Selective attention modulates feature accumulation speed during perception and memory retrieval

Yu Zhou[1], Liang Zhang[1], Yujian Dai[1], Xiaojing Peng[1], Gui Xue[1,2*]

[1]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China.
[2]Chinese Institute for Brain Research, Beijing 102206, PR China

*Corresponding author. Email: gxue@bnu.edu.cn

Abstract

Mounting evidence has elucidated the sequential information processing trajectory, progressing from low-level perceptual to higher-level conceptual features during visual object recognition and reversing during memory retrieval. However, the extent to which the processing stream can be modulated by selective attention, as dictated by task requirements, remains unclear. The present study combined convolutional neural network (AlexNet), drift-diffusion model and multivariate decoding analysis to address this question. By using single-trial-based multivariate decoding analysis, we tracked the onset and peak time of feature representations across various visual perceptual and mnemonic categorization tasks, encompassing color, animacy, and size feature categorization. The findings of this study revealed several key outcomes: (1) A significant increase in reaction time was observed for the visual perception of color, animacy, and size features. (2) The findings from AlexNet and the drift-diffusion model revealed that, despite a parallel processing nature in these three features, those with larger decision boundaries necessitate subjects to spend more reaction time in making decisions about them. (3) In perceptual tasks, selective attention induced a reversal in the onset time of perceptual and conceptual features in the occipital and parietal lobe, resulting in the earlier detection of conceptual features compared to perceptual features. Conversely, in mnemonic tasks, selective attention led to the precedence of the peak time of perceptual features over conceptual features in the frontal lobe, signifying the reactivation of perceptual features with the highest fidelity prior to conceptual features. (4) Phase coupling analysis indicated a reversed information flow in the beta band during the task preparation stage in visual perception and memory retrieval tasks. Additionally, interactions between the occipital and temporal lobes in the theta band during the task preparation stage were found to promote an earlier onset time of target features in the visual perception stage. Collectively, these results underscore the parallel processing nature of features and demonstrate that the speed of feature information accumulation can be modulated by task requirements, with a more pronounced effect observed during perceptual compared to mnemonic stages.

Introduction

Many research found that the representational time courses varied across different features (Carlson et al., 2013; Cichy et al., 2014; Khaligh-Razavi et al., 2018; Wang et al., 2022a). Specifically, low-level perceptual features were discriminated faster behaviorally, and could be decoded from brain activity earlier, than high-level conceptual features when seeing an object (Clarke et al., 2013; Linde-Domingo et al., 2019; Mirjalili et al., 2021). Different from the visual perception stage, the sequence in which the perceptual and conceptual features were retrieved was reversed during the memory retrieval stage, with reaction times and brain activity patterns indicating that conceptual information was reconstructed more rapidly than perceptual details (Linde-Domingo et al., 2019; Mirjalili et al., 2021).

There were two perspectives on the timing of object categorization at different levels to explain this phenomenon. By one account, the relative timing suggests a sequence of stages in visual processing linked to specific levels of object categorization: fast categorizations are fast because they precede other categorizations within the visual processing hierarchy (Brincat & Connor, 2004; Serre et al., 2007; Yau et al., 2013). By another account, the relative timing reflects when perceptual features are available over time and the quality of perceptual evidence used to drive a perceptual decision process: fast simply means fast, it does not mean first. Alternatively, fast may also denote the presence of narrow decision boundaries. In instances where decision boundaries are narrow, the temporal duration required to reach such boundaries is anticipated to be shorter (Ratcliff, 1978; Ratcliff & McKoon, 2008a). The perceived or retrieved feature information can be seen as the outcomes of decisions generated by a noisy process that accumulates information over time. Our mainly focus was on investigating whether selective attention could influence the evidence accumulation process, thereby altering the representational timing of target and non-target features. This included the intriguing possibility of detecting conceptual features earlier than perceptual features when subjects observed an object and vice versa during the memory retrieval stage.

In our study, we first employed the deep neural network AlexNet (Krizhevsky et al., 2017) and the drift-diffusion model (DDM) (Ratcliff, 1978; Ratcliff & McKoon, 2008a) to examine the two perspectives on the timing of object categorization at different levels, respectively. Then we investigated whether selective attention as imposed by task requirement could modulate the processing stream during perception and memory retrieval. Selective attention is the prioritization of the neural representations that are most relevant to one's current behavioral goals. Such prioritization is necessary because the brain is a limited-capacity information system. Representations of external stimuli and internal thoughts compete for access to these limited processing resources, and attention helps to resolve that competition in favor of the information that is currently task-relevant. Mounting evidence has shown that selective attention could enhance target information and suppress non-target information. For example, the activation or representational strength of the target

information was stronger than that of the non-target information (Baldauf & Desimone, 2014; Grootswagers et al., 2021; Long & Kuhl, 2018; Maunsell & Treue, 2006). Besides, selective attention could increase the representational distance among target items, allowing for a finer discrimination (Nastase et al., 2017). Despite the weaker representational strength and fidelity of the non-target information compared to the target information, both were still incorporated into the processing procedure (Chen, 2012; Scholl, 2001). However, it is still uncertain whether the representational timing of target and non-target information would be modulated by selective attention, especially when these pieces of information are inherently bound together but exist at different levels, such as the coexistence of inherent perceptual and conceptual features within the same object.

To investigate this, we manipulated both perceptual (color) and conceptual features (animacy and real-world size) of objects in a magnetoencephalography (MEG) experiment. Two tasks were conducted: a learning-recall task and a perception task. In the learning-recall task, participants acquired object-location associations and subsequently recalled the target feature of the cued object. In the perception task, participants categorized objects based on the target feature. Initially, we characterized the three features using both the deep neural network AlexNet (Krizhevsky et al., 2017) and the drift-diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008a). Interestingly, we observed a increasing patterns in the boundary separation for color, animacy, and real-world size, despite they were processed in parallel. Subsequently, by employing single-trial-based decoding analysis, we were able to track the temporal dynamics of representations for the perceptual and conceptual features of the same object. Our study demonstrated that selective attention expedites the detection of conceptual features over perceptual features in visual perception, while during memory retrieval, selective attention made perceptual features reconstructed with the highest fidelity before conceptual features. Our findings provide new insights into the modulation effects of selective attention on temporal dynamics in feature processing.

Results

Behavioral performance

We used three tasks, i.e., color, animacy and real-word size tasks to examine how task requirement could modulate feature processing stream during perceptual and retrieval stages. The mnemonic task preceded the perception task to mitigate potential influences of stimulus familiarity on the memory retrieval process. For the mnemonic tasks, participants were first asked to learning the picture-location associations, and then make either color, animacy or size categorization of the object associated with the cued location (Figure 1A). To further probe the verbatim memory, a recognition task followed, wherein participants were tasked with selecting the target items from a set of three similar pictures. For the perceptual tasks, participants were asked to categorize the object based on a specific feature (Figure 1B. also See Methods).

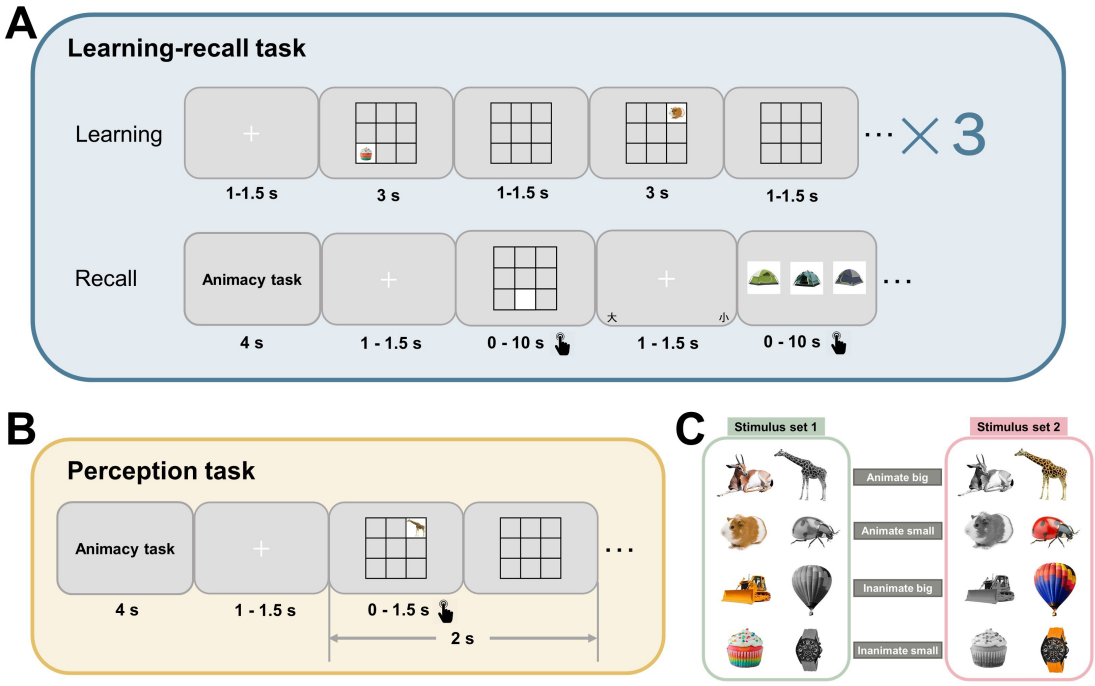

Figure 1. Stimuli examples and experimental design. A. Learning-recall task. During this task, participants acquired object-location associations, subsequently retrieved the target features of cued objects, and recognized the target objects presented in the learning stage. **B. Perception task.** In this task, participants categorized objects based on the target feature indicated during the task cue period. **C. Object dimensions.** A total of 192 objects varied along three dimensions: one perceptual dimension (Objects were presented in either colorful or grayscale formats.); two conceptual dimensions, one is the animacy dimension (Objects belonged to either the animate or inanimate.), and the other is the real-world size dimension (Objects could be categorized as small or big in the real world).

In the first analysis, we aimed to examine the processing stream of different features by reaction time. According to the previous findings (Clarke et al., 2013; Linde-Domingo et al., 2019; Mirjalili et al., 2021), we hypothesized that, during the visual perception stage, the accuracy for perceptual features would be higher than that for conceptual features, and reaction times for perceptual features would be faster than those for conceptual features. Our results aligned with previous research. In the feature perception task, accuracy in the experiment was close to the ceiling for the color and animacy task (color task: *Mean* = 97.05%, *SD* = 2.24%; animacy task: *Mean* = 96.30%, *SD* = 2.69%), and high in the size task (size task: *Mean* = 93.58%, *SD* = 3.15%). Only correct trials were used for all further reaction time (RT) analyses. We used the linear mixed-effect model to directly test the difference in accuracy and reaction time across three tasks. The task type (color, animacy and size) was used to predict the accuracy or the reaction time. Participant IDs were included as a random effect. The fixed factor task type significantly predicted participants' accuracy ($F_{2, 52}$ = 24.435, $p < 0.001$). Post-hoc tests indicated that both the color and animacy tasks showed significantly higher accuracy than the size task (color task vs. size task: t_{52} = 6.644, $p_{corrected} < 0.001$; animacy task vs. size task: t_{52} = 5.205, $p_{corrected} < 0.001$). The

accuracy of the color and animacy tasks were comparable (t_{52} = 1.440, $p_{corrected}$ = 0.328) (Figure 2A). The mean reaction times were 0.699 s (*SD* = 0.092 s) for the color task, 0.770 s (*SD* = 0.104 s) for the animacy task, and 0.798 s (*SD* = 0.088 s) for the size task. The fixed factor task type significantly predicted reaction times ($F_{2, 52}$ = 48.738, $p < 0.001$). Post-hoc tests showed that the color task was significantly faster than the animacy task (t_{52} = -6.876, $p_{corrected} < 0.001$) and size task (t_{52} = -9.574, $p_{corrected} < 0.001$), and the animacy task was significantly faster than the size task (t_{52} = -2.698, $p_{corrected}$ = 0.025) (Figure 2B).

During the feature retrieval stage, we did not observe reversed patterns for accuracy and reaction times compared to the feature perception stage. The mean accuracy was 89.06% (*SD* = 9.37%) for the color task, 90.39% (*SD* = 7.99%) for the animacy task, and 86.23% (*SD* = 7.02%) for the size task. The fixed factor task type significantly predicted participants' accuracy ($F_{2, 52}$ = 5.913, $p < 0.01$). Post-hoc testing indicated that the accuracy of the animacy task was significantly higher than that of the size task (animacy task vs. size task: t_{52} = 3.366, $p_{corrected} < 0.01$), while the accuracy of the color task did not exhibit a significant difference compared to the other two tasks. (color task vs. animacy task: t_{52} = -1.075, $p_{corrected}$ = 0.533; color task vs. size task: t_{52} = 2.291, $p_{corrected}$ = 0.066) (Figure 2C). The mean reaction times were 2.083 s (*SD* = 0.837 s) for the color task, 2.048 s (*SD* = 0.682 s) for the animacy task, and 2.508 s (*SD* = 0.719 s) for the size task. The fixed factor task type significantly predicted reaction times ($F_{2, 52}$ = 16.124, $p < 0.001$). Post-hoc testing showed that the reaction time of the size task was significantly longer than that of the color and animacy task (size task vs. color task: t_{52} = 4.713, $p_{corrected} < 0.001$; size task vs. animacy task: t_{52} = 5.100, $p_{corrected} < 0.001$) (Figure 2D). The behavioral results of recognition task were presented in Figure S1.

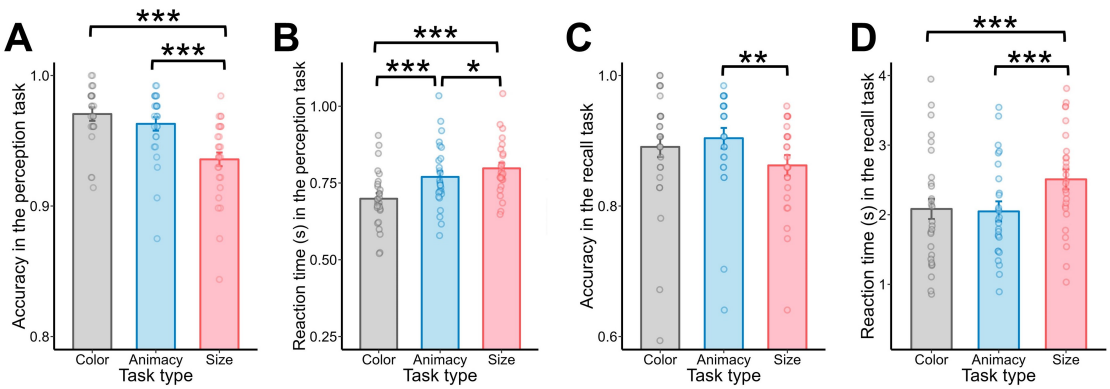

Figure 2. Behavioral results during the perception and retrieval stage. A. The accuracy of both color and animacy perception tasks surpassed that of the size task. **B.** The reaction time (RT) of color, animacy and size perception tasks showed a significantly increasing pattern: $RT_{(color)} < RT_{(animacy)} < RT_{(size)}$, and all of the comparisons were statistically significant. **C.** The accuracy of the animacy retrieval task was significantly higher than that of the size retrieval task. **D.** The reaction time for the size retrieval task was significantly longer than that of both the color and animacy retrieval tasks. Error bars reflect within-subject SEM (the standard error of the mean).

The features were represented in parallel in the deep neural network

To explore whether the progressively slower reaction times for the color, animacy, and size features were attributable to their positions at different stages within the visual processing hierarchy, we utilized the deep neural network (DNN) AlexNet (Krizhevsky et al., 2017) to address this question. The AlexNet comprises eight layers, encompassing five convolutional layers and three fully connected layers, mirroring the hierarchical structure of neurons along the ventral visual stream. Broadly, the early layers of the AlexNet are involved in processing early visual features such as colors, contrasts, and frequencies. In contrast, the deeper layers are responsible for handling higher-order visual representations, such as the surface structure of objects or body parts of animals. We operated under the assumption that if the layers containing the color, animacy, and size features were positioned higher and higher along the hierarchy, it might suggest a progressively later order for processing color, animacy, and size features. This could explain the progressively slower reaction times for color, animacy, and size features.

We extracted representational patterns for each picture in each layer, and then calculated Spearman's correlations between the representational patterns of all pairs of pictures in each layer. This process yielded a total of eight representational similarity matrices (Figure 3A). To discern feature-specific representations, we contrasted between-level similarities with within-level similarities using a t-test, employing the resulting t-value as a measure of sensitivity for the feature-specific representations. For instance, color-specific representations were derived by contrasting between-level similarity (similarity between colorful and grayscale pictures) with within-level similarity (similarity within colorful pictures or grayscale pictures) via a t-test (see Methods). We also use a permutation test to validate the statistical significance of the feature-specific representations. Remarkably, we identified significant representations for all three features across all eight layers (Figure 4B) (Stimulus set 1: color feature: $t_s = 5.362 \sim 16.296$, $p_s \leq 0.001$; animacy feature: $t_s = 1.634 \sim 82.416$, $p_s \leq 0.051$; size feature: $t_s = 4.537 \sim 25.242$, $p_s \leq 0.001$. Stimulus set 2: color feature: $t_s = 5.697 \sim 18.803$, $p_s \leq 0.001$; animacy feature: $t_s = 1.437 \sim 85.461$, $p_s \leq 0.075$; size feature: $t_s = 3.950 \sim 26.723$, $p_s \leq 0.001$). The results of the permutation test indicated that the t-values for all three features, obtained from the actual labels, consistently surpassed the corresponding threshold in each layer (Thresholds for stimulus set 1: color feature: $t_s = 0.743 \sim 1.756$; animacy feature: $t_s = 0.753 \sim 1.890$; size feature: $t_s = 0.707 \sim 1.712$. Thresholds for stimulus set 2: color feature: $t_s = 0.792 \sim 1.740$; animacy feature: $t_s = 0.776 \sim 1.707$; size feature: $t_s = 0.779 \sim 1.701$) (Figure 4B). These findings suggest that, contrary to our hypothesis, the color, animacy, and size features are processed in parallel within the deep neural network framework.

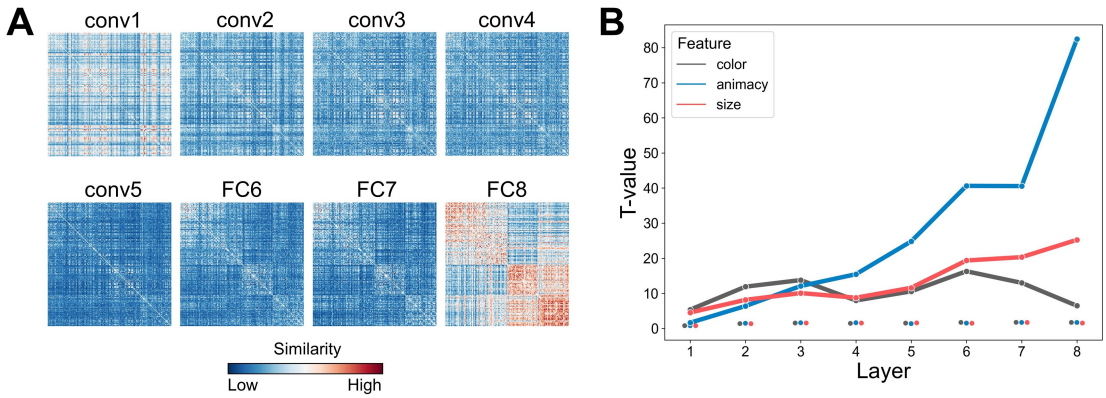

Figure 3. Representations of three features (color, animacy and size) in the deep neural network AlexNet (stimulus set 1). A. Representational similarity matrix. The matrix was generated by correlating the activations of artificial neurons in each DNN layer. The diagonal in each matrix was set to NaN. **B. Sensitivity of feature information.** This plot illustrates the sensitivity of feature information in each DNN layer. The dots, not connected by lines, represent the threshold of feature-specific representations in each layer. The figure was generated using data from stimulus set 1, and the figure for stimulus set 2 is available in the supplementary materials (Figure S2). conv: convolutional layer; FC: fully connected layer.

The evidence accumulation process of three features

In the feature perception task, we observed a gradual increase in reaction time for color, animacy, and size features based on behavioral performance results. From the AlexNet, we found that these features don't lie in different stages of the visual hierarchy, so our investigation shifted towards exploring whether the three features exhibited differences in the evidence accumulation process. To answer this question, we employed the drift-diffusion model (DDM) (Ratcliff, 1978; Ratcliff & McKoon, 2008a), which utilizes several parameters to characterize the evidence accumulation process. In this model, participants are assumed to accumulate evidence for one choice over another, and the rate of evidence accumulation within a trial is referred to as the drift rate (δ). The non-decision time (τ) accounts for time not dedicated to the decision-making process within the reaction time of each trial. Typically, non-decision time includes preprocessing time before the evidence accumulation process and motor response time after a decision has been reached. The boundary separation (α) represents the distance between the two decision boundaries, such as animate and inanimate (Figure 4A). In general, larger boundary separation, slower drift rate and logner non-decision time is longer would lead to longer reaction times.

The mean boundary separation for the color, animacy and size task was 3.39 (*SD* = 0.41), 3.71 (*SD* = 0.53) and 3.85 (*SD* = 0.26). The results of the linear mixed-effect model showed that the task type predicted the boundary separation significantly ($F_{2, 52}$ = 46.47, $p < 0.001$). Post-hoc testing showed that the boundary separation of the size task was significantly larger than that of the color (t_{52} = 9.42, $p_{corrected} < 0.001$) and animacy task (t_{52} = 2.91, $p_{corrected} < 0.05$), and the boundary separation of the animacy

task was significantly larger than that of the color task (t_{52} = 6.50, $p_{corrected}$ < 0.001) (Figure 4B). In addition, the mean drift rate for the color, animacy and size task was 4.16 (*SD* = 0.07), 4.39 (*SD* = 0.05) and 4.87 (*SD* = 0.05). The results of the linear mixed-effect model showed that the task type predicted the drift rate significantly ($F_{2, 52}$ = 3115.56, p < 0.001). The drift rate (DR) of color, animacy and size tasks showed a significantly increasing pattern: DR(color) < DR (animacy) < DR (size), and all of the comparisons were statistically significant (animacy task vs. color task: t_{52} = 25.59, $p_{corrected}$ < 0.001; size task vs. color task: t_{52} = 77.46, $p_{corrected}$ < 0.001; size task vs. animacy task: t_{52} = 51.88, $p_{corrected}$ < 0.001). The mean non-decision time for the color, animacy and size task was 265.47 ms (*SD* = 48.91 ms), 285.77 ms (*SD* = 52.06 ms) and 277.72 ms (*SD* = 23.83 ms). The results of the linear mixed-effect model showed that the task type predicted the non-decision time significantly ($F_{2, 52}$ = 3.50, p < 0.05). Post-hoc testing showed that the non-decision time of the animacy task was significantly longer than that of the color task (t_{52} = 2.63, $p_{corrected}$ < 0.05). Considering the results of these three parameters collectively, the differences in reaction times among the three tasks are likely primarily due to variations in boundary separation between the tasks. A larger boundary separation requires individuals to spend more time making decisions.

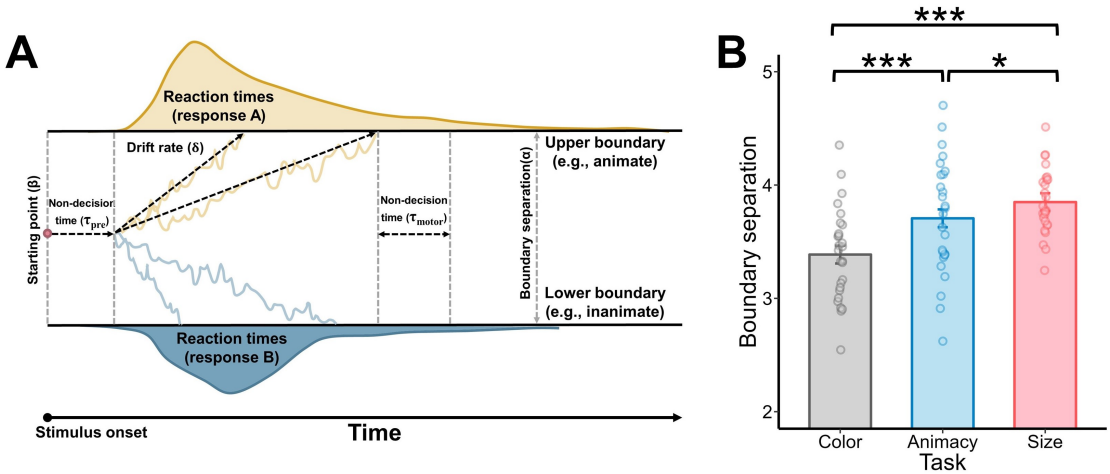

Figure 4. Results of the drift-diffusion model. A. Graphical illustration of the drift-diffusion model. At the onset of a trial, a participant's information level commences at β. Over time, feature information accumulates until it reaches either the upper boundary (e.g., animate) or lower boundary (e.g., inanimate), resulting in response A or B. The accumulation rate of feature information is represented by the drift rate δ. The time utilized for preprocessing before evidence accumulation and motor response after decision-making is the non-decision time τ. **B. Pattern of increasing boundary separation.** A discernible pattern showcased increasing boundary separation along the color, animacy, and size tasks, and all of the comparisons were statistically significant.

Selective attention modulated the onset time of feature perception

The drift-diffusion model revealed a sequential increase in boundary separation for color, animacy, and size features, aligning with prolonged reaction times. However,

the precise timing of feature representations at the neural level and the potential modulation by selective attention remained unclear. To address this, we employed single-trial-based multivariate classification analyses to identify feature onset and peak times in each task (color, animacy, and size). We hypothesized that target features would be represented earlier than non-target features during the perception stage.

Recognizing that attention may modulate different brain regions differently, all decoding analyses were carried out separately in four regions of interest (ROIs): the occipital, temporal, parietal, and frontal lobes. The sensors encompassed within these four regions were precisely defined using the MNE toolbox (Gramfort et al., 2013) (Figure S3). We used the learning stage as an independent localizer task to find the most informative time window (i.e., the peak bin) of each feature in each ROI (Liu et al., 2019; Wimmer et al., 2020). For each participant, four distinct classifiers were trained at each time point and ROI using data from the learning stage: one to classify the color category (colorful or gray), one for the animacy category (animate or inanimate), and two for the real-world size category (big or small) – one for animate objects and another for inanimate objects (Figure 5A). The decision to classify the real-world size category separately for animate and inanimate objects was informed by research indicating that real-world size drives differential responses only in the object domain, not the animate domain (Konkle & Caramazza, 2013). For the learning stage, we found significant clusters of color, animacy and size features in the occipital, temporal and parietal lobes ($P_{corrected} < 0.05$). In the frontal lobe, we only found significant clusters of the animacy feature ($P_{corrected} < 0.05$) (Figure S4). It's important to note that significant clusters for the size feature in animate objects were not identified during the learning stage. Consequently, the onset and peak times for the size feature during the perception and recall stage were exclusively calculated for inanimate objects. The peak latencies of significant clusters of the color, animacy and size features during the learning stage were 305 ms, 515 ms and 315ms in the occipital lobe; 365 ms, 485 ms and 445 ms in the temporal lobe; 565 ms, 495 ms and 445 ms in the parietal lobe. The peak latency of the animacy feature in the frontal lobe was 595 ms.

To identify the most informative time window containing feature information, a time window of 50 ms was selected, which included the peak time point and four surrounding time points in each condition (i.e., peak bin). Utilizing these peak bins ensured that classifiers learned activation patterns with the highest signal-to-noise ratio. Trained classifiers were then tested on data from each time point in each task type during the perception stage. The resulting time course of decoding performance for each feature in each categorization trial, ROI, and participant was obtained. Permutation and bootstrapping methods were employed to construct a null distribution for each time point. The 95th percentile of the null distribution served as the threshold for the corresponding time point. A cluster with a duration of at least 20 ms, wherein all time points surpassed the corresponding thresholds, was considered to

significantly contain feature information. The onset time of feature representation in a trial was the first time point of the first significant cluster in this trial. This onset time represented the earliest time point at which feature information was detected in the brain. The peak time point, corresponding to when feature information was represented with the highest fidelity, was determined by selecting the time point with the highest d-value across all significant clusters in each trial. To investigate whether the onset or peak time of target features was earlier than that of non-target features during the perception stage, single-trial-based generalized linear mixed models (GLMM) were employed. The GLMM included task type (color, animacy, and size tasks), feature type (color, animacy, and size features), and their interactions as fixed effects. Participant ID, including the intercept, was modeled as a random factor.

For the onset times, we found that the interaction between task and feature were significant in the occipital ($F_{4, 17104} = 10.659$, $p < 0.001$), temporal ($F_{4, 16809} = 6.663$, $p < 0.001$), and parietal lobe ($F_{4, 17066} = 4.173$, $p = 0.002$). Then we examined whether the onset time of 1) conceptual features (i.e., animacy vs. size) or in a further step, 2) the perceptual and conceptual features (i.e., color vs. animacy; color vs. size) was modulated by selective attention. We anticipated that the onset time of target features would precede that of non-target features. Additionally, if conceptual features were designated as target features, their onset times might precede those of perceptual features when the latter were non-target features.

For features belonging to the conceptual category, i.e., the animacy and size feature, post-hoc analyses showed that in the occipital lobe, the onset time of the animacy feature (target feature) was significantly earlier than that of the size feature (non-target feature) in the animacy task (animacy feature vs. size feature: $t_{17104} = -4.019$, $p_{corrected} < 0.001$), while in the size task, the onset time of these two features didn't show a significant difference (animacy feature vs. size feature: $t_{17104} = -0.013$, $p_{corrected} = 0.989$) (Figure 5B, left panel). More importantly, in the parietal lobe, the onset time of the animacy feature was significantly earlier than that of the size feature in the animacy task (animacy feature vs. size feature: $t_{17066} = -1.969$, $p_{corrected} = 0.049$), whereas a reversed pattern was found in the size task (animacy feature vs. size feature: $t_{17066} = 2.352$, $p_{corrected} = 0.019$) (Figure 5B, right panel). These results showed that selective attention made the onset time of the target feature earlier than that of the non-target feature.

For features belonging to different categories, i.e., the perceptual feature color and the conceptual feature animacy, post-hoc comparisons indicated that in the occipital lobe, the onset time of the color feature was significantly earlier than that of the animacy feature in the color task (color feature vs. animacy feature: $t_{17104} = -2.114$, $p_{corrected} = 0.035$). While in the animacy task, the onset time of the color feature was significantly later than that of the animacy feature (color feature vs. animacy feature: $t_{17104} = 4.901$, $p_{corrected} < 0.001$) (Figure 5C). These findings suggest that selective attention indeed influenced the onset time of both perceptual and conceptual features.

Moreover, a similar pattern was observed for the color and size features. In the occipital lobe, the onset time of the color feature occurred significantly earlier than that of the size feature in the color task (color feature vs. size feature: t_{17104} = -3.930, $p_{corrected}$ < 0.001), while in the size task, the onset time of the color feature occurred significantly later than that of the size feature (color feature vs. size feature: t_{17104} = 2.916, $p_{corrected}$ = 0.004) (Figure 5D, left panel). We also found that in the temporal lobe, the onset time of the color feature was significantly earlier than that of the size feature in the color task (color feature vs. size feature: t_{16809} = -2.956, $p_{corrected}$ = 0.003), and in the size task, the onset time of the color and size feature had no significant difference (color feature vs. size feature: t_{16809} = 1.138, $p_{corrected}$ = 0.255) (Figure 5D, right panel). The aforementioned results highlight that selective attention played an important role in modulating the onset time of features at different levels, and intriguingly, it even led to an earlier onset time for conceptual features compared to perceptual features.

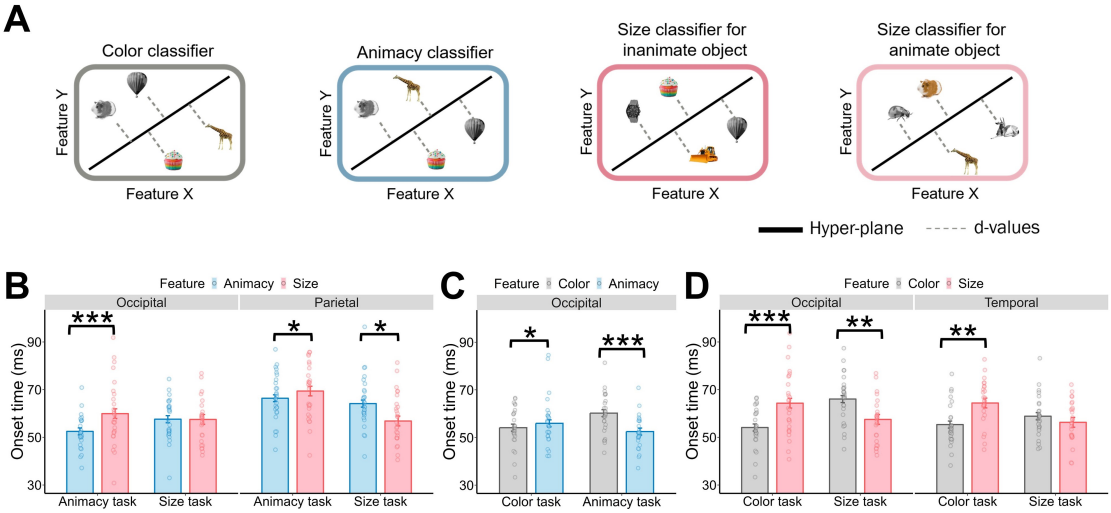

Figure 5. Modulation effects of selective attention during the perception stage. A. Classifier training. Four classifiers were trained in each ROI using data from the learning stage for each participant. These included one color classifier (discriminating colorful from gray), one animacy classifier (discriminating animate from inanimate), and two real-world size classifiers (discriminating small from big) for inanimate and animate objects. All classifiers were tested at each time point of each trial, generating a time series of classification performance (d-value) for each feature in every trial. **B, C, D. Modulation effects.** The effects of selective attention on the onset time of **B.** the animacy and size feature, **C.** the color and animacy feature, and **D.** the color and size feature are depicted. Error bars represent within-subject SEM (standard error of the mean).

The correlation between the phase coupling strength during the task preparation stage and the onset time of feature representations during the perception stage

The above results showed that selective attention had a discernible impact on the onset time of both target and non-target features. Remarkably, this modulation effect was particularly pronounced in the occipital lobe, a sensory cortex. This observation

raised the intriguing question of why the occipital lobe exhibited such a robust modulation effect. Previous research has posited that visual attention relies on brain networks comprised of higher-order regions, such as the temporal cortex attention network (Long & Kuhl, 2018; Ramezanpour & Fallah, 2022; Yeo et al., 2011). These networks are believed to support visual attention by influencing and assessing sensory representations within visual cortical areas (Desimone & Duncan, 1995; Egner & Hirsch, 2005; Gazzaley & Nobre, 2012; Serences & Yantis, 2006). Moreover, synchronization between higher- and lower-order regions has been correlated with the strength of sensory responses (Baldauf & Desimone, 2014). Consequently, we assumed that the occipital lobe might not solely execute the entire modulation process on its own; instead, it might receive assistance from higher-order regions to achieve this modulation effect. To explore if the occipital lobe was influenced by other regions and when this influence occurred, we utilized the phase slope index (PSI) to estimate phase coupling patterns between the occipital lobe and the higher-order brain regions in the theta (4-8 Hz), alpha (9-13 Hz), beta (14-30 Hz) and gamma bands (31-100 Hz) during the task preparation stage in the perception task, with the occipital lobe as the target ROI and higher-order regions as seed ROIs. Our analysis revealed significant phase coupling between the temporal and occipital lobes, with two significant clusters identified in the theta band for each task type ($p_{corrected} < 0.001$) (Figure 6A). In each task type, PSI values in the first significant cluster were positive, indicating a top-down flow of information from the temporal to the occipital lobe. Conversely, PSI values in the second significant cluster were negative, signifying a reversed information flow compared to the first cluster. This suggested that, in the theta band, the information flow between the temporal and occipital cortex underwent a sequential top-down and bottom-up process during the task preparation stage.

Having obtained the phase coupling between the seed and target ROIs, we computed Pearson's correlation between the phase coupling strength and the onset times of target and non-target features in the occipital lobe, respectively. This analysis aimed to explore whether a more robust phase coupling between the occipital lobe and higher-order regions during the preparation stage would be correlated with an earlier onset time of target features rather than non-target features during the subsequent perception stage. Our findings revealed a significant negative correlation between the strength of phase coupling in the theta band between the temporal and occipital lobe and the onset times of the target features ($R = -0.39$, $p_{corrected} = 0.042$; Figure 6B, left panel) instead of non-target features ($R = -0.15$, $p_{corrected} = 0.46$; Figure 6B, right panel) in the occipital lobe. This suggests that a stronger phase coupling between the temporal and occipital lobe in the theta band during the task preparation stage is associated with earlier onset times of target features in the occipital lobe during the subsequent perception stage.

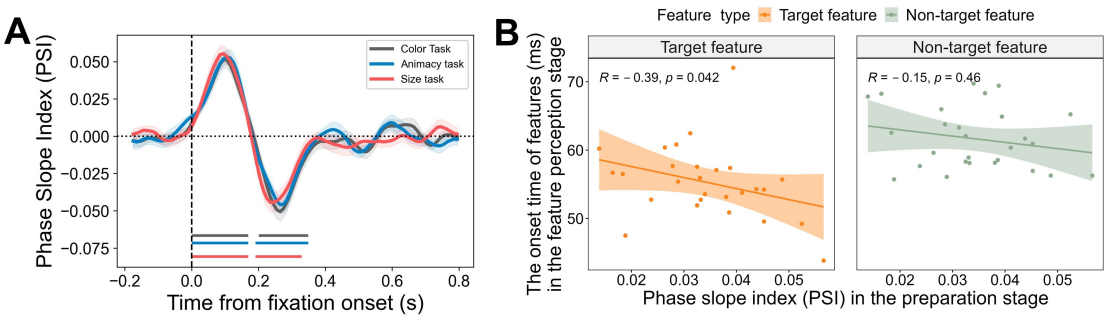

Figure 6. Phase coupling results during the feature perception stage. A. Phase coupling patterns. Significant phase coupling was observed between the temporal and occipital lobes in the theta band for each task type ($p_{corrected} < 0.001$). Two distinct directions of coupling were identified: one from the temporal to the occipital lobe, and the other in the opposite direction. **B. Correlation with onset times.** A significant negative correlation was found between the strength of temporal-occipital phase coupling in the theta band during the task preparation stage and the onset times of target features (R = -0.39, $p_{corrected}$ = 0.042) instead of non-target features (R = -0.15, $p_{corrected}$ = 0.46) in the occipital lobe during the perception stage.

Selective attention modulation effect on the peak time of feature retrieval

The previous analysis unveiled the modulation of selective attention on the onset time of features in the perception task, even leading to an earlier onset time for conceptual features compared to perceptual features. In a subsequent analysis, our aim is to investigate whether selective attention can also influence the onset or peak time of feature representations during the memory retrieval stage. This inquiry is intriguing due to the distinct nature of visual perception and memory retrieval as two separate mental processes. Visual perception involves an outside-in process where individuals process visual inputs from the external world, while memory retrieval is an inside-out process where individuals retrieve information stored in their minds. Additionally, past research has indicated that conceptual information is reconstructed more swiftly than perceptual details during the memory retrieval stage, showcasing an inverted pattern compared to the object recognition process (Linde-Domingo et al., 2019; Mirjalili et al., 2021). The question of how selective modulation effects on feature representational time vary across these two different mental processes remains an enigma. Employing the same decoding and statistical analysis methods utilized in the feature perception stage, our findings reveal selective attention modulation effects on the peak time rather than the onset time of feature retrieval. Specifically, the results from Generalized Linear Mixed Models (GLMM) indicate that the interaction between task type (color, animacy, and size tasks) and feature type (color, animacy, and size features) was significant in the frontal lobe ($F_{4, 8674}$ = 5.697, p < 0.001), and marginally significant in the temporal lobe ($F_{4, 8611}$ = 2.189, p = 0.068).

For the conceptual features (animacy and size), post-hoc comparisons indicated that in the temporal lobe, the peak time of the animacy feature was significantly earlier than that of the size feature in the animacy task (animacy feature vs. size feature: t_{8611} = -3.128, $p_{corrected}$ = 0.046), while the peak time of these two features had no significant

difference in the size task (animacy feature vs. size feature: t_{8611} = -0.996, $p_{corrected}$ = 0.986) (Figure 7A). For the perceptual feature color and conceptual feature animacy, post-hoc comparisons indicated that in the frontal lobe, the peak time of the color feature was significantly earlier than that of the animacy feature in the color task (color feature vs. animacy feature: t_{8674} = -3.414, $p_{corrected}$ = 0.019), while this pattern was reversed in the animacy task (animacy feature vs. color feature: t_{8674} = -3.121, $p_{corrected}$ = 0.047) (Figure 7B, right panel). In the temporal lobe, the peak time of the animacy feature was significantly earlier than that of the color feature in the animacy task (animacy feature vs. color feature: t_{8611} = -3.708, $p_{corrected}$ = 0.007), while in the color task, the peak time of these two features didn't show a significant difference (animacy feature vs. color feature: t_{8611} = -1.430, $p_{corrected}$ = 0.886) (Figure 7B, left panel). These results indicated that the perceptual feature (color) can be reconstructed with the highest fidelity earlier than the conceptual feature (animacy) in the frontal lobe. For the color and size feature in the temporal cortex, the peak time of the size feature was significantly earlier than that of the color feature (size feature vs. color feature: t_{8611} = -3.455, $p_{corrected}$ = 0.016), while there was no significant difference in the peak time between these two features in the color task (size feature vs. color feature: t_{8611} = -1.885, $p_{corrected}$ = 0.624) (Figure 7C).

In contrast to the perception stage, our observations during the retrieval stage indicate that attention modulates the peak time rather than the onset time of feature representations. Notably, this modulation results in the target features being reconstructed with highest fidelity earlier than non-target features. The comparison between color and animacy features further suggests that attention modulations may facilitate a faster reconstruction of perceptual features compared to conceptual features.

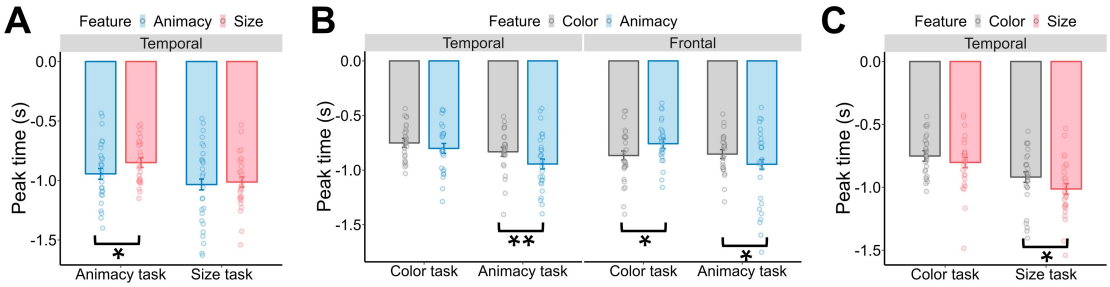

Figure 7. Modulation effects of selective attention on the peak time of A. the animacy and size feature, **B.** the color and animacy feature, **C.** the color and size feature. Error bars reflect within-subject SEM (the standard error of the mean).

Intriguingly, in contrast to the perception task where modulation effects primarily manifested in the occipital lobe, the retrieval task revealed modulation effects predominantly situated in the temporal and frontal lobes. We posit that information exchange between the temporal and frontal lobes may commence as early as the task preparation stage, specifically during the fixation period preceding the presentation of the spatial cue. Participants likely engage in preemptive object or target feature retrieval, facilitating rapid response upon the spatial cue's appearance. To scrutinize

interactions between these lobes, we computed the phase slope index (PSI) across four frequency bands between the temporal (target ROI) and frontal (seed ROI) lobes, utilizing fixation period data before the spatial cue stage. Remarkably, significant clustering between these regions emerged in the beta band for each task type (color task: 225-305 ms, 365-515 ms; animacy task: 295-525 ms; size task: 415-535 ms) ($p_{corrected} < 0.05$) (Figure 8A). And the values in each significant cluster were positive, signifying a prevailing direction of information flow from the frontal to the temporal lobe. A striking departure from this pattern became evident when scrutinizing the results of the perception preparation stage (Figure 8B). Notably, during this stage, no significant cluster featuring the same direction of information flow emerged. Instead, we uncovered a significant cluster with information flow directed from the temporal to the frontal lobe, observed in the animacy (175-255 ms) and size task (115-195 ms) ($p_{corrected} < 0.05$). These nuanced results underscore the specificity of the frontal-to-temporal information flow during the task preparation stage, discernibly characteristic of the retrieval task rather than the perception task.

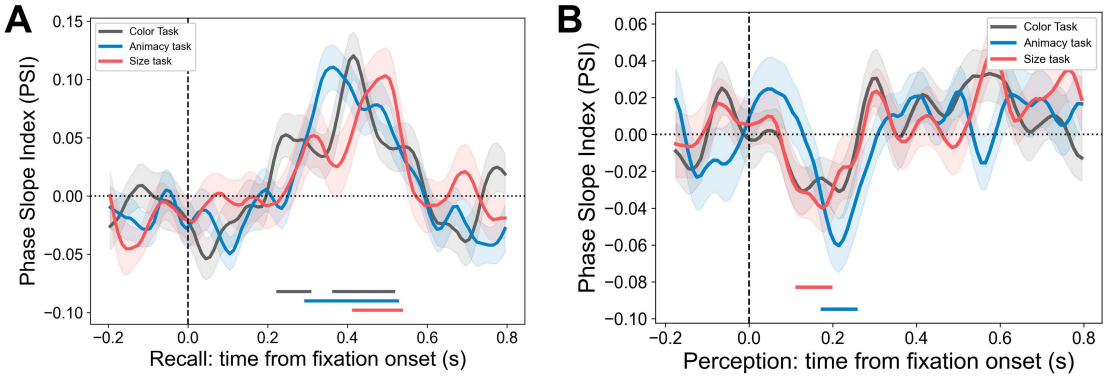

Figure 8. The phase coupling between the temporal and frontal lobes in the beta band during the preparation stage. A. In the recall task, there existed a significant cluster of phase coupling from the frontal to temporal lobe in each task type (color, animacy and size task) ($p_{corrected} < 0.05$). **B.** In the perception task, there existed a significant cluster of phase coupling from the temporal to the frontal lobe in the animacy and size task ($p_{corrected} < 0.05$).

Discussion

In our investigation utilizing single-trial-based multivariate decoding analysis, we delved into the modulation effects of selective attention on the representational times of both perceptual and conceptual features during the visual perception and retrieval stages. Intriguingly, during the feature perception stage, selective attention emerged as a catalyst, advancing the onset times of target features over non-target features. Notably, it even orchestrated a scenario where conceptual features took precedence over perceptual features under the conceptual task. In contrast, within the feature retrieval task, selective attention exerted its influence on the peak times, rather than the onset times, advancing the peak times of target features over non-target features. Furthermore, in the feature retrieval task, it orchestrated a situation where perceptual features were reactivated with the highest fidelity before conceptual features. These

groundbreaking findings challenge the conventional notions that perceive visual perception as a feedforward process (Carlson et al., 2013; Cichy et al., 2014; Clarke & Tyler, 2015; Lehky & Tanaka, 2016; Martin et al., 2018; Serre et al., 2007) and memory retrieval as a feedback-driven process (Linde-Domingo et al., 2019; Mirjalili et al., 2021).

The representational times of perceptual and conceptual features can be reversed by selective attention

Models of object categorization suggested a simultaneous processing of object information across different categorization levels, with varied time courses of evidence accumulation (Kravitz et al., 2013; Mack & Palmeri, 2011). Previous empirical research supported these models, revealing early onset latencies of decoding accuracy for various categorization levels (48-70 ms), with higher-level features peaking later than lower-level features (Carlson et al., 2013; Cichy et al., 2014). Our study, utilizing the deep neural network AlexNet (Krizhevsky et al., 2017), demonstrated parallel processing of color, animacy, and real-world size features, validating the sequential order-free nature of their processing. Furthermore, employing the drift-diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008b), we confirmed the variability in boundary separation across these features, and revealed a sequential increase in boundary separation for color, animacy, and size features, aligning with their prolonged reaction times.

Building on these insights, we investigated whether the onset or peak times of perceptual and conceptual features can be modulated by selective attention in a feature perception task. Leveraging a single-trial-based decoding analysis, a more sensitive approach (Linde-Domingo et al., 2019), we aimed to discern the nuanced dynamics of feature processing. We found that, in a visual perceptual task, when attention was directed toward perceptual features (e.g., color), their onset times preceded those of conceptual features. Intriguingly, when attention shifted to conceptual features (e.g., animacy and real-world size), the onset times of conceptual features significantly outpaced those of perceptual features (see Figure 5). This suggests a dynamic and flexible nature of visual perception, challenging the notion of a fixed sequence where perceptual features invariably precede conceptual features (Clarke et al., 2013; Linde-Domingo et al., 2019; Mirjalili et al., 2021).

In contrast to the typical observation in memory retrieval studies where conceptual features peak earlier than perceptual features (Linde-Domingo et al., 2019; Mirjalili et al., 2021), our study introduces a novel finding. By directing participants' attention specifically to perceptual features during the memory retrieval stage, we discovered that the peak time of the perceptual feature, color, could precede that of the conceptual feature, animacy, in the frontal lobe (see Figure 7B, right panel). In tasks focused on conceptual feature retrieval, we observed that the peak times of conceptual features led those of perceptual features in the temporal and frontal lobes (see Figure 7). These findings suggest that during memory retrieval, conceptual features are not

invariably prioritized, and perceptual features can be reactivated with highest fidelity before conceptual features. It is noteworthy that our study did not identify selective modulation effects on the onset times of feature retrieval. This implies that selective attention may not alter the moment features are initially reactivated in the brain. Instead, it seems to exert influence on the timing of features being reconstructed with the utmost fidelity.

The interactions between brain regions during the preparation stage promoted the evidence accumulation speed of target features in the visual perception task

It is noteworthy that the selective attention modulation effect occurred in the occipital lobe for all three feature pairs during the visual perception stage. Previous research has demonstrated that top-down factors can influence stimulus representations in early visual areas (Baldauf & Desimone, 2014; Ester et al., 2016; Jehee et al., 2011; Sprague & Serences, 2013). Moreover, cognitive intentions, such as expectations and mental readiness, embedded in top-down processes, may precede the presentation of a stimulus. This intentional processing guides bottom-up sensory processing by reallocating attention based on subjective expectations, enhancing the speed and efficiency of perceptual identification (Baldauf & Desimone, 2014; Córdova et al., 2016; Sakai, 2008; Siegel et al., 2008). To understand how this top-down modulation was achieved, we explored the phase coupling between the occipital lobe and higher-order brain regions (temporal, parietal, and frontal lobes) during the task preparation stage when stimuli were not presented. In the theta band, we observed an information flow from the temporal to occipital lobe during the initial 200 ms, followed by a backward information flow in the subsequent 200 ms. Further analysis revealed that the strength of phase coupling in this first top-down and then bottom-up cluster during the task preparation stage correlated with the onset times of target rather than non-target features in the occipital lobe during the perception stage. Specifically, stronger phase coupling between the occipital and temporal lobes in the theta band was associated with earlier onset times of target rather than non-target features, emphasizing the role of interactions between brain regions during the preparation stage in promoting the evidence accumulation speed for target features in visual perception tasks.

The frontal-to-temporal information flow was specific to the memory preparation stage

In our study, participants are required to categorize objects in both visual perception and memory retrieval tasks. However, these tasks differ in their foundations: one relies on visual input from the external environment, while the other draws upon information retrieved from internal mental representations. Extensive research has established that object recognition and representations of visual features at different levels involve a hierarchical network of regions in the ventral and dorsal visual pathways (Cichy et al., 2014; Felleman & Van Essen, 1991; Freud et al., 2016; Grill-Spector & Malach, 2004; Kourtzi & Connor, 2011). Neural reactivation associated

with different feature levels not only encompasses the dorsal and ventral visual streams but extends into the frontal cortex (Bone et al., 2020; Linde-Domingo et al., 2019), indicating the involvement of higher-order regions in memory retrieval compared to visual perception. Previous research has highlighted the crucial role of frontal regions in prioritizing items in visual working memory, selecting and integrating task-relevant features, and maintaining working memory flexibility while avoiding interference (Buschman & Kastner, 2015; Chatham et al., 2014; Nee & Jonides, 2008; Nobre et al., 2004), the selection and integration of task-relevant features (Mante et al., 2013), the flexibility of working memory and the avoidance of interference due to competing representations (Bouchacourt & Buschman, 2019). In our study, the observed reversal of the peak time of feature retrieval in the frontal lobe also underscores the significant role of frontal regions in the selective attention modulation effect. In addition to frontal lobe, prior research indicates that regions within the temporal cortex, such as the inferior temporal cortex (ITC), house neurons responsive to visual features and objects (Kravitz et al., 2013), and neurons involved in visual associative memory maintenance and retrieval (Naya et al., 2001). Our results indicate that the temporal lobe demonstrates a selective attention modulation effect, with the retrieval of target features occurring before non-target features. Additionally, the frontal cortex, especially the prefrontal cortex, has reciprocal anatomical connections with the ITC (Gerbella et al., 2010; Saleem et al., 2014), exerting top-down signals that modulate the visual neural responses of the ITC via top-down signals (Fuster et al., 1985) and, in particular, induce activity related to memory retrieval (Tomita et al., 1999; Zhou et al., 2023). These findings prompted an investigation into potential top-down modulations from the frontal lobe to the temporal lobe. Phase coupling analysis revealed a significant positive cluster in the beta band, indicating information flow from the frontal to temporal lobe. Notably, this directional information flow was absent during the preparation stage in the feature perception task. Previous research also found that beta has been associated with top-down information. Task rules are reflected in different patterns of beta synchrony in the frontal (Buschman et al., 2012) and visual cortex (Richter et al., 2018) as if the beta was helping form ensembles for the rules. Such content-specific ‘‘beta ensembles’’ have also been found for other types of top-down information, like learned categories (Antzoulatos & Miller, 2016; Stanley et al., 2018; Wutz et al., 2018). Our findings extend this knowledge by demonstrating that top-down beta oscillations can occur as early as the task preparation stage, even preceding the presentation of the memory cue. Although we identified a specific frontal-to-temporal information flow in the beta band during the preparation stage in the memory retrieval task, no correlations were found between the strength of this information flow and the peak times of feature representations.

The question of whether the information flow involving different levels of features bound within objects follows a fixed hierarchical order during both visual perception and retrieval stages is intriguing. Our findings suggest that the representational times of various levels of features, such as perceptual and conceptual features, can indeed

be modulated by selective attention. Specifically, conceptual features can be detected earlier than perceptual features during the visual perception stage, while perceptual features can be reconstructed with the highest fidelity before conceptual features during the memory retrieval stage. This study provides valuable insights into the temporal modulation effects of selective attention, contributing to a more comprehensive and mechanistic understanding of its impact.

Methods

Participants

Thirty healthy volunteers with normal or corrected-to-normal visual acuity, with no history of psychiatric or neurological disorders, took part in the study. Three of them were too drowsy to complete the experiment. The analysis of the remaining 27 participants was reported (8 female; mean age = 21.70±2.37 years). All experiments were carried out following the Declaration of Helsinki. All participants provided written informed consent before the start of the experiment, which was approved by the Research Ethics Committee at Peking University and the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University in China.

Stimuli

To scrutinize the processing hierarchy during both the perception and retrieval stages, it was crucial to independently manipulate the perceptual and conceptual features of the presented objects. In our experiment, objects exhibited variations along three orthogonal dimensions: one perceptual dimension (color), where the object was presented either in a colorful or grayscale format; and two conceptual dimensions (animacy and real-world size), where the object was categorized as either animate or inanimate (animacy dimension), and big or small (real-world size dimension) (Figure 1C).

We utilized 192 distinct images of everyday objects and common animals in the primary experiment. An additional 24 images were employed in the practice session. Three types of features were introduced: color, animacy, and real-world size. Color was designated as the perceptual feature, while animacy and real-world size were considered conceptual features. For the color feature, half of the pictures were presented in grayscale, and the remaining half were in color. To control for potential color-related biases in the results, two sets of pictures were created, with color assignments reversed between the sets (e.g., a colorful cupcake in stimulus set 1 and a gray cupcake in stimulus set 2) (Figure 1C). Half of the participants were exposed to stimulus set 1, while the other half encountered stimulus set 2. The two animacy levels (animate and inanimate) and size levels (big and small) were evenly distributed within each level of the color feature (i.e., colorful and gray). The criteria for determining big or small real-world size were based on a reference box measuring approximately 40 (length) × 18 (width) × 17 (height) centimeters in the experiment room. Objects that fit into this box were classified as small, typically handheld items, while those that didn't fit were deemed big and could support a human (e.g., chair-sized or larger). We labeled the categories colorful, animate, and big as level 0, while gray, inanimate, and small were labeled as level 1. The 192 objects were divided into three groups (group 1, group 2, and group 3) for the three tasks (color, animacy, and size categorization tasks). Six assignment plans were implemented to ensure all combinations of stimuli groups and tasks were covered. Levels of each feature were equally distributed across the three stimuli groups. All images were presented at the center of the screen, with a rescaled size of 198×198 pixels.

Procedure

The experiment comprised two parts: the learning-recall task and the perception task. Prior to each task, participants received verbal instructions and practiced to become familiar with the experimental procedure. The learning-recall task consisted of eight blocks, each containing three mini-blocks (24 mini-blocks in total). Each mini-block consisted of a learning and recall stage. During the learning stage, a jittered fixation cross (1-1.5 s from a uniform distribution) was followed by a series of objects presented in pseudorandom order within the eight cells surrounding the central cell (no object was shown in the center cell). Each object was displayed for three seconds, and the inter-stimulus interval was jittered around 1-1.5 s from a uniform distribution. Participants were instructed to learn the object-location associations in preparation for a memory test, with no responses required during the learning stage. The learning process in each mini-block was repeated three times consecutively. Objects of different types occurred with equal probability across all cells during the entire learning-recall task. Following the learning stage, participants engaged in the feature retrieval and object recognition tasks, where they were prompted to recall the features of the cued objects and distinguish target items from lures (Figure 1A). At the beginning of the retrieval stage, a task cue (indicating the color, animacy, or size task) was displayed for 4 seconds, instructing participants on which feature of the cued object they needed to retrieve (the target feature). A centrally presented fixation cross appeared for 1-1.5 seconds before the initiation of each retrieval trial. Subsequently, a white square, identical in size to the pictures presented during the learning stage, appeared in any one of the eight cells. Participants were required to retrieve the target feature of the cued object and press the button to provide their answers once retrieval was successful. For instance, in the animacy task, participants had to categorize whether the object in the corresponding location was animate or inanimate. Following this, three pictures were presented on the screen: a target picture shown during the learning stage and two lure pictures resembling the target picture. Participants were tasked with choosing the target picture by pressing the button. Lure trials were included to ensure that participants remembered the visual details of the pictures, and the mental process of feature retrieval was based on the retrieved images. Each feature retrieval and object recognition trial allowed participants up to 10 seconds to make a response. Rest breaks were provided after each block, and the learning-recall task lasted approximately one hour.

Following the learning-recall task, participants took a break before commencing the feature perception task, which also utilized a block design (Figure 1B). Each block initiated with a task cue displayed for 4 seconds, directing participants on which feature to categorize (the target feature). Within each trial, a jittered fixation cross appeared centrally for 1-1.5 seconds, after which a random picture was presented in any one of the eight cells. Participants were required to rapidly and accurately categorize the target feature, pressing the button to provide their answers within 1.5 seconds upon seeing the picture. Once the key response was made, the picture vanished from the screen. If the key response was not submitted within 1.5 seconds

after the picture onset, the picture would also disappear. The combined duration of the picture and the blank nine-cell grid was 2 seconds. Objects of different types occurred with equal probability in each cell across the entire perception task. The entire perception task spanned approximately half an hour.

Throughout both the learning-recall and perception tasks, key responses were counterbalanced across participants. For instance, the left key could be designated for the colorful level in one participant, while the right key could serve this purpose in another participant. However, within participants, the key response assignments remained consistent. This approach aimed to mitigate the potential switching cost of key responses, which might otherwise prolong reaction times in feature perception. Our objective was to obtain accurate reaction times for the feature categorization task and minimize the influence of task-switching costs. All stimuli were presented using PsychoPy (Peirce et al., 2019).

MEG data acquisition

The MEG Data were acquired from a 306-sensor (204 planar gradiometers, 102 magnetometers) Elekta Neuromag MEG system (Helsinki, Finland) at Peking University, Beijing, China. Participants completed the MEG experiments inside a sound-attenuated, dimly lit, and magnetically shielded room. The head position inside the MEG helmet was continuously monitored utilizing four head position indicator (HPI) coils. A 3D digitizer was used to record the location of the HPI coils and the general head shape relative to three anatomical fiducials (nasion, left and right preauricular points). Both horizontal and vertical electrooculograms (EOGs) were recorded. MEG data were recorded at 1000 Hz sampling frequency and bandpass filtered between 0.1 and 330 Hz.

Deep neural network representations

To investigate whether the three features were processed in parallel or some features were processed ahead of others, we used the deep neural network "AlexNet" (Krizhevsky et al., 2017) to solve this question. The AlexNet consists of eight layers, five convolutional layers and three fully connected layers, which simulate the hierarchical structure of neurons along the ventral visual stream. In general, the early layers of the AlexNet process early visual features, including colors, contrasts, and frequencies, while the deeper layers process higher-order visual representations, e.g., the surface structure of objects or body parts of animals. The AlexNet used in this study had been well pretrained via using the ImageNet dataset (Deng et al., 2009). Here, we extracted the features of each picture from each layer of AlexNet, and calculated the similarity between features of every two pictures. The similarity between the same picture was not included. All correlation values were Fisher Z-transformed before further analyses. To obtain the feature-specific representations, we contrasted the between-level similarities versus within-level similarities via t-test and

used the t-value as the sensitivity of the feature-specific representations. For example, the color-specific representations were calculated by contrasting the between-level similarity (the similarity between colorful and gray pictures) versus the within-level similarity (the similarity within colorful pictures or gray pictures). Besides the p-values from the t-test, we also used the permutation test to double-check the statistical significance of the feature-specific representations. In each condition, a null distribution was created through 1000 permutations. The threshold for feature-specific representations was determined by selecting the 95th percentile from this null distribution. During each permutation, the correspondence between the picture labels and representational patterns in each layer was shuffled. Subsequently, the feature-specific representations were computed using the same procedure as for the actual labels.

The drift-diffusion model

To explore the evidence accumulation process of three features (color, animacy and real-world size) during the perception stage, we used the drift-diffusion model to extract the parameters characterizing the evidence accumulation process of features from the rection times. The drift-diffusion models are a widely-used class of models for jointly predicting participants' choices and reaction times (RT) during two-choice decision-making tasks (Ratcliff, 1978; Ratcliff & McKoon, 2008b; Stone, 1960). In the drift-diffusion model, it is assumed that participants accumulate evidence for one choice over another. The rate of evidence accumulation within a trial is called the drift rate (δ) (Figure 4A). The evidence is accumulated toward the upper boundary (e.g., animate) or lower boundary (e.g., inanimate). The upper and lower boundaries correspond to two responses respectively. The boundary separation (α) is the distance between the two decision boundaries. The parameter that encodes the starting position of evidence (β) reflects a measure of response bias towards one of the two response boundaries. The non-decision time (τ) is equal to the amount of time within the reaction time of each trial that is not dedicated to the decision-making process. Typically, non-decision time is assumed to be equal to the sum of preprocessing time before the evidence accumulation process and motor response time after a decision has been reached. The drift-diffusion model was implemented in *brms* package in R (Bürkner, 2017). Only correct trials were used to fit the model. After finishing the model fitting, we extracted the boundary separation, drift rate and non-decision time of each task for each participant.

MEG pre-processing

Raw MEG signals were cleaned with the signal space separation (Taulu & Simola, 2006) method provided by MaxFilter to (1) suppress magnetic interferences and (2) interpolate bad MEG sensors. MEG data were pre-processed using the MNE toolbox (version 1.1) (Gramfort et al., 2013). To remove line noise the data were band-stop filtered at 50 Hz and its harmonics. The muscle artifacts were automatically detected

by MNE function *annotate_muscle_zscore* with default parameters and the annotations of muscle artifacts were then visually inspected by the experimenters. The signal which was annotated as muscle artifacts was not included in the further analysis.

Data recorded during the learning stage was epoched between -500 ms and +3000 ms relative to the picture onset. During the perception stage, data of feature categorization was epoched between -500 ms and +1000 ms relative to the picture onset. The epochs were baseline-corrected based on the pre-stimulus signal (-500 ms to onset). During the retrieval stage, data of the feature retrieval was epoched between -4000 ms and +1000 ms relative to the key response. Since the post-response signal during retrieval might still contain task-relevant (i.e., feature-specific) information, we baseline-corrected the signal based on the whole trial. The data of fixation cross before the feature perception (or retrieval) was epoched between -200 ms and 800 ms relative to the fixation onset. These epochs were baseline-corrected based on the pre-fixation signal (-200 ms to onset). In addition, epochs with excessive noise were discarded utilizing the autoreject toolbox (Jas et al., 2017). Independent component analysis was then used to remove eye-blink, horizontal eye movement and electrocardiogram (ECG) artifacts. To increase the signal-to-noise ratio (SNR) for multivariate decoding, the pre-processed MEG time courses were smoothed using a Gaussian kernel with a full width at a half-maximum of 10 ms and down-sampled to 200 Hz.

Time-resolved multivariate decoding

Before conducting the decoding analysis, we used a smooth window with 50 ms length and 10 ms step and averaged signals in each window to further increase the SNR. Our primary emphasis was on investigating the selective attention modulation effect on feature representations during the perception and retrieval stages. The learning task functioned as a functional localizer task, enabling the training of independent feature classifiers without any confounding effects related to motion. For each participant, four distinct classifiers were trained at each time point and ROI, utilizing data from the learning stage. These classifiers included one for color category classification (colorful or gray), one for animacy category classification (animate or inanimate), and two for real-world size category classification—one for animate objects and another for inanimate objects. The latter two classifiers distinguished between big and small real-world size categories (Figure 5A). The reason for classifying the real-world size category separately for the animate and inanimate objects was that research found real-world size drives differential responses only in the object domain, not the animate domain (Konkle & Caramazza, 2013), so we assumed that the representational patterns of real-world size for animate and inanimate objects might be different. The feature perception epochs of the same object were averaged. For both the epochs of feature perception and retrieval, the pre-processed amplitudes of channels in each ROI at each time point were whitened and then used as features for classifiers. The whitening process used a standard scaler that

z-scored each channel at each time point across trials. Then the classification using logistic regression models was performed separately for each time point, ROI and participant, with a stratified 8-fold cross-validation approach. The logistic regression model parameters were set to their default values as provided by the Scikit-learn package (King et al., 2016), setting the penalty term as L2 regularization, tolerance for stopping criteria as 1e-4, the inverse of regularization strength C as 1, algorithm to use in the optimization problem as 'lbfgs' solver (Limited-memory Broyden–Fletcher–Goldfarb–Shanno). The decoding performance during the learning stage was indicated by the Area Under the Receiver Operating Characteristic Curve (ROC_AUC). We used cluster-based permutation analyses (Maris & Oostenveld, 2007), which intrinsically correct for multiple comparison issues, to test the significance of feature representations in each ROI.

After obtaining the significant cluster of each feature in each ROI at the group level (averaged across participants) during the learning stage, the time point of peak ROC_AUC in the significant cluster and its surrounding four time points (e.g., if the peak time point was the 78th time point, then its surrounding four time points were the 76th, 77th, 79th and 80th time point) were considered as the most informative time window (i.e., peak bin) carrying feature information. The averaged and standardized amplitudes in each peak bin were used to train the new feature classifier which was then tested at each time point in each task during the perception and retrieval stage to detect the feature information without any impact of motor confounding, for example, the animacy classifier trained during the learning stage in each ROI was tested in the three tasks (the color, animacy and size task) respectively during the perception and retrieval stage in the corresponding ROI. The cross-validation is not used for across-stage analyses as no inferences are made based on the decoding performance in the learning stage itself.

To detect the onset and peak time of feature representations during the perception stage, following Linde-Domingo et al., 2019, we conducted the classification analyses at a single trial level. As it was emphasized in Linde-Domingo et al., 2019, many factors could obscure differences between times of different conditions when average latency across trials for each condition, such as variance in processing speed across trials. Therefore, it was believed that the single-trial-based method was more sensitive to differences in timing between conditions. So we test the trained classifiers at each time point of each trial. As a result, for each participant and each trial, there were 4 (ROIs) * 4 (features) = 16 decoding performance time courses in total (4 ROIs: the occipital, temporal, parietal and frontal cortex; 4 features: the features of color for all objects, animacy for all objects, size for inanimate and animate objects respectively). Different from Linde-Domingo et al., 2019, we not only used the peak latency of significant clusters of each trial as the time index to measure the time lag between different features, but also used the onset (i.e., the first time point) of the first significant cluster of each trial. We posit that the onset time signifies the initiation of feature representations, capturing the very inception of when these features begin to

manifest in our brain. Detecting this initial time point is crucial for understanding when these features first come into representation. On the other hand, the peak latency signifies the time at which a feature is represented with the utmost fidelity. In this context, the onset time allows us to discern whether tasks induce the earlier detection of target features compared to non-target features in our brain. Meanwhile, the peak time helps us understand whether target features are represented with highest fidelity compared to non-target features.

The decision value (i.e., d-value) is an appropriate index to characterize the decoding performance at the single-trial level (Linde-Domingo et al., 2019; Ritchie et al., 2015b). It should be noticed that the d-value has negative and positive signs. The sign indicates in which feature level the observation had been classified (e.g., - for animate and + for inanimate). The absolute value of d-value at each time point is the distance to the hyper-plane that divided the two levels (e.g., animate and inanimate), with the hyper-plane being 0 (Figure 5A), indicating how confident the classifier was at assigning a given object to a given feature level at this time point. Regarding getting significant clusters carrying feature information for every single trial, the cluster-based permutation analysis based on multiple trials doesn't work in this case, so we compared d-values at each time point in each trial with the 95th percentile of the null distribution (the threshold) of the corresponding time point, and the time window of at least T ms in which all the time points surpassed the corresponding thresholds was considered a significant cluster (details about null distributions are in the Generating empirical null distributions for classifiers section).

Generating empirical null distributions for classifiers

To detect the significant clusters of feature representations on the single-trial level, following the method in Linde-Domingo et al., 2019, we used permutation and bootstrapping to generate null distributions for each time point of each feature level in each task type, then chose the 95th percentile of each null distribution as the threshold for the corresponding time point. In each trial, the time window of at least T ms in which all the time points must surpass the corresponding thresholds was considered a significant cluster, and the first time point of the first significant cluster was considered the earliest time point when the feature was represented in this trial, i.e., the onset time. The time point with the highest d-value across all significant clusters in each trial was considered the time point when the feature was represented with the highest fidelity, i.e., the peak time.

For the permutation analyses, we randomly shuffled the feature level labels (e.g., shuffled the colorful and gray labels) of trials during the learning stage at each permutation and carried out the same training procedure as for real data. The classifiers were then tested on the data of three tasks during the perception and retrieval stage using the test procedures as for real data. This procedure was conducted 150 times independently per participant. For each participant, ROI, task and feature type, there was a total of 151 classification outputs for each trial of the

perception and retrieval stage respectively, one using the real labels and the remaining using the randomly shuffled labels when training the classifiers. The real data was included to make our next bootstrapping analyses more conservative since under the null hypothesis, the real classifier output could have been obtained just by chance. Then the bootstrapping approach was used to estimate the classification chance distribution. For each ROI, task and feature type, we randomly select one of the 151 classification outputs per participant in each bootstrapped repetition. Trials in this classification output were divided into two groups based on their feature levels, and the d-value time courses of each feature level were averaged across trials. As a result, for each ROI, task and feature type in each bootstrapped repetition, each participant had two d-value time courses, one feature level each (e.g., one for the animate level and the other for the inanimate level). Then for each feature level, the d-value time courses were averaged across participants. This procedure was repeated with replacement 10000 times. Therefore, for each feature level, each time point had a null distribution consisting of 10000 values. We chose the 95th percentile of each null distribution as the threshold for the corresponding time point. In each trial, the time window of at least T ms in which all the time points surpassed the corresponding thresholds was considered a significant cluster. The minimum duration (T) of a significant cluster was tested from 20 ms to 100 ms with a step size of 10 ms, to see the impact of T on the detected onset times and the relative time lag among different features in each task. In general, when T was set to 20, 30, and 40 ms, the onset times of all features were within 100 ms, which was consistent with previous research (Teichmann et al., 2019, 2020, 2021; Cichy et al., 2014; Cichy & Pantazis, 2017; Ritchie et al., 2015). When T was adjusted to more than 40 ms, the onset times of features exceeded 100 ms. As T increased, the onset times were delayed and the patterns of the time lags among features remained although obscured (examples in Figure S5). After a thorough comparison of the onset times in our results with those in previous research, we determined that a time window of at least 20 ms, where all time points exceeded the corresponding threshold, was considered a significant cluster containing feature information. In each trial, the first time point of the first significant cluster was considered the earliest time point at which the feature was represented. The time point with the highest d-value across all significant clusters in each trial was considered the peak time point. It should be noted that the peak time within the reaction time of each trial was considered effective and used for further analysis. Ultimately, for each participant, ROI, and task type, we obtained onset and peak times for color, animacy, and size features, respectively, for each trial.

GLMM analyses

Following Linde-Domingo et al., 2019, we employed generalized linear mixed models (GLMMs) to examine our hypotheses regarding the relative onset or peak time lags of d-values among different features within the same task during both the perception and retrieval stages. GLMMs were chosen instead of more commonly used GLM-based models (i.e., ANOVAs or t-tests) because they make fewer assumptions about the distribution of the data and are better suited to model d-value onsets. The

task type (color, animacy and size tasks), feature type (color, animacy and size features) and their interactions were modeled as fixed effects in the GLMM. Through the interaction between the factors of task and feature type, we assessed whether selective attention had a modulatory effect on the onset or peak times of different features. Participant ID (including the intercept) was modeled as a random factor. All models for analyzing d-value onset or peak times used a gamma probability distribution and an identity link function as Linde-Domingo et al., 2019 used.

Phase coupling estimation

If selective attention can modulate the times of feature representations, how is this process achieved? We hypothesized that during the fixation period (i.e., the task preparation stage) when stimuli (or spatial cues) were not presented, the brain networks were configured to be prepared for the appropriate execution of the instructed task. The stronger the connectivity between brain regions, the more prepared the configured brain networks are, and the faster the target features might be processed. We used the phase slope index (PSI) to characterize the connectivity patterns between the occipital and the other three higher-order regions (the temporal, parietal, and frontal lobes), and between the temporal and frontal lobes (a control analysis to contrast with the retrieval task) during the preparation stage in the feature perception task. For the preparation stage in the feature retrieval task, we calculated the PSI between the temporal and frontal lobes. The PSI is a method to estimate the phase coupling patterns between brain regions and to detect the direction of frequency-specific neural interactions from time series, that is, which one of the two phase-coupled brain areas is the 'leader' (Nolte et al., 2008). Compared to the Granger causality, the phase slope index is insensitive to mixtures of independent sources, thus would not yields highly significant false detections for mixtures of independent sources (Nolte et al., 2008). The PSI between two time series for channels *i* and *j* is defined as

$$\tilde{\Psi}\, ij = \Im\left(\sum_{f \in F} C_{ij}^{*}(f) C_{ij}(f + \delta f)\right),$$

where $C_{ij}(f) = S_{ij}(f) / \sqrt{S_{ii}(f) S_{jj}(f)}$ is the complex coherency, *S* is the cross-spectral matrix, δf is the frequency resolution, and $\Im(\cdot)$ denotes taking the imaginary part. *F* is the set of frequencies over which the slope is summed. We used the continuous Morlet wavelet transform to estimate the cross-spectral matrix. Before calculating PSI, to reduce the computation time and increase the SNR, the principal component analysis (PCA), a.k.a. spatial filters, was used to transform the channel data to new sources/virtual channels, for each participant and ROI. The PCA analysis was conducted for the preparation stage in the feature perception and retrieval task separately. The representative time courses in each ROI were defined as the smallest set of signals that explained at least 90% of the raw amplitude of the pre-processed MEG time courses.

PSI values were calculated for all pairs of virtual channels of each pair of ROIs with

frequency resolution 1 Hz for each frequency band (theta: 4-8 Hz; alpha: 9-13 Hz; beta: 14-30 Hz; low gamma1: 31-40 Hz; low gamma2: 41-60 Hz; high gamma: 61-100 Hz). In each pair of ROIs, the higher-order ROIs were set as seed ROIs and the occipital lobe was set as target ROI. For example, participant 1 had n and m virtual channels in the temporal (the seed ROI) and occipital (the target ROI) lobe respectively. The shape of the obtained PSI values for the temporal-occipital pair would be n*m for each time point. In addition, the sign of PSI values indicates the information direction. For example, when the temporal and occipital cortex were set as the seed and target ROI respectively, a positive PSI value means the information direction is from the temporal to the occipital lobe, while a negative PSI value indicates the opposite direction. We employed cluster-based permutation analyses to investigate in which time ranges the PSI values of each task, frequency band, and pair of ROIs differed significantly from zero. This indicated the presence of phase coupling between ROIs.

After identifying the significant clusters of PSI on the group level for the preparation stage in the feature perception task, we extracted the PSI values from the corresponding time range of significant clusters for each participant and averaged them across the time points in each cluster, respectively. To be noticed, when the PSI values of a significant cluster were negative, they were converted to positive, with larger values indicating stronger phase coupling. If there were more than one significant cluster in one condition, the PSI values were averaged across clusters. We then correlated (Pearson's correlation) the averaged PSI values and the onset times of features in the target ROI to investigate whether stronger phase coupling between brain regions during the task preparation stage would lead to faster processing of target instead of non-target features during the perception stage. For each participant and ROI, the onset time of target features was calculated as the average of the onset times of the color feature in the color task, animacy feature in the animacy task, and size feature in the size task. The onset time of non-target features was calculated as the average of the onset times of the color feature in the animacy and size tasks, the animacy feature in the color and size tasks, and the size feature in the color and animacy tasks.

References

1. Antzoulatos, E. G., & Miller, E. K. (2016). Synchronous beta rhythms of frontoparietal networks support only behaviorally relevant representations. *eLife*, *5*, e17822. https://doi.org/10.7554/eLife.17822
2. Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. *Science*, *344*(6182), 424–427. https://doi.org/10.1126/science.1247003
3. Bone, M. B., Ahmad, F., & Buchsbaum, B. R. (2020). Feature-specific neural reactivation during episodic memory. *Nature Communications*, *11*(1), 1945. https://doi.org/10.1038/s41467-020-15763-2
4. Bouchacourt, F., & Buschman, T. J. (2019). A Flexible Model of Working Memory. *Neuron*, *103*(1), 147-160.e8. https://doi.org/10.1016/j.neuron.2019.04.020
5. Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, *7*(8), 880–886. https://doi.org/10.1038/nn1278
6. Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01
7. Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., & Miller, E. K. (2012). Synchronous Oscillatory Neural Ensembles for Rules in the Prefrontal Cortex. *Neuron*, *76*(4), 838–846. https://doi.org/10.1016/j.neuron.2012.09.029
8. Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144. https://doi.org/10.1016/j.neuron.2015.09.017
9. Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–1. https://doi.org/10.1167/13.10.1
10. Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal Output Gating during Selection from Working Memory. *Neuron*, *81*(4), 930–942. https://doi.org/10.1016/j.neuron.2014.01.002
11. Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, *74*(5), 784–802. https://doi.org/10.3758/s13414-012-0322-z
12. Cichy, R. M., & Pantazis, D. (2017). Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *NeuroImage*, *158*(July), 441–454. https://doi.org/10.1016/j.neuroimage.2017.07.023
13. Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635
14. Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From Perception to Conception: How Meaningful Objects Are Processed over Time. *Cerebral Cortex*, *23*(1), 187–197. https://doi.org/10.1093/cercor/bhs002
15. Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends in Cognitive Sciences*, *19*(11), 677–687. https://doi.org/10.1016/j.tics.2015.08.008
16. Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755
17. Córdova, N. I., Tompary, A., & Turk-Browne, N. B. (2016). Attentional modulation of

background connectivity between ventral visual cortex and the medial temporal lobe. *Neurobiology of Learning and Memory*, *134*, 115–122. https://doi.org/10.1016/j.nlm.2016.06.011

18. Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848
19. Desimone, R., & Duncan, J. (1995). NEURAL MECHANISMS OF SELECTIVE VISUAL ATTENTION. *Annu. Rev. Neurosci*, *18*, 193–222.
20. Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. https://doi.org/10.1016/j.tics.2008.01.001
21. Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790. https://doi.org/10.1038/nn1594
22. Ester, E. F., Sutterer, D. W., Serences, J. T., & Awh, E. (2016). Feature-Selective Attentional Modulations in Human Frontoparietal Cortex. *Journal of Neuroscience*, *36*(31), 8188–8199. https://doi.org/10.1523/JNEUROSCI.3935-15.2016
23. Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1
24. Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'What' Is Happening in the Dorsal Visual Pathway. *Trends in Cognitive Sciences*, *20*(10), 773–784. https://doi.org/10.1016/j.tics.2016.08.003
25. Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional Interactions Between Inferotemporal and Prefrontal Cortex in a Cognitive Task. *Brain Research*, *330*(1985), 299–307.
26. Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135. https://doi.org/10.1016/j.tics.2011.11.014
27. Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., & Luppino, G. (2010). Cortical Connections of the Macaque Caudal Ventrolateral Prefrontal Areas 45A and 45B. *Cerebral Cortex*, *20*(1), 141–168. https://doi.org/10.1093/cercor/bhp087
28. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(7 DEC), 1–13. https://doi.org/10.3389/fnins.2013.00267
29. Grill-Spector, K., & Malach, R. (2004). THE HUMAN VISUAL CORTEX. *Annual Review of Neuroscience*, *27*(1), 649–677. https://doi.org/10.1146/annurev.neuro.27.070203.144220
30. Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2021). The neural dynamics underlying prioritisation of task-relevant information. *Neurons, Behavior, Data Analysis, and Theory*, *5*(1). https://doi.org/10.51628/001c.21174
31. Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*(December 2016), 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030
32. Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-

Relevant Features in the Human Visual Cortex. *The Journal of Neuroscience*, *31*(22), 8210–8219. https://doi.org/10.1523/JNEUROSCI.6153-09.2011

33. Khaligh-Razavi, S.-M., Cichy, R. M., Pantazis, D., & Oliva, A. (2018). Tracking the Spatiotemporal Neural Dynamics of Real-world Object Size and Animacy in the Human Brain. *Journal of Cognitive Neuroscience*, *30*(11), 1559–1576. https://doi.org/10.1162/jocn_a_01290
34. King, J. R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134. https://doi.org/10.1016/j.neuron.2016.10.051
35. Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013
36. Kourtzi, Z., & Connor, C. E. (2011). Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. *Annual Review of Neuroscience*, *34*(1), 45–67. https://doi.org/10.1146/annurev-neuro-060909-153218
37. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011
38. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386
39. Lehky, S. R., & Tanaka, K. (2016). Neural representation for object recognition in inferotemporal cortex. *Current Opinion in Neurobiology*, *37*, 23–35. https://doi.org/10.1016/j.conb.2015.12.001
40. Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), 179. https://doi.org/10.1038/s41467-018-08080-2
41. Liu, J., Zhang, H., Yu, T., Ren, L., Ni, D., Yang, Q., Lu, B., Zhang, L., Axmacher, N., & Xue, G. (2021). Transformative neural representations support long-term episodic memory. *Science Advances*, *7*(41), eabg9715. https://doi.org/10.1126/sciadv.abg9715
42. Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. (2019). Human Replay Spontaneously Reorganizes Experience. *Cell*, *178*(3), 640-652.e14. https://doi.org/10.1016/j.cell.2019.06.012
43. Long, N. M., & Kuhl, B. A. (2018). Bottom-Up and Top-Down Factors Differentially Influence Stimulus Representations Across Large-Scale Attentional Networks. *The Journal of Neuroscience*, *38*(10), 2495–2504. https://doi.org/10.1523/JNEUROSCI.2724-17.2018
44. Mack, M. L., & Palmeri, T. J. (2011). The Timing of Visual Object Categorization. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00165
45. Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742
46. Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

47. Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife*, *7*, e31873. https://doi.org/10.7554/eLife.31873
48. Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001
49. Mirjalili, S., Powell, P., Strunk, J., James, T., & Duarte, A. (2021). Context Memory Encoding and Retrieval Temporal Dynamics are Modulated by Attention across the Adult Lifespan. *Eneuro*, *8*(1), ENEURO.0387-20.2020. https://doi.org/10.1523/ENEURO.0387-20.2020
50. Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention Selectively Reshapes the Geometry of Distributed Semantic Representation. *Cerebral Cortex*, *27*(8), 4277–4291. https://doi.org/10.1093/cercor/bhx138
51. Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward Spreading of Memory-Retrieval Signal in the Primate Temporal Cortex. *Science*, *291*(5504), 661–664. https://doi.org/10.1126/science.291.5504.661
52. Nee, D. E., & Jonides, J. (2008). Neural correlates of access to short-term memory. *Proceedings of the National Academy of Sciences*, *105*(37), 14228–14233. https://doi.org/10.1073/pnas.0802081105
53. Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting Attention to Locations in Perceptual Versus Mental Representations. *Journal of Cognitive Neuroscience*, *16*(3), 363–373. https://doi.org/10.1162/089892904322926700
54. Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T., & Müller, K.-R. (2008). Robustly Estimating the Flow Direction of Information in Complex Physical Systems. *Physical Review Letters*, *100*(23), 234101. https://doi.org/10.1103/PhysRevLett.100.234101
55. Nosofsky, R. M., & Palmeri, T. J. (1997). An Exemplar-Based Random Walk Model of Speeded Classification. *Psychological Review*, 266–300. https://doi.org/10.1037/0033-295X.104.2.266
56. Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y
57. Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*(1), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525
58. Ramezanpour, H., & Fallah, M. (2022). The role of temporal cortex in the control of attention. *Current Research in Neurobiology*, *3*, 100038. https://doi.org/10.1016/j.crneur.2022.100038
59. Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59
60. Ratcliff, R., & McKoon, G. (2008a). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

61. Ratcliff, R., & McKoon, G. (2008b). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420
62. Richter, C. G., Coppola, R., & Bressler, S. L. (2018). Top-down beta oscillatory signaling conveys behavioral context in early visual cortex. *Scientific Reports*, *8*(1), 6991. https://doi.org/10.1038/s41598-018-25267-1
63. Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015a). Emerging Object Representations in the Visual System Predict Reaction Times for Categorization. *PLOS Computational Biology*, *11*(6), e1004316. https://doi.org/10.1371/journal.pcbi.1004316
64. Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015b). Emerging Object Representations in the Visual System Predict Reaction Times for Categorization. *PLoS Computational Biology*, *11*(6), 1–19. https://doi.org/10.1371/journal.pcbi.1004316
65. Sakai, K. (2008). Task Set and Prefrontal Cortex. *Annual Review of Neuroscience*, *31*(1), 219–245. https://doi.org/10.1146/annurev.neuro.31.060407.125642
66. Saleem, K. S., Miller, B., & Price, J. L. (2014). Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *522*(7), 1641–1690. https://doi.org/10.1002/cne.23498
67. Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1–2), 1–46. https://doi.org/10.1016/S0010-0277(00)00152-9
68. Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45. https://doi.org/10.1016/j.tics.2005.11.008
69. Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104
70. Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A. K. (2008). Neuronal Synchronization along the Dorsal Visual Pathway Reflects the Focus of Spatial Attention. *Neuron*, *60*(4), 709–719. https://doi.org/10.1016/j.neuron.2008.09.010
71. Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. https://doi.org/10.1038/nn.3574
72. Stanley, D. A., Roy, J. E., Aoi, M. C., Kopell, N. J., & Miller, E. K. (2018). Low-Beta Oscillations Turn Up the Gain During Category Judgments. *Cerebral Cortex*, *28*(1), 116–130. https://doi.org/10.1093/cercor/bhw356
73. Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260. https://doi.org/10.1007/BF02289729
74. Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology*, *51*(7), 1759–1768. https://doi.org/10.1088/0031-9155/51/7/008
75. Teichmann, L., Grootswagers, T., Carlson, T. A., & Rich, A. N. (2019). Seeing versus knowing: The temporal dynamics of real and implied colour processing in the human brain. *NeuroImage*, *200*(October 2018), 373–381. https://doi.org/10.1016/j.neuroimage.2019.06.062
76. Teichmann, L., Grootswagers, T., Moerel, D., Carlson, T. A., & Rich, A. N. (2021). Temporal dissociation of neural activity underlying synesthetic and perceptual colors. *Proceedings of*

the National Academy of Sciences of the United States of America, *118*(6), 2–4. https://doi.org/10.1073/pnas.2020434118

77. Teichmann, L., Quek, G. L., Robinson, A. K., Grootswagers, T., Carlson, T. A., & Rich, A. N. (2020). Yellow strawberries and red bananas: The influence of object-colour knowledge on emerging object representations in the brain. *The Journal of Neuroscience*, *40*(35), 6779–6789. https://doi.org/10.1101/533513
78. Tijl Grootswagers, Susan G. Wardle, and T. A. C. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, 1–10. https://doi.org/10.1162/jocn
79. Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*(6754), 699–703. https://doi.org/10.1038/44372
80. Wang, R., Janini, D., & Konkle, T. (2022a). Mid-level Feature Differences Support Early Animacy and Object Size Distinctions: Evidence from Electroencephalography Decoding. *Journal of Cognitive Neuroscience*, *34*(9), 1670–1680. https://doi.org/10.1162/jocn_a_01883
81. Wang, R., Janini, D., & Konkle, T. (2022b). Mid-level feature differences underlie early animacy and object size distinctions: Evidence from EEG decoding. *bioRxiv*, 2022.01.12.475180. https://doi.org/10.1101/2022.01.12.475180
82. Wimmer, G. E., Liu, Y., Vehar, N., Behrens, T. E. J., & Dolan, R. J. (2020). Episodic memory retrieval success is associated with rapid replay of episode content. *Nature Neuroscience*, *23*(8), 1025–1033. https://doi.org/10.1038/s41593-020-0649-z
83. Wutz, A., Melcher, D., & Samaha, J. (2018). Frequency modulation of neural oscillations according to visual task demands. *Proceedings of the National Academy of Sciences*, *115*(6), 1346–1351. https://doi.org/10.1073/pnas.1713318115
84. Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. *The Journal of Neuroscience*, *37*(11), 2986–2998. https://doi.org/10.1523/JNEUROSCI.2324-16.2017
85. Xue, G. (2018). The Neural Representations Underlying Human Episodic Memory. *Trends in Cognitive Sciences*, *22*(6), 544–561. https://doi.org/10.1016/j.tics.2018.03.004
86. Yau, J. M., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2013). Curvature Processing Dynamics in Macaque Area V4. *Cerebral Cortex*, *23*(1), 198–209. https://doi.org/10.1093/cercor/bhs004
87. Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*, *106*.
88. Zhou, T., Kawasaki, K., Suzuki, T., Hasegawa, I., Roe, A. W., & Tanigawa, H. (2023). Mapping information flow between the inferotemporal and prefrontal cortices via neural oscillations in memory retrieval and maintenance. *Cell Reports*, *42*(10), 113169. https://doi.org/10.1016/j.celrep.2023.113169

Supplementary materials

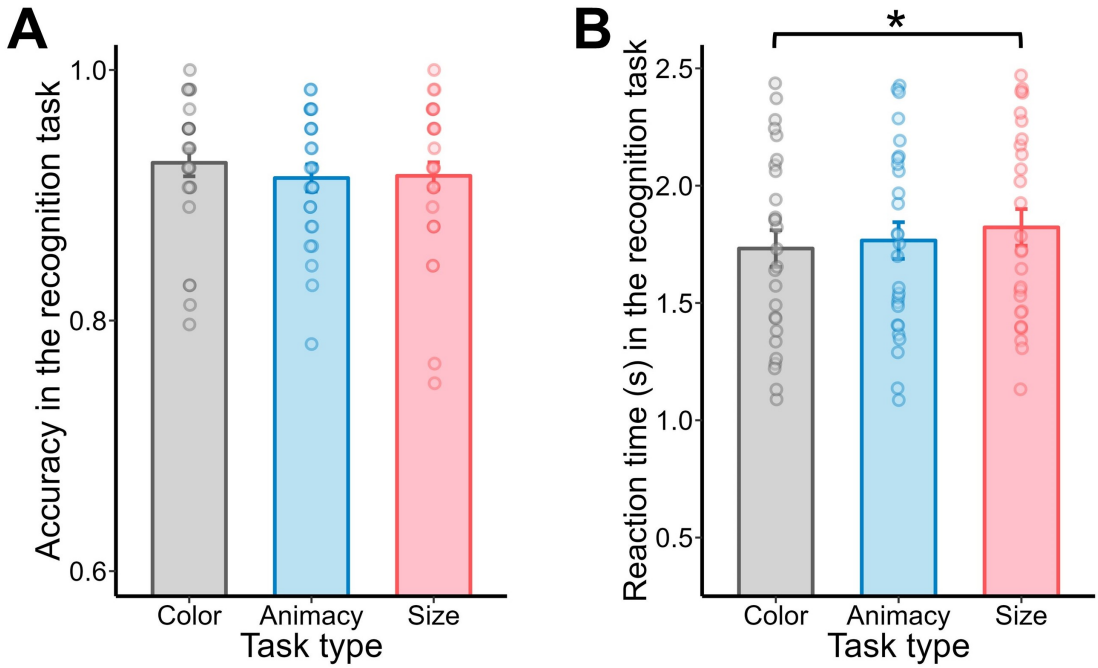

Figure S1. The behavioral results in the recognition task. A. Accuracy. The mean accuracy of the object recognition under the color, animacy and size task conditions was 92.59% (*SD* = 5.44%), 91.38% (*SD* = 5.12%) and 91.55% (*SD* = 6.21%). The task type didn't predict the accuracy of object recognition under three task conditions ($F_{2, 52} = 1.181$, $p = 0.315$). **B. Reaction time.** The mean reaction times of the object recognition under the color, animacy and size task conditions were 1.732 s (*SD* = 0.403 s), 1.766 s (*SD* = 0.402 s), and 1.822 s (*SD* = 0.412 s). The task type predicted the reaction times ($F_{2, 52} = 4.255$, $p = 0.019$). Post-hoc testing showed that the reaction time of the object recognition under the color task condition was significantly faster than that under the size task condition ($t_{52} = -2.889$, $p_{corrected} = 0.015$), while the difference between the animacy task condition and the other two task conditions was not significant (animacy condition vs. color condition: $t_{52} = 1.097$, $p_{corrected} = 0.520$; animacy condition vs. size condition: $t_{52} = -1.793$, $p_{corrected} = 0.182$).

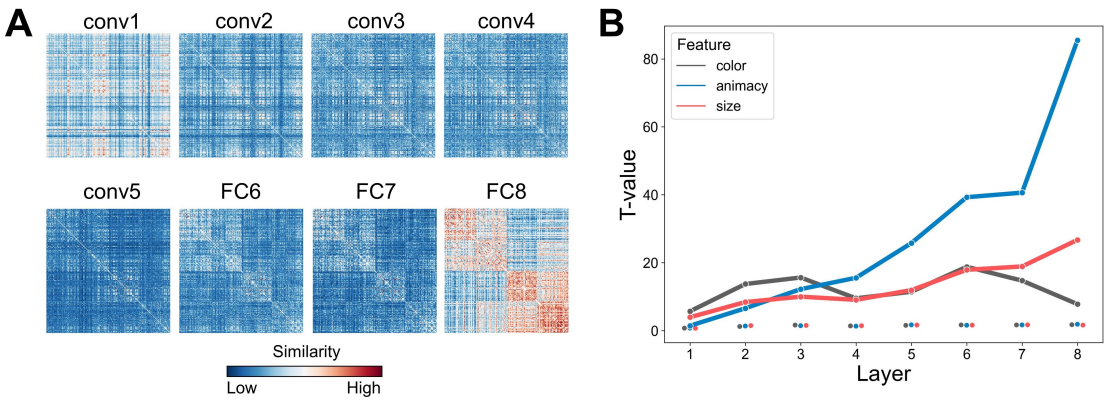

Figure S2. Representations of three features (color, animacy and size) in the deep neural

network AlexNet (stimulus set 2). A. Representational similarity matrix. The matrix was generated by correlating the activations of the artificial neurons in each DNN layer. The diagonal in each matrix was set to NaN. **B. Sensitivity of feature information.** The threshold of feature-specific representations in each layer was indicated by the dots which were not connected by lines. This figure was generated by the data of the stimulus set 2. conv: convolutional layer; FC: fully connected layer.

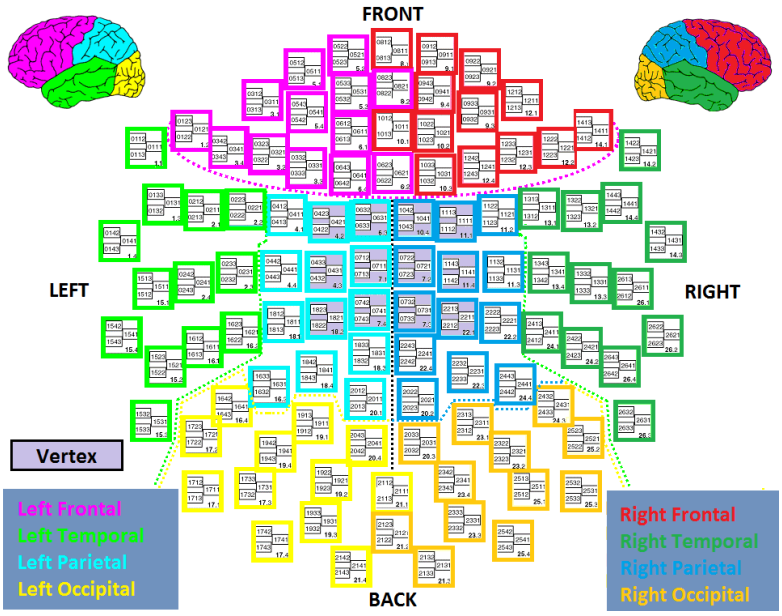

Figure S3. Elekta Neuromag MEG channel positions. Channels corresponding to different lobes are color-coded (figure adapted from www.megwiki.org)

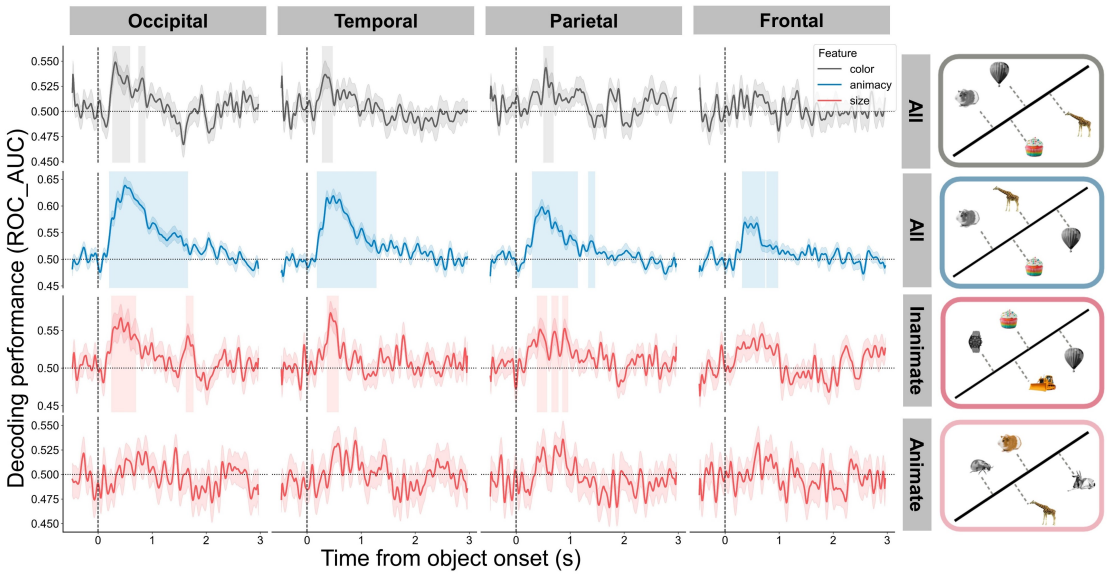

Figure S4. The decoding performance of three features during the learning stage. Significant clusters of color, animacy and size features in the occipital, temporal and parietal cortex were found ($P_{corrected} < 0.05$). In the frontal cortex, only significant clusters of the animacy feature were

found ($P_{corrected} < 0.05$). There were no significant clusters of size feature for the animate objects. The shaded areas surrounding the classification performance time courses indicated standard error across participants. The vertical shaded areas marked the significant clusters carrying the feature information.

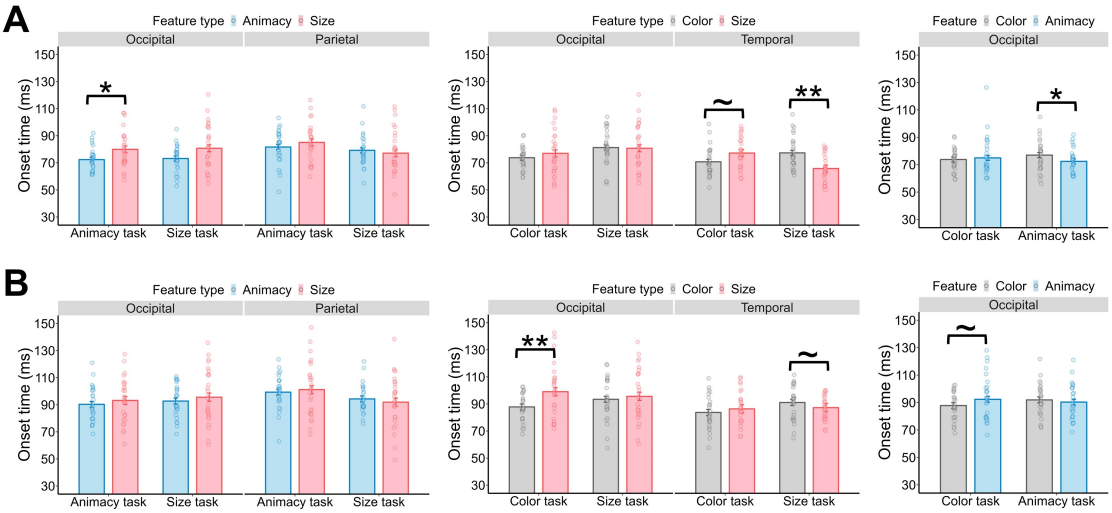

Figure S5. The modulation effects of selective attention when setting T=30 ms (A) or 40 ms (B), i.e., the time window of at least 30 ms or 40 ms in which the d-values of all time points exceeding the corresponding threshold was considered as a significant cluster containing feature information. As shown in the figures, when the minimum duration (T) of a significant cluster increased, the overall onset times of features were delayed accordingly, and the patterns of time lags between different features remained remained discernible, although some of them became obscured. Error bars reflect within-subject SEM (the standard error of the mean).