Dual-format attentional template during preparation in human visual cortex

Yilin Chen[1], Taosheng Liu[2], Ke Jia[3,4,5*], Jan Theeuwes[6], Mengyuan Gong[1*]

[1] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, Zhejiang, China

[2] Department of Psychology, Michigan State University, East Lansing, Michigan, USA

[3] Liangzhu Laboratory, MOE Frontier Science Center for Brain Science and Brain-machine Integration, State Key Laboratory of Brain-machine Intelligence, Zhejiang University, Hangzhou, Zhejiang, China

[4] Department of Neurobiology, Affiliated Mental Health Center & Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[5] NHC and CAMS Key Laboratory of Medical Neurobiology, Zhejiang University, Hangzhou, Zhejiang, China

[6] Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

***Corresponding Author**

Dr. Mengyuan Gong

Department of Psychology and Behavioral Sciences, Zhejiang University

gongmy426@zju.edu.cn

Dr. Ke Jia

MOE Frontier Science Center for Brain Science and Brain-machine Integration, Zhejiang University

kjia@zju.edu.cn

Abstract

Goal-directed attention relies on forming internal templates of key information relevant for guiding behavior, particularly when preparing for upcoming sensory inputs. However, evidence on how these attentional templates is represented during preparation remains controversial. Here, we combine functional magnetic resonance imaging (fMRI) with an orientation cueing task to isolate preparatory activity from stimulus-evoked responses. Using multivariate pattern analysis, we found decodable information of the to-be-attended orientation during preparation; yet preparatory activity patterns were different from those evoked when actual orientations were perceived. When perturbing the neural activity by means of a visual impulse ('pinging' technique), the preparatory activity patterns in visual cortex resembled those associated with perceiving these orientations. The observed differential patterns with and without the impulse perturbation suggest a predominantly non-sensory format and a latent, sensory-like format of representation during preparation. Furthermore, the emergence of the sensory-like template coincided with enhanced information connectivity between V1 and frontoparietal areas and was associated with improved behavioral performance. By engaging this dual-format mechanism during preparation, the brain is able to encode both abstract, non-sensory information and more detailed, sensory information, potentially providing advantages for adaptive attentional control. For example, consistent with recent theories of visual search, a predominantly non-sensory template can support the initial guidance and a latent sensory-like format can support prospective stimulus processing.

Keywords

Attentional selection, representational format, fMRI, decoding, visual cortex, impulse perturbation

Introduction

To address the challenge of processing the overwhelming amounts of sensory inputs from the external environment, the brain must allocate attentional resources to prioritize the processing of task-relevant information. Importantly, humans can proactively prepare for stimulus selection before the arrival of sensory inputs (Summerfield & De Lange, 2014). For example, when preparing to hail a taxi on the road, we tend to form a mental representation of the defining features of a taxi (e.g., yellow with a car-like shape). This ability relies on the formation of attentional templates — mental representations of the target — to accelerate stimulus selection and resolve perceptual competition by enhancing task-relevant information and suppressing irrelevant information (Desimone & Duncan, 1995; Kastner et al., 1999). While most attentional models posit that attentional templates during stimulus processing reflect veridical representations of the target (Jigo et al., 2018; Malcolm & Henderson, 2009), the nature of the template during preparation remains less understood.

A classical view suggests that attentional template during preparation may reflect veridical target features, analogous to the representational format during stimulus selection. However, evidence supporting this account has been mixed. For example, while some previous fMRI studies have demonstrated that preparatory activity contains target information similar to the sensory responses to the corresponding targets (Kok et al., 2014; Lewis-Peacock et al., 2015; Stokes et al., 2009), more recent electrophysiological studies suggest that, if anything, this template is engaged only shortly before the expected arrival of sensory input rather than being continuously active (Grubert & Eimer, 2018; Myers et al., 2015). Notably, in some cases the template is even largely undetectable during preparation (Wen et al., 2019). Alternatively, an emerging view suggests a non-veridical template suffices for guiding attention during preparation, where precise processing of stimuli may be unnecessary at this stage. Support for this notion comes from the identification of attentional signals during preparation that differ from neural signals observed during perceptual target processing (Gong et al., 2022). Recent theories of visual search also propose a non-veridical, "good-enough"

template for early attentional guidance (Wolfe, 2021; Yu et al., 2023). However, it remains unclear whether a "good-enough" template for search also applies to preparatory attention.

The notion that there may be a sensory and non-sensory attentional template might not be as far-fetched as it seems. Indeed, it is feasible that during preparation, following stimulus presentation, attentional signals undergo a transformation from a non-sensory to a sensory-like template. Previous behavioral (Hamblin-Frohman & Becker, 2021; Yu et al., 2022) and neural studies (Gong et al., 2022; Jigo et al., 2018; Wen et al., 2019) are generally consistent with this idea of coarse-to-fine transitions, suggesting that during preparation, a sensory-like template may not be initially necessary but only becomes relevant when the stimulus needs to be identified. However, if and in what way the brain coordinates these non-sensory and sensory-like templates remains unclear. Here, we propose that during preparation, a sensory-like template may be stored in a latent (e.g., activity-silent) state concurrently with a non-sensory template. This idea parallels recent findings from working memory studies, which suggest that information intended for proactive use is kept in activity-silent traces to support future behavior (Stokes, 2015; Wolff et al., 2015, 2017). The present study seeks to determine the possibility of the latent, sensory-like template during the preparation for discriminating an upcoming stimulus.

To test these hypotheses, participants engaged in a cuing task in which they prepared during an extended period of time for the presentation of a compound stimulus grating containing the cued orientation and a distractor orientation. In addition, in order to be able to construct the sensory-format representations (leftward and rightward orientation), single orientations were presented during the perception task. Critically, we used a "pinging" technique combined with multivariate decoding methods, which has been shown to be effective in retrieving information from latent brain states (Duncan et al., 2023; Wolff et al., 2015, 2017; Zhang & Luo, 2023). In the standard condition (*No-Ping session*), the preparation period was devoid of visual impulses. During preparation, the neural activity patterns in visual and frontoparietal areas could discriminate between the orientations that participants were preparing for. Yet, neural activity patterns evoked by the perception of orientations were distinct from those evoked during preparation for upcoming orientations, suggesting a predominantly non-sensory template during preparation. By contrast,

when we presented a high-contrast, task-irrelevant impulse stimulus during preparation (*Ping session*), neural activity patterns evoked by the perception of orientations were similar to those activated by the preparation for orientations in the visual cortex, suggesting the existence of a latent, sensory-like format of representation during preparatory attention. Furthermore, the emergence of sensory-like template coincided with enhanced information connectivity between V1 and frontoparietal areas and was associated with improved behavioral performance. Our findings provide evidence for the co-existence of two formats of attentional templates (non-sensory vs. sensory-like) during preparation, as well as a novel neural mechanism for their maintenance in different functional states (active vs. latent). We propose that this dual-format representation may serve to increase flexibility of attentional control.

Results

Behavioral performance during the attention tasks

In an orientation cueing task, participants were shown a color cue indicating the reference orientation (45° or 135°) to attend to during preparation period (a delay of 5.5 s or 7.5 s) with or without the impulse perturbation (Fig 1A). This was followed by the presentation of a compound stimulus consisting of two oriented gratings. During the stimulus selection period (after the gratings appeared), participants were tasked with discriminating a small angular offset of the cued grating from the cued reference orientation. The angular offset was individually thresholded before the scanning sessions (mean offset = 2.50° in the No-Ping session and 2.52° in the Ping session) without significant difference between two sessions (independent t-test: $t(19) = 0.085$, $p = 0.932$, Cohen's d = -0.027). Participants' discrimination performance showed no significant difference between two attended orientations in either the No-Ping (paired t-test: $t(19) = 1.439$, $p = 0.166$, Cohen's d = -0.321) or the Ping session (paired t-test: $t(19) = 0.494$, $p = 0.627$, Cohen's d = 0.122; Fig 1C). A two-way mixed ANOVA (attended orientation × session) revealed neither significant main effects (attended orientation: $F(1,38) = 0.392$, $p = 0.535$, $\eta_p^2 = 0.01$; session: $F(1,38) = 0.001$, $p = 0.970$, $\eta_p^2 < 0.001$) nor interaction effect ($F(1,38) = 1.811$, $p = 0.186$, $\eta_p^2 = 0.045$). Bayesian

analyses provided moderate evidence to support the null hypothesis ($BF_{excl} > 3.633$), suggesting comparable performance levels between two sessions and two attended orientations.

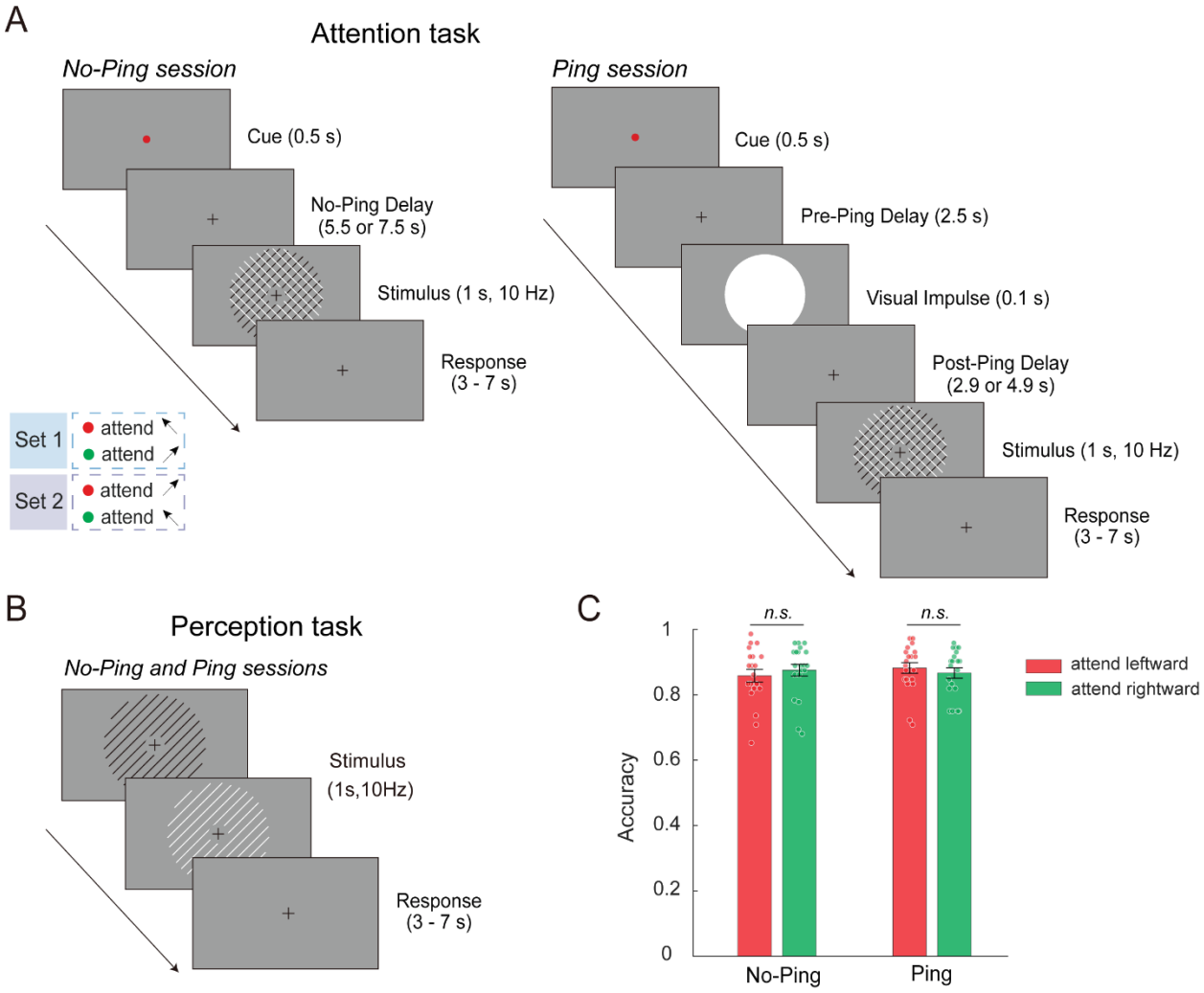


Fig 1. Experiment procedure and behavioral performance. (A) Attention tasks in the No-Ping and Ping sessions. Note that only long-delay trials are shown. A small proportion of short-delay trials (20%, with a delay of 1.5 s or 3.5 s) were included to create temporal uncertainty and encourage consistent active preparation during the delay. Both component gratings were flickering at 10 Hz between white and black, so that luminance could not confound either the task strategy (e.g., attending to luminance) or neural measures. The inset shows two sets of color-orientation mapping, which were reversed halfway through the experiment to minimize the impact of cue-induced sensory difference on neural activity. A high-contrast impulse was presented during preparation period in the Ping session. (B) Perception task. Similar to the attention task, the single-orientation grating also flickered at 10 Hz between white and

black. (C) Behavioral accuracy in the attention tasks in the No-Ping and Ping sessions. Each dot represents one subject's data. Error bars denote standard error of the means (SEM).

A default, non-sensory representation of attentional template during preparation

The first aim of this study was to determine whether attentional signals during preparation are encoded in a sensory-like or non-sensory format. To address this, we first examined whether, in the attention task during the No-Ping condition, the distributed neural pattern contained feature-specific information. We trained and tested separate classifiers to predict the attended orientation during the preparation and stimulus selection periods (Fig 2A, "Attention decoding"; see Materials and Methods for details). This analysis was performed for each of the four regions along the visual hierarchy, including primary visual cortex (V1), extrastriate visual cortex (EVC), intraparietal sulcus (IPS) and prefrontal cortex (PFC). The average decoding accuracies for both preparation and stimulus selection periods were significantly above chance level in each region (Permutation analyses: $ps < 0.004$ across regions, Fig 2B), indicating that the brain maintained reliable information about the attended feature both before and after the onset of the compound grating. Next, we examined whether the preparatory activity reflected a sensory-like format of attentional template (Fig 2A, "Cross-task generalization", see Materials and Methods). We trained a classifier using data from the perception task (leftward vs. rightward orientation; Fig 1B) and tested its performance on data from the preparation period in the attention task (attend leftward vs. attend rightward). However, this cross-task generalization analysis yielded no significant effects ($ps > 0.132$ across the regions). In contrast, we observed above-chance generalization from the perception task to the stimulus selection period ($ps < 0.001$ across regions, Fig 2C), confirming previous findings of the sensory-like attentional template following stimulus presentation (Gong et al., 2022; Jigo et al., 2018; Wen et al., 2019).

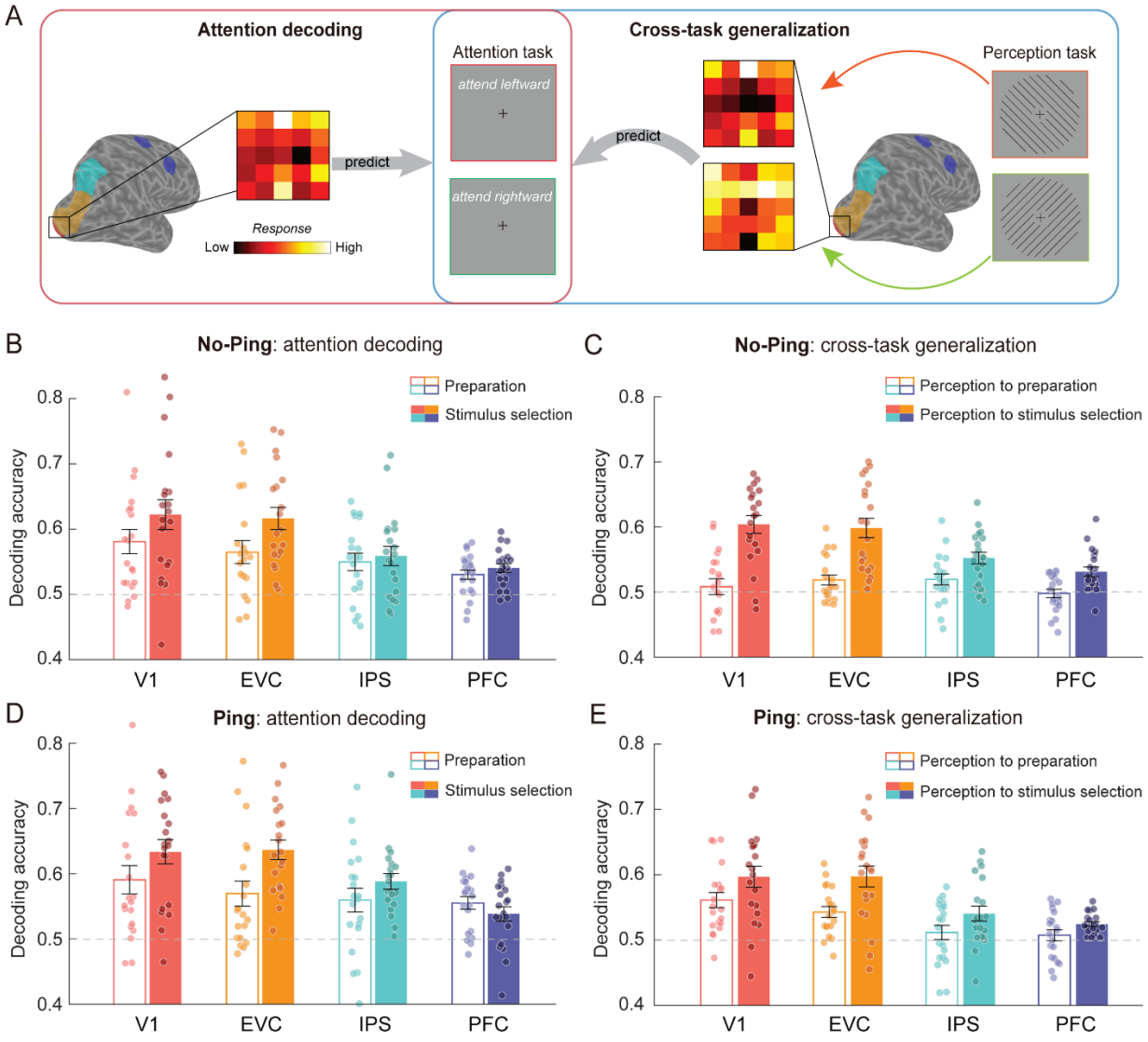


Fig 2. MVPA for the No-Ping session and Ping session. (A) Schematic illustration of the decoding of attended orientation (attend leftward vs. attend rightward) in the attention task (left panel) and the cross-task generalization analysis from perception task to the attention task (right panel). The four regions are shown on a representative right hemisphere as colored areas: V1 is marked in red, EVC in yellow, IPS in cyan and PFC in purple. (B) Decoding accuracy during preparation and stimulus selection periods across regions in the No-Ping and (D) Ping session. (C) Cross-task generalization performance from the perception task to the preparatory periods and the stimulus selection periods across regions in the No-Ping and (E) Ping session. The dashed lines represent the theoretical chance level (0.5). Each dot represents one subject's data. Error bars denote SEM.

Before drawing conclusions based on the lack of generalization from the perception task to preparatory attention, we considered two alternative explanations to rule out potential confounds. First, the robust attention decoding during preparation ruled out the possibility that participants were not actively engaged in the task during preparation (Fig 2B, unfilled bars). Second, the generalizable effect from the perception task to the stimulus selection period across regions (Fig 2C, filled bars) argues against the possibility of low statistical power. Overall, these findings suggest that the preparatory attention and sensory processing of features have distinct formats, presumably reflecting a non-sensory format of representation during the preparation. These results replicate those of a previous fMRI study using motion stimuli with a similar design (Gong et al., 2022). Furthermore, consistent with previous studies (Gong et al., 2022; Jigo et al., 2018), univariate analysis did not reveal any reliable difference in overall BOLD responses between attention orientations (Supplementary Information and S1 Fig).

A latent, sensory-like attentional template during preparation revealed by visual impulse

The second aim of our study was to examine whether a latent, sensory-like template exists during preparation. While this precise template may not be necessary for preparation, it is relevant for subsequent target selection and discrimination (i.e., select the cued grating from the compound stimulus and discriminate a small angular offset between the cued grating and the reference orientation). To test this hypothesis, we perturbed the neural activity by means of a visual impulse during the preparation period in the Ping session (Fig 1A, right panel). Using the same analyses as those performed in the No-Ping session, robust attentional signals were observed during both preparation and stimulus selection periods (Permutation analyses: $ps < 0.001$ across regions; Fig 2D). Importantly, the cross-task generalization analyses indicated that the visual impulse led to above-chance generalization from the perception task to preparation period (Fig 2E, unfilled bars) in V1 and EVC ($ps < 0.001$), but not in IPS and PFC ($ps > 0.584$), along with generalizable effects from the perception task to the stimulus selection periods ($ps < 0.036$ across regions; Fig 2E; filled bars). These results suggest that different brain areas are involved in coding for sensory-like templates. To further evaluate whether the cross-task generalization from the perception task

188 to the preparation period was statistically different with and without visual impulse, we conducted a two-
189 way mixed ANOVA (session × region) on the generalization performance. The analysis revealed main
190 effects of region ($F(3,114) = 5.220$, $p = 0.002$, $\eta_p^2 = 0.121$), session ($F(1,38) = 7.321$, $p = 0.010$, $\eta_p^2 = 0.162$),
191 and importantly, a significant interaction effect ($F(3,114) = 3.964$, $p = 0.010$, $\eta_p^2 = 0.094$), supporting the
192 observation that the visual impulse led to significantly increased decoding accuracy in V1 (independent t-
193 test: $t(38) = 3.145$, $p = 0.003$, Cohen's d = 0.995) and EVC (independent t-test: $t(38) = 2.153$, $p = 0.038$,
194 Cohen's d = 0.681), but not in the frontoparietal regions (*ps* > 0.374). This dissociable result between the
195 two sessions further supports the activation of a latent, sensory-like template by the visual impulse during
196 preparatory attention.

197 To further solidify this conclusion, following analyses were used to examine several alternative
198 possibilities. First, we examined whether the impulse-driven generalization resulted from stronger feature
199 information in the Ping compared to No-Ping session during the perception task. This was not the case, as
200 evidenced by comparable levels of decodable orientation information between the Ping and No-Ping
201 sessions (see Supplementary Information and S2 Fig). Next, we asked whether the increased generalization
202 was due to generally stronger attentional signals in the Ping session during the attention tasks – for example,
203 if visual impulses simply refocused attention during long delays. This was not the case, as the two-way
204 mixed ANOVAs (session × region) on attention decoding accuracy revealed neither a significant main
205 effect of session nor an interaction effect during both the preparation (*ps* > 0.519; $BF_{excl} > 3.247$) and
206 stimulus selection periods (*ps* > 0.336; $BF_{excl} > 3.297$), suggesting comparable amount of attentional
207 information between the two sessions. Therefore, the findings of impulse-driven sensory-like template in
208 the visual cortex during preparation cannot be explained by general differences between two sessions.
209

210 **Matching preparatory attention to sensory template: impact on neural representation and**
211 **behavior**

212 The reported decoding accuracy from the cross-task generalization analysis quantifies the degree to which
213 differences in neural activity pattern between two conditions are shared across attention and perception

214 tasks. However, it does not directly measure how similar the neural patterns are when attending to an
215 orientation compared to perceiving that orientation. Unlike decoding accuracies, Mahalanobis distance
216 provides a continuous measure for characterizing representational geometries between different conditions
217 (Mahalanobis, 1936). To further corroborate our findings of the impulse-driven sensory-like template, we
218 calculated the Mahalanobis distance between each attention condition during preparation and each
219 perception condition (see Materials and Methods). If the patterns of activity reflect a sensory-like template,
220 we would expect greater pattern similarity (smaller distance) between "attend leftward" and "perceive
221 leftward" than between "attend leftward" and "perceive rightward", and vice versa for the "attend rightward"
222 conditions (see Fig 3A for a schematic of the four pair-wise distance measures), leading to an interaction
223 between attended and perceived orientation conditions.

224 We used a two-way repeated-measures ANOVA (attended orientation × perceived orientation) on
225 the Mahalanobis distance, separately for each session and each region. During the preparatory period in the
226 No-Ping session, no significant interaction effects were observed across regions (*ps* > 0.443; Fig 3B). In
227 contrast, the same analyses applied to the Ping session revealed significant interaction effects in visual areas
228 (V1: $F(1,19) = 9.335$, $p = 0.007$, $\eta_p^2 = 0.329$; EVC: $F(1,19) = 8.563$, $p = 0.009$, $\eta_p^2 = 0.311$; Fig 3C; also see
229 Supplementary Table 1), but not for frontoparietal regions (*ps* > 0.213). This cross-region difference is
230 consistent with the function of sensory areas in encoding precise neural representations for basic visual
231 features. Next, we directly compared whether attentional modulation of Mahalanobis distance was
232 statistically different with and without the visual impulse. We defined a new condition label based on
233 orientation consistency between attended and perceived orientations: (1) same orientation: averaging
234 "attend leftward/perceive leftward" and "attend rightward/perceive rightward"; and (2) different orientation:
235 averaging "attend leftward/perceive rightward" and "attend rightward/perceive leftward". A two-way
236 mixed ANOVA (session × orientation consistency) on Mahalanobis distance revealed a main effect of
237 orientation consistency in V1 ($F(1,38) = 4.21$, $p = 0.047$, $\eta_p^2 = 0.100$), indicating that activity patterns were
238 more similar when attended and perceived orientations matched. No significant main effect of session was
239 found ($p = 0.923$). Importantly, a significant interaction was found in V1 ($F(1,38) = 5.00$, $p = 0.031$, $\eta_p^2 =$

0.116), suggesting that visual impulse enhanced the similarity between preparatory attentional template and the perception of corresponding orientation. In EVC, the same analysis revealed only a main effect of orientation consistency ($F(1,38) = 5.87$, $p = 0.020$, $\eta_p^2 = 0.134$), with no other significant effects (*ps* > 0.240). We also calculated the Mahalanobis distance between neural patterns evoked by superimposed gratings during the stimulus selection period and each condition in the perception task, finding similar results (see Supplementary Information and S4 Fig). This result was expected, as feature-based attention is known to selectively enhance task-relevant features while filtering out task-irrelevant ones.

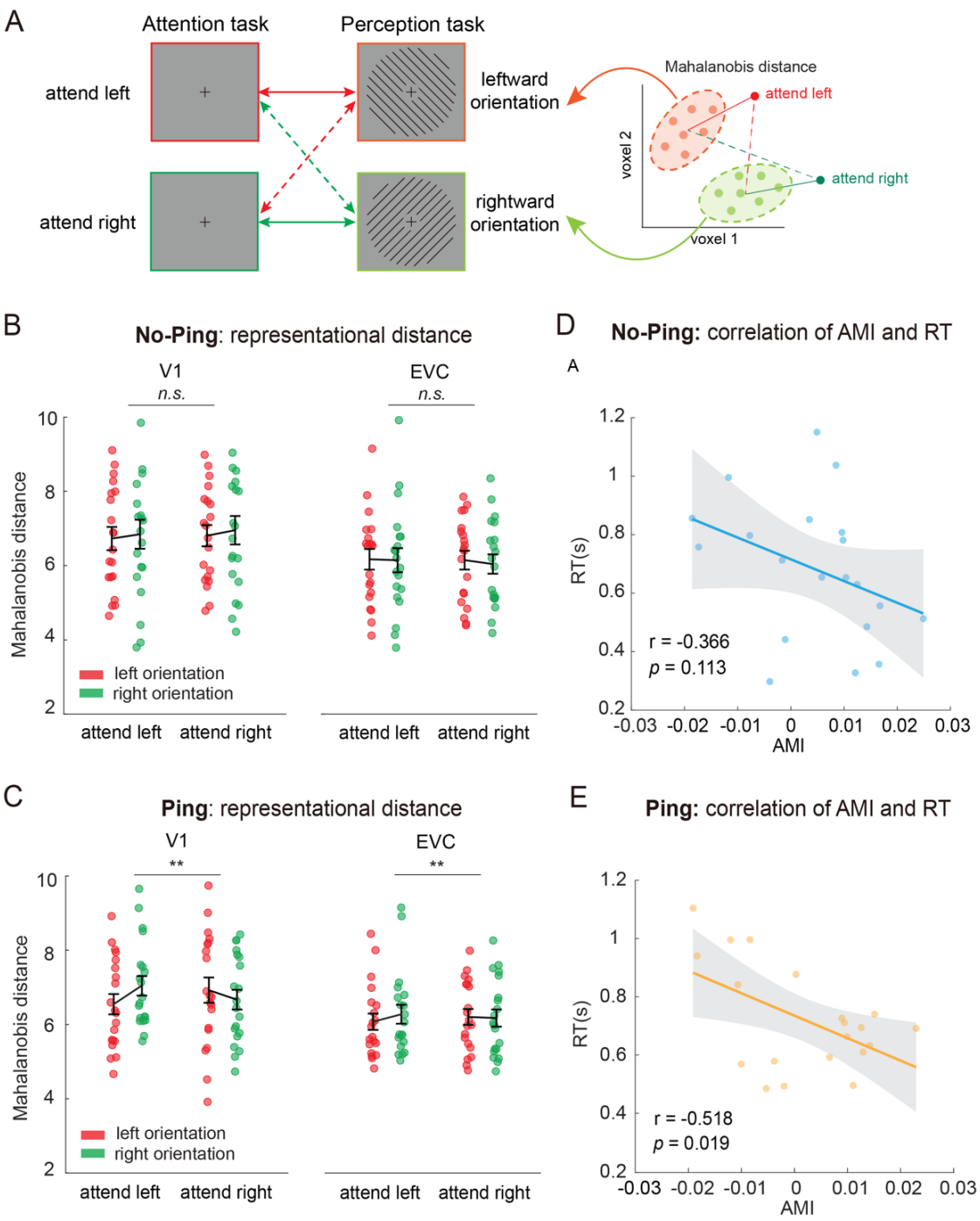


Fig 3. Orientation-selective attentional modulations on neural pattern distances during preparation. (A) Schematic illustration of the representational distance (mean Mahalanobis distances) between each of the attention conditions and each of the perception conditions. Colored arrows indicate measures of the pair-wise Mahalanobis distance. The right panel shows two attention trials (red indicates attend-to-leftward and green indicates attend-to-rightward) to the distribution of each perception condition (shown in a cloud of light-colored dots). (B) Mahalanobis

distance between preparatory attention condition and perceived orientation condition in the No-Ping and (C) Ping sessions. Error bars denote SEM. (D) Correlations between attentional modulation index (AMI) and reaction time (RT) in the No-Ping and (E) Ping sessions. Each dot represents one subject's data. The shaded area represents the confidence intervals of the regressed lines. * $p < 0.05$, ** $p < 0.01$.

The continuous nature of the Mahalanobis distance also made it possible to further investigate potential neural-behavioral correlations. We examined whether activating a sensory-like template during preparation would benefit subsequent orientation processing. In particular, we calculated attentional modulation indices (AMIs) based on trialwise Mahalanobis distance in V1. The index was calculated as follows: $AMI = (D_{different} - D_{same})/(D_{different} + D_{same})$, where D_{same} and $D_{different}$ are the measured distance (D) in the Same (e.g., attend and perceive the same orientation) and Different (e.g., attend and perceive different orientations) orientation condition, respectively (see Methods and Materials). Then, we calculated the correlation between AMI and both reaction time (RT) and accuracy across participants, separately for each session. In the No-Ping session, we observed no significant correlation between AMI in V1 and RT ($r = -0.366$, $p = 0.113$; Fig. 3D). By contrast, the same analysis in the Ping condition revealed a significantly negative correlation ($r = -0.518$, $p = 0.019$; Fig. 3E). These results indicate that the attentional modulations evoked by visual impulse was associated with faster RTs. These effects were not observed for accuracy ($ps > 0.550$). Furthermore, we also performed within-subject analysis by sorting trials as "strong modulation" and "weak modulation" trials based on each individual's AMI values, facilitated RTs were observed in "strong modulation" trials during the Ping session (Supplementary Information and S3 Fig). These results suggest that the impulse-driven sensory-like template in primary visual cortex is functionally relevant to subsequent attentional selection, providing evidence for the prospective use of sensory-like template in this task. In addition, we did not observe such behavioral differences in analogous analyses using data from the stimulus selection period in either session ($ps > 0.230$), which might due to the potential dilution by strong stimulus-evoked responses during the stimulus selection period.

Activating sensory-like template strengthens the informational connectivity between sensory and frontoparietal areas

Selective attention is generally believed to rely on coordinated network activity (Corbetta & Shulman, 2002). In particular, studies have shown that functional connectivity between sensory and frontoparietal areas was modulated by attentional control (Bressler et al., 2008; Rosenberg et al., 2020). Given that the impulse-driven sensory-like template facilitated behavior, we reasoned that it may also enhance network communication. Thus, we examined informational connectivity measures to explore how the impulse altered network function during the attention task.

We used a method that allows inference based on multivoxel pattern information rather than univariate BOLD response (Jia et al., 2020; Ng et al., 2021). For each ROI, we calculated the cross-validated Mahalanobis distance from each attention trial (from one left-out run) to the distribution of each attended orientation (all trials from remaining runs) during preparation (see Methods and Materials). To quantify the degree of attentional modulation during preparation, we calculated the AMI based on trial-wise Mahalanobis distance and generated a time course of AMI values across trials. Pearson correlation was used to estimate the covariation between each pair of ROIs, the resulting correlation coefficients were transformed using Fisher's z-transform for statistical inference (Fig 4A). The analysis revealed numerically higher levels of connectivity in Ping than in No-Ping session, this impulse-driven increase in connectivity reached statistical significance in two pairs (Fig 4B): V1-IPS (independent t-test: $t(38) = 2.566$, $p = 0.014$; Cohen's d = 0.812) and V1-PFC (independent t-tests: $t(38) = 3.158$, $p = 0.003$; Cohen's d = 0.999). The enhanced functional connectivity between V1 and frontoparietal areas driven by the impulse may potentially facilitate information flow among areas to improve attentional control, as implicated by a trend of ping-enhanced correlations between V1-PFC and RTs (Supplementary Information and S5 Fig). Additionally, the same analysis of AMI based on cross-validated Mahalanobis distance during the stimulus selection period showed no significant differences in information connectivity between No-Ping and Ping sessions (*ps* > 0.224; Supplementary Information and S6A Fig). The lack of changes in long-range

connectivity during stimulus selection period may be attributed to a general rise in connectivity caused by strong sensory inputs in this period, which could have attenuated any potential impacts of visual impulses. Furthermore, connectivity analysis based on mean BOLD response over time did not reveal any significant changes in inter-cortical connections between the two sessions (*ps* > 0.136; S6B Fig), suggesting that the impulse-driven increased information connectivity between V1 and higher-order areas was unlikely contributed by the overall changes of BOLD response.

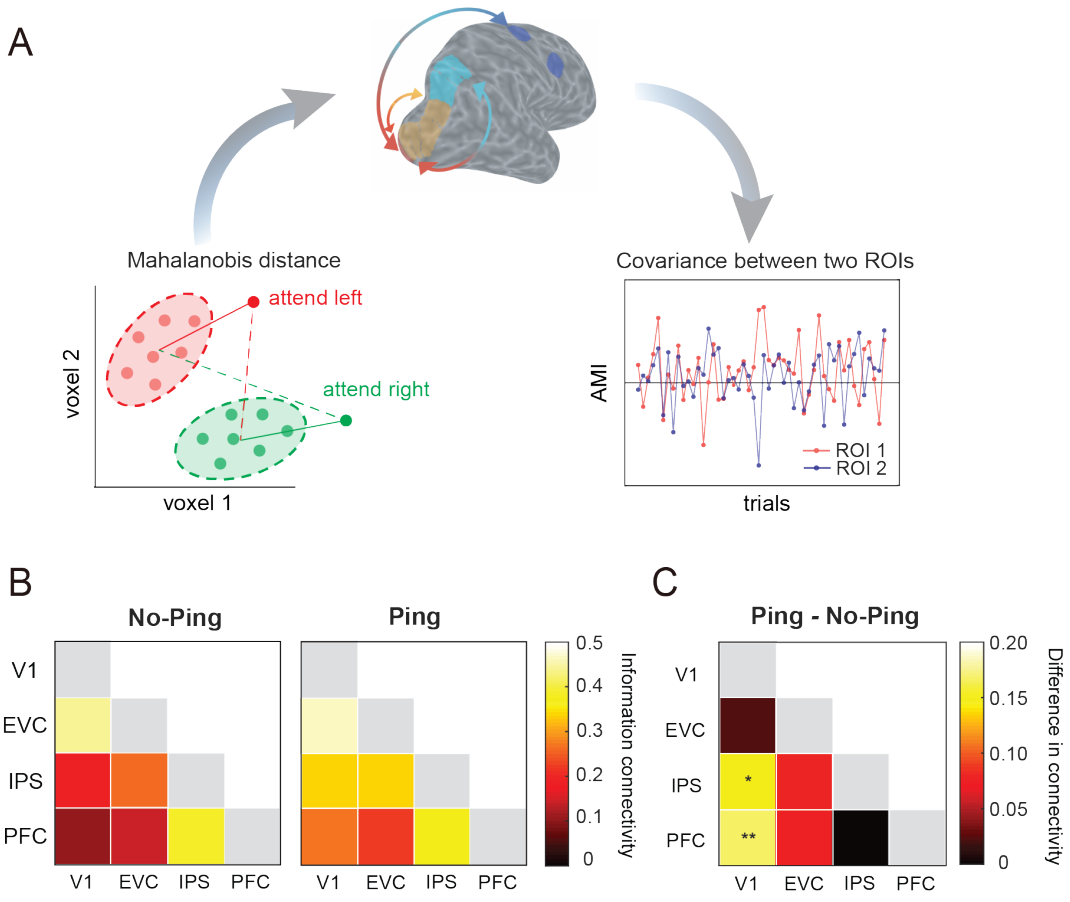


Fig 4. Information connectivity analysis. (A) Schematic illustration of the procedure for the information connectivity analysis in the space of two hypothetical voxels. For each region, we calculated the Mahalanobis distance of the attention trial (from one left-out run) from two attention distributions (all trials from remaining runs). Red and green dots indicate activity patterns from two trials (right panel). The brain image shows an example pair of intercortical information connectivity between V1 and PFC. The time series (lower-left panel) consisted of attentional modulation

index (AMI) based on the Mahalanobis distance. (B) Between-region information connectivity in the No-Ping and Ping sessions. (C) The differences in connectivity between the Ping and No-Ping sessions. * $p < 0.05$, ** $p < 0.01$.

Discussion

While there is ample evidence that the brain can maintain an attentional template of an upcoming target before sensory information is presented (Desimone & Duncan, 1995; Kastner et al., 1999; Summerfield & De Lange, 2014), its representational format remains unclear. To address this, we used an orientation cueing paradigm with separated preparation and stimulus selection periods and applied MVPA to decode neural activity patterns associated with feature-specific attentional information of the upcoming target. The analyses showed robust attentional information both before and after the presentation of the compound grating, indicating a sustained maintenance of attentional templates throughout a trial. Importantly, while the decoders trained on the perception of single orientations could not generalize to preparation until the stimulus selection period (*No-Ping session*), perturbing the brain with a visual impulse resulted in generalizable activity patterns during preparation in V1 and EVC (*Ping session*). These results suggest a predominantly non-sensory format of representation, with a sensory-like template in a latent state during feature-based preparation in the visual cortices. Furthermore, impulse-driven sensory-like template was accompanied by enhanced information connectivity between V1 and frontoparietal areas, as well as enhanced orientation-specific neural modulations of neural distances in the visual areas that predicted levels of behavioral performance. We observed similar response profiles in V1 and EVC, with V1 exhibiting more robust ping-evoked changes compared to EVC, consistent with its primary role in orientation processing (Priebe, 2016). The differences between the Ping and No-Ping sessions could not be attributed to differences in sensory information from the perception task, overall strength of preparatory attention, or differences in eye position (S7 Fig). Therefore, our findings suggest a dual-format neural representation scheme (non-sensory vs. sensory-like) operating in different functional states (active vs. latent). This mechanism may give rise to flexible attentional control, allowing effective transition from coarse to fine attentional templates at various processing stages (initial guidance vs. precise stimulus discrimination).

342 Recent advances in theories of visual search differentiated between the "guiding template" and
343 "target template" based on measures of behavioral performance and eye movements (Wolfe, 2021; Yu et
344 al., 2023). According to these theories, early attentional guidance typically depends on a non-veridical
345 codes that represent only the most diagnostic information, whereas later, target-match processes utilize
346 more precise codes to optimize decision accuracy (Kerzel, 2019; Scolari et al., 2012; Yu et al., 2022). Our
347 study reveals a parallel coding mechanism in the context of feature-based attention, expanding upon these
348 theoretical notions in three key aspects.

349 First, we provide neural evidence for a default, predominantly non-sensory template during
350 preparation, indicating that the concept of a "guiding template", as proposed by visual search theories
351 (Wolfe, 2021; Yu et al., 2023), also applies to preparatory attention in a non-search context. This highlights
352 a shared functional role of a non-veridical attentional template in early guidance across different scenarios.
353 Second, despite the theoretical notion that the brain maintains a more veridical template with detailed target
354 information than is typically utilized to form the guiding template (Wolfe, 2021; Yu et al., 2023), neural
355 evidence supporting this hypothesis is currently lacking. We provide evidence for this notion and propose
356 a plausible neural implementation for preserving a more veridical, sensory-like template in the latent state.
357 A natural question is why the sensory-like template remains latent during preparation. We note that our task
358 requires both coarse and fine featural information. During preparation, a coarse, non-veridical guiding
359 template suffices for target-distractor discrimination, while during stimulus selection, a precise template is
360 needed for the fine discrimination task (reporting the tilted direction of a small angular offset). Maintaining
361 a latent sensory-like template during preparation is thus efficient, as it facilitates future sensory processing
362 while conserving resources. Finally, information connectivity between visual and higher-order
363 frontoparietal regions was enhanced by visual impulse during preparation, which correlated with improved
364 behavioral performance in feature selection. This result suggests that improved information flow across the
365 relevant areas leads to enhanced attentional control, which in turn contributes to refined sensory
366 representations of target in early visual cortex, facilitating transitions from a non-sensory to sensory-like
367 template. Future studies may adopt layer-specific fMRI to infer the direction of this improved information

368 flow (Jia et al., 2023; Jia et al., 2024) and explore the relationship between long-range connections and the
369 utilization of different formats of target templates.

370 It could be argued that preparatory attention relies on the same mechanisms as working memory
371 maintenance (Bettencourt & Xu, 2016; Sheremata et al., 2018). While these functions are intuitively similar
372 and likely overlap, there is also evidence indicating that they can be dissociated (Battistoni et al., 2017). In
373 particular, we note that in our task, attention is guided by symbolic cues (color-orientation associations),
374 while working memory tasks typically present the actual visual stimulus as the memorandum. A central
375 finding in working memory studies is that neural signals during WM maintenance are sensory in nature, as
376 demonstrated by generalizable neural activity patterns from stimulus encoding to maintenance in visual
377 cortex (Harrison & Tong, 2009; Serences et al., 2009; Rademaker et al., 2019). However, in our task, neural
378 signals during preparation were non-sensory, as demonstrated by a lack of such generalization in the No-
379 Ping condition (see also Gong et al., 2022). We believe that the differences in cue format and task demand
380 in these studies may account for such differences. In addition to the difference in the sensory nature of the
381 preparatory versus delay-period activity, our ping-related results also exhibited divergence from working
382 memory studies (Wolff et al., 2015, 2017). While these studies used the visual impulse to differentiate
383 active and latent representations of *different items* (e.g., attended vs. unattended memory item), our study
384 demonstrated the active and latent representations of *a single item in different formats* (i.e., non-sensory vs.
385 sensory-like). Moreover, unlike our study, the impulse did not evoke sensory-like neural patterns during
386 memory retention (Wolff et al., 2017). These observations suggest that the cognitive and neural processes
387 underlying preparatory attention and working memory maintenance could very well diverge. Future studies
388 are necessary to delineate the relationship between these functions both at the behavioral and neural level.

389 While we found that the ping allowed us to detect a sensory-like template during preparation, the
390 underlying neural mechanism of such effects remains unclear. One possibility, as informed by theoretical
391 studies of working memory, is that the sensory-like template could be maintained via an "activity-silent"
392 mechanism through short-term changes in synaptic weights (Mongillo et al., 2008). In this framework, a
393 visual impulse may function as nonspecific inputs that momentarily convert latent traces into detectable

activity patterns (Rademaker & Serences, 2017). Related to our findings, it is unlikely that the orientation-specific templates observed during the Ping session emerged *de novo* from purely non-sensory representations and were entirely induced by an exogenous ping, which was devoid of any orientation signal. Instead, the more parsimonious explanation is that visual impulse reactivated pre-existing latent sensory signals, consistent with the models of "activity-silent" working memory. However, the detailed circuit-level mechanism of such reactivation is still unclear, as well as whether this effect is modality-specific. Prior work shows that only visual, but not auditory, impulses reactivate latent visual working memory (Wolff et al., 2020), suggesting some degree of modality-specificity. However, this finding warrants direct investigation in future studies. Furthermore, we acknowledge that whether pinging identifies an activity-silent mechanism is currently debated (Barbosa et al., 2021; Schneegans & Bays, 2017). An alternative possibility is that the visual impulse amplified a subtle but active representation of the sensory template during preparation. Distinguishing between these alternatives likely requires future studies with more detailed neurophysiological measurements. Regardless of the precise neural mechanism for the observed latent, sensory representation, our results suggest that both sensory and non-sensory templates likely co-exist.

The non-generalizable activity patterns from perception to preparatory attention, in the absence of visual impulse, suggests a default, predominantly non-sensory template during preparation. This finding is largely consistent with electrophysiological studies (Myers et al., 2015; Wen et al., 2019) and our prior fMRI work on preparatory attention to motion directions (Gong et al., 2022), but differs from some previous neuroimaging studies that demonstrated sensory-like templates during preparation (Kok et al., 2014; Peelen & Kastner, 2011; Stokes et al., 2009). One potential account for these discrepancies is that those studies used cue-only trials where the target was expected but not actually presented, in contrast to our task where the target was shown on every trial with temporally separated preparation and stimulus selection periods. This seemingly subtle difference may significantly impact the formats of the neural representations. Because cue-only trials increased the likelihood of target appearance at the subsequent time point, sensory template may be activated due to modulations of temporal expectations (Grubert & Eimer, 2018). This

explanation is consistent with theories suggesting differential influences of expectation and attention on neural activity: expectation reflects visual interpretations of stimuli due to sensory uncertainty, whereas attention is guided based on the task relevance of sensory information (Rungratsameetaweemana & Serences, 2019; Summerfield & Egner, 2009, 2016). Our finding of a predominantly non-sensory format may indicate an optimized coding strategy employed by the brain to effectively and robustly represent information for future use. This aligns with the proposed role of attention in modulating sensory representations to encode only currently relevant information at a minimal cost (Yu et al., 2023).

While our findings cannot pinpoint the exact format of this non-sensory template, we consider categorical coding a plausible candidate based on previous findings. For instance, visual search studies demonstrate that categorical attributes (e.g., steep vs. shallow; left-tilted vs. right-tilted) efficiently guide attention for simple features, such as an orientation or a color (Kong et al., 2017; Wolfe et al., 1992), particularly when features are consistent and predictable (Hout et al., 2017). In our task, the angular relations between the target and distractor orientation were defined by categorical attributes (e.g., left-tilted vs. right-tilted) and remained consistent across trials, making a categorical template feasible during preparatory attention. Furthermore, categorical template allows for greater tolerance of stimulus variability, which is also useful given the trial-by-trial variations in target orientation around the reference orientation in our task. Future studies are needed to address the nature of the non-sensory template during preparation as well as task parameters that might modulate them.

In summary, the current study suggests that there are two formats of attentional templates each having a distinct functional state: a default, non-sensory format and a latent, sensory-like format. This dual-format representation aligns with theories on the dual-function of attentional template for different task goals (Hout & Goldinger, 2015; Yu et al., 2023). The current findings provide a plausible neural implementation for these theories by demonstrating different formats in different functional states. This mechanism likely reflects an optimized coding scheme that effectively balances processing efforts and demands, particularly well-suited for flexible control and transitions from coarse to fine task demands in visually guided behavior.

Materials and Methods

Participants

Twenty individuals participated in the No-Ping session (11 females, mean age = 22.9) and twenty individuals participated in the Ping session (14 females, mean age = 23.7). Among them, 14 participants took part in both sessions, while 12 of them took part in only one session. The sample size was comparable to previous studies using similar attention tasks (Baldauf & Desimone, 2014; Gong & Liu, 2020a; 2020b; Guo et al., 2012; Jigo et al., 2018; Liu & Hou, 2013). Because our primary interest is the generalization from the perception task to attention task, we used the minimal effect size of decoding accuracy across regions (one-sample t-tests: d = 0.868) from our previous study with a similar design (Gong et al., 2022), and used G*Power (Version 3.1) (Faul et al., 2007) to confirm that this sample size is sufficient to detect a cross-task generalization effect with a power greater than 95% (a = 0.05). All participants were right-handed and had a normal or corrected-to-normal vision. Participants provided written informed consent according to the study protocol approved by the Institutional Review Board at Zhejiang University (2020-06-001). They were paid ¥200 (~$27.4) for their participation in each session.

Stimuli and Apparatus

Stimuli were generated using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007) implemented in MATLAB. The stimuli were presented on an LCD monitor (resolution: 1920 × 1080, refresh rate: 60 Hz) during behavioral training, at a viewing distance of 90 cm in a dark room. During the fMRI scans, stimuli were projected to a screen via a MR-compatible LCD projector (PT-011, Jiexin Technology Co., Ltd., Shenzhen, China) with the same resolution and refresh rate as the LCD monitor during behavioral training. Participants viewed the screen via an angled mirror attached to the head coil at a viewing distance of 115 cm. Angular stimulus size was the same across behavioral and fMRI sessions.

The orientation stimuli were square-wave gratings (1.3 cycles per deg, duty cycle: 10%) in a circular aperture (inner radius: 1.5°; outer radius: 6°). The gratings flashed on a gray background at 10 Hz, alternating between black and white. There were two types of stimuli: two overlapping gratings orientated

leftward (~135°) and rightward (~45°), or a single grating with one of the two orientations (~135° or ~45°). Here, we refer to the 45° and 135° orientations as the reference orientations. The impulse stimulus was a high-contrast, white (at the maximum projector output level) circular disk that covered the same area as the orientation stimulus (radius: 6°).

Experimental Procedures and Tasks

Each participant completed at least two fMRI sessions on different days. One session was used for defining ROIs (see Definition of Regions of Interests), while the remaining sessions were used for the main experiment (see Attention task and Perception task). Before the scanning sessions, participants were trained to familiarize with the tasks in a separate behavioral session. The procedures and tasks were similar to our previous work (Jigo et al., 2018; Gong et al., 2022).

Attention task. We used a cueing paradigm (Fig 1A). Each trial began with a color cue (red or green) for 0.5 s to indicate the reference orientation of the upcoming target (leftward vs. rightward orientation). In the No-Ping session, the cue was followed by a blank display during the preparation period; in the Ping session, a task-irrelevant, high-luminance visual impulse ("ping", 0.1 s) occurred at either 0.5 s (for short delays of 1.5 s and 3.5 s) or 2.5 s (for long delays of 5.5 s and 7.5 s) after the onset of the cue display. The orders of these sessions were counterbalanced across participants who completed both. Following the preparatory period, two superimposed gratings were then shown for 1 s. The target grating was shown with a small angular offset with respect to the cued reference orientation, whereas the distractor grating was shown in the uncued reference orientation (e.g., if rightward orientation was cued, the rightward grating was shown in 45°± d and the leftward grating was shown in 135°). Note that the angular offset was determined individually based on the threshold obtained during the training session (at least 3 blocks, 30 trials/block), using a staircase procedure (Best Parameter Estimation by Sequential Testing, Best PEST), as implemented in the Palamedes Toolbox (Prins & Kingdom, 2009). Participants used a keypad to report whether the attended orientation was more leftward or rightward relative to the reference orientation. Each trial was separated by an inter-trial interval of 3 s to 7 s (2 s per step). Trial-by-trial feedback ("correct" or "incorrect") was provided in the training session but not during scanning. Instead, the percentage of correct

497 responses was provided at the end of each run in the scanning session to avoid the impact of trial-level
498 feedback on neural activity.

499 Given the need to maximize the number of trials for fMRI-based MVPA, we could not accomodate
500 additional conditions (e.g., neutral cue) to measure the behavioral effects of attention. However, our prior
501 work using similar feature cueing paradigms (Liu et al., 2007; Jigo et al., 2018) found that attentional cueing
502 improved behavioral performance relative to a neutral condition. Thus, it is highly likely that our well-
503 trained participants used the cue to direct their attention in the fMRI experiment. Furthermore, our neural
504 measures of attentional signals revealed feature-specific attentional modulations, further validating our
505 approach. To prevent the cue-related sensory difference from contributing to neural activity, we reversed
506 the mapping between colors and orientations halfway through the experiment (e.g., red indicated "attend
507 leftward orientation" and green indicated "attend rightward orientation" in the first half of the runs, and
508 vice versa for the second half of the runs), with the order counterbalanced across subjects. The mapping of
509 colors and orientations was reversed only once in the middle of the experiment to prevent misremembering
510 of the color-orientation associations. To reduce temporal expectancy over a fixed period, the preparatory
511 period (i.e., cue-to-stimulus interval) varied from 1.5 s to 7.5 s with different probabilities (10% for 1.5 s
512 or 3.5 s each, 40% for 5.5 s or 7.5 s each). The long-delay trials (5.5 s or 7.5 s) were selected for subsequent
513 analyses, as they allow the separation of the preparatory activity from the grating-evoked response during
514 fMRI scanning. The short-delay trials were included to encourage a sustained maintenance of attention
515 throughout the entire preparation period.

516 *Perception task.* On each trial of the perception task (Fig 1B), a single grating was shown for 1 s,
517 followed by an inter-trial interval between 3 s to 7 s. To equate the sensory inputs between attention and
518 perception tasks, the orientation was shifted away from the reference orientation by the same angular offset
519 as that used in the attention task with each individual participant's own threshold. Participants performed
520 the same orientation discrimination task by comparing the single orientation to the reference orientation.
521 We provided the percentage of correct responses at the end of each run as feedback.

522 **Eye Tracking and Analysis**

To evaluate the stability of visual fixation, we used Eyelink Portable Duo system (SR Research, Ontario, Canada) to monitor each observer's eye position during the training session at a sampling rate of 500 Hz. One participant's data was not used due to the unstable recording of the eye. The data were then analyzed using custom Matlab code.

To examine whether participants adopted a space-based strategy during the preparatory period in the attention task, such as directing their gaze leftward in attend leftward trials, and vice versa for the attend rightward trials, we analyzed the participants' eye positions recorded during the training session. Horizontal and vertical eye positions were analyzed separately. Paired t-tests were performed to compare horizontal and vertical eye positions between two attention conditions. A two-way mixed ANOVA (2 sessions × 2 attended orientations) was applied to the horizontal and vertical positions, respectively.

fMRI Data Acquisition

Imaging was performed on a Siemens 3T scanner (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) using a 20-channel coil at Zhejiang University (Hangzhou, China). For each participant, we acquired high-resolution T1-weighted anatomical images (field of view, 240 × 240 mm, 208 sagittal slices; 0.9 mm^3 resolution), T2*-weighted echo-planar functional images consisting of 46 slices (TR, 2 s; TE, 34 ms; flip angle, 50°; matrix size, 80 × 80; in-plane resolution, 3 × 3 mm; slice thickness, 3 mm, interleaved, no gap) and a 2D T1-weighted anatomical image (0.8 × 0.8 × 3 mm) for aligning functional data to high-resolution anatomical data.

fMRI Data Preprocessing

Data analyses were performed using mrTools (Gardner et al., 2018) and custom code in Matlab. For each run, functional data were preprocessed with head motion correction, linear detrending and temporal high pass filtering at 0.01 Hz. Data were converted to percentage signal change by dividing the time course of each voxel by its mean signals in each run. We concatenated the 6 runs of the attention task and the 3 runs of the perception task separately for further analysis. One of the attention runs in one subject was excluded due to low accurate performance (<50%).

Definition of Regions of Interests (ROI)

Visual and parietal ROIs. Following previous work (Jigo et al., 2018; Gong et al., 2020b; Gong et al., 2022), for each observer, we ran a separate retinotopic mapping session to obtain ROIs in occipital and parietal areas. Observers viewed four runs of rotating wedges (i.e., clockwise and counterclockwise) and two runs of rings (i.e., expanding and contracting) to map the polar angle and radial components, respectively (DeYoe et al., 1996; Engel, 1997; Sereno et al., 1995). Borders between areas were defined as the phase reversals in a polar angle map of the visual field. Phase maps were visualized on computationally flattened representations of the cortical surface, which were generated from the high-resolution anatomical image using FreeSurfer (http://surfer.nmr.mgh.harvard.edu) and custom Matlab code.

To help identify the topographic areas in parietal areas, we ran 2 runs of memory-guided saccade task modeled after previous studies (Konen & Kastner, 2008; Schluppeck et al., 2006; Sereno et al., 2001). Observers fixated at the screen center while a peripheral (~10° radius) target dot was flashed for 500 ms. The flashed target was quickly masked by a ring of 100 distractor dots randomly positioned within an annulus (8.5° – 10.5°). The mask remained on screen for 3 s, after which participants were instructed to make a saccade to the memorized target position, then immediately saccade back to the central fixation. The position of the peripheral target shifted around the annulus from trial to trial in either a clockwise or counterclockwise order. Data from the memory-guided saccade task were analyzed using the same phase encoding method as the wedge and ring data. Therefore, the following regions of interest (ROIs) in each hemisphere were identified after the completion of this session: V1, V2, V3, V3A/B, V4, V7/IPS0, IPS1 to IPS4.

Frontal ROIs. Following previous work (Jigo et al., 2018; Gong et al., 2020b; Gong et al., 2022), we used a deconvolution approach by fitting each voxel's time series from the attention task with a general linear model (GLM) to determine the event-related activations in the brain (see Supplementary Materials: Deconvolution). For each voxel, we computed the goodness of fit measure (r^2 value), which indicates the amount of variance explained by the deconvolution model (Gardner et al., 2018). The r^2 value represents the degree to which the voxel's time series is correlated with the task events, regardless of any differential

574 responses among conditions. Based on the task-related activation (as indexed by r^2 value) and anatomical
575 criteria, we defined two frontal areas in each hemisphere that were active during the attention task: one is
576 located superior to the precentral sulcus and near the superior frontal sulcus (FEF) and the other is located
577 towards the inferior precentral sulcus, close to the junction with the inferior frontal sulcus (IFJ).

578 **Groups of Region.** To characterize the patterns of neural response across cortical hierarchy and
579 streamline data presentation, we grouped results from the nine areas into four groups based on functional
580 and anatomical considerations: primary visual cortex (V1); extrastriate visual cortex (EVC), consisting of
581 V2, V3, V3ab and V4; IPS, consisting of IPS0 to IPS4; prefrontal cortex (PFC), consisting of FEF and IFJ.
582 Individual areas within each group exhibited qualitatively similar results.

583 Note that we analyzed V1 separately for two reasons. First, previous studies consistently identify
584 V1 as the main locus of sensory-like templates during feature-specific preparatory attention (Kok et al.,
585 2014; Aitken et al., 2020). Second, V1 shows the strongest orientation selectivity within the visual hierarchy
586 (Priebe, 2016). In contrast, the extrastriate visual cortex (EVC; comprising V2, V2, V3AB and V4)
587 demonstrates broader selectivity for complex features (Grill-Spector & Malach, 2004). Therefore, it would
588 be particularly informative to analyze V1 separately for our orientation-based attention paradigm.

589 **Multivoxel Pattern Analysis (MVPA)**

590 **Decoding of attended orientation.** To test if multivariate patterns of activity represent information
591 of the attended orientation, separate MVPA analyses were applied on the activity patterns for the
592 preparation and stimulus selection periods. Following previous work (Jigo et al., 2018; Gong et al., 2020b;
593 Gong et al., 2022), for this analysis, we extracted fMRI signals from raw time series in the long delay trials
594 with correct behavioral responses (~72 trials per attention condition); short delay trials were excluded as
595 they could not provide enough data points to measure preparatory activity. We then obtained averaged
596 BOLD response in a 2 s window for each voxel and each trial in a given ROI, separately for preparatory
597 activity (4 to 6 s after the onset of the cue) and stimulus-evoked activity (4 to 6 s after the onset of the
598 gratings). The response amplitudes across two attention conditions in each ROI were further z-normalized,
599 separately for the preparation and stimulus-related activity. These normalized single-trial BOLD responses

were used for the MVPA. We trained a classifier using the Fisher linear discriminant (FLD) analysis to discriminate between two attended orientations (leftward vs. rightward) and tested its performance with a leave-one-run-out cross-validation scheme. This process was repeated until each run was tested once and the decoding accuracy (i.e., the proportion of correctly classified trials) was averaged across the cross-validation folds. The statistical significance of decoding accuracy was evaluated by comparing it to the chance level obtained from a permutation test (see Permutation test). To assess if the decoding accuracy differed between No-Ping and Ping experiments, we performed two-way mixed ANOVAs (2 sessions × 4 regions) on the decoding accuracy.

Cross-task generalization from the perception task to attention task. Following previous work (Jigo et al., 2018; Gong et al., 2022), to test whether the neural patterns in the preparatory and stimulus selection periods from the attention task reflected sensory processing of isolated features, we trained an FLD classifier using the normalized BOLD responses from the perception task (4 to 6 s after the trial onset) to discriminate leftward vs. rightward orientation. Then, we tested this classifier on the normalized response from the independent runs of attention task to discriminate between attend leftward vs. attend rightward orientations, separately for preparation and stimulus selection periods. The significance of decoding accuracy was compared to the chance level obtained from a permutation test (see Permutation test). To assess if the generalization performance differed between No-Ping and Ping sessions, we performed two-way mixed ANOVAs (2 sessions × 4 regions) on the decoding accuracy.

Neural distance between attended and perceived orientations. The decoding accuracy from the cross-task generalization test reflects a discretized readout of the pattern similarity between different conditions. However, employing continuous similarity measures, such as Mahalanobis distance (Mahalanobis, 1936), could be more reliable compared to decoding accuracy (Walther et al., 2016). Therefore, we calculated the Mahalanobis distance to quantify the pattern similarity between two attended orientations and two perceived orientations. For each participant and each ROI, we have M points (i.e., M trials for each attended orientation) in the N-dimensional space (N = 100, number of voxels). For each data point in the attended orientation condition, we computed its distance to each of the orientation distributions

626 (from the perception task). Averaged distance values were then calculated for each combination of attended
627 orientation and perceived orientation pairs. A sensory-like hypothesis would predict smaller distance
628 between the distribution of the attended orientation (e.g., attend leftward) and the distribution of
629 corresponding orientation (e.g., perceive leftward) compared to the alternative orientation (e.g., perceive
630 rightward). Two-way repeated-measures ANOVA (2 attended orientations × 2 perceived orientations) was
631 applied on the Mahalanobis distance, separately for each region and each session.

632 **Neural-behavioral relationships.** We tested if the representation format during preparatory
633 attention was associated with subsequent behavior. For each trial, we calculated the Mahalanobis distance
634 between the attention conditions (attend leftward and attend rightward) and the perceived orientations
635 (leftward and rightward orientation). We estimated the attentional modulation index (AMI) based on these
636 distance values. This index measures how much attention modulated the pattern similarity for the Same
637 orientation condition (e.g., attend and perceive the same orientation) relative to the Different orientation
638 condition (e.g., attend and perceive different orientations). The index was calculated as follows: AMI =
639 $(D_{different} - D_{same})/(D_{different} + D_{same})$, where D_{same} and $D_{different}$ are the measured distance (D) in the Same and
640 Different orientation condition, respectively. Next, we tested the behavioral relevance of AMI in two ways:
641 (1) Inter-subject analysis: correlated AMI with RT and accuracy across participants, separately for each
642 session; (2) Within-subject analysis: for each participant, we sorted the single-trial AMI values in
643 descending order and selected top-ranked 25% trials and bottom-ranked 25% trials to represent "strong
644 modulation" and "weak modulation" trials, respectively. We then extracted behavioral responses on these
645 selected trials and calculated RT and accuracy for each trial type. Paired t-tests were used to compare
646 between "strong modulation" and "weak modulation" trials in each session.

647 **Informational connectivity analysis.** We used Informational Connectivity (IC) to examine shared
648 changes in pattern discriminability over time, a method that allows inference based on multivoxel pattern
649 information rather than overall BOLD response (Jia et al., 2020; Ng et al., 2021). To track the flow of
650 multivariate information across time (i.e., across trials), we measured the fluctuations (covariance) in
651 pattern-based discriminability by calculating the Mahalanobis distance of each trial to the two attended

652 orientations, using a leave-one-run-out cross-validation scheme. For each ROI, we calculated the
653 Mahalanobis distance between the pattern of activity for each attention trial from one left-out run and the
654 distribution of each attended orientation of the remaining runs. To quantify the degree of attentional
655 modulation, we calculated the AMI using the same formula as mentioned above, where D_{same} and $D_{different}$
656 are the measured distance (D) in the Same and Different condition. This index measures how much the
657 pattern similarity increased for the same attention condition (e.g., attend leftward to attend leftward) relative
658 to the different attention condition (e.g., attend leftward to attend rightward). A positive AMI indicates
659 relative proximity to the same attention condition, whereas a negative AMI indicates relative proximity to
660 the different attention condition. A time course of AMI values was generated across runs and pairwise
661 correlated between ROIs using Pearson correlation analysis and Fisher z-transformed. Independent t-tests
662 were used to compare the connectivity between No-Ping and Ping sessions. To assess the relationship
663 between ICs and behavior, we correlated ICs with RT and accuracy across participants, separately for each
664 session.

665 **Permutation test to evaluate classifier performance**

666 Following previous work (Jigo et al., 2018; Gong et al., 2022), for each brain area, we evaluated the
667 statistical significance of the observed decoding accuracy using a permutation test scheme. We first shuffled
668 the trial labels in the training data and trained the same FLD classifier on the shuffled data. We then tested
669 the classifier on the (unshuffled) test data to obtain decoding accuracy. For each ROI and each participant,
670 we repeated this procedure 1000 times to compute a null distribution of decoding accuracy. To compute
671 the group-level significance, we averaged the 20 null distributions to obtain a single null distribution of
672 1000 values for each ROI. To determine if the observed decoding accuracy significantly exceeds the chance
673 level, we compared the observed value to the 95 percentiles of this group-level distribution (corresponding
674 to $p = 0.05$). Note that these ROIs were pre-defined with strong priors as their activation in attention tasks
675 has been consistently reported in the literature. Nevertheless, for those analyses where multiple comparisons
676 were performed across regions, we applied a Bonferroni-correction to adjust the *p*-values.

Bayesian analysis

To evaluate the strength of evidence for the null hypothesis, we conducted Bayesian analyses (Wagenmakers, 2007) using standard priors as implemented in JASP Version 0.17.1 (JASP Team, 2023). We performed Bayesian t-tests and computed Bayes factor (BF_{01}) to compare between two attention conditions (attend leftward vs. attend rightward). Additionally, we used Bayesian repeated-measures ANOVA and computed the exclusion Bayes factors (BF_{excl}) to assess the evidence for excluding specific effects across all models. A Bayes factor (BF) greater than 1 provides support for the null hypothesis. Specifically, a BF between 1 and 3 indicates weak evidence, a BF between 3 and 10 indicates moderate evidence, and a BF greater than 10 indicates strong evidence (Van Doorn et al., 2021).

Approach to handle partially overlapped samples

Our study used partially overlapping samples, with 14 out of 20 participants completing both No-Ping and Ping sessions, while the remainder completed one of the two sessions. The most important analyses entailed assessing whether decoding accuracy was above chance, for which we used the permutation-based method (see above) within each session. Thus, these analyses were unaffected by the partially overlapping samples. In a few analyses where we compared across sessions, we used statistical tests treating "session" as a between-subject factor. We believe this is a reasonable approach, as a between-subject test is more conservative than a within-subject test, such that any significant effect emerged should be a genuine effect. To be certain, we also conducted additional analyses with "session" as a within-subject factor on the subset of data from the 14 participants who completed both sessions in a counterbalanced order. The results were highly similar to those reported in the main text.

References

Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R. S., Corbin, N., Callaghan, M. F., & Kok, P. (2020). Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. *PLOS Biology*, *18*(12), e3001023. https://doi.org/10.1371/journal.pbio.3001023

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, 19(1), 150-157. https://doi.org/10.1038/nn.4174

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*(6182), 424–427. https://doi.org/10.1126/science.1247003

Barbosa, J., Lozano-Soldevilla, D., & Compte, A. (2021). Pinging the brain with visual impulses reveals electrically active, not activity-silent, working memories. *PLOS Biology*, *19*(10), e3001436. https://doi.org/10.1371/journal.pbio.3001436

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*(1), 92-107. https://doi.org/10.1111/nyas.13320

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of Neuroscience*, *28*(40), 10056–10061. https://doi.org/10.1523/JNEUROSCI.1776-08.2008

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., & Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences*, *93*(6), 2382–2386. https://doi.org/10.1073/pnas.93.6.2382

Duncan, D. H., van Moorselaar, D., & Theeuwes, J. (2023). Pinging the brain to reveal the hidden attentional priority map using encephalography. *Nature Communications*, *14*(1), 4749. https://doi.org/10.1038/s41467-023-40405-8

Engel, S. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gardner, J. L., Merriam, E. P., Schluppeck, D., Besle, J., & Heeger, D. J. (2018). mrTools: Analysis and visualization package for functional magnetic resonance imaging data (Version 4.7) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.1299483

Gong, M., Chen, Y., & Liu, T. (2022). Preparatory attention to visual features primarily relies on non-sensory representation. *Scientific Reports*, *12*(1), 21726. https://doi.org/10.1038/s41598-022-26104-2

Gong, M., & Liu, T. (2020a). Biased neural representation of feature-based attention in the human frontoparietal network. *Journal of Neuroscience*, *40*(43), 8386-8395. https://doi.org/10.1523/JNEUROSCI.0690-20.2020

Gong, M., & Liu, T. (2020b). Continuous and discrete representations of feature-based attentional priority in human frontoparietal network. *Cognitive Neuroscience*, *11*(1–2), 47–59. https://doi.org/10.1080/17588928.2019.1601074

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

Grubert, A., & Eimer, M. (2018). The time course of target template activation processes during preparation for visual search. *The Journal of Neuroscience*, *38*(44), 9527–9538. https://doi.org/10.1523/JNEUROSCI.0409-18.2018

Guo, F., Preston, T. J., Das, K., Giesbrecht, B., & Eckstein, M. P. (2012). Feature-independent neural coding of target detection during search of natural scenes. *Journal of Neuroscience*, *32*(28), 9499–9510. https://doi.org/10.1523/JNEUROSCI.5876-11.2012

Hamblin-Frohman, Z., & Becker, S. I. (2021). The attentional template in high and low similarity search: Optimal tuning or tuning to relations? *Cognition*, *212*, 104732. https://doi.org/10.1016/j.cognition.2021.104732

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632-635. https://doi.org/10.1038/nature07832

Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*(1), 128–149. https://doi.org/10.3758/s13414-014-0764-6

Hout, M. C., Robbins, A., Godwin, H. J., Fitzsimmons, G., & Scarince, C. (2017). Categorical templates are more useful when features are consistent: Evidence from eye movements during search for societally important vehicles. *Attention, Perception, & Psychophysics*, *79*(6), 1578–1592. https://doi.org/10.3758/s13414-017-1354-1

JASP Team. (2023). JASP (Version 0.17.1) [Computer software]. https://jasp-stats.org/

Jia, K., Goebel, R., & Kourtzi, Z. (2023). Ultra-high field imaging of human visual cognition. *Annual Review of Vision Science*, *9*(1), 479–500. https://doi.org/10.1146/annurev-vision-111022-123830

Jia, K., Zamboni, E., Kemper, V., Rua, C., Goncalves, N. R., Ng, A. K. T., Rodgers, C. T., Williams, G., Goebel, R., & Kourtzi, Z. (2020). Recurrent processing drives perceptual plasticity. *Current Biology*, *30*(21), 4177-4187.e4. https://doi.org/10.1016/j.cub.2020.08.016

Jia, K., Wang, M., Steinwurzel, C., Ziminski, J. J., Xi, Y., Emir, U., & Kourtzi, Z. (2024). Recurrent inhibition refines mental templates to optimize perceptual decisions. *Science Advances*, *10*(31), eado7378. https://doi.org/10.1126/sciadv.ado7378

Jigo, M., Gong, M., & Liu, T. (2018). Neural determinants of task performance during feature-based attention in human cortex. *eNeuro*, *5*(1), 1–15. https://doi.org/10.1523/ENEURO.0375-17.2018

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761. https://doi.org/10.1016/S0896-6273(00)80734-5

Kerzel, D. (2019). The precision of attentional selection is far worse than the precision of the underlying memory representation. *Cognition*, *186*, 20–31. https://doi.org/10.1016/j.cognition.2019.02.001

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.

Kok, P., Failing, M. F., & De Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554. https://doi.org/10.1162/jocn_a_00562

Konen, C. S., & Kastner, S. (2008). Representation of eye movements and stimulus motion in topographically organized areas of human posterior parietal cortex. *Journal of Neuroscience*, *28*(33), 8361–8375. https://doi.org/10.1523/JNEUROSCI.1930-08.2008

Kong, G., Alais, D., & Van Der Burg, E. (2017). Orientation categories used in guidance of attention in visual search can differ in strength. *Attention, Perception, & Psychophysics*, *79*(8), 2246–2256. https://doi.org/10.3758/s13414-017-1387-5

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2015). Neural evidence for the flexible control of mental representations. *Cerebral Cortex*, *25*(10), 3303–3313. https://doi.org/10.1093/cercor/bhu130

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron*, 55(2), 313-323. https://doi.org/10.1016/j.neuron.2007.06.030

Liu, T., & Hou, Y. (2013). A hierarchy of attentional priority signals in human frontoparietal cortex. *Journal of Neuroscience*, *33*(42), 16606–16616. https://doi.org/10.1523/JNEUROSCI.1780-13.2013

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India, 2*, 49-55.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 8–8. https://doi.org/10.1167/9.11.8

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, *319*(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Myers, N. E., Walther, L., Wallis, G., Stokes, M. G., & Nobre, A. C. (2015). Temporal dynamics of attention during encoding versus maintenance of working memory: complementary views from event-related potentials and alpha-band oscillations. *Journal of Cognitive Neuroscience*, *27*(3), 492–508. https://doi.org/10.1162/jocn_a_00727

Ng, A. K. T., Jia, K., Goncalves, N. R., Zamboni, E., Kemper, V. G., Goebel, R., Welchman, A. E., & Kourtzi, Z. (2021). Ultra-High-Field neuroimaging reveals fine-scale processing for 3D perception. *The Journal of Neuroscience*, *41*(40), 8362–8374. https://doi.org/10.1523/JNEUROSCI.0065-21.2021

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12125–12130. https://doi.org/10.1073/pnas.1101042108

Priebe, N. J. (2016). Mechanisms of orientation selectivity in the primary visual cortex. *Annual Review of Vision Science*, *2*(1), 85-107. https://doi.org/10.1146/annurev-vision-111815-114456

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. http://www.palamedestoolbox.org

Rademaker, R. L., & Serences, J. T. (2017). Pinging the brain to reveal hidden memories. *Nature Neuroscience*, 20(6), 767-769. https://doi.org/10.1038/nn.4560

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*(8), 1336-1344. https://doi.org/10.1038/s41593-019-0428-x

Rosenberg, M. D., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., Ramani, R., Qiu, M., Constable, R. T., & Chun, M. M. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proceedings of the National Academy of Sciences*, *117*(7), 3797–3807. https://doi.org/10.1073/pnas.1912226117

Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology*, *29*, 181–186. https://doi.org/10.1016/j.copsyc.2019.03.014

Schluppeck, D., Curtis, C. E., Glimcher, P. W., & Heeger, D. J. (2006). Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. *The Journal of Neuroscience*, *26*(19), 5098–5108. https://doi.org/10.1523/JNEUROSCI.5330-05.2006

Schneegans, S., & Bays, P. M. (2017). Restoration of fMRI decodability does not imply latent working memory states. *Journal of Cognitive Neuroscience*, *29*(12), 1977–1994. https://doi.org/10.1162/jocn_a_01180

Scolari, M., Byers, A., & Serences, J. T. (2012). Optimal deployment of attentional gain during fine discriminations. *The Journal of Neuroscience*, *32*(22), 7723–7733. https://doi.org/10.1523/JNEUROSCI.5558-11.2012

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207-214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893. https://doi.org/10.1126/science.7754376

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545), 1350–1354. https://doi.org/10.1126/science.1063695

Sheremata, S. L., Somers, D. C., & Shomstein, S. (2018). Visual short-term memory activity in parietal lobe reflects cognitive processes beyond attentional selection. *Journal of Neuroscience*, 38(6), 1511-1519. https://doi.org/10.1523/JNEUROSCI.1716-17.2017

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition*, *28*(5–8), 420–432. https://doi.org/10.1080/13506285.2020.1825141

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences*, *106*(46), 19569–19574. https://doi.org/10.1073/pnas.0905306106

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., & Egner, T. (2016). Feature-based attention and feature-based expectation. *Trends in Cognitive Sciences*, *20*(6), 401–404. https://doi.org/10.1016/j.tics.2016.03.008

Van Doorn, J., Van Den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wen, T., Duncan, J., & Mitchell, D. J. (2019). The time-course of component processes of selective attention. *NeuroImage*, *199*, 396–407. https://doi.org/10.1016/j.neuroimage.2019.05.067

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34–49. https://doi.org/10.1037/0096-1523.18.1.34

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, *9,* 123. https://doi.org/10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546

Wolff, M. J., Kandemir, G., Stokes, M. G., & Akyürek, E. G. (2020). Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *Journal of Neuroscience*, 40(3), 671-681. https://doi.org/10.1523/JNEUROSCI.1194-19.2019

Yu, X., Hanks, T. D., & Geng, J. J. (2022). Attentional guidance and match decisions rely on different template information during visual search. *Psychological Science*, *33*(1), 105–120. https://doi.org/10.1177/09567976211032225

Yu, X., Zhou, Z., Becker, S. I., Boettcher, S. E. P., & Geng, J. J. (2023). Good-enough attentional guidance. *Trends in Cognitive Sciences*, *27*(4), 391–403. https://doi.org/10.1016/j.tics.2023.01.007

Zhang, H., & Luo, H. (2023). Feature-specific reactivations of past information shift current neural encoding thereby mediating serial bias behaviors. *PLOS Biology*, *21*(3), e3002056. https://doi.org/10.1371/journal.pbio.3002056

Acknowledgements

This work was supported by National Science and Technology Innovation 2030—Major Project 2021ZD0200409, National Natural Science Foundation of China (32371087, 32300855, 3200784), Fundamental Research Funds for the Central University (226-2024-00118), a grant from the MOE Frontiers Science Center for Brain Science & Brain- Machine Integration at Zhejiang University and Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences 2023-PT310-01.

Author contributions

Original idea by K.J. and M.G. Experiment programmed by Y.C. Data collected by Y.C. Analysis programmed by Y.C., K.J. and M.G. Manuscript drafted by Y.C., T.L., and M.G. Manuscript edited and revised by T.L., J. T., K.J., and M.G.

Data Availability

All data, analyses, and task codes have been made publicly available via the Open Science Framework at https://osf.io/ghaxv/

Competing interests

The authors declare no competing interests.

Supplementary Materials

Table of Content

Supplementary Methods

S1 Fig. Univariate BOLD results

S2 Fig. Sensory decoding in the perception task

Table 1. Descriptive statistics of Mahalanobis distance during preparatory attention

S3 Fig. Relationship between attentional modulation on Mahalanobis distance and RT

S4 Fig. Mahalanobis distance during stimulus selection period

S5 Fig. Relationship between information connectivity and RT

S6 Fig. Functional connectivity analysis

S7 Fig. Analysis of eye positions

933 **Supplementary Methods**

934 **Univariate Analysis**

935 **Deconvolution**. For the data in the attention task, we used a deconvolution approach by fitting each
936 voxel's time series with a general linear model (GLM) with seven regressors, four corresponding to the
937 long delay (5.5 s and 7.5 s) trials with correct responses (attended orientation × delay), two corresponding
938 to the short delay (1.5 s and 3.5 s) trials with correct responses, and the remaining four regressors
939 corresponding to incorrect trials for each length of the delay. Each trial was modelled by a set of twelve
940 finite impulse response functions (FIR) after the trial onset. For the perception task data, we used a GLM
941 with two regressors, corresponding to two features (leftward vs. rightward). Each trial was modelled by a
942 set of eight FIR after the trial onset. The design matrix was pseudo-inversed and multiplied by the time
943 series to obtain an estimate of the hemodynamic response (HRF) evoked by each condition. Correct trials
944 from the attention task and all trials from the perception task were entered into further univariate and
945 multivariate analysis.

946 **Attentional modulations on BOLD response.** To select active voxels during the task, we
947 discarded noisy voxels with mean responses larger than 10% signal change. Because the stimuli were
948 presented at the center of the display, areas across two hemispheres were combined. Using the r^2 values
949 from the deconvolution analysis to index the relevance of each voxel to the task, voxels in each brain area
950 were sorted in descending order of their r^2 values. The top-ranked 100 voxels in each combined ROI were
951 then selected for further analyses. The results remained highly similar when using different numbers of
952 voxels (e.g., 80 or 120).

953 To assess whether feature-based attention modulated overall BOLD response during the
954 preparatory and stimulus selection periods, we averaged the fMRI response across voxels for long delay
955 trials of the attention task, separately for each ROI, each attention condition and each time point. Then, we
956 averaged a time window of 4 to 6 s after the onset of the cue to index the preparatory activity, and averaged
957 a time window of 4 to 6 s after the onset of superimposed gratings to index the stimulus-evoked response.

To test the influence of attention in different brain regions, we applied two-way repeated-measures ANOVAs (attended orientation × region) on BOLD response, separately for activity during preparatory and stimulus selection periods.

Univariate Results

During the No-Ping session, when examining the group-level fMRI time courses, we observed slightly elevated activity during the preparation period, particularly in higher-order frontoparietal areas, followed by a robust response to the superimposed gratings (S1A Fig). To examine whether feature-based attention modulated overall BOLD in the attention task across regions, we conducted two-way repeated-measures ANOVAs (attended orientation × region) on the BOLD response during preparation and stimulus periods, separately. The results showed a main effect of region in the preparation period ($F(3,57) = 6.748$; $p = 0.001$; $\eta_p^2 = 0.262$) and stimulus periods ($F(3,57) = 38.230$; $p < 0.001$; $\eta_p^2 = 0.668$). No other effects were significant in both periods (preparation: $ps > 0.526$; stimulus: $ps > 0.135$). These results demonstrate stronger response in higher-order frontoparietal areas than in sensory areas. The lack of attention modulation on BOLD response suggests that attending to specific features did not influence mean neural activity during the preparation and stimulus selection periods. These results are consistent with our previous studies using similar paradigms.

The presence of visual impulse during preparation in the Ping session yielded a similar time course of neural activity (S1B Fig). Consistent with the observation in No-Ping session, two-way repeated-measures ANOVAs revealed a main effect of region in both the preparation ($F(3,57) = 3.445$; $p = 0.023$; $\eta_p^2 = 0.153$) and stimulus periods ($F(3,57) = 29.911$; $p < 0.001$; $\eta_p^2 = 0.612$). No other effects were significant in both periods (preparation: $ps > 0.462$; stimulus: $ps > 0.063$). These results suggest that a briefly presented impulse stimulus caused little changes to the neural response amplitude and promoted us to further investigate multivariate neural activity patterns.

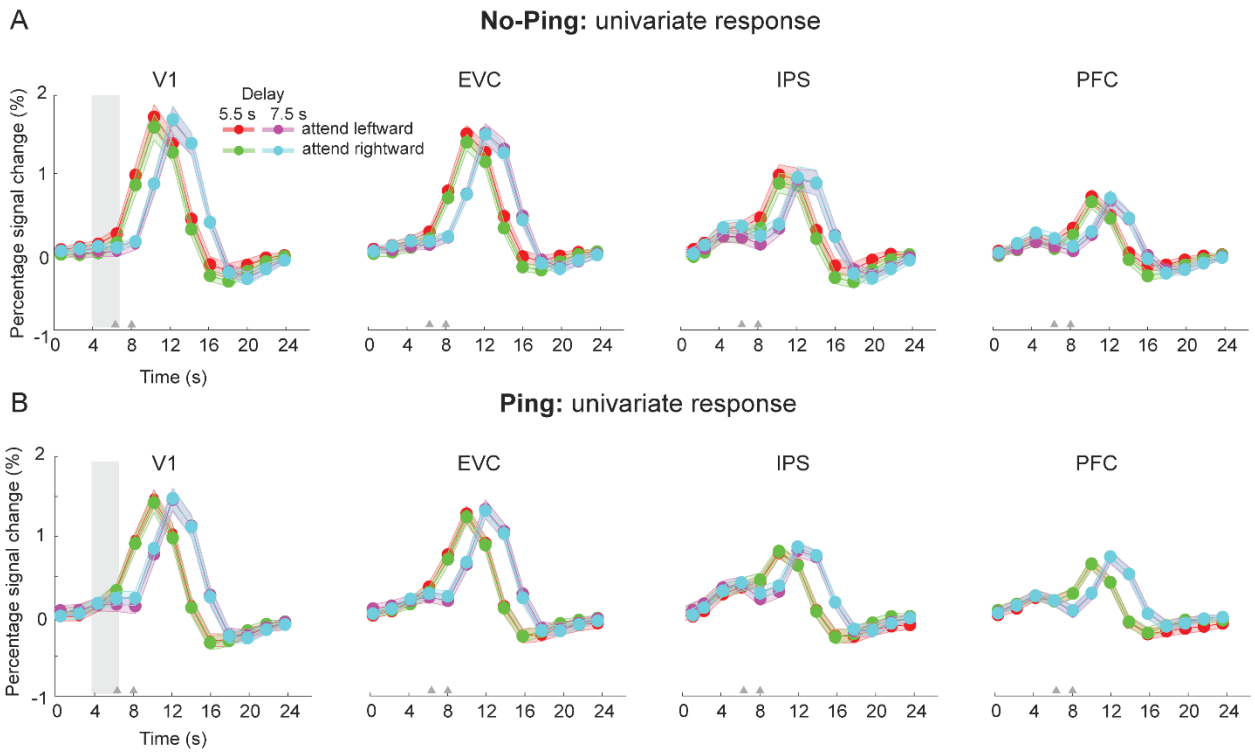


Fig S1. Univariate BOLD results. (A). Mean fMRI time course from four regions (V1, EVC, IPS, PFC) for No-Ping session and (B) Ping session. The gray triangle indicates the onset of superimposed gratings. The gray area indicates a slightly elevated response in the presence of visual impulse. Error bars denote standard error of the means.

Multivariate Results

Orientation decoding in the perception task. The decoding accuracy was significantly above-chance across regions in both sessions (*ps* < 0.004, S2 Fig). A two-way mixed ANOVA (sessions × region) revealed a main effect of region ($F(3,114) = 84.274$, $p < 0.001$, $\eta_p^2 = 0.689$), Neither the main effect of session ($F(1,38) = 1.103$, $p = 0.300$, $\eta_p^2 = 0.028$; $BF_{excl} = 2.605$) nor an interaction effect was significant ($F(3,114) = 0.592$, $p = 0.621$, $\eta_p^2 = 0.015$; $BF_{excl} = 5.809$).

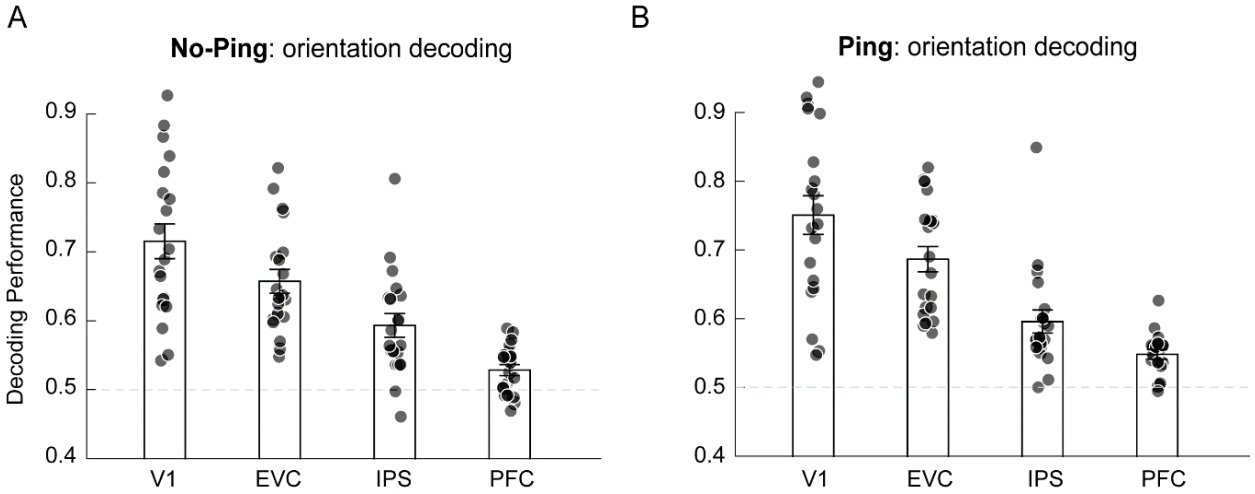


Fig S2. Sensory decoding in the perception task. (A) Decoding of orientations in No-Ping and (B) Ping sessions. The dashed lines represent theoretical chance level (0.5). Each dot represents one participant. Error bars denote standard error of the means.

Attentional modulation on the Mahalanobis distance during preparatory period and its relationship with behavior. Analysis of Mahalanobis distance showed greater pattern similarity between "attend leftward" and "perceived leftward" in the Ping session in comparison to the No-Ping session during preparatory attention in visual cortex. To enhance the understanding of the data, we summarized the descriptive statistics of Mahalanobis distance (data from Fig. 3B and 3C in the main text)) in Table 1.

Table 1. Descriptive statistics of Mahalanobis distance during preparatory attention

	No-Ping session		Ping session	
V1	Attend leftward	Attend rightward	Attend leftward	Attend rightward
Perceive leftward	6.73	6.81	6.55	6.92
Perceive rightward	6.84	6.95	7.04	6.67
EVC				
Perceive leftward	6.17	6.15	6.07	6.20
Perceive rightward	6.14	6.04	6.27	6.17

Using the attentional modulation index (AMI = $(D_{different} - D_{same})/(D_{different} + D_{same})$, see Methods) calculated based on trial-wise Mahalanobis distance in V1, we sorted the AMI values in a descending order and selected the top-ranked 25% of trials (i.e., 36 trials) and bottom-ranked 25% of trials to represent "strong modulation" and "weak modulation" trials, respectively. We then extracted behavioral responses on these selected trials and calculated RT and accuracy for each trial type in each session. The analysis revealed faster responses in the "strong modulation" than "weak modulation" trials in Ping session (paired t-test: t(19) = -2.746, p = 0.013, Cohen's d = -0.614; S3B Fig), but not in the No-Ping session (paired t-test: t(19) = -1.487, p = 0.154, Cohen's d = -0.332; S3A Fig). These effects were not observed for accuracy (ps > 0.163).

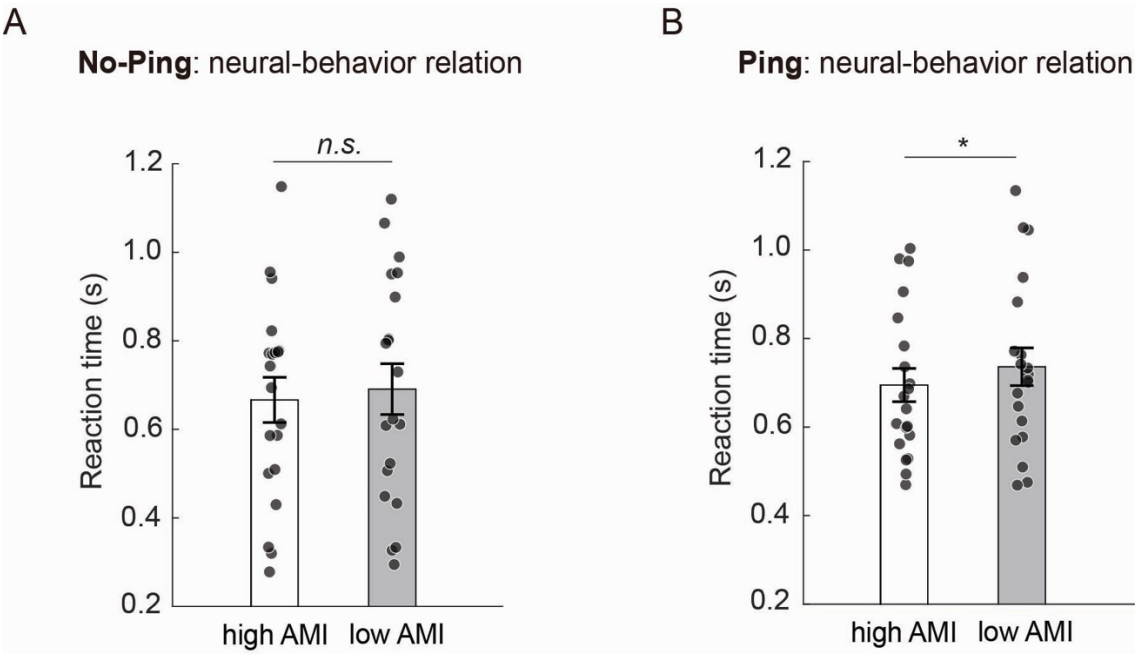


Fig S3. Relationship between attentional modulation on Mahalanobis distance and RT. (A) Behavioral performance (as indexed by RT) for "strong modulation" (high AMI) trials and "weak modulation" (low AMI) trials, sorted by V1 preparatory activity in the No-Ping and (B) Ping sessions. Each dot represents one subject's data. Error bars denote SEM. * $p < 0.05$

Attentional modulation on the Mahalanobis distance during stimulus selection. To examine whether the neural representations during stimulus selection periods reflect a sensory-like template, we

used two-way repeated-measures ANOVAs (attended orientation × orientation) on the Mahalanobis distance during stimulus selection. As shown in S4 Fig, we found significant interaction effects in both the No-Ping session (V1: $F(1,19) = 12.470$, $p = 0.002$, $\eta_p^2 = 0.396$; EVC: $F(1,19) = 26.538$, $p < 0.001$, $\eta_p^2 = 0.583$; IPS: $F(1,19) = 4.250$, $p = 0.053$, $\eta_p^2 = 0.183$; PFC: $F(1,19) = 1.557$, $p = 0.227$, $\eta_p^2 = 0.076$) and Ping session (V1: $F(1,19) = 11.698$, $p = 0.003$, $\eta_p^2 = 0.381$; EVC: $F(1,19) = 12.742$, $p = 0.002$, $\eta_p^2 = 0.401$; IPS: $F(1,19) = 5.227$, $p = 0.034$, $\eta_p^2 = 0.216$; PFC: $F(1,19) = 0.087$, $p = 0.771$, $\eta_p^2 = 0.005$). These results support the predicted role of attention in selectively enhancing the representation of task-relevant features while filtering out task-irrelevant ones, leading to neural representations that closely resembled the perception of single features.

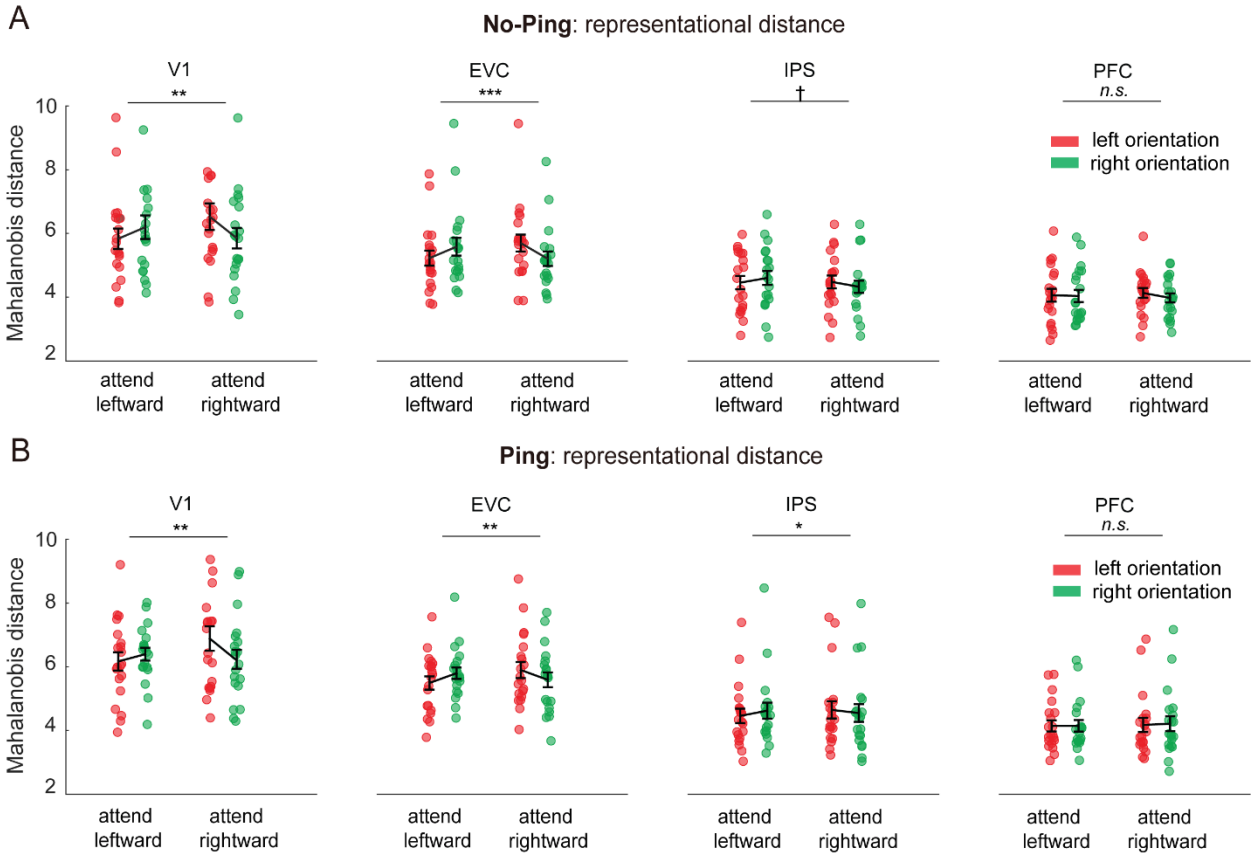
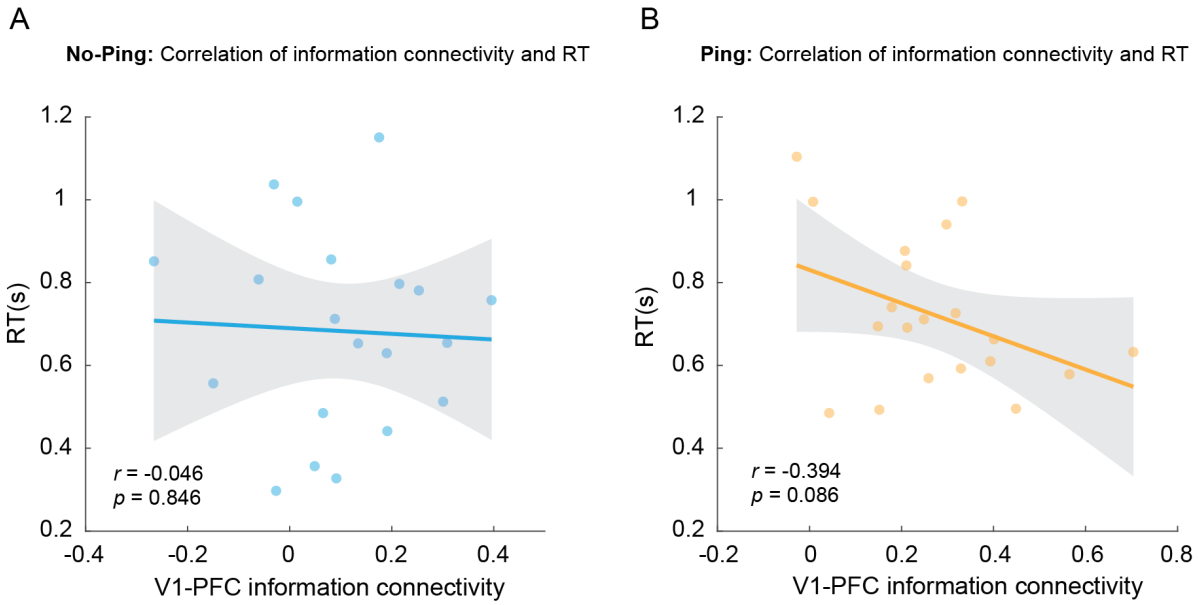


Fig S4. Mahalanobis distance during stimulus selection period. (A) Attentional modulation on Mahalanobis distance in No-Ping and (B) Ping sesions. Each dot represents one participant. Error bars denote standard error of the means. † $p < 0.06$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1035 **Correlation between information connectivity (IC) and behavior.** To assess whether ICs relate
1036 to behavior, we computed Pearson correlations between ICs and both RT and accuracy across participants,
1037 separately for each session (S5 Fig). We observed a trend of correlations between V1-PFC connectivity and
1038 RTs in the Ping session (r = -0.394, p = 0.086), but not in the No-Ping session (r = -0.046, p = 0.846). No
1039 significant correlations were found between V1-IPS and RTs (ps > 0.400) or between ICs and accuracy
1040 (ps > 0.399).



1041

1042 **Fig S5. Relationship between information connectivity and RT.** (A) Correlations between V1-PFC connectivity
1043 and RTs in the No-Ping and (B) Ping sessions. Each dot represents one subject's data. The shaded area represents the
1044 confidence intervals of the regressed lines.

1045

1046 **Pattern-based and mean BOLD-based functional connectivity.** To examine whether impulse
1047 altered network function during the stimulus selection period, we conducted a pattern-based information
1048 connectivity analysis (as Methods and Materials), using AMI based on cross-validated Mahalanobis
1049 distance during the stimulus selection period. We found no significant differences in functional connectivity
1050 between No-Ping and Ping sessions (ps > 0.224; S6A Fig). To further explore whether the impulse-driven
1051 connectivity changes during preparation were influenced by overall changes in BOLD response, we

conducted a standard functional connectivity analysis based on mean BOLD response over time. This analysis revealed no significant changes in inter-cortical connections between the two sessions (*ps* > 0.136; S6B Fig).

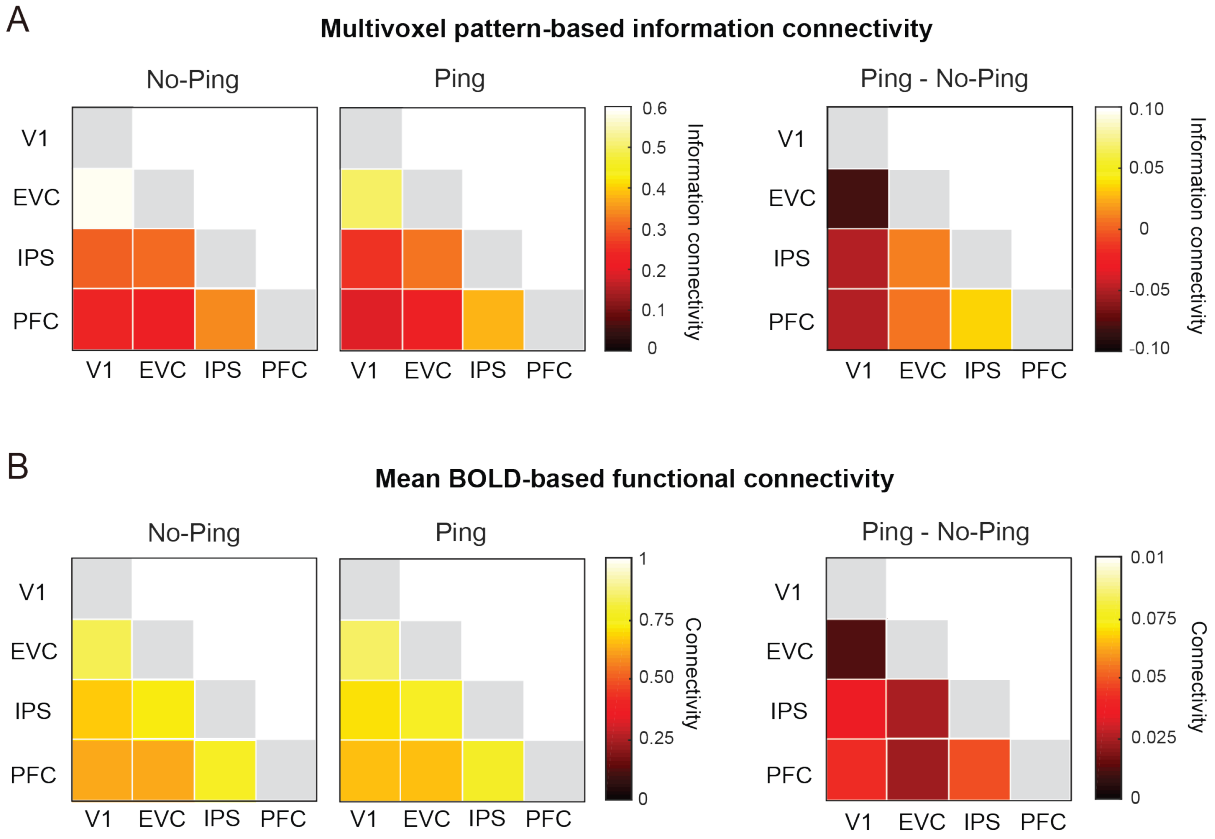


Fig S6. Functional connectivity analysis. (A) Connectivity based on multivoxel patterns during stimulus selection period: between-region connectivity in the No-Ping session and Ping session based on cross-validated Mahalanobis distance (left panel) and the differences in connectivity between the two sessions (right panel). (B) Connectivity based on mean BOLD responses over time: between-region connectivity in the No-Ping session and Ping session based on raw time series (left panel) and the differences in connectivity between the two sessions (right panel).

Analysis of Eye Movement

Although participants were required to maintain stable fixation in our experiment, they might have shifted their spatial attention differently when preparing to attend different features (e.g., eyes moved towards

leftward or rightward). To assess the potential influence of overt spatial attention in our results, we analyzed the participants' eye position recorded during the training session and observed no significant difference between attended features during preparation in the No-Ping ($p = 0.328$, $BF_{01} = 2.760$ for horizontal; $p = 0.231$, $BF_{01} = 2.210$ for vertical; S7A Fig) and Ping sessions ($p = 0.450$, $BF_{01} = 3.229$ for horizontal; $p = 0.588$, $BF_{01} = 3.675$ for vertical; paired t-tests; S7B Fig). To examine whether there were any differences in eye movement between two sessions, a two-way mixed ANOVA (sessions × attended orientation) was applied to the eye position data. The analyses revealed neither main effects nor an interaction effect for the horizontal ($ps > 0.222$, $BF_{excl} > 1.593$) and vertical eye movements ($ps > 0.158$, $BF_{excl} > 1.185$). We note that we only collected eye-tracking data during behavioral training and not during scanning due to technical limitations. However, it seems unlikely that the participants would adopt a different eye fixation strategy after training, especially given the centrally presented stimulus configuration that made eye movement an ineffective strategy.

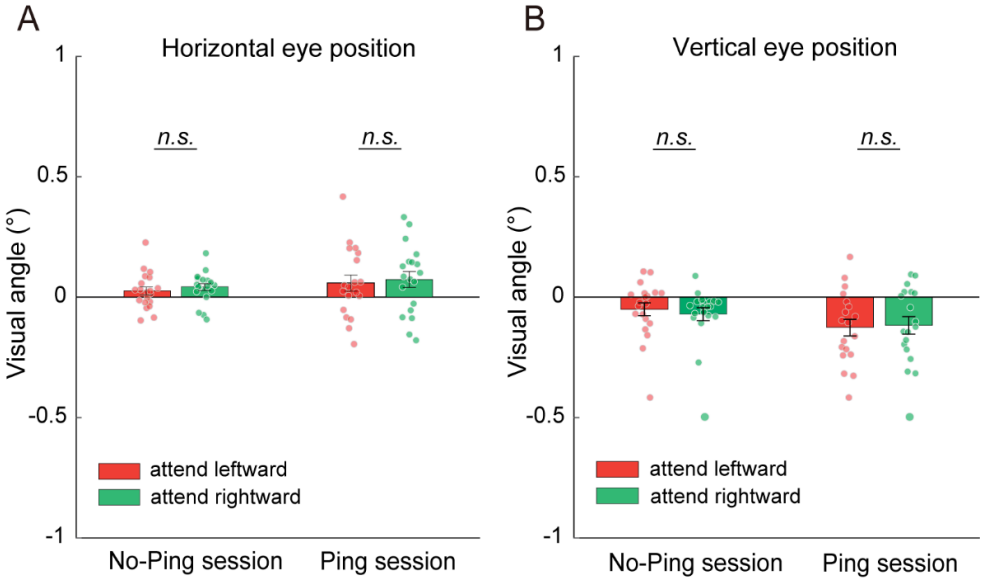


Fig S7. Analysis of eye positions. (A) Group-level horizontal eye position and (B) vertical eye position for the preparation period during behavioral training in No-Ping and Ping sessions. Each dot represents one participant. Error bars denote standard error of the means.