 (2007).
62. Pisauro MA, Fouragnan E, Retzler C, Philiastides MG. Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications* **8**, 15808 (2017).
63. Wongtrakun J, Zhou SH, O'Connell RG, Chong TT, Bellgrove MA, Coxon JP. The role of human intraparietal sulcus in evidence accumulation revealed by EEG and model-informed fMRI: IPS accumulates evidence during decision-making. *bioRxiv*, (2025).
64. Kaneko K. *Life: an introduction to complex systems biology*. Springer (2006).
65. Conrad M. *Adaptability: The significance of variability from molecule to ecosystem*. Springer Science & Business Media (2012).
66. Brezinski C, Zaglia MR. *Extrapolation methods: theory and practice*. Elsevier (2013).
67. Arora S, Barak B. *Computational complexity: a modern approach*. Cambridge University Press (2009).

68. Gladwell M. *Blink: The power of thinking without thinking*. Back bay books (2007).
69. Dijksterhuis A, Bos MW, Nordgren LF, van Baaren RB. On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science* **311**, 1005-1007 (2006).
70. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier (2011).
71. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping* **2**, 189-210 (1994).
72. Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. Event-Related fMRI: Characterizing Differential Responses. *NeuroImage* **7**, 30-40 (1998).
73. Holmes AP, Friston KJ. Generalisability, Random Effects & Population Inference. *NeuroImage* **7**, (1998).
74. Mumford JA, Nichols T. Simple group fMRI modeling and inference. *NeuroImage* **47**, 1469-1475 (2009).
75. Pedregosa F*, et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825-2830 (2011).
76. Virtanen P*, et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272 (2020).
77. Park J. pyEDM [Computer software]: Python/Pandas toolset for Empirical Dynamic Modeling.). Version 2.2.1 edn (2025).

Acknowledgements

We are grateful to Dr. Tetsuya Shimokawa for his considerable encouragement. This study was supported in part by MIC under a grant entitled “R&D of ICT Priority Technology (JPMI00316)” and JSPS KAKENHI Grant Number JP25H01365.

Author contributions

TM, TY, and KH designed the study and developed the concept. IO, KK (Kunihiko), and TY provided supervision for the study. TM and KK (Komaki) collected the data. KH and SS analyzed the data. IO, KK (Kunihiko), and MH provided insights regarding data interpretation. TM, MH, and KH drafted the manuscript. All authors contributed critical revisions and approved the final manuscript for publication.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary information

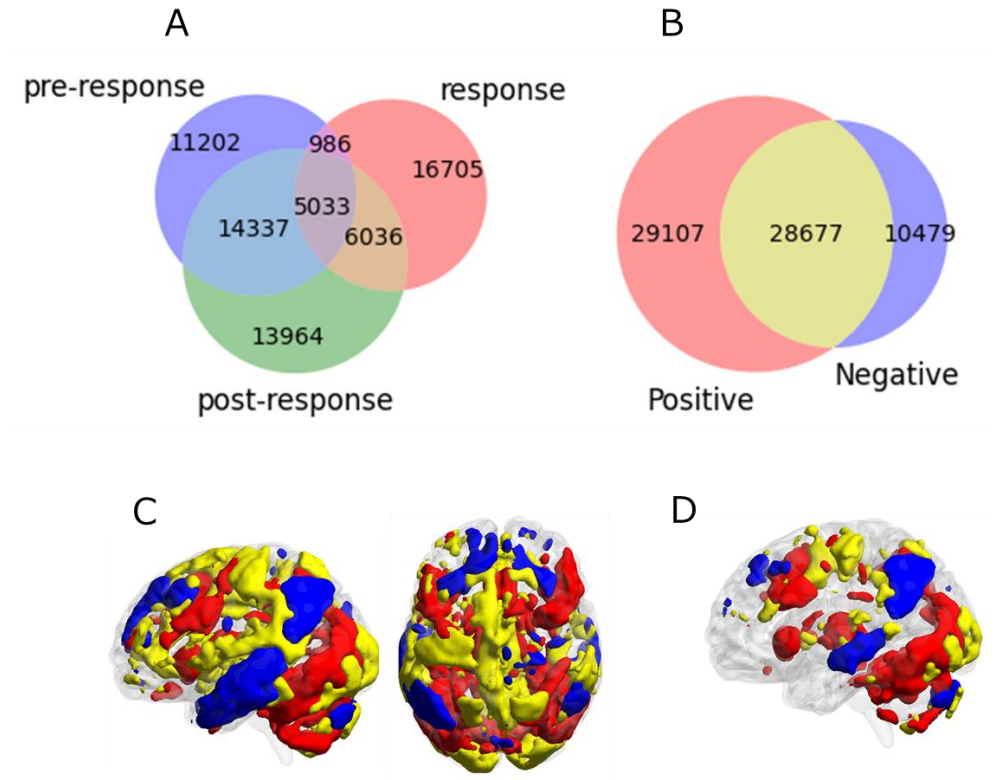

Figure S1. Pre-screening of task-associated regions. **A**, Venn diagram showing the number of voxels significantly associated with each of the three distinct temporal phases: pre-response (stimulus onset to response), response (a transient event at response), and post-response (response to stimulus offset). Each of the three temporal regressors was associated with significant activation in roughly 30,000 voxels, with substantial overlap observed among them. **B**, Venn diagram illustrating the overlap of voxels that showed significant effects for at least one of the three regressors, classified by whether they exhibited a positive or negative beta value. **C**, Anatomical distribution of the voxel clusters depicted in panel B, color-coded according to beta polarity. **D**, Spatial distribution of voxels that survived a more stringent multiple-comparisons correction using the Benjamini–Yekutieli method (shown as an illustrative example). Even when applying such a more stringent multiple-comparisons correction, the overall conclusions remained unchanged.

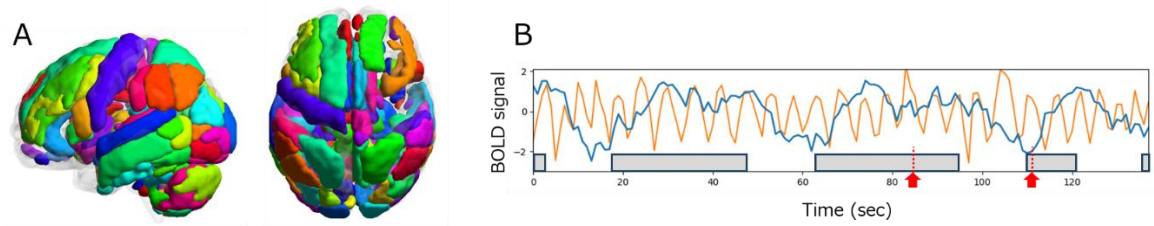

Figure S2. Pre-functional parcellation and extraction of representative subregion BOLD time courses. **A**, A total of 148 subregions identified via pre-functional parcellation and retained based on the GLM-based prescreening. Each color represents a different subregion; colors are used for visualization purposes only and do not carry specific meaning. For example, the left precentral gyrus, which is expected to be involved in right-hand button presses, was preserved (purple), whereas the corresponding region in the right hemisphere, which was not task-related, was excluded as expected. **B**, Representative BOLD time series from a subregion included in the prescreening (blue) and from one excluded by the prescreening (orange). Gray boxes indicate stimulus presentation periods, and red arrows mark the time points of response. The blue trace appeared to exhibit stimulus-related responses. In contrast, the orange trace displayed oscillatory fluctuations that appeared unrelated to stimulus presentation.

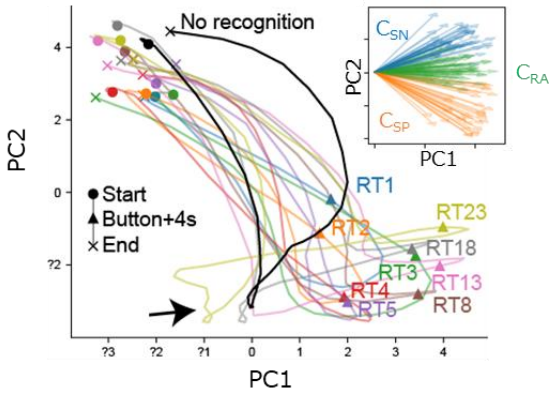

Figure S3 Trajectories in a state space. We applied principal component analysis (PCA) to reduce the dimensionality of the BOLD signals from the 148 subregions to two principal components. Unlike the analysis in Figure 2A, which focused on task-locked segments, this PCA was performed on the entire time series comprising 41,446 time points, and the first two components together accounted for 34% of the total variance. Note that this PCA differs from the one used earlier for the Gaussian mixture model clustering: in the current PCA, the goal was to project spatial activity patterns (across 148 subregions) into a low-dimensional state space at each time point; in contrast, the earlier PCA reduced the temporal dimension, from 41,446 time points to 148 time features, to capture major temporal patterns for clustering. Using the transformation obtained from this PCA, we visualized the trajectories in the PC1–PC2 space for trials with different RTs. A notable distinction emerged: in trials with RT = 1 or RT = 2, the neural state at the moment of response occupies a different region of the state space compared to trials with longer RTs. In trials with longer RTs, the response moment consistently converges at the lower-right edge of the space. This pattern appears to reflect the simultaneous increase in activity within both the C_{SP} and C_{RA} functional clusters, as indicated by the subregion loadings on the PC axes (inset, orange and blue, respectively). The state-space trajectories also suggest that, in trials with longer reaction times, including those in which recognition failed, a characteristic intermediate neural state is consistently traversed (black arrow). This state corresponds to the plateau observed in the cluster C_{SP} during the middle portion (10 to 20 s) of the time course shown in Figure 2A (v). The cluster C_{SP} is characterized by sustained activity throughout the task and is likely to reflect brain processes that contribute directly to the perceptual interpretation of Mooney images. A similar intermediate state may also be present in trials with RT = 8; however, in those cases, the peak response at the response moment may obscure this plateau, making it less apparent in the observed data. The presence of this plateau-like state appears more consistent with a CEM than with a DDM (see Discussion).

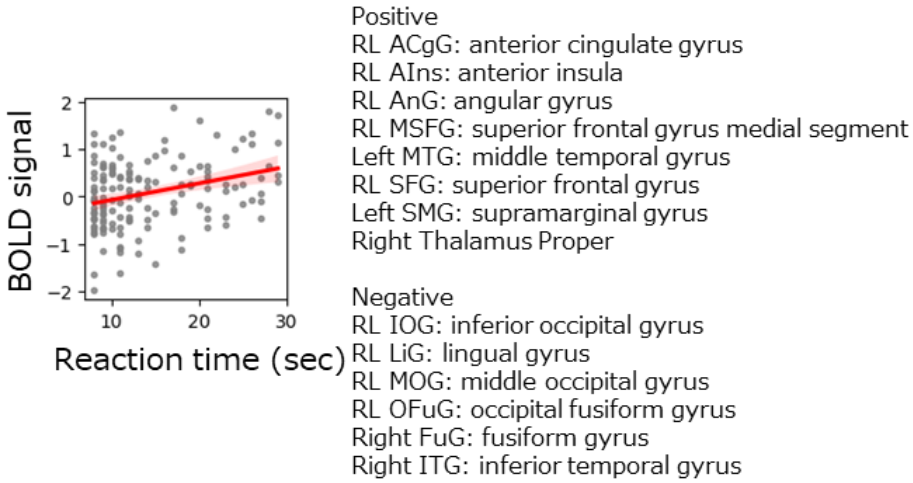

Figure S4. Subregions whose activation levels systematically varied with Reaction Time (RT). Specifically, we performed linear regression analyses between RT and the BOLD signal at 3–5 second after response as a HRF peak for each subregion (an example is shown), and obtained p-values for the slope coefficients to determine whether they were significantly greater or less than zero. The resulting p-values were corrected for multiple comparisons using the false discovery rate (FDR) method (Benjamini–Hochberg procedure, $q < 0.05$). Contrary to the peak-based analysis, most of the subregions exhibiting significant RT dependence were located in the cerebral cortex, with the exception of the right thalamus proper. Interestingly, all subregions with positive slopes were found in clusters C_{SN} or C_{RA}, while those with negative slopes were exclusively part of cluster C_{SP}. This observation may relate to prior findings suggesting a role for the DMN in Mooney image recognition and insight, given that cluster C_{SN}, which includes many DMN regions, exhibited increasing activation with longer RTs. This pattern is also consistent with the global time series shown in Figure 2A(v), where activity in C_{SN} and C_{RA} gradually increased with longer viewing durations, while activity in C_{SP} decreased slightly.