Behavioral/Cognitive

Time Course of Orientation Ensemble Representation in the Human Brain

Xizi Gong,[1,2,3,4] Tao He,[5] Qian Wang,[1,2,3,4] Junshi Lu,[1,2] and Fang Fang[1,2,3,4]

[1]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, People's Republic of China, [2]IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, People's Republic of China, [3]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, People's Republic of China, [4]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, People's Republic of China, and [5]Center for the Cognitive Science of Language, Beijing Language and Culture University, Beijing 100083, People's Republic of China

Natural scenes are filled with groups of similar items. Humans employ ensemble coding to extract the summary statistical information of the environment, thereby enhancing the efficiency of information processing, something particularly useful when observing natural scenes. However, the neural mechanisms underlying the representation of ensemble information in the brain remain elusive. In particular, whether ensemble representation results from the mere summation of individual item representations or it engages other specific processes remains unclear. In this study, we utilized a set of orientation ensembles wherein none of the individual item orientations were the same as the ensemble orientation. We recorded magnetoencephalography (MEG) signals from human participants (both sexes) when they performed an ensemble orientation discrimination task. Time-resolved multivariate pattern analysis (MVPA) and the inverted encoding model (IEM) were employed to unravel the neural mechanisms of the ensemble orientation representation and track its time course. First, we achieved successful decoding of the ensemble orientation, with a high correlation between the decoding and behavioral accuracies. Second, the IEM analysis demonstrated that the representation of the ensemble orientation differed from the sum of the representations of individual item orientations, suggesting that ensemble coding could further modulate orientation representation in the brain. Moreover, using source reconstruction, we showed that the representation of ensemble orientation manifested in early visual areas. Taken together, our findings reveal the emergence of the ensemble representation in the human visual cortex and advance the understanding of how the brain captures and represents ensemble information.

Key words: ensemble coding; inverted encoding model; MEG; visual cortex

---

Significance Statement

Ensemble coding, a cognitive process of extracting summary statistical information from groups of similar items, stands as a pivotal strategy enabling humans to efficiently process complex natural scenes with limited sensory capacities. However, the neural mechanisms of ensemble coding remain largely unknown. Recent modeling studies have predominantly highlighted the importance of the summed activation across all items in ensemble coding. Intriguingly, here, we show that ensemble orientation representation differed from the summed representation of all component item orientations, suggesting that ensemble coding incorporates additional processes beyond mere summation. Additionally, we explore how the ensemble orientation representation per se evolved in the human visual cortex. Our findings significantly extend our understanding of ensemble coding.

Introduction

Despite our capacity-limited visual system, humans continuously process richly detailed natural scenes (Cohen et al., 2016; Whitney and Leib, 2018; Fu et al., 2021). The apparent gap between our subjective rich perceptual experience and our objective limited sensory capacity has been extensively discussed. One way to reconcile this gap is through ensemble coding, a process that leverages the regularity and redundancy in natural scenes (Cohen et al., 2016; Baek and Chong, 2020). Ensemble coding

allows the brain to condense a large amount of useful information into summary statistical descriptors, such as mean (Haberman and Whitney, 2009; Leib et al., 2016), variance (Solomon, 2010; Michael et al., 2014), and outliers (Cant and Xu, 2020; Epstein et al., 2020) from a group of similar items, remarkably enhancing the efficiency of information processing. The pervasiveness of ensemble coding in perception (Whitney and Leib, 2018; Corbett et al., 2023) further emphasizes its importance in shaping our rich perceptual experience. Notably, ensemble coding has been reported for stimuli at various levels of visual processing, ranging from low-level features, such as orientation (Dakin and Watt, 1997; Parkes et al., 2001; Michael et al., 2014) and hue (Maule et al., 2014; Webster et al., 2014), to high-level semantics, such as emotion (Haberman and Whitney, 2009; Im et al., 2017) and lifelikeness (Leib et al., 2016).

Despite the fundamental role of ensemble coding in information processing of the brain, its neural mechanisms remain elusive. Previous neuroimaging studies have primarily focused on identifying the neural substrates involved in ensemble representations. For example, Cant and Xu (2012, 2017, 2020) found that the parahippocampal place area and retrosplenial cortex were highly sensitive to changes in object ensembles. In addition, Im and colleagues reported enhanced activation in the intraparietal sulcus (IPS) and the superior frontal gyrus in response to emotion ensembles compared with individual facial expressions (Im et al., 2017). However, two key aspects of the neural mechanisms underlying ensemble coding remain unclear. First, the neural representation of ensemble information per se is still not clear. It is possible that activations in the aforementioned regions might result from summation of the neural responses to all individual items rather than a process specific to ensemble coding (Corbett et al., 2023). Second, few studies have examined the temporal aspects (e.g., time courses) of the neural mechanisms of ensemble coding. Both issues are crucial for gaining a comprehensive picture of ensemble coding and its relationship with the coding of individual items.

In this study, our aims were two-fold. First, we aimed to distinguish the neural responses to ensemble information from the summed neural responses to individual items (i.e., simple summation hypothesis). To this end, we utilized a set of orientation ensembles where none of the individual item orientations were the same as the ensemble orientation (i.e., mean orientation; Fig. 1B). This differed from most previous studies (Dakin and Watt, 1997; Dakin, 2001; Solomon, 2010; Attarha and Moore, 2015; Epstein and Emmanouil, 2021; Tark et al., 2021), where the ensemble information of the stimuli was often very similar or even identical to the feature information of some individual items. Second, we aimed to unveil the time course of ensemble representation by recording high temporal-resolution magnetoencephalography (MEG) signals while participants performed an ensemble orientation discrimination task. Participants could estimate the ensemble orientations and perform an ensemble orientation discrimination task with our stimuli. The ensemble orientation could be reliably decoded from the recorded MEG signals, and the peak decoding accuracy was highly correlated with the behavioral accuracy. Furthermore, by virtue of the inverted encoding model (IEM), we converted the recorded MEG signals to orientation channel responses, thereby investigating how the ensemble orientation representation emerges over time. We found that orientation channel responses were modulated ∼370 ms after stimulus onset, suggesting that ensemble coding engages specific neural processes. Finally, using source reconstruction, we showed that the representation of ensemble orientation manifested in early visual areas.

Materials and Methods

Participants. Twenty-three healthy participants (five males; mean age, 22.86) were recruited for the main experiment and fifteen (nine males; mean age, 22.27) for the control experiment. All participants reported normal or corrected-to-normal vision and had no known neurological or visual disorders. Each participant provided written informed consent prior to the study in accordance with the procedures and protocols approved by the human subject review committee of Peking University.

Stimuli and design. In our study, the visual stimulus consisted of 32 items (oriented bars, 0.2° × 0.7°). These items were randomly but evenly distributed on three invisible concentric circles (radius = 1.5°, 2.8°, and 4.1°, respectively) centered at fixation (item number = 5, 11, and 16). In addition, each item was further jittered <0.2°. This stimulus design aimed to disrupt any stimulus configuration and discourage participants from paying attention to particular fixed locations. An item could be at one of nine possible orientations, from 5° to 165° in steps of 20° (i.e., 5°, 25°, 45°, 65°, 85°, 105°, 125°, 145°, and 165°; Fig. 1A). The visual stimulus could be either homogeneous or heterogeneous in terms of item orientation (Fig. 1B). In a homogeneous stimulus, all item orientations ($\alpha°$) were identical. Therefore, the ensemble orientation ($\theta°$, i.e., the mean orientation of all items) equates to its component item orientation ($\alpha° = \theta°$). In a heterogeneous stimulus, its ensemble orientation ($\theta°$) could be one of the nine possible orientations. Here, the ensemble orientation was $\theta°$, and its component item orientations were $\theta° \pm 20°$ and $\theta° \pm 40°$ (eight items for each orientation). Note that there was no item with the orientation of $\theta°$ in heterogeneous stimuli.

Participants were required to perform an ensemble orientation discrimination task in the MEG (Fig. 1C). They were instructed to maintain fixation on the central dot, pay attention to all items, and estimate the ensemble orientation. In a trial, a homogeneous or heterogeneous stimulus was displayed for 500 ms, followed by a 500 ms interstimulus interval (ISI), during which only a fixation point was presented. A black probe line (length, 9°; orientation, $\theta° \pm 10°$ or $\theta° \pm 20°$) was then presented for 500 ms. Participants pressed a key to indicate whether the probe line was oriented clockwise or counterclockwise relative to the ensemble orientation of the first stimulus.

All stimuli were generated and controlled using Psychtoolbox-3 (Matlab; Pelli, 1997) and were projected (spatial resolution, 1,024 × 768; refresh rate, 60 Hz) onto a translucent screen inside a dimly lit, magnetically shielded room. Participants viewed the stimuli from a fixed distance of 85 cm. Throughout the main experiment, participants were instructed to maintain fixation, and their eye movements were monitored using an EyeLink 1000 Plus eye tracker (SR Research). Before the experiment, participants practiced the task along with feedback to ensure a clear understanding of the task. In the main experiment, they completed at least five runs with homogeneous stimuli and at least eight runs with heterogeneous stimuli. Each run consisted of 108 trials (12 trials for each of the nine ensemble orientations) in a randomized order.

In addition to the main experiment described above, we also performed a control experiment in which trials with homogeneous and heterogeneous stimuli were mixed within the same run. All other experimental settings were identical. Participants completed at least 10 runs, each consisting of 108 trials. The composition of these runs varied: some included 36 homogeneous and 72 heterogeneous trials, while others included 72 homogeneous and 36 heterogeneous trials, all arranged in a randomized order.

MEG acquisition and preprocessing. MEG data were collected using a 306-channel (204 planar gradiometer sensors and 102 magnetometer sensors) Elekta Neuromag TRIUX system at a sampling rate of 1,000 Hz. Raw data were first preprocessed offline with Maxfilter Software (Elekta) using the temporal extension of the signal space separation method for noise reduction. The data were then processed using MNE-python (Gramfort et al., 2013). Data were bandpass filtered between 0.1 and 45 Hz. To remove eyeblink artifacts, an independent component analysis was performed. Trials were extracted from 120 ms before to 1,000 ms after stimulus onset. Data were checked via visual
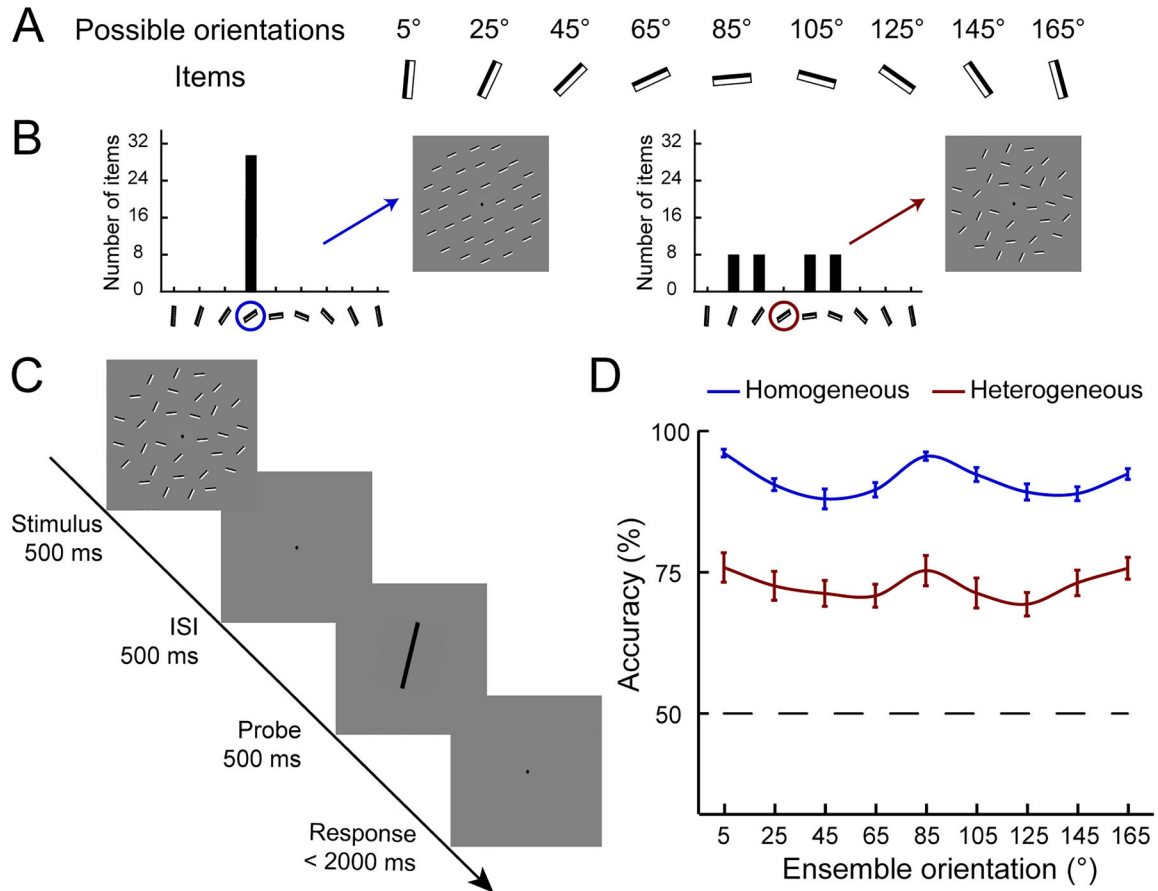
Figure 1. Stimuli and behavioral task. **A**, The nine possible orientations in the visual stimuli ranging from 5° to 165° in steps of 20° and schematic descriptions of the component items at the corresponding item orientations (*a*°). **B**, Schematic descriptions of the two types of visual stimuli: homogeneous (left, blue) and heterogeneous (right, red) stimuli. Colored circles indicate the ensemble orientation (i.e., mean orientations, θ°; here, θ° = 65°) of the visual stimuli. Black bars in the graph indicate the number of items corresponding to each of the nine orientations. **C**, Schematic description of the ensemble orientation discrimination task. Participants performed the ensemble orientation discrimination task at all nine possible orientations with both heterogeneous and homogenous stimuli. **D**, Behavioral performance in discriminating ensemble orientations with heterogeneous and homogeneous stimuli. Accuracies are plotted as a function of the ensemble orientation; the dashed line indicates a 50% chance level.

inspection for artifacts, and trials with artifacts were excluded from further analysis. All the remaining trials were baseline corrected and resampled to 250 Hz to increase the signal-to-noise ratio.

Sensor-level decoding. To decode the ensemble orientation information in the stimuli, we applied multivariate pattern analysis (MVPA) to MEG sensor signals. The decoding analysis was based on linear support vector machines using MNE-python and was conducted separately for each participant and stimulus type. First, we randomly discarded some trials to guarantee that an equal number of trials was available for each of the nine ensemble orientations. Then, all the remaining trials were split into four subsets with random assignment and no replacement. To reduce trial-to-trial noise, trials with the same orientation in each subset were averaged and normalized (Guggenmos et al., 2018). Seventy-two sensors, labeled as "Occipital" in the MEG data acquisition system, were selected for the decoding analyses. Therefore, for each timepoint, the MEG data were arranged in the form of 72-dimensional vectors, yielding 36 (4 subsets × 9 ensemble orientations) vectors per timepoint. Next, a nine-way decoder was trained to classify these vectors into one of the nine ensemble orientations using a four-fold cross-validation procedure. The aforementioned procedure was repeated 100 times, each time with a new random trial assignment. Finally, the resulting decoding accuracies were averaged over repetitions, yielding an overall decoding accuracy time course for each participant and stimulus type.

We further investigated how persistent the orientation representations were (Cichy et al., 2014; Isik et al., 2014; King and Dehaene,

2014; Dobs et al., 2019) by extending the decoding procedure with a temporal generalization approach; i.e., the decoders were trained to distinguish ensemble orientation representations at each timepoint but were tested on data at all other timepoints, thereby generating a temporal generalization matrix for each participant and stimulus type (Fig. 2B,E,H). The decoder would exhibit above-chance decoding accuracies at test timepoints when the orientation representations were similar to those at the training timepoint. Based on the temporal generalization matrix of homogeneous stimuli, we aimed to identify an optimal period during which sensor signals contained the ensemble orientation information most generalizable to other timepoints. We calculated the generalization index for each training timepoint by counting the number of test timepoints (up to 500 ms after stimulus onset) at which the decoder exhibited significant decoding accuracies. We defined the optimal timepoint as that with the highest diagonal decoding accuracy among the top 20 most generalized timepoints. The 80 ms optimal period was centered at the optimal timepoint.

Next, we trained cross-decoders between the data elicited by homogeneous and heterogeneous stimuli and generated a temporal generalization matrix for the cross-decoding (Oosterhof et al., 2012). Last, we extracted and averaged the decoding accuracies for the training timepoints in the optimal period, yielding one cross-decoding accuracy time course for each participant.

IEM. In addition to the decoding analyses, we implemented the IEM to reconstruct the ensemble orientation representation in heterogeneous stimuli. In our study, MEG responses to the visual stimulus were

expressed as a weighted sum of the responses of the nine hypothetical orientation channels (5°–165°, in steps of 20°):

$$B_1 = W\ C_1,\qquad(1)$$

where B_1 (n sensors $\times\ m$ trials) is the matrix of sensor signals at a given timepoint across trials, W (n sensors $\times$ 9 channels) is the matrix of linear weights mapping the orientation channel space to the sensor space, and C_1 (9 channels $\times\ m$ trials) is a design matrix of channel responses.

The IEM analysis consisted of two stages. In the first stage, we performed a model-based encoding. Data elicited by homogeneous stimuli was used to construct an encoding model and to estimate the weight matrix W. To this end, we first extracted the orientation pattern (i.e., B_1) from the MEG sensor signals at each timepoint in the optimal period. Based on previous studies (Mo et al., 2019, 2022; Rademaker et al., 2019), we modeled the idealized orientation tuning functions as half-sinusoidal functions raised to the eighth power peaked at the nine possible ensemble orientations. Hence, for each trial in the model training sessions, channel responses (i.e., C_1) could be predicted from these idealized tuning functions. Accordingly, we could estimate the weight matrix with the least-square linear regression:

$$\hat{W} = B_1\ C_1^T\ (C_1\ C_1^T)^{-1}.\qquad(2)$$

Next, in the model-based reconstruction stage, we applied the estimated weight matrix $\hat{W}$ with its Moore–Penrose pseudoinverse to the data elicited by heterogeneous stimuli at all timepoints in order to acquire instantaneous channel responses. To reconstruct the orientation representation of heterogeneous stimuli, we assumed that this weight matrix was invariant across timepoints and tasks, an assumption supported by the results from the temporal generalization matrixes:

$$C_2 = (\hat{W}^T\ \hat{W})^{-1}\ \hat{W}^T B_2.\qquad(3)$$

The equation above provides model-based reconstructions of the instantaneous channel responses on a trial-to-trial basis for individual participants. Note that to increase the signal-to-noise ratio, the MEG data underwent similar random assignment and averaging procedures as those in the decoding analysis (see above, Sensor-level decoding, paragraph 1). The encoding stage was repeated 20 times, and the reconstruction stage was repeated 10 times per encoding stage. The reconstructions were recentered at the ensemble orientation and averaged across trials and are stated as "channel response profile" hereafter. Hence, we converted the time-resolved MEG signals into orientation channel responses (0°, ±20°, ±40°, ±60°, ±80°; 0° indicates the ensemble orientation). Note that both channel response profiles—whether predicted or reconstructed—were obtained using encoding models estimated with the same data elicited by homogeneous stimuli. Therefore, the two profiles stand on an equal footing and they are comparable (Sprague et al., 2018, 2019).

To quantify the quality of the orientation information encoded, we computed the representational fidelity of the channel response profiles as in previous studies (Oh et al., 2019; Rademaker et al., 2019; Tark et al., 2021). For each timepoint, the recentered channel responses were taken as nine vectors in the orientation space (0°–180°), each pointing to the corresponding channel orientation. These vectors were then projected to a new orientation space spanning 0°–360°. The representational fidelity was the sum of these vectors at 0° (i.e., the ensemble orientation). A larger value of fidelity indicates a preferentially stronger representation at the ensemble orientation.

Bayesian probabilistic decoding. To further validate our findings, we utilized a probabilistic decoding approach known as TAFKAP (Li et al., 2021; van Bergen and Jehee, 2021) to estimate the probability of different ensemble orientations producing the measured MEG responses. This approach integrates a generative model with Bayesian inference. To our knowledge, this is the first application of TAFKAP to MEG data. According to the generative model, MEG responses to the ensemble orientation stimuli could be represented as a weighted sum of the responses of the nine hypothetical orientation channels (5°–165°, in steps of 20°), combined with noise. This noise consisted of two components: a channel-specific component that was shared among sensors with similar channel tuning functions and a sensor-specific component that was unique to each sensor. Both components followed zero-mean Gaussian distributions. This led to a simple form of noise covariance matrix Ω as follows:

$$\Omega = \sigma^2 \hat{W}\hat{W}^T +\ \rho\tau\tau^T + (1-\rho)I \circ \tau\tau^T.\qquad(4)$$

The first term captured the covariance of the channel-specific component, where σ was the standard deviation of orientation channels, $\hat{W}$ was the estimated weight matrix from Equation 2. The last two terms approximated the covariance of the sensor-specific component, where I is the identity matrix, $\circ$ denotes element-wise multiplication, τ is a vector containing each sensor's independent standard deviation, and ρ is a scalar between 0 and 1, which controlled the amount of shared variability among sensors irrespective of their channel tuning functions (for detailed derivations, see van Bergen et al., 2015; van Bergen and Jehee, 2021).

For each participant, the generative model was trained using data elicited by homogeneous stimuli during the optimal period. Note that every five MEG trials were averaged to generate a new trial for further analysis, following procedures similar to those in the decoding and IEM analyses. In TAFKAP, the training data were resampled with replacement to generate multiple bootstrap datasets. For each resampled training set j, the Ω_j was further regularized using the empirical covariance matrix derived from that training set. Then, the free parameters $\theta_j = \{\hat{W}_j, \Omega_j\}$ were estimated using ordinary least squares. The obtained generative model was then applied to the test dataset. Using the Bayes' rule and a flat prior, the posterior probability of a stimulus s at a certain ensemble orientation given the MEG responses b was computed as follows:

$$p(s|b; \theta_j) = \frac{p(b|s; \theta_j)p(s)}{\int p(b|s; \theta_j)p(s)ds}.\qquad(5)$$

The obtained posterior probabilities were averaged over resamples, yielding the posterior probability function for all possible orientations based on the measured MEG signals.

To evaluate the TAFKAP's performance with MEG data, we conducted a cross-validation analysis using data elicited by homogeneous stimuli. First, as expected, the obtained probability function was flat before stimulus onset and peaked at the ensemble orientation during the optimal period, indicating that the TAFKAP effectively captures orientation information from MEG data. Second, we benchmarked the TAFKAP's performance against that of the MVPA by transforming its probability function into classification accuracy. We found a significant positive correlation between the accuracies of the two methods, further validating the TAFKAP. These findings strongly supported the applicability of the TAFKAP to MEG data, while future studies with simulations could provide additional insights into its underlying assumptions and further validate its robustness.

Importantly, in our study, the trained generative model was applied to two independent test datasets: one dataset was MEG signals elicited by heterogeneous stimuli, and the other dataset was synthesized data using MEG signals elicited by four homogeneous stimuli. The synthesized data were the simple summation of the MEG signals elicited by each of the four item orientations, which were measured separately using the homogeneous stimuli with corresponding orientations and then multiplied with 0.25 (MacEvoy and Epstein, 2009; Baeck et al., 2013). According to the simple summation hypothesis, the two test datasets should be indistinguishable. For both test datasets, this Bayesian probabilistic decoding procedure was repeated 100 times, and the obtained posterior probability functions were recentered at the ensemble orientation and averaged across trials.

Source reconstruction. To better understand how the ensemble orientation representation emerges along the visual hierarchy, we performed source reconstruction to estimate the responses in different regions of interest (ROIs). Source reconstruction was performed using

MNE-python and the FreeSurfer toolbox. For each participant, structural magnetic resonance images (MRI) were collected with a 3 T Siemens Prisma (T1-weighted; 3D MPRAGE; $0.5 \times 0.5 \times 1$ mm^3 resolution). Then individual cortical surfaces were reconstructed using the default *recon-all* process and segmented using the *watershed* algorithm in FreeSurfer. Based on the reconstructed surfaces (inner skull surface), individual volume conductions were estimated using single-layer boundary element models (BEMs). We then set up individual source spaces comprising 4,096 points per hemisphere (corresponding to 4.9 mm source spacing). After manually coregistering the MEG data to the MRI coordinate system using the head-digitized shape and fiducials (Fischl, 2012; Gramfort et al., 2013), we calculated the forward model. Sensor noise covariance matrices were estimated across all trials from the baseline period (–120 to 0 ms with respect to stimulus onset; Engemann and Gramfort, 2015). Next, the inverse operators were generated with default MNE parameters and applied at the single-trial level (method dSPM, lambda = 1/3). Hence, the sensor-level data covering the whole brain (i.e., 306 sensors) were projected into the source space.

We extracted source-level data in four predetermined ROIs, namely V1, V2, V3, and IPS (Tark et al., 2021). Each ROI was derived from a surface-based atlas (Wang et al., 2015). Next, the atlas files were mapped onto individual cortical surfaces using the Neuropythy toolbox. For each ROI, we applied the same MVPA and IEM procedures to the source-level data.

Statistical analysis. To assess the statistical significance of decoding accuracies and channel responses while controlling for multiple comparisons, we performed cluster-based permutation tests (Maris and Oostenveld, 2007). The null hypothesis was a one-ninth chance level for decoding accuracy and 0 for channel response. We first defined clusters as temporally consecutive significant timepoints (cluster-defining threshold $p < 0.05$, uncorrected). Next, we obtained a summed cluster-level test statistic for each cluster (i.e., t-scores), which was compared with a permutation-based null distribution. We permuted the labels by randomly multiplying accuracies or responses by +1 or –1 (i.e., sign permutation test) and searched for the cluster with the highest statistic. This procedure was repeated 5,000 times, yielding a null distribution. Lastly, the p value for each cluster (i.e., $p_{corrected}$) was calculated as the proportion of cluster-level statistic in the null distribution exceeding the observed cluster-level statistic.

In addition, we estimated the onset latencies (i.e., earliest significant timepoint after stimulus onset) of decoding accuracies using a jackknife-based approach (Kiesel et al., 2008; Zhang et al., 2023). From the data, we obtained a jackknife sample consisting of n data resamples. For each resample, one participant's data was omitted, and an onset latency was estimated using data from the remaining $n–1$ participants. The onset latencies were then compared across the jackknife samples. We computed 5,000 permuted samples of differences between two onset latencies after randomly recording the condition from which each onset latency was taken, yielding a null distribution. The p value for the observed difference between the two conditions was calculated as the proportion of differences in the null distribution exceeding the observed difference.

Results

Behavioral performance in the ensemble orientation discrimination task
In the main experiment, 23 participants performed an ensemble orientation discrimination task with homogeneous or heterogeneous stimuli in the MEG. With homogeneous stimuli, participants' responses were highly accurate at all ensemble orientations, reaching an average accuracy of 91.39% ($t_{(22)} = 47.339$; $p = 1.24 \times 10^{-23}$; Cohen's $d = 9.871$). With heterogeneous stimuli, participants were still able to discriminate the ensemble orientations (72.82%; $t_{(22)} = 12.393$; $p = 2.14 \times 10^{-11}$; Cohen's $d = 2.585$), though with a significantly lower accuracy compared with that with homogeneous stimuli ($t_{(22)} = 11.739$; $p = 6.06 \times 10^{-11}$; Cohen's $d = 2.448$; Fig. 1D). These results demonstrate that participants successfully estimated the ensemble orientations in both

homogeneous and heterogeneous stimuli, suggesting that our brain could represent the ensemble orientation.

MVPA decoding analysis
We employed a time-resolved MVPA to examine whether ensemble orientations could be decoded from MEG sensor signals. We first decoded the ensemble orientation in homogeneous stimuli. Here, the ensemble orientation was the same as the item orientation. Figure 2A showed the time course of decoding accuracy. A cluster-based permutation test showed that the ensemble orientation information could be decoded from the MEG signals after 100 ms with respect to stimulus onset ($p_{corrected} < 0.001$), with a peak at 512 ms. In addition, we extended the MVPA by a temporal generalization approach and found that the decoders generalized well after ~170 ms (cluster-based permutation test, $p_{corrected} < 0.001$; Fig. 2B), indicating stable neural representations of the ensemble orientation. Next, we decoded the ensemble orientation in heterogeneous stimuli where the ensemble orientation differed from its component item orientations. A cluster-based permutation test showed that the ensemble orientation could be reliably decoded from MEG signals after 168 ms with respect to stimulus onset ($p_{corrected} < 0.001$; Fig. 2D), with a peak at 637 ms. For heterogeneous stimuli, the decoders generalized well after ~210 ms (cluster-based permutation test, $p_{corrected} < 0.001$; Fig. 2E). Importantly, the latency of ensemble orientation decoding for heterogeneous stimuli (112 ms; t test: $p_{corrected} < 0.05$, FDR corrected, ≥ 5 consecutive significant timepoints) was significantly later (permutation test, $p < 0.001$; Fig. 2G) than that for homogeneous stimuli (100 ms; t test: $p_{corrected} < 0.05$, FDR corrected, ≥ 5 consecutive significant timepoints). Moreover, for heterogeneous stimuli, the peak decoding accuracy was highly correlated with the behavioral accuracy across individual participants (0.594, $p = 0.001$; Fig. 2F). In contrast, no such correlation was found for homogeneous stimuli (0.021, $p = 0.462$; Fig. 2C), likely due to a ceiling effect, as participants consistently achieved high behavioral accuracy.

Although item orientations in heterogeneous stimuli differed from those in homogeneous stimuli, both stimuli had the same ensemble orientations. Therefore, we performed cross-decoding analyses to investigate whether the two stimuli had a similar representation of the ensemble orientation in the brain. We found that the cross-decoder trained with data from one stimulus type could successfully decode ensemble orientations in the other stimulus type and demonstrate good generalization between ~250 and ~700 ms (cluster-based permutation test, $p_{corrected} < 0.001$; Fig. 2H). Further, we defined an optimal period for the training data elicited by homogeneous stimuli and averaged the decoding accuracy time courses for heterogeneous stimuli based on the training timepoints in the optimal period (420–500 ms; see Materials and Methods). The resulting time courses exhibited successful decoding of the ensemble orientations in heterogeneous stimuli (364–624 ms, cluster-based permutation test, $p_{corrected} < 0.001$; Fig. 2I).

These findings demonstrate that our brain can represent ensemble orientation in a reliable and persistent way. Moreover, such ensemble representation can be independent of local item orientations.

IEM analysis
Using IEM, we decomposed the sensor-level MEG signals to an ensemble stimulus into weighted responses of a set of orientation channels, each preferring one of the nine possible orientations. Using the data elicited by homogeneous stimuli during the optimal
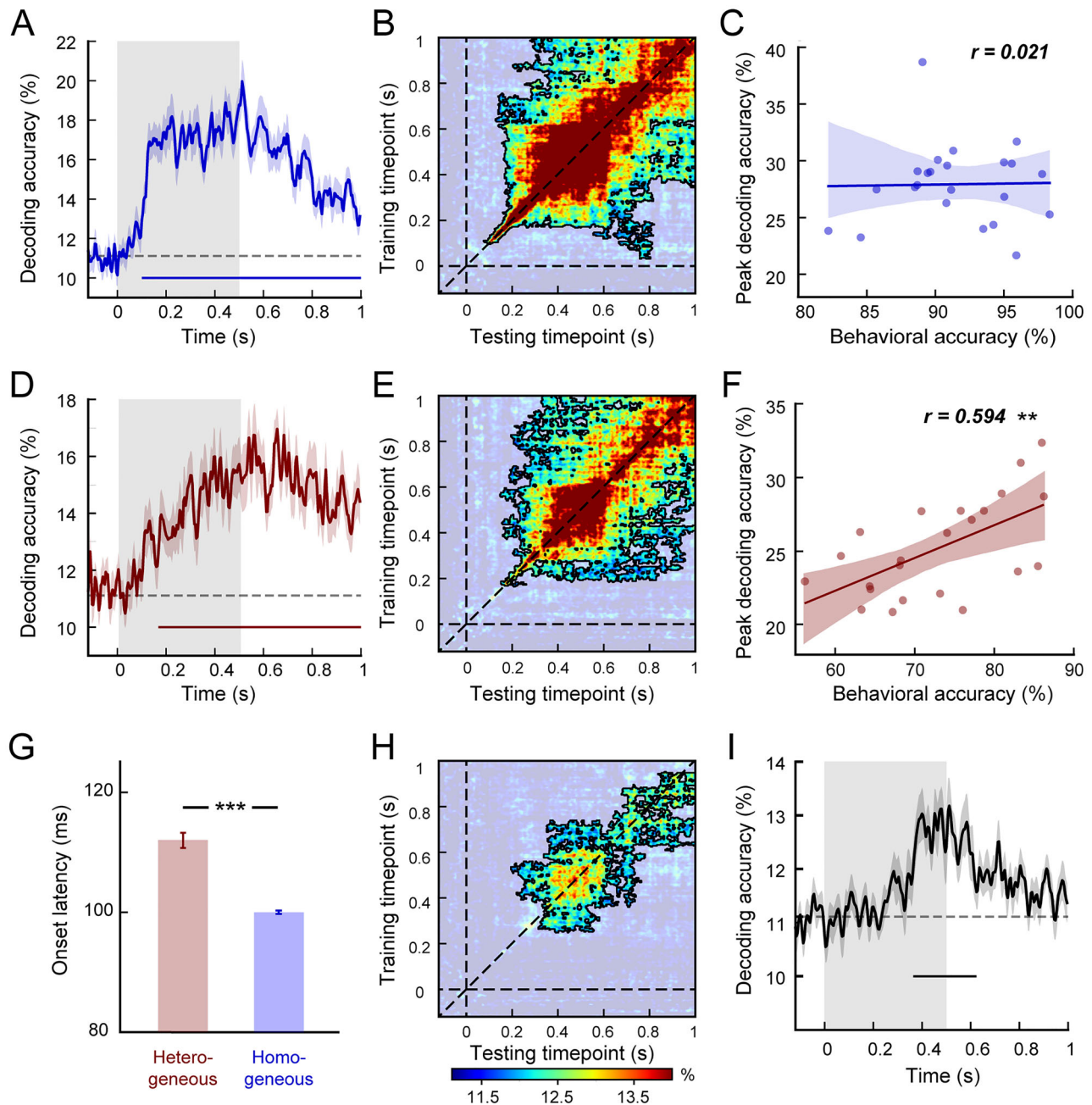
Figure 2. Time-resolved ensemble orientation decoding analysis. The decoding analysis was performed separately for each participant and each stimulus type using MEG occipital sensor signals. Panels *A–C* in the first row illustrate results with homogeneous (blue) stimuli, while panels *D–F* illustrate results with heterogeneous (red) stimuli. *A, D*, Time courses of ensemble orientation decoding accuracy. The ensemble orientation decoders were trained using MEG data at a specific timepoint and tested on the leftout data at the same timepoint. A gray bar indicates the visual stimulus presentation interval, and the lines below indicate significant decoding accuracies using the cluster-based permutation test ($p_{corrected} < 0.001$). *B, E*, Temporal generalization matrices of the ensemble orientation decoding. The decoders were trained using data at a single timepoint and tested on all timepoints. Vertical and horizontal dot lines mark the onset of visual stimuli. The black contour indicates significant decoding accuracies using the cluster-based permutation test ($p_{corrected} < 0.001$). *C, F*, Correlations between behavioral accuracy and peak decoding accuracy. *G*, Onset latencies for ensemble orientation decoding with heterogeneous and homogeneous stimuli. Error bars indicate SEM. Asterisks indicate significant correlations or differences (**$p < 0.01$; ***$p < 0.001$). *H–I*, Temporal generalization matrix and extracted time courses for coss-decoding analysis. The cross-decoders were trained using MEG data from one stimulus type and tested on the data from the other stimulus type.

period, we constructed encoding models and estimated the respective contribution (i.e., weight) of each orientation channel to the MEG sensor signals. With these weights, we could reconstruct the channel responses of the nine orientation channels (i.e., channel response profile) for each timepoint when participants viewed either homogeneous or heterogeneous stimuli (Fig. 3A).

We first examined the reconstructed channel response profiles for both homogeneous and heterogeneous stimuli. The

reconstructed channel response profiles were circularly shifted to align their ensemble orientations to a common 0°. In the absence of ensemble orientation information, such as before stimulus onset, the profile should appear flat, without peaks. However, after stimuli are processed in the brain, peaks at their respective ensemble orientations (i.e., 0°) should emerge in the channel response profiles. For homogeneous stimuli, we reconstructed the channel response profiles using leftout data. As
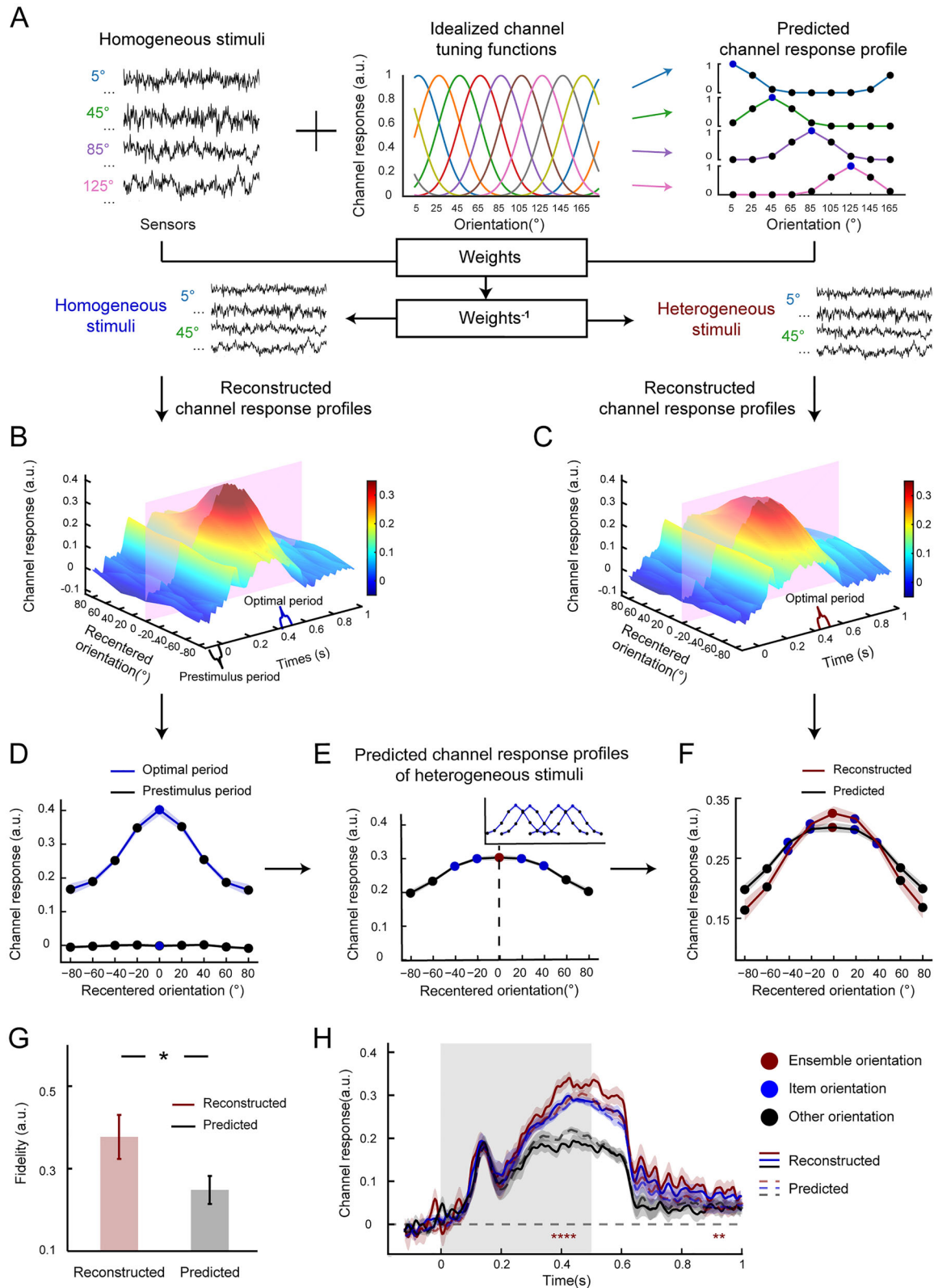
Figure 3. IEM-based time-resolved reconstruction of the ensemble orientation representation. These analyses were performed separately for each participant using MEG occipital sensor signals. ***A***, Schematic overview of the IEM analysis. Participants performed the ensemble orientation discrimination task at nine ensemble orientations (5°–165° in steps of 20°). Each ensemble orientation corresponded to the preferred orientation of a hypothetical orientation channel. The idealized tuning functions of these nine channels were characterized as Gaussian-like functions centered at their respective preferred orientation (coded in different colors). The predicted channel responses to a given stimulus were derived from these idealized tuning functions. Using data from homogeneous stimuli and predicted channel responses, we estimated the weights of each orientation channel for each MEG sensor. Next, to acquire instantaneous channel response profiles, these estimated weights were inverted and applied to the leftout data elicited by homogenous stimuli and all data elicited by heterogeneous stimuli at all timepoints. ***B***, ***C***, Time-resolved reconstructions of homogeneous (***B***) and heterogeneous (***C***) stimuli at all timepoints. ***D***, Average of the reconstructed channel response profiles in a prestimulus period (black) or the optimal poststimulus period (blue). To predict the channel response profiles of a heterogeneous stimulus, we adopted the channel response profile obtained from the data elicited by homogeneous

shown in Figure 3B, we obtained the instantaneous channel response profile for each timepoint. We further averaged the reconstructed channel response profiles across all timepoints in an 80 ms time window for both a prestimulus period (−100 to −20 ms relative to the stimulus onset) and the optimal period (420–500 ms). We found a bell-shaped average channel response profile, with the highest response located at 0° in the optimal period (repeated-measures ANOVA: $F_{(1.308,28.767)} = 49.482$; $p = 1.40 \times 10^{-8}$; $\eta_p^2 = 0.692$) but not in the prestimulus period ($F_{(3.100,68.190)} = 0.548$; $p = 0.657$; Fig. 3D). Similarly, for heterogeneous stimuli, we obtained the instantaneous channel response profile for each timepoint (Fig. 3C). We then calculated the average channel response profile and found that they peaked at 0° in the optimal period ($F_{(2.177,47.883)} = 30.107$; $p = 1.43 \times 10^{-9}$; $\eta_p^2 = 0.578$; Fig. 3F). However, in the prestimulus period, the average channel response profile was flat ($F_{(3.407,74.959)} = 0.599$; $p = 0.639$).

Next, we examined whether the ensemble representation is a mere summation of individual item representations or whether it engages other specific processes. Note that in a heterogeneous stimulus, there were four possible item orientations, all different from its ensemble orientation. To predict the channel response profile under the mere summation hypothesis, we adopted the channel response profiles estimated with the data elicited by homogeneous stimuli as response tunings to the four component item orientations in heterogeneous stimuli and computed the summed responses with an equal weight (0.25) assigned to each component item orientation (Fig. 3E). We speculated that if the ensemble representation did engage other specific processes, the reconstructed channel response profile should differ from the predicted one. Therefore, we compared the reconstructed channel response profile with the predicted one by performing repeated-measures ANOVAs with approach (reconstructed and predicted channel responses) and orientation channel (9 orientation channels) as two within-subject factors. Interestingly, we found a significant interaction between approach and orientation channel during the optimal period ($F_{(2.168,47.692)} = 3.416$; $p = 0.038$; $\eta_p^2 = 0.134$; Fig. 3F) but not in the prestimulus period ($F_{(3.320,73.048)} = 0.445$; $p = 0.741$), suggesting that orientation representations in the brain were further modulated by ensemble coding. To assess the quality of the encoded ensemble orientation information, we computed the representational fidelity from channel response profiles following the method in previous studies (Mo et al., 2019; Rademaker et al., 2019). For both reconstructed and predicted channel response profiles, we computed the average representational fidelity for the ensemble orientation during the optimal period. We found a significant higher fidelity from the reconstructed channel response profile compared with that from the predicted one ($t_{(22)} = 2.337$; $p = 0.016$; Fig. 3G). Together, these findings indicate that our brain engages in additional processes beyond a mere summation to represent the ensemble orientation.

To further examine the modulation of ensemble coding on orientation representations, we split the nine orientation channels into three groups (Fig. 3E,H): (1) the ensemble orientation group (i.e., 0°); (2) the item orientation group (i.e., ±20°, ±40°); and (3) the other orientation group (i.e., ±60°, ±80°). The channel responses were averaged within groups. To compare the reconstructed and predicted channel responses, for each timepoint, we performed a repeated-measures ANOVA with approach (reconstructed and predicted channel responses) and orientation group (ensemble, item, and other orientation groups) as two within-subject factors. We found significant interactions between orientation group and approach (intervals: 372–432, 912–932 ms, $p_{corrected} < 0.05$, FDR corrected, ≥5 consecutive significant timepoints). As shown in Figure 3H, for item orientation, no significant difference was observed between the reconstructed and predicted channel responses averaged at 372–432 ms ($t_{(22)} = 1.647$; $p_{corrected} = 0.341$; Bonferroni's correction level: 3). In stark contrast, the reconstructed channel response to the ensemble orientation was significantly elevated than the predicted one ($t_{(22)} = 3.397$; $p_{corrected} = 0.008$), while that to the other orientation were lower than the predicted one ($t_{(22)} = -2.629$; $p_{corrected} = 0.046$).

Finally, we performed source reconstruction to estimate the responses in different ROIs in order to investigate the neural substrates of ensemble orientation representation. We extracted source-level data from four predetermined ROIs: V1, V2, V3, and IPS. For each ROI, we applied the same MVPA procedure to decode the ensemble orientation. All four ROIs exhibited significant decoding of the ensemble orientation in heterogeneous stimuli (V1: 332–608, 612–768 ms; V2: 332–1,000 ms; V3: 328–948 ms; IPS: 212–1,000 ms; cluster-based permutation test, $p_{corrected} < 0.001$; Fig. 4A, right). However, no difference in the peak decoding accuracy was found between ROIs, except for the peak decoding accuracy in IPS, which was higher than that in V1 ($t_{(22)} = 4.216$; $p_{corrected} = 0.002$; Bonferroni's correction level: 6). Interestingly, only V2 exhibited a significant correlation between individual peak decoding and behavioral accuracies ($r = 0.471$; $p_{corrected} = 0.046$; Bonferroni's correction level: 4; Fig. 4B). The onset latencies of ensemble orientation decoding in V2 and V3 were significantly earlier than that in IPS ($ps_{corrected} < 0.01$; permutation test, Bonferroni's correction level: 6; Fig. 4A, left), with no significant difference between V2 and V3 ($p_{corrected} = 0.054$).

We also conducted IEM analyses for each ROI and compared the reconstructed and predicted channel response profiles for heterogeneous stimuli. During the optimal period, both V2 and V3 exhibited significantly higher fidelity of the ensemble orientation from the reconstructed profiles (V2: $t_{(22)} = 2.664$, $p_{corrected} = 0.028$; V3: $t_{(22)} = 2.644$, $p_{corrected} = 0.030$; Bonferroni's correction level: 4; Fig. 4C). However, we observed significant interactions between orientation group and approach only in V2 (intervals: 392–420, 440–472, 508–528, 756–776, 788–808 ms; $p_{corrected} < 0.05$; FDR corrected; ≥5 consecutive significant timepoints; Fig. 4D). Together, these results indicate that early visual areas, including

---

stimuli as response tuning to each of the four component item orientations in the heterogeneous stimulus and computed the summed responses (i.e., a summation hypothesis). **E**, Predicted channel response profile to a heterogeneous stimulus based on the summation hypothesis. **F**, Average of the predicted (black) and reconstructed (red) channel response profiles during the optimal period. **G**, Representational fidelities of the ensemble orientation from the reconstructed (red) or the predicted (black) channel response profiles during the optimal period. Further, we split the nine orientation channels into three groups: the ensemble orientation (i.e., 0°; red), the component item orientation (i.e. ±20°, ±40°; blue), and the other orientation (i.e., ±60°, ±80°; black). **H**, Time courses of reconstructed (solid curves) or predicted (dashed curves) channel responses for the three orientation groups. A gray bar indicates visual stimulus presentation interval. For each timepoint, we performed a repeated-measures ANOVA with approach (reconstructed and predicted channel responses) and orientation group (ensemble, item, and other orientation groups) as two within-subject factors. Asterisks indicate significant interactions between approach and orientation group ($p_{corrected} < 0.05$; FDR corrected; ≥5 consecutive significant timepoints).
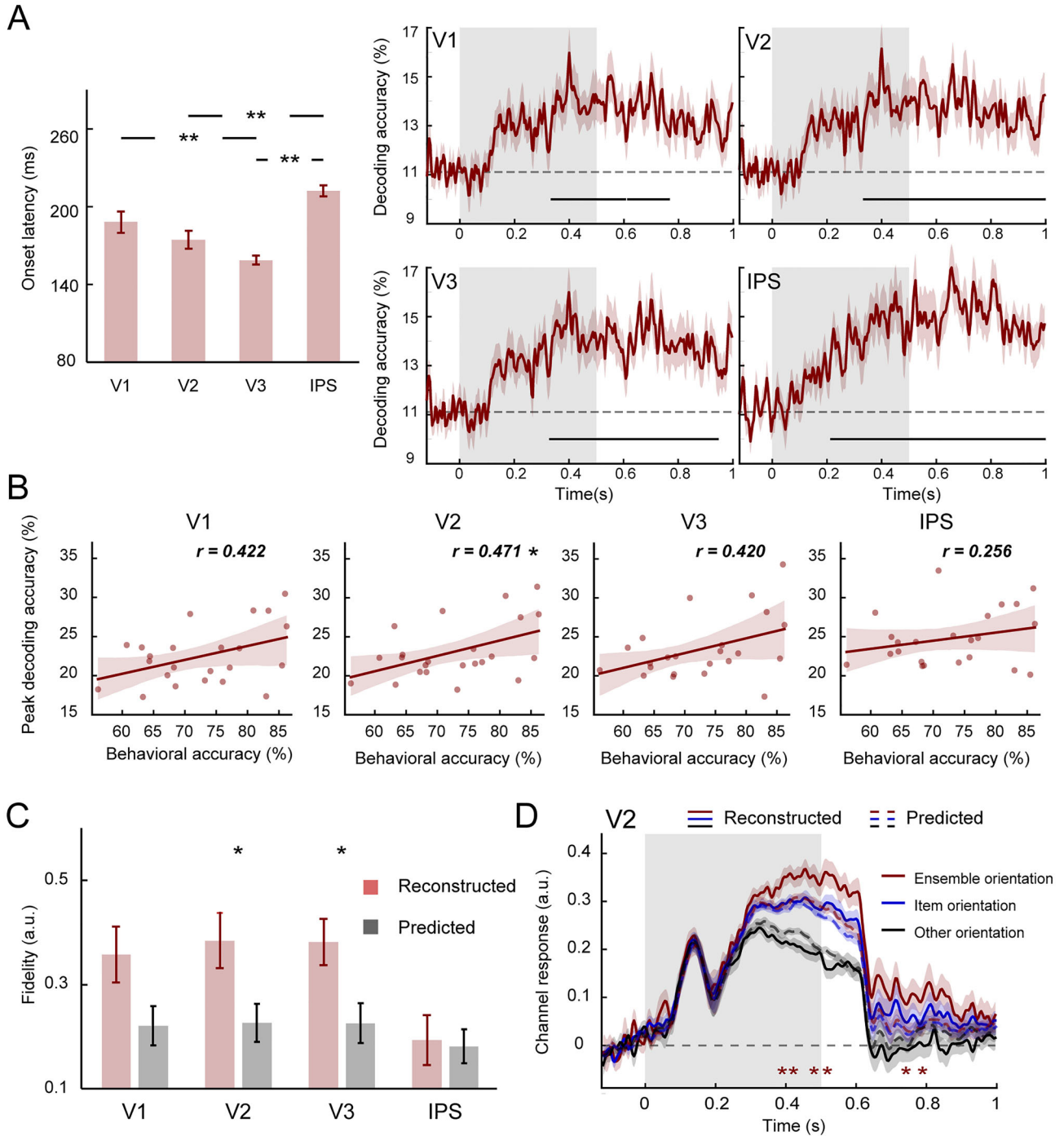
Figure 4. Results of source reconstruction analysis in V1, V2, V3, and IPS. ***A***, Onset latencies and time courses of the ensemble orientation decoding accuracy with heterogeneous stimuli. ***B***, Correlations between individual peak decoding accuracy and behavioral accuracy. Asterisks indicate significant correlations after Bonferroni's correction (*$p_{corrected} < 0.05$). ***C***, Representational fidelities of the ensemble orientation from either the reconstructed (red) or the predicted (black) channel responses during the optimal period. Error bars indicate SEM across participants. Asterisks indicate significant correlations after Bonferroni's correction (*$p_{corrected} < 0.05$). ***D***, Time courses of reconstructed or the predicted (dashed curves) channel responses for the three orientation groups in V2. A gray bar indicates visual stimulus presentation interval. Asterisks indicate significant interactions between approach and orientation group ($p_{corrected} < 0.05$; FDR corrected; ≥5 consecutive significant timepoints).

V1, V2, and V3, play an important role in the ensemble orientation representation, rather than high-level areas.

Bayesian analysis and control experiment
To further validate our findings, we reanalyzed our data in the main experiment with a Bayesian probabilistic decoding algorithm to estimate how likely different orientations were to

produce the observed MEG signals, i.e., probability functions. We trained the Bayesian model with data elicited by homogeneous stimuli and then applied the trained model to two independent datasets: one containing data elicited by heterogeneous stimuli and the other that was synthesized based on the mere summation hypothesis (see Materials and Methods). Similar to the logic in our IEM analysis, we hypothesized that if the
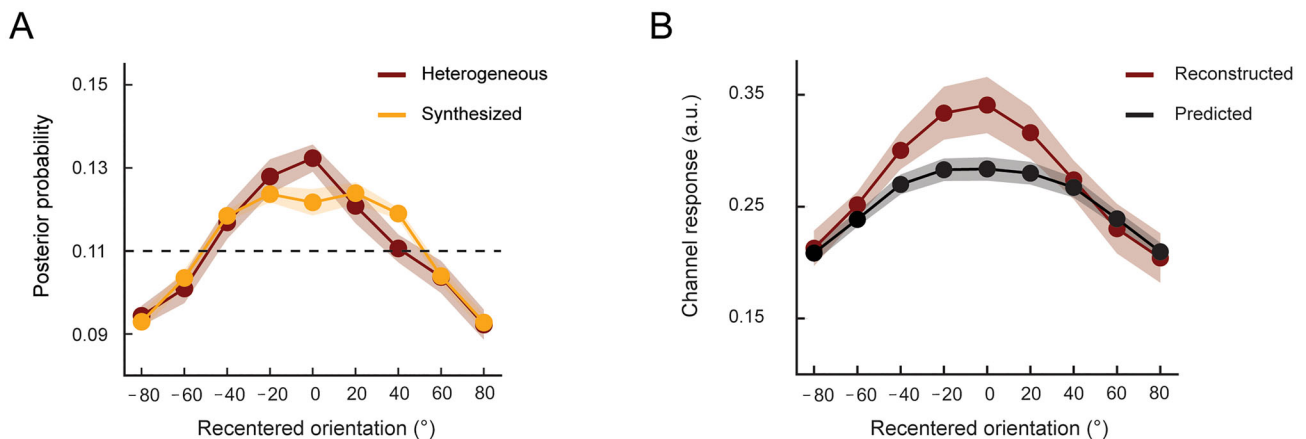
Figure 5. Results of Bayesian probabilistic decoding in the main experiment and IEM analysis in the control experiment. *A*, Posterior probabilities of nine possible ensemble orientations for the heterogeneous (red) and synthesized (orange) datasets, during the optimal period. *B*, Predicted (black) and reconstructed (red) channel response profiles during the optimal period.

ensemble representation does engage an integration process beyond a mere summation, the probability function estimated from the heterogeneous dataset should differ from that estimated from the synthesized one, especially during the optimal period. Indeed, for the heterogeneous dataset, the probability function peaked at the ensemble orientation (i.e., 0°), with the probability of the ensemble orientation significantly higher than the averaged probability of the four item orientations (±20° and ±40°; $t_{(22)} = 4.11$; $p = 4.5 \times 10^{-4}$; Fig. 5A). In contrast, the synthesized data did not show such a peak in the probability function, showing no difference between probabilities of the ensemble and item orientations ($t_{(22)} = 0.15$; $p = 0.88$). Furthermore, during the optimal period, the probability of the ensemble orientation in the heterogeneous dataset was significantly higher than that in the synthesized dataset ($t_{(22)} = 2.51$; $p = 0.019$), a distinction not observed during the prestimulus period ($t_{(22)} = 0.74$; $p = 0.47$). Consistent with the results of the IEM analysis, these results also suggest that our brain engages in additional processes beyond a mere summation to represent the ensemble orientation.

Note that in the main experiment, participants viewed homogeneous and heterogeneous stimuli in separate runs. With this design, it could be argued that different perceptual or cognitive processes might be engaged in different runs. Therefore, we conducted a control experiment where homogeneous and heterogeneous trials were mixed within the same run. With 15 participants, we successfully replicated the findings in the main experiment. Participants were able to discriminate the ensemble orientations, achieving 90.14% ($t_{(14)} = 24.32$; $p = 7.48 \times 10^{-13}$) accuracy for homogeneous and 71.44% ($t_{(14)} = 9.03$; $p = 3.23 \times 10^{-7}$) for heterogeneous stimuli. MVPA analysis confirmed that the ensemble orientation could be decoded from MEG sensor signals, both in heterogeneous (144–804 ms; cluster-based permutation test; $p_{\text{corrected}} < 0.001$) and homogeneous trials (84–800; 804–1,000 ms). Notably, for heterogeneous stimuli, the peak decoding accuracy significantly correlated with the behavioral accuracy ($r = 0.52$; $p = 0.025$). Furthermore, significant cross-decoding (324–388 ms; $p_{\text{corrected}} < 0.05$) of the ensemble orientation was also observed. Most importantly, as demonstrated in Figure 5B, the IEM analysis revealed that during the optimal period, the reconstructed channel response profile was evidently elevated at the ensemble orientation compared with the predicted one, showing significantly higher fidelity ($t_{(14)} = 1.83$; $p = 0.045$) and higher channel response ($t_{(14)} = 2.92$; $p = 0.0058$) to the ensemble orientation.

Taken together, converging evidence from both the Bayesian analysis and the control experiment supports that our brain engages in more complex processes beyond a mere summation to represent the ensemble orientation.

Discussion

We applied time-resolved MVPA and IEM to MEG data to investigate the neural representation of ensemble orientation and track its time course. We reliably decoded ensemble orientations of both homogeneous and heterogeneous stimuli using MEG signals covering the occipital lobe and significant cross-decoding between the two types of stimuli was also observed. Additionally, the peak decoding accuracy for heterogeneous stimuli strongly correlated with the behavioral accuracy. Furthermore, we demonstrated that the ensemble orientation representation, which emerged ∼370 ms after stimulus onset and primarily manifested in early visual areas, is not just the sum of the representations of component item orientations. Together, our findings revealed how the neural representation of ensemble orientation evolved in the human visual cortex, contributing to a more comprehensive understanding of ensemble orientation coding and its relationship with the coding of individual items.

An important and unique contribution of our study is that we identify the neural representation specific to the ensemble orientation. In previous studies, the mean of all item feature values was very similar, if not identical, to individual item feature value, rendering the ensemble and item representations largely indistinguishable. However, in our heterogeneous stimuli, instead of randomly sampling the item orientation from a narrow normal or uniform feature distribution (Dakin, 2001; Solomon, 2010; Attarha and Moore, 2015; Epstein and Emmanouil, 2021; Tark et al., 2021), we ensured a minimal difference of 20° between each individual item orientation and the ensemble orientation. Intriguingly, despite such substantial orientation difference, we still achieved successful cross-decoding between heterogeneous and homogeneous stimuli, indicating that the ensemble orientation representation was, at least to some extent, independent of local feature representations. This is consistent with previous behavioral studies showing that ensemble perception remained robust when individual items were noisy, unrecognized, or neglected (Parkes et al., 2001; Demeyere et al., 2008; Fischer and Whitney, 2011; Haberman and Whitney, 2011; Hochstein et al., 2015). Further, although recent modeling studies have highlighted the importance of the summed activation across all

items in ensemble perception (Haberman and Whitney, 2012; Brezis et al., 2018; Robinson and Brady, 2023; Utochkin et al., 2023), our results suggest that the human brain engages in additional processes beyond simple summation, to represent the ensemble orientation. This view is further supported by the significant differences between the predicted orientation representation (based on the summation hypothesis) and the reconstructed orientation representation in both main and control experiments, particularly the higher response/probability to the ensemble orientation.

Our findings characterize how the ensemble orientation representation temporally evolved in the brain. Previous studies have reported the onset latency for enhanced activation or decoding accuracy of ensemble oddball as well as the identity of face ensemble is roughly 100 ms (Roberts et al., 2019; Epstein and Emmanouil, 2021; Im et al., 2021; Sama et al., 2024). In line with these studies, we revealed a decoding latency of 112 ms for the ensemble orientation in heterogeneous stimuli (Fig. 2D). However, whether these results reflect the onset latency of the summed neural responses to all individual items or if they are specific to the ensemble information remains unclear, as both the activation and decoding analysis may not be able to differentiate the two hypotheses. In stark contrast, using IEM, we found the channel response to the ensemble orientation elevated ∼370 ms after stimulus onset. This suggested that the ensemble orientation representation per se may emerge at a relatively late stage. In behavioral studies, ensemble information can be reliably reported or reproduced with an exposure duration of as short as 50 ms (Ariely, 2001; Dakin, 2001; Haberman and Whitney, 2007, 2009; Leib et al., 2016; Li et al., 2016), evidently shorter than the 370 ms latency we found. What leads to such a discrepancy? One possible explanation is that the processing of visual stimuli in the brain is not halted by removing them from the display because of the persistence of vision. On the other hand, increasing exposure duration can improve the accuracy of the estimated ensemble information (Haberman and Whitney, 2009; Whiting and Oriet, 2011; Li et al., 2016), suggesting that ensemble coding may operate at longer temporal scales. Another factor we should consider is task difficulty that may contribute to the relatively late emergence of ensemble orientation representation. In our study, estimating the ensemble orientation from heterogeneous stimuli was challenging due to the large variance of item orientations (Dakin, 2001; Solomon, 2010), which may require a longer time to integrate information and form the ensemble representation (Mazurek et al., 2003; Palmer et al., 2005).

Furthermore, the findings regarding the specific ensemble orientation representation and its temporal evolution collaboratively point toward an integration process in ensemble coding beyond a mere summation. Contrary to recent computational studies that consider ensemble coding as a straightforward feedforward process with uniform weighting (Robinson and Brady, 2023; Utochkin et al., 2023), our results suggest a more intricate integration process. This process likely involves assigning uneven weights to different components based on their feature proximity to the ensemble mean, as supported by previous behavioral (Epstein et al., 2020) and computational studies (de Gardelle and Summerfield, 2011; Li et al., 2017; Ni and Stocker, 2023). Specifically, as hinted at by Figure 3F, participants appear to assign lower weights to the component orientation that differ by 40° from the ensemble orientation. Note that such uneven weight assignment would necessitate acquiring the distribution of component feature values (Utochkin et al., 2023), thereby

challenging the perspective of a pure feedforward process. Instead, it points to a recursive or potentially iterative process in the brain, aiming at refining ensemble representation, which likely demands longer processing time, as reflected in our time course results. Supporting this, Epstein et al. (2020) demonstrated how outliers were progressively discounted from the mean and how the noise they introduced decreased over time. Taken together, our findings suggest a more complex process in the formation and refinement of ensemble representation than previously believed. Moreover, while our results highlighted the importance of feature proximity, other factors, such as feature saliency, spatial location, and feature uncertainty (Kanaya et al., 2018; Tiurina et al., 2024), may also contribute to an uneven weighting process, pointing to a more complex interplay during the integration process in ensemble coding.

In our study, the ensemble orientation representation emerged predominantly in early visual areas. Notably, although only the correlation between the behavioral accuracy and the peak decoding accuracy in V2 remained significant after Bonferroni/FDR correction, the results in V1, V2, and V3 were highly similar, suggesting all these early visual areas contribute to the ensemble coding. In contrast, IPS exhibited a different pattern, with the fidelity of the reconstructed channel response profile being very similar to the predicted one, unlike the significant differences observed in V2 and V3 (Fig. 4C). Notably, recent studies have highlighted the role of the dorsal visual stream, such as IPS, in ensemble coding (Im et al., 2017, 2021; Tark et al., 2021). For example, Tark et al., (2021) found a gradual response increase to the ensemble orientation along the visual hierarchy and pinpointed ensemble orientation representation in V3 and IPS. Such observed discrepancies might be due to many factors, such as differences in ensemble stimulus design, task, and technical method. Given the relatively small receptive field sizes of neurons in early visual areas compared with the size of our stimuli (Rosa et al., 1988; Dumoulin and Wandell, 2008; He et al., 2019; Klink et al., 2021; Luo et al., 2024), how should we interpret our results? One possible explanation is that the ensemble orientation representation in early visual areas might result from the feedback signals from high-level areas where the receptive fields are much larger, such as IPS. If this is the case, the peak decoding accuracy in IPS should also correlate with the behavioral accuracy; however, this was not observed. Alternatively, the emergence of ensemble orientation representation in early visual areas could be due to abundant horizontal connections in this area to facilitate interactions among individual items (Roelfsema, 2006). Future research, preferably utilizing high spatial resolution techniques, is warranted to further differentiate these areas and elucidate the specific role of V1–V3 in orientation ensemble representation.

The current study focused on the neural mechanisms of ensemble orientation coding. Given the pervasiveness of ensemble coding in perception, two important issues in ensemble coding need to be addressed in the future. First, future research may investigate whether our findings could be generalized to other ensemble stimuli processed at different levels of the visual hierarchy, for example, face ensembles. Second, while we focused on the representation of the ensemble mean—the most studied signature of ensemble perception, ensemble representation encompasses other summary information, such as variance, outlier, and distribution. How they are represented in the brain is still largely unknown.

In summary, our findings demonstrated the emergence of the neural representation specific to ensemble orientation in the human visual cortex. By distinguishing neural responses to

ensemble information from the summed neural responses to individual items, we provide novel insights into the neural basis of ensemble coding.

Data Availability

Data and codes are available upon request by contacting the corresponding author.

References

Ariely D (2001) Seeing sets: representation by statistical properties. Psychol Sci 12:157–162.

Attarha M, Moore CM (2015) The capacity limitations of orientation summary statistics. Atten Percept Psychophys 77:1116–1131.

Baeck A, Wagemans J, Op de Beeck HP (2013) The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. Neuroimage 70:37–47.

Baek J, Chong SC (2020) Ensemble perception and focused attention: two different modes of visual processing to cope with limited capacity. Psychon Bull Rev 27:602–606.

Brezis N, Bronfman ZZ, Usher M (2018) A perceptual-like population-coding mechanism of approximate numerical averaging. Neural Comput 30:428–446.

Cant JS, Xu Y (2012) Object ensemble processing in human anterior-medial ventral visual cortex. J Neurosci 32:7685–7700.

Cant JS, Xu Y (2017) The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. J Cogn Neurosci 29:398–412.

Cant JS, Xu Y (2020) One bad apple spoils the whole bushel: the neural basis of outlier processing. Neuroimage 211:116629.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Cohen MA, Dennett DC, Kanwisher N (2016) What is the bandwidth of perceptual experience? Trends Cogn Sci 20:324–335.

Corbett JE, Utochkin I, Hochstein S (2023) *The pervasiveness of ensemble perception: not just your average review*. Cambridge: Cambridge University Press.

Dakin SC (2001) Information limit on the spatial integration of local orientation signals. J Opt Soc Am A 18:1016–1026.

Dakin SC, Watt RJ (1997) The computation of orientation statistics from visual texture. Vision Res 37:3181–3192.

De Gardelle V, Summerfield C (2011) Robust averaging during perceptual judgment. Proc Natl Acad Sci U S A 108:13341–13346.

Demeyere N, Rzeskiewicz A, Humphreys KA, Humphreys GW (2008) Automatic statistical processing of visual properties in simultanagnosia. Neuropsychologia 46:2861–2864.

Dobs K, Isik L, Pantazis D, Kanwisher N (2019) How face perception unfolds over time. Nat Commun 10:1258.

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647–660.

Engemann DA, Gramfort A (2015) Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. Neuroimage 108:328–342.

Epstein ML, Emmanouil TA (2021) Ensemble statistics can be available before individual item properties: electroencephalography evidence using the oddball paradigm. J Cogn Neurosci 33:1056–1068.

Epstein ML, Quilty-Dunn J, Mandelbaum E, Emmanouil TA (2020) The outlier paradox: the role of iterative ensemble coding in discounting outliers. J Exp Psychol 46:1267–1279.

Fischer J, Whitney D (2011) Object-level visual information gets through the bottleneck of crowding. J Neurophysiol 106:1389–1398.

Fischl B (2012) Freesurfer. Neuroimage 62:774–781.

Fu Y, Yan W, Shen M, Chen H (2021) Does consciousness overflow cognitive access? Novel insights from the new phenomenon of attribute amnesia. Sci China Life Sci 64:847–860.

Gramfort A, et al. (2013) MEG and EEG data analysis with MNE-Python. Front Neuroinform 7:267.

Guggenmos M, Sterzer P, Cichy R (2018) Multivariate pattern analysis for MEG: a comparison of dissimilarity measures. Neuroimage 173:434–447.

Haberman J, Whitney D (2007) Rapid extraction of mean emotion and gender from sets of faces. Curr Biol 17:R751–R753.

Haberman J, Whitney D (2009) Seeing the mean: ensemble coding for sets of faces. J Exp Psychol Hum Percept Perform 35:718–734.

Haberman J, Whitney D (2011) Efficient summary statistical representation when change localization fails. Psychon Bull Rev 18:855–859.

Haberman J, Whitney D (2012) *From perception to consciousness: searching with Anne Treisman* (Wolfe J, Robertson L, eds), pp 339–349. New York: Oxford University Press.

He D, Wang Y, Fang F (2019) The critical role of V2 population receptive fields in visual orientation crowding. Curr Biol 29:2229–2236. e3.

Hochstein S, Pavlovskaya M, Bonneh YS, Soroker N (2015) Global statistics are not neglected. J Vis 15:7–7.

Im HY, Albohn DN, Steiner TG, Cushing CA, Adams RB, Kveraga K (2017) Differential hemispheric and visual stream contributions to ensemble coding of crowd emotion. Nat Hum Behav 1:828–842.

Im HY, Cushing CA, Ward N, Kveraga K (2021) Differential neurodynamics and connectivity in the dorsal and ventral visual pathways during perception of emotional crowds and individuals: a MEG study. Cog Affect Behav Neurosci 21:776–792.

Isik L, Meyers EM, Leibo JZ, Poggio T (2014) The dynamics of invariant object recognition in the human visual system. J Neurophysiol 111:91–102.

Kanaya S, Hayashi MJ, Whitney D (2018) Exaggerated groups: amplification in ensemble coding of temporal and spatial features. Proc R Soc B 285:20172770.

Kiesel A, Miller J, Jolicœur P, Brisson B (2008) Measurement of ERP latency differences: a comparison of single-participant and jackknife-based scoring methods. Psychophysiology 45:250–274.

King J-R, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18:203–210.

Klink PC, Chen X, Vanduffel W, Roelfsema PR (2021) Population receptive fields in nonhuman primates from whole-brain fMRI and large-scale neurophysiology in visual cortex. Elife 10:e67304.

Leib A, Kosovicheva A, Whitney D (2016) Fast ensemble representations for abstract visual impressions. Nat Commun 7:13186.

Li V, Herce Castañón S, Solomon JA, Vandormael H, Summerfield C (2017) Robust averaging protects decisions from noise in neural computations. PLoS Comput Biol 13:e1005723.

Li H, Ji L, Tong K, Ren N, Chen W, Liu CH, Fu X (2016) Processing of individual items during ensemble coding of facial expressions. Front Psychol 7:1332.

Li HH, Sprague TC, Yoo AH, Ma WJ, Curtis CE (2021) Joint representation of working memory and uncertainty in human cortex. Neuron 109:3699–3712.

Luo L, et al. (2024) Local field potentials, spiking activity, and receptive fields in human visual cortex. Sci China Life Sci 67:543–554.

MacEvoy SP, Epstein RA (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. Curr Biol 19:943–947.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Maule J, Witzel C, Franklin A (2014) Getting the gist of multiple hues: metric and categorical effects on ensemble perception of hue. J Opt Soc Am A 31:A93–A102.

Mazurek ME, Roitman JD, Ditterich J, Shadlen MN (2003) A role for neural integrators in perceptual decision making. Cereb Cortex 13:1257–1269.

Michael E, de Gardelle V, Summerfield C (2014) Priming by the variability of visual information. Proc Natl Acad Sci U S A 111:7873–7878.

Mo C, Lu J, Shi C, Fang F (2022) Neural representations of competing stimuli along the dorsal and ventral visual pathways during binocular rivalry. Cereb Cortex 33:2734–2747.

Mo C, Lu J, Wu B, Jia J, Luo H, Fang F (2019) Competing rhythmic neural representations of orientations during concurrent attention to multiple orientation features. Nat Commun 10:5264.

Ni L, Stocker AA (2023) Efficient sensory encoding predicts robust averaging. Cognition 232:105334.

Oh B-I, Kim Y-J, Kang M-S (2019) Ensemble representations reveal distinct neural coding of visual working memory. Nat Commun 10:5665.

Oosterhof NN, Tipper SP, Downing PE (2012) Visuo-motor imagery of specific manual actions: a multi-variate pattern analysis fMRI study. Neuroimage 63:262–271.

Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. J Vis 5:376–404.

Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M (2001) Compulsory averaging of crowded orientation signals in human vision. Nat Neurosci 4:739–744.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Rademaker RL, Chunharas C, Serences JT (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. Nat Neurosci 22:1336–1344.

Roberts T, Cant JS, Nestor A (2019) Elucidating the neural representation and the processing dynamics of face ensembles. J Neurosci 39:7737–7747.

Robinson MM, Brady TF (2023) A quantitative model of ensemble perception as summed activation in feature space. Nat Hum Behav 7:1–14.

Roelfsema PR (2006) Cortical algorithms for perceptual grouping. Annu Rev Neurosci 29:203–227.

Rosa MGP, Sousa APB, Gattass R (1988) Representation of the visual field in the second visual area in the Cebus monkey. J Comp Neurol 275:326–345.

Sama MA, Nestor A, Cant JS (2024) The neural dynamics of face ensemble and central face processing. J Neurosci 44:e1027232023.

Solomon JA (2010) Visual discrimination of orientation statistics in crowded and uncrowded arrays. J Vis 10:19–19.

Sprague TC, Adam KC, Foster JJ, Rahmati M, Sutterer DW, Vo VA (2018) Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. eNeuro 5:1–5.

Sprague TC, Boynton GM, Serences JT (2019) The importance of considering model choices when interpreting results in computational neuroimaging. eNeuro 6:ENEURO.0196-19.2019.

Tark K-J, Kang M-S, Chong SC, Shim WM (2021) Neural representations of ensemble coding in the occipital and parietal cortices. Neuroimage 245: 118680.

Tiurina NA, Markov YA, Whitney D, Pascucci D (2024) The functional role of spatial anisotropies in ensemble perception. BMC Biol 22:28.

Utochkin IS, Choi J, Chong SC (2023) A population response model of ensemble perception. Psychol Rev 131:36–57.

van Bergen RS, Jehee JFM (2021) TAFKAP: an improved method for probabilistic decoding of cortical activity. *BioRxiv*, 2021-03.

van Bergen RS, Ji Ma W, Pratte MS, Jehee JF (2015) Sensory uncertainty decoded from visual cortex predicts behavior. Nat Neurosci 18:1728–1730.

Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic maps of visual topography in human cortex. Cereb Cortex 25:3911–3931.

Webster J, Kay P, Webster MA (2014) Perceiving the average hue of color arrays. J Opt Soc Am A 31:A283–A292.

Whiting BF, Oriet C (2011) Rapid averaging? Not so fast! Psychon Bull Rev 18:484–489.

Whitney D, Leib A (2018) Ensemble perception. Annu Rev Psychol 69:105–129.

Zhang Y, Bi K, Li J, Wang Y, Fang F (2023) Dyadic visual perceptual learning on orientation discrimination. Curr Biol 33:2407–2416.