Simplifying the Complex: The Neural Mechanisms of Visual Ensemble Perception

Authors: Sun Huanxiang, Zhang Fan, Li Sijia, Zhang Xiuling, Jiang Yi, Zhang Xiuling

Date: 2025-10-23T21:49:20+00:00

Abstract

Ensemble perception refers to the process through which the visual system efficiently extracts summary statistics such as mean and variance from the complex external world, which holds significant importance for human adaptation to the environment. Investigating its neural mechanisms contributes to understanding how the visual system achieves efficient abstract representation. The present review summarizes the temporal course of ensemble perception, reviews the theoretical models and empirical evidence for this integration mechanism, and distinguishes the functions and neural bases of ensemble coding versus member or individual coding. Based on existing research findings, we propose a "coarse-fine-calibration" integration model: during the processing of visual features at different levels, the brain may sequentially recruit domain-general and domain-specific mechanisms, with early coarse processing relying on the general magnocellular pathway, followed by relatively fine, specific representations that depend on the parvocellular pathways of feature-specific brain regions, and finally calibrated through iterative feedforward-feedback loops. Future research may focus on the neural pathways and specific brain regions of visual ensemble perception, the roles of feedforward and feedback processes, the generality and specificity of information coding, and the influence of development and experience on ensemble perception.

Full Text

Preamble

Simplify Complexity: The Neural Mechanisms of Visual Ensemble Perception

SUN Huanxiang[1]#, ZHANG Fan[1],[2]#, LI Sijia[1], ZHANG Xiuling[1], JIANG Yi[3],[4]

[1] School of Psychology, Northeast Normal University, Changchun, 130024
[2] Jinzhong College of Information, Jinzhong, 030800
[3] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101
[4] Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049

Abstract
Ensemble perception is the process by which the visual system efficiently extracts summary statistics (such as mean and variance) from complex visual scenes at a glance, which is crucial for human adaptation to the environment. Investigating its neural mechanisms helps us understand how the visual system achieves efficient abstract representation. This article summarizes the temporal dynamics of ensemble perception, reviews theoretical models and empirical evidence for this integration mechanism, and distinguishes the functional and neural bases of ensemble coding from member or individual coding. Based on existing research, we propose a "Coarse-Fine-Refine" integration model: when processing visual features at different levels, the brain may sequentially employ domain-general and domain-specific mechanisms. Early processing relies on coarse analysis via the general magnocellular pathway, followed by relatively fine representations that are domain-specific and depend on parvocellular pathways in feature-selective brain regions, and finally undergoes calibration through iterative feedforward-feedback loops. Future research should focus on the neural pathways and specific brain regions underlying visual ensemble perception, the roles of feedforward and feedback processes, the generality versus specificity of information coding, and the influence of development and experience on ensemble perception.

Keywords: ensemble perception, statistical summary representation, perceptual integration, temporal dynamics, neural mechanism
Classification Code: B842

1 Introduction

Imagine walking into a kitchen and glancing at a plate of apples. Without deliberately scrutinizing each apple' s details, you can instantly grasp their overall size and ripeness. This "at-a-glance" ability is not coincidental but rather reflects an efficient processing strategy of the visual system—ensemble perception. When confronted with a group of similar stimuli, the visual system can rapidly and accurately extract the group's overall statistical properties (such as mean and variance) rather than processing the details of individual stimulus attributes (Ariely, 2001; Whitney & Yamanashi Leib, 2018). This mechanism fully demonstrates the efficiency of the human visual system when handling large amounts of information.

Ensemble perception is also known as ensemble coding. This field has developed rapidly since Ariely' s (2001) seminal work, with early researchers often using

the term "statistical summary representation" to describe this phenomenon (e.g., Tong et al., 2015). In vision science, both "ensemble perception" and "statistical summary representation" refer to the cognitive process of statistical processing and integration of large stimulus sets, and they are frequently used interchangeably in many research reports (Alvarez, 2011; Corbett et al., 2023). However, some researchers have noted subtle differences between the two concepts: ensemble perception emphasizes that the research object is a large, clustered group of stimuli, whereas statistical summary representation highlights the statistical computation process during representation. The former can include complex attributes that are not amenable to statistical description, making it a broader, more general concept (Li & Chen, 2022). Notably, there is not yet a unified translation for "Ensemble Perception" in Chinese academia. Some researchers translate it as "cluster perception" (Li & Chen, 2022), "ensemble perception" (Liu, 2021), "global perception" (e.g., Hao et al., 2023), or "global awareness" (e.g., Zhao et al., 2024). However, to avoid confusion with "cluster" in computer science and industrial economics, and to distinguish it from the important concept of "holistic processing" in cognitive science, this article recommends using "ensemble perception." Furthermore, if future research reports only involve the "mean" or "variance" of stimulus sets, the terms "average representation" and "variance representation" can be used directly.

Central tendency (e.g., mean) and dispersion tendency (e.g., variance) are two core metrics of ensemble perception (Tong et al., 2015; Norman et al., 2015). Average representation is a key indicator describing the central tendency of stimulus sets. Ariely (2001) used a mean identification task with sets of circles of different sizes and found that individuals could quickly and accurately represent the set mean, and this representation was unaffected by set size (i.e., the number of items), while accurate representation of individual set members was difficult. Subsequent research has found that average representation exists widely across multiple visual features, from low-level to high-level, including stimulus size (Haberman & Suresh, 2021), spatial location (Sun et al., 2021), depth (Wardle et al., 2012), orientation (Parkes et al., 2001), color and contrast (Rajendran et al., 2021), brightness (Bauer, 2009), as well as high-level features such as facial emotion (Haberman et al., 2009), facial gender (Haberman & Whitney, 2007), facial identity (de Fockert & Wolfenstein, 2009; Davis et al., 2021), gaze direction (Florey et al., 2016), biological motion (Sweeny et al., 2013; Nguyen et al., 2021), and trustworthiness (Marini et al., 2023; Chwe & Freeman, 2024).

Variance representation is an important indicator describing the dispersion tendency of sets. Haberman, Lee, and Whitney (2015) were the first to explore ensemble perception of facial emotion from a variance perspective. Using an adjustment-matching method, they asked participants to adjust the variance of one set of faces with different emotions to match another set. The results showed that participants could accurately represent the set variance, unaffected by set size. Subsequent research has further demonstrated that, similar to average representation, the visual system can represent variance across multiple visual features, including color (Maule & Franklin, 2020; Ward et al., 2016), brightness

(Tong et al., 2015), size (Tokita et al., 2016), as well as facial gender and race (Phillips et al., 2018). These studies show that the visual system is also efficient in processing set dispersion. However, research on dispersion remains limited, and whether it has cross-level, cross-feature perceptual universality similar to average representation remains unclear.

In summary, visual ensemble perception is an efficient processing strategy for coping with complex visual input. Investigating its neural mechanisms helps us understand the neural computation of statistical information such as mean values. This article delves into the early and late temporal dynamics of its processing, the complex mechanisms of information integration (such as representation hierarchy and computational mechanisms), clarifies the differences in neural substrates between ensemble and single-stimulus coding and related theoretical models, and proposes an integrated hypothesis of visual statistical processing. Finally, we outline future research directions, hoping to provide ideas and inspiration for further revealing the neural mechanisms of ensemble perception.

2.1 Evidence for Early Automatic Processing in Temporal Dynamics

Regarding attentional demands, recent Event-Related Potential (ERP) studies have examined whether average representation of facial emotion depends on attentional resources. Ji et al. (2018) used a cue-target paradigm, asking participants to judge whether the average emotion of a face set was positive or negative under valid or invalid cue conditions. The results showed that even under invalid cue conditions with limited attention, participants could effectively extract average emotion, and no N2pc (N2-posterior-contralateral) component —reflecting spatial selective attention—was observed, indicating that average emotion extraction relies minimally on attentional resources. Notably, under valid cue conditions, the Sustained Posterior Contralateral Negativity (SPCN) component—whose amplitude typically reflects visual short-term memory load —did not differ significantly between ensemble and individual tasks, suggesting that multiple face items were compressed into a "single" object stored in visual short-term memory (Ji et al., 2018).

Under invalid cue conditions, although attentional resources were reduced, attentional leakage remained possible. Building on this, Ji et al. (2024) further investigated whether average emotion information could still be automatically extracted under stricter attentional manipulation. Using an oddball paradigm, participants performed a central fixation task unrelated to peripheral face emotion sets. The results showed that despite full attention being focused on the central task, participants could automatically detect changes in peripheral face average emotion. This automatic detection showed significant valence differences: deviant stimuli with negative average emotion evoked a rapid and sustained Visual Mismatch Negativity (vMMN) at approximately 92 ms, while oddball stimuli with positive average emotion evoked a mismatch positivity

(MP) at a later 168 ms. Multivariate Pattern Analysis (MVPA) results further showed that the brain could distinguish standard from oddball stimuli around 90 ms. These findings indicate that average emotion extraction can proceed automatically without attentional engagement, and that positive and negative average emotions may involve different neural mechanisms.

Regarding processing speed and temporal dynamics, another ERP study using the oddball paradigm explored the time course of ensemble perception composed of multiple line segments. Researchers distinguished between oddball stimuli defined by a single member versus those defined by the entire set. They found that compared to the member condition, the P3b component latency was significantly earlier for ensemble oddball stimuli. Additionally, MVPA results showed that neural signals could classify standard versus oddball stimuli earlier in the ensemble condition (starting around 102 ms post-stimulus). These results suggest that ensemble perception occurs rapidly and that ensemble representation precedes representation of individual member attributes (Epstein & Emmanouil, 2021).

However, a recent study on average orientation has challenged the fully automatic mechanism of ensemble perception. Lukashevich et al. (2025) also used an oddball paradigm with two experiments to test the necessity of attention in average orientation identification. Experiment one was an attention condition where participants explicitly attended to and reported changes in ensemble average orientation; clear P3 components were observed, typical markers of conscious change detection and attentional engagement. Experiment two was a non-attention condition where participants focused entirely on changes at a central fixation point, ignoring background ensemble stimuli. Even when ensemble average orientation changed identically to experiment one, no vMMN component was detected. MVPA results aligned with ERP findings, failing to decode average orientation changes under unattended conditions. The authors concluded that at least for moderate differences in average orientation, identification and change detection are not fully automatic but require attentional resources. This finding contrasts interestingly with Ji et al.' s (2024) results on automatic processing of average emotion, suggesting that different types of statistical summary information—such as facial emotion versus line orientation —may rely on different cognitive mechanisms. Facial emotion has greater evolutionary adaptive significance and may be linked to rapid emotional processing pathways involving the amygdala, making average representation more likely under unattended conditions. Additionally, the degree of automatic processing may be influenced by the salience of differences; for example, the difference between negative and positive emotions may be greater than that between 40° and 60° orientations.

Furthermore, while studies by Im et al. (2021) and Roberts et al. (2019) primarily explored the neural basis of ensemble versus individual coding, they also analyzed the temporal dynamics of ensemble perception, with results involving neural activity characteristics at different processing stages (see Section 4.1). It

is important to emphasize that previous reviews based on behavioral research have extensively discussed the relationship between automatic processing and attention in ensemble perception (see Tong et al., 2015; Alvarez, 2011; Whitney & Yamanashi Leib, 2018), so this article will not reiterate those points.

2.2 Evidence for Late Precision Processing in Temporal Dynamics

Recent studies have also used neurophysiological techniques to explore the late temporal dynamics of ensemble perception, revealing that its construction may be a gradually refined dynamic process.

Yashiro et al. (2024) combined the inverted encoding model (IEM) with Electroencephalography (EEG) data to reconstruct neural representations of average orientation at different time points and correlated them with behavioral responses (reported average orientation perception). They found that although neural representations of homogeneous orientation conditions could be significantly decoded relatively early (174–828 ms), the strength of neural representations for heterogeneous condition average orientation was enhanced only in a later time window (400–700 ms), and this strength correlated significantly with behavioral performance only at an even later stage (600–700 ms). The researchers thus speculated that ensemble perception may combine parallel and serial processing: ensemble information is coarsely extracted quickly via rapid feedforward processing, while more precise representations require iterative feedforward-feedback mechanisms for refinement (Yashiro et al., 2024).

Gong et al. (2025) recently used Magnetoencephalography (MEG) combined with MVPA and IEM to reveal similar temporal dynamics. Investigating ensemble orientation (where no member orientation matched the average orientation), they found that although ensemble orientation information could be decoded around 112 ms post-stimulus, only the decoding peak at 600 ms correlated significantly with behavioral accuracy. IEM analysis showed that neural activity truly representing the ensemble average orientation appeared at an even later time point (around 370 ms) and exhibited a unique neural representation pattern different from simple summation of member information.

Together, these studies suggest that while initial extraction of ensemble information may be rapid, forming precise ensemble perception that differs from simple summation of individual items requires longer processing time, likely through recursive, iterative, or more complex processes to refine the representation (Yashiro et al., 2024; Gong et al., 2025).

3 Integration Mechanisms of Ensemble Perception

How does the visual system integrate ensemble perception from complex individual stimulus information? This is a core scientific question. Early theoretical models provided a basic framework for understanding this process, while sub-

sequent research has revealed more complex neural computational mechanisms that go beyond simple linear averaging.

3.1 Signal Pooling Hypothesis: A Descriptive Theoretical Assumption

Whitney and his team have conducted extensive research in ensemble perception and made outstanding contributions. Based on substantial behavioral evidence in this field, they proposed a theoretical framework regarding the neural pathways and representation processes underlying ensemble perception, which we call the Signal Pooling Hypothesis (Haberman & Whitney, 2012; Whitney et al., 2014).

As an important early theoretical framework, Signal Pooling provides an explanation at the computational principle level for how the visual system efficiently processes large amounts of information. However, this hypothesis also has limitations; aspects such as the specific integration mechanisms of ensemble perception, the influence of attention and awareness, and the processing mechanisms for cross-domain stimuli require further investigation. The main points are as follows: (1) The neural mechanism of ensemble perception may involve pooling of neural signals through linear integration or population coding. For example, for a set of line segments with different orientations, neurons sensitive to each orientation are activated in the primary visual cortex (V1). Neural signals are pooled layer by layer along the visual hierarchy, ultimately forming a perception of the entire set. This framework is based on hierarchical representation in the visual system, relying on the characteristic that neuronal receptive fields increase progressively at each level (Rousselet et al., 2004), and suggests that pooling can effectively reduce noise and improve representation precision through averaging. (2) Ensemble perception relies on implicit, parallel processing mechanisms. Its representation process is unaffected by the number of items in the set (Chong & Treisman, 2003) and can proceed with limited attentional resources (Wolfe et al., 2015). (3) Different ensemble perceptions occur at different hierarchical levels and pathways of visual processing. For example, pooling of low-level features such as average brightness, color, and orientation may occur in V1 or even subcortical regions, while high-level features like shape and faces may be localized in the ventral pathway, and average motion and position may be related to the dorsal pathway.

Recent neuroimaging studies support the view of progressive pooling along visual processing pathways. For example, researchers used functional Magnetic Resonance Imaging (fMRI) and IEM to investigate average orientation representation in grating ensembles. They found that although overall Blood-Oxygen-Level-Dependent (BOLD) signals did not differ significantly between ensemble and individual tasks, neural responses obtained through IEM showed differences. Under task-relevant conditions, both occipital and frontoparietal regions showed selective responses to average orientation (ensemble task) and individual orientation (individual task). Notably, selective responses to average orientation

were not significant in V1 but were significant in V2 and V3, showing a significant linear increase from V1 to V2 to V3. These results indicate that the visual system pools information at multiple levels to form neural representations of statistical summary information, supporting the progressive pooling view (Tark et al., 2021).

However, evidence also suggests that the pooling process is not a simple linear average of low-level signals at high levels but rather a high-order (Allik et al., 2022) or high-dimensional (Jia et al., 2022) integration. Additionally, when forming average representations of high-level visual stimuli (e.g., faces), the brain needs advanced neurons that receive input from multiple faces. This means the brain requires not only numerous neurons representing individual faces but also an even larger number of "statistical summary neurons." Whitney' s team later acknowledged that simple hierarchical pooling models struggle to explain ensemble perception at the object level (Whitney & Yamanashi Leib, 2018). Therefore, when addressing the relationship between wholes and individuals (especially when explaining crowding effects), sparse selection models (Chaney et al., 2014) and reverse hierarchy theory (Hochstein et al., 2015) may be more applicable. In these cases, statistical information processing may rely more on population coding, where multiple neurons representing different faces are activated simultaneously, forming an ensemble representation at the population level.

Furthermore, the Signal Pooling Hypothesis emphasizes that ensemble perception is an automatic process relying on parallel processing and is not limited by attentional resources or set size. While there is some supporting evidence—for instance, crowding effect studies show that ensemble perception performance remains good even when attention and awareness of individual members are impaired (Parkes et al., 2001; Fischer & Whitney, 2011), and studies of unilateral spatial neglect (USN) patients show that average representation of contralateral visual information is relatively preserved despite right parietal lobe damage causing left spatial neglect (Yamanashi Leib et al., 2012; Pavlovskaya et al., 2015)—this view remains controversial. Some studies suggest that ensemble perception precision may decline with increasing set size, possibly due to capacity limitations or early encoding noise (Marchant et al., 2013; Utochkin & Tiurina, 2014), and that attention may modulate ensemble perception by increasing the weighting of individual items (Ying, 2022).

Finally, the Signal Pooling Hypothesis posits that multiple rather than single mechanisms exist across different levels of visual processing. However, current empirical results on this issue remain inconsistent (Chang, Cha & Gauthier, 2024; Haberman, Brady et al., 2015), making it an active research area (see Section 6.3).

3.2 Computational Models of Integration

Although hierarchical pooling or traditional linear averaging assumptions provide a basic framework for understanding ensemble perception, they are merely descriptive theoretical overviews. To better explain integration mechanisms, advances in computational modeling have provided more specific neural computational pathways and hypotheses.

The population response model proposed by Utochkin et al. (2024) can be understood as a computational version of the Signal Pooling Hypothesis, suggesting that ensemble perception results from pooling population responses of feature-selective neurons. This can be simulated through a two-layer neural network model. The first layer is a feature layer with relatively small receptive fields, where signals from low-level neurons encoding individual stimuli are fed forward through weighted synaptic connections to the second layer. The second layer is a pooling layer with larger receptive fields that integrates incoming signals to form an overall population response tuning curve reflecting the statistical properties (e.g., mean, variance, distribution shape) of the stimulus set. This pooling mechanism based on population coding can effectively represent statistical information at the neural level through feedforward mechanisms without requiring arithmetic operations of "summing then dividing by number." This model received further support from Iakovlev and Utochkin (2023), who used skewed distributions of orientation sets and found that participants' estimated average orientation was systematically biased toward the distribution mode. The degree of deviation depended not only on the distance between mean and mode but also on the entire distribution shape. Importantly, in the population response model, neural signals pool in feature space, naturally causing the response peak to shift toward denser regions of the distribution (near the mode), with deviation degree influenced by the width of neuronal tuning curves. This provides a neurally plausible explanation for "robust averaging" that down-weights outliers, suggesting that the population response model can serve as a neural basis for understanding non-uniform integration phenomena including saliency weighting and attentional modulation (Iakovlev & Utochkin, 2023; Utochkin et al., 2024).

Additionally, Robinson and Brady's (2023) Perceptual Summation model reveals ensemble representation mechanisms based on the Target Confusability Competition (TCC) framework, validated by experimental data. They argue that ensemble perception reflects direct summation of distributed activation patterns evoked by each item' s features, with representation based on probability rather than point estimates. The model has only one signal-to-noise ratio parameter ( ) estimated from individual item memory tasks and no free parameters. The summation process is thought to occur at early encoding stages rather than memory stages, relying on individual representations rather than being an automatic process independent of individual representations (Robinson & Brady, 2023).

$$r_{ENS} = \text{argmax}\left(\left(\sum f(x)_i^{d'}\right) + \sigma_{noise}\right)$$

Note: r_{ENS} represents the model' s predicted feature value for the ensemble perception task (Ensemble Task, abbreviated as ENS), i.e., the value representing the ensemble average feature that the model predicts participants will select. d' (d prime) is the signal-to-noise ratio parameter quantifying the strength or clarity of each individual memory representation, adjusting the intensity of activation patterns $f(x)_i$ according to memory task demands (e.g., memory load, encoding time, memory interval). (Quoted from Robinson & Brady, Nature Human Behavior (pp. 1638-1651) (2023))

Computational models provide a potential neural computational framework for ensemble perception. The population response model simulates how statistical information (including robust estimation of skewed distributions) can be efficiently generated from individual information through neural networks. The perceptual summation model emphasizes simple linear summation of early perceptual responses.

3.3 Evidence Against Simple Linear Averaging

Increasing evidence shows that the brain' s integration of individual information to form ensemble perception is far more complex than expected. Some studies have challenged the linear pooling mechanism held by Whitney' s pooling hypothesis and the linear summation model without free parameters, suggesting that integration of set items may involve more complex nonadditive integration mechanisms or non-uniform weighting (Choi & Chong, 2020; Jia et al., 2022; Wang et al., 2023; Gong et al., 2025; Tiurina et al., 2024; Kanaya et al., 2018).

On one hand, the integration process may not be simple linear summation based on member representations. To investigate whether specialized neural mechanisms for ensemble perception exist in the brain, Jia et al. (2022) used Steady-State Visual Evoked Potential (SSVEP) to track neural responses to ensemble representations of average size in periodically changing ring sets. They observed neural responses to overall average size at parieto-occipital electrodes and used Time Response Function (TRF) to separate neural responses to individual size and inter-item interactions (including global and local interactions). Results showed that only global interactions directly contributed to overall size perception, indicating the existence of specific neural mechanisms in the brain dedicated to representing ensemble size. This mechanism does not involve simple linear averaging of individual information but represents relative relationships between individuals, manifesting as high-dimensional integration (Jia et al., 2022). Another study from the same team used regression-based Event-Related Potentials (rERPs) to further propose that ensemble perception relies on nonadditive integration mechanisms (Wang et al., 2023). They separated neural responses at three information levels from EEG signals (individual items,

local interactions, global interactions) and found that only neural representations of global interactions correlated significantly with behavioral perceptual precision, and distributed attention could enhance early neural responses of global interactions. Therefore, linear averaging is insufficient to capture interactions between stimuli and their contribution to the whole, failing to accurately reflect complex relationships within stimulus sets (Jia et al., 2022; Wang et al., 2023). Gong et al. (2025) also provided direct neural evidence for this integration mechanism beyond simple linear summation. Using IEM to analyze MEG data, they found that for heterogeneous orientation sets, the brain' s average orientation representation reconstructed via IEM differed significantly from that predicted by linearly summing responses to individual member orientations. Specifically, actual representations showed higher fidelity and response strength at the average orientation. Additionally, Bayesian probabilistic decoding of source-reconstructed neural activity also confirmed that neural activity patterns evoked by heterogeneous sets differed from patterns simulating simple summation, with the former showing higher orientation response strength at the average orientation. These findings strongly suggest that when forming ensemble representations, the brain does not simply perform linear summation of individual information but involves additional, more complex integration processes (Gong et al., 2025).

On the other hand, weighting during integration may be non-uniform, with attention, item saliency, and spatial location playing important roles. For example, more salient items (e.g., larger, higher frequency) may be weighted more heavily, causing perceived averages to be biased toward these items and producing an "amplification effect" (Kanaya et al., 2018). Selective attention can also actively modulate this weighting process, not only directly increasing weights of attended items but also influencing final average estimates through "perceptual enlargement" (i.e., increasing weights of attended items), particularly when attention is pre-cued to specific items (Choi & Chong, 2020). Moreover, this weighting process is not uniform across visual space. Tiurina et al. (2024) found that when estimating ensemble average orientation, participants showed clear spatial anisotropies, specifically a strong bias toward stimuli presented at the fovea, with this bias increasing as stimulus uncertainty increased (e.g., when stimulus orientation was oblique or ensemble variance increased).

Human cognitive processes are often more complex than models. Behavioral and neuroscientific evidence reveals that nonadditive integration, saliency- and attention-based weighting, and stimulus spatial location all play important roles in forming ensemble perception. These empirical studies reveal the complexity of integration mechanisms in ensemble perception, challenging simple linear averaging or uniform weighting assumptions. Future research needs to integrate findings from these different approaches (e.g., computational methods of integration, temporal stages, attentional modulation, spatial weighting) to build more comprehensive neural integration models of ensemble perception.

4 Neural Basis Specific to Single-Stimulus Coding

Whether ensemble coding and single-item coding are fundamentally different seems to be the most famous debate in the ensemble perception field (Ariely, 2001; Corbett et al., 2023). Are there differences in neural substrates and processing characteristics between processing ensemble wholes and single stimuli? How does the brain distinguish and process ensemble global attributes versus member information? This section explores evidence for potential dissociation between neural mechanisms of ensemble coding and individual coding, including theoretical frameworks of underlying mechanisms, processing priority, involved neural pathways, sensitivity to different information, and representation in working memory.

4.1 Neural Basis of Ensemble and Individual Coding

Research has shown that specific neural mechanisms exist in the brain for representing ensemble attributes (Roberts et al., 2019; Jia et al., 2022). Recent studies have used various neuroimaging techniques to investigate how the brain processes information from ensemble versus single stimuli, analyzing activation patterns and functional division of labor in relevant brain regions. For example, ensemble coding and individual coding show functional dissociation between dorsal and ventral pathways (Im et al., 2017; Im et al., 2021). Spatial distribution of stimuli (Sama et al., 2024) and different spatial frequencies (Zhao, Shen et al., 2023) also differentially affect ensemble versus individual stimulus information. Additionally, research has examined brain regions involved in working memory storage of ensemble representations (e.g., prefrontal and occipitoparietal cortex) and their pattern differences (Oh et al., 2019).

First, Roberts et al. (2019) used EEG to explore the neural basis of facial identity ensemble perception, investigating whether specific neural mechanisms exist for holistic processing of face ensemble information. The experiment presented participants with face sets or single faces. The ensemble condition showed smaller P1 amplitudes and shorter N1 and N2 latencies. Multivariate pattern analysis using linear Support Vector Machine (SVM) for neural decoding revealed that neural signals could distinguish not only different single faces but also face sets with different average identities. Interestingly, neural signals struggled to distinguish between two sets with the same average identity but different members (results marginally significant). The temporal course of ensemble representation decoding differed from that of single-face representation, reflecting a gradual accumulation of perceptual information over time. Neural signal-based image reconstruction could accurately visualize ensemble representation content. In another article from this research group (using face set data from Roberts et al., 2019), multivariate feature selection based on linear SVM and recursive feature elimination found that neural signals (in time and frequency domains) could distinguish different face sets. Compared to single faces and words, face sets relied not only on temporal-occipital electrode channels but also on central channels (Nemrodov et al., 2020).

Additionally, to examine perceptual differences between ensemble emotion coding and single-face emotion coding, researchers used fMRI to explore dissociation in activated brain regions. They found that the intraparietal sulcus and superior frontal gyrus in the dorsal pathway participated in holistic ensemble emotion perception, while the fusiform cortex in the ventral pathway participated in individual face emotion perception. This study also found right hemisphere lateralization advantage for facial emotion ensemble coding (Im et al., 2017).

In 2021, this team replicated their previous findings using MEG, showing that the dorsal pathway participated in holistic processing of ensemble facial emotion while the ventral pathway participated in identifying and discriminating individual facial emotion. Importantly, MEG revealed that the dorsal pathway could respond very rapidly to ensemble facial emotion (68 ms post-stimulus), likely relying on fast information input from the magnocellular pathway to form ensemble perception of facial emotion (Im et al., 2021).

Zhao et al. (2023) further explored how different spatial frequencies affect ensemble facial emotion perception. They generated high spatial frequency (HSF), low spatial frequency (LSF), and broadband (BSF) face ensemble stimuli through different filtering and tested participants' recognition of happy, neutral, and fearful emotions in face ensembles. They found that LSF conditions facilitated recognition of fearful ensemble faces. fMRI results revealed differential roles of spatial frequency in group emotion perception: HSF information primarily activated the ventral visual pathway (e.g., fusiform gyrus, bilateral middle and inferior occipital gyri), processing detailed information but poorly supporting holistic emotion identification; broadband information conditions activated the right inferior frontal gyrus, indicating that integrating multi-frequency information requires more cognitive resources.

To further explore division of labor within ensemble information processing, Sama et al. (2024) used a novel center-surround paradigm (one central face surrounded by six peripheral faces) combined with EEG and MVPA to reveal the dynamic temporal course of neural representations for central versus peripheral ensemble faces. They found that neural decoding of the central face occurred significantly earlier than decoding of peripheral ensemble faces. Interestingly, the temporal course of central face decoding was very similar to that of single faces presented alone, while decoding patterns differed extensively between single faces and ensemble faces. This suggests that the brain may process attended individual information (even when within an ensemble) and broader ensemble information through relatively independent neural mechanisms. This finding provides direct evidence for temporal dynamic dissociation between ensemble coding and individual coding and between peripheral and central focus information, emphasizing the potential role of attention in distinguishing these two processing types (Sama et al., 2024). Additionally, by independently manipulating face shape and surface attributes, the study found both attributes played important roles in ensemble representation, with the latter contributing more, as reflected in higher classification accuracy.

Finally, some researchers have examined whether working memory representations are structured like visual hierarchical representations, with higher-level brain regions representing statistical summary information and lower-level regions representing complex sensory information. Using line orientation ensemble stimuli and IEM to decode EEG signals, results supported the structured representation hypothesis. Specifically, both stable coding of simple features (stable coding, generalizable over time) and stable coding of ensemble averages existed in frontocentral regions. Importantly, frontocentral activity even correlated significantly with behavioral ensemble judgments. In parieto-occipital regions, dynamic coding of features existed (dynamic coding, not generalizable over time), with stable coding of ensemble averages only present in target adjustment tasks, not in old-new judgment tasks, indicating modulation by task demands. In summary, the higher the brain region in the visual hierarchy, the more abstract the content of working memory representation (Oh et al., 2019).

4.2 Reverse Hierarchy Theory

The Reverse Hierarchy Theory (RHT) proposed by Hochstein and colleagues (Hochstein & Ahissar, 2002; Hochstein et al., 2015) provides a fundamental theoretical framework for the internal mechanisms of ensemble perception. Some argue that RHT, as a broad and comprehensive theoretical framework explaining rapid processing of ensemble information (e.g., "seeing the forest before the trees"), is currently the most effective model for explaining ensemble perception (Corbett et al., 2023).

In contrast to traditional "bottom-up" models, RHT emphasizes that visual perception first grasps the "gist" of the ensemble. Its core view is that visual processing follows a "fast feedforward and detailed feedback" pattern along the visual cortical hierarchy (see Figure 1 [Figure 1: see original paper]). Initial visual processing is implicit and feedforward, with information about local features transmitted bottom-up and forming summary representations of the ensemble at higher brain regions (i.e., "vision with a glance"). Subsequently, conscious perception follows a reverse hierarchical path: when local detail information needs to be processed, conscious perception returns to lower-level brain regions via feedback connections in a top-down manner to form precise representations of individual members (i.e., "vision with scrutiny"). Local processing can then correct and supplement details to the initial visual representation approximation (Hochstein & Ahissar, 2002). Hochstein et al. (2015) further noted that this "global-first, local-refinement" processing progression is closely related to the rapid, holistic processing characteristics of ensemble perception.

Figure 1. Traditional hierarchical structure and Reverse Hierarchy Theory (RHT).
Note: The traditional view holds that visual input is first received and processed by neurons in low-level visual cortical areas that respond to simple geometric shapes of stimuli. Information is then transmitted bottom-up layer by layer, representing global features. Finally, information is integrated in higher cortical

areas to form representations of abstract shapes, objects, and categories, without feedback involvement. RHT proposes that feedforward pathways only implicitly process feature information, while conscious visual holistic perception forms in high-level cortex as summary representations of scene gist. Conscious perception then returns to corresponding low-level cortex via feedback connections to form more detailed representations. (Adapted with permission from Hochstein & Ahissar, Neuron (pp. 791-804) (2002))

Empirical studies also support RHT' s assumption that ensemble representation precedes local representation. For example, Tian et al. (2021) indirectly revealed this holistic priority through average discrimination and attractiveness rating tasks. They found that participants' perception of average attractiveness depended on average face representation. When facing large-capacity sets (e.g., 12 faces), the dominant role of synthesized average representation was more significant, while with smaller sets (e.g., 4 faces), more detail processing resources were allowed, individual face information influence increased, and the dominance of synthesized average representation decreased accordingly. Recent ERP studies have also reached supportive conclusions. Epstein and Emmanouil (2021) found that compared to member oddballs (defined by specific line segment orientation), participants responded faster to ensemble oddballs (defined by ensemble average orientation), with earlier EEG neural response latencies, earlier decoding, higher accuracy, and stronger generalization patterns. Liu et al. (2023) further revealed the time-dependency of this processing priority by manipulating presentation duration and item number of ensemble faces: at short durations (e.g., 50 ms), participants' judgments of average facial emotion were better than single-face emotion, while at long durations (e.g., 750 ms) the opposite was true. Additionally, ERP results showed that the N2pc component reflecting item individuation was unaffected by face number in the set under short-duration conditions but increased with face number under long-duration conditions.

Therefore, RHT provides an explanatory framework for understanding different processing mechanisms and priorities when the brain processes ensemble versus member information. The dissociation between ensemble and member information in neural mechanisms and temporal course not only explains the global precedence effect observed at the behavioral level but also implies differences in neural substrates between ensemble coding and individual coding. However, we must recognize that the complex mechanisms of ensemble perception may not be fully covered by a single theory. While viewing RHT as an effective perspective for explaining ensemble perception, we should also integrate other theoretical models and related empirical research to build a more comprehensive understanding (Corbett et al., 2023).

5 Initial Exploration of Neural Mechanisms for Variance Representation

Variance representation helps humans identify and evaluate information diversity, uncertainty, and reliability, and also guides attentional bias based on statistical saliency, helping us quickly locate "anomalies"to flexibly cope with complex external worlds. Deeply investigating the neural mechanisms of variance representation will reveal how the brain encodes the dispersion or heterogeneity of stimulus sets, helping to clarify the internal mechanisms by which the visual system represents group diversity and regularity information. However, current neuroscience research on variance representation remains limited.

Visual variance representation may occur in early visual pathways, perhaps even beginning in subcortical structures. For example, Norman et al. (2015) found that variance adaptation depends on retinal coordinates (relative to fixation position) rather than spatial visual coordinates. To further explore the neural basis of variance coding, this study tested a patient with visual cortex damage and found no significant difference from normal participants in accuracy of grating orientation variance representation. The patient only had intact left hemisphere V1, with damage to many visual areas including bilateral ventromedial occipitotemporal cortex, right hemisphere V1, and partial bilateral V2, V3, and V4 regions, suggesting that variance representation may occur at early stages of the visual system, such as V1 (Norman et al., 2015). Animal studies also provide evidence that variance representation depends on early visual processing pathways; for example, the lateral geniculate nucleus (LGN) in cats can produce specific responses to brightness standard deviation (Bonin et al., 2006). Behavioral studies also suggest that subcortical structures may play important roles in variance representation. Using a stereoscope to present stimuli separately to participants' eyes, researchers found that participants' accuracy in variance judgment was higher when two stimulus sequences were presented to the same eye (monocular presentation) than when presented to different eyes (dichoptic presentation). This "same-eye advantage effect" suggests that visual variance information may be preliminarily processed in subcortical structures before signals reach V1 (Zeng et al., 2024).

Recent electrophysiological studies provide direct evidence for automatic processing mechanisms of variance. Chen and Ji (2025) used a reverse oddball paradigm in an ERP study, requiring participants to focus attention on a central fixation discrimination task while peripheral face ensembles unrelated to the task were presented. They manipulated consistency of emotional intensity within sets to distinguish high versus low emotional variance (which served as standard and oddball stimuli). Results showed that when average emotion was neutral, only high-emotion-variance face ensembles as oddball stimuli evoked vMMN components. When average emotion was angry, both high- and low-emotion-variance oddball stimuli could evoke late visual mismatch positivity (vMMP) and vMMN components, respectively, indicating that the brain can automatically process variance information of facial emotion ensembles under

non-attentional conditions. However, when average emotion was happy, no significant effects were observed, suggesting that social information such as emotion type can modulate the brain' s sensitivity to variance changes. In terms of temporal course, MVPA results further showed that across all conditions, the brain could begin discriminating stimuli with different variance levels at a very early stage (approximately 70–90 ms post-stimulus).

Furthermore, different variance levels in stimulus sets may lead to differential activation strength in brain regions. On one hand, activity in some brain regions positively correlates with variance level. For example, Michael et al.' s (2015) fMRI study found that BOLD signals in visual cortex and superior parietal lobule increased with higher stimulus set variance. Similarly, Allard et al. (2021) distinguished high- versus low-variance orientation groups by varying orientation differences in grating sets, asking participants to judge whether the average orientation was clockwise or counterclockwise relative to vertical. fMRI results showed that high-variance orientation grating sets activated right superior frontal gyrus and left middle frontal gyrus significantly more than low-variance sets. Activation in these high-level brain regions observed in the study may be related to participants needing to complete ensemble average orientation discrimination tasks (Allard et al., 2021). On the other hand, Michael et al. (2015) revealed the opposite activity pattern: BOLD signals in dorsomedial Prefrontal Cortex (dmPFC) and anterior insula actually decreased with increasing variance, showing stronger activation when stimuli were more homogeneous (low variance). Overall, different brain regions may play different roles in variance representation, with activation patterns related not only to task demands and stimulus physical salience (Allard et al., 2021; Michael et al., 2015) but also closely linked to the social significance of stimuli themselves (e.g., emotion) (Chen & Ji, 2025).

However, from early animal experiments and single-patient brain damage studies to later behavioral, brain imaging, and electrophysiological studies, we cannot draw definitive conclusions about human neural mechanisms for dispersion representation from such limited research. Future studies could adopt different experimental paradigms and techniques to provide new insights. For example, combining explicit judgment paradigms that require participants to directly compare or estimate set variance with fMRI can identify brain regions whose activity changes systematically with stimulus variance. Additionally, applying the classic neural adaptation paradigm to variance representation tasks combined with fMRI adaptation suppression analysis is an effective means to test for the existence of variance-specific tuned neuronal populations. Meanwhile, MVPA methods can decode distributed neural activity patterns to provide further insights into specific representation modes of variance information. Finally, applying high temporal resolution EEG/MEG techniques to these paradigms could reveal the complete temporal course of variance information extraction and processing. Notably, future research addressing this issue needs to distinguish potential effects of different visual-level stimuli while also attending to interactions between average and variance representations (see Section 6.3

"Domain-General or Domain-Specific Mechanisms").

6 Research Outlook

This article systematically reviews research on neural mechanisms of visual ensemble perception, covering its basic characteristics, temporal dynamics, integration mechanisms, and dissociation between ensemble and individual coding. Future research should not only focus on the temporal course of ensemble coding and the significance of feedback signals but also need to address the functions of different visual pathways and potential influences of development and experience. We also recommend that researchers distinguish between neural mechanisms of ensemble perception and statistical summary representation. Finally, based on existing research, we propose that ensemble perception follows a "Coarse-Fine-Refine" multi-stage processing model and discuss its similarities and differences with major theories in the field (see Section 6.6).

6.1 Ensemble Perception and Feedback Signals

Different theoretical models hold different views on whether ensemble perception involves feedback signals. On one hand, feedforward-dominant models (such as Signal Pooling Hypothesis and population response model) advocate that ensemble perception is based on a feedforward process. Neural signals pool progressively along the visual hierarchy from lower to higher brain regions, ultimately forming stable representations of ensemble statistical properties. Feedback roles are limited to relatively minor functions, such as marking or identifying extreme values in ensembles (Whitney et al., 2014; Utochkin et al., 2024). On the other hand, stage-processing models argue that top-down feedback is crucial. For example, Reverse Hierarchy Theory proposes that visual perception follows a "forest before trees" global precedence principle, where feedforward forms overall awareness of the forest and feedback forms awareness of member trees (Hochstein et al., 2015). Similarly, Predictive Coding (PC) theory also emphasizes the important role of feedback (Gilbert & Sigman, 2007; He et al., 2012; Shipp, 2016). According to predictive coding, ensemble perception may involve feedback processes: for sets or groups of multiple objects, we may first form predictions about ensemble statistical properties, which are then fed back to visual cortex to compare with actual visual input. Differences between them constitute prediction errors, which the visual system minimizes by continuously updating predictions. Both Reverse Hierarchy Theory and predictive coding can explain the previously observed phenomenon of ensemble precedence, where people' s average representations of ensembles are better than their representations of members (Chong & Treisman, 2003; Li et al., 2016).

Although direct neural evidence confirming the importance of feedback in ensemble perception is still needed, research from related fields such as perceptual integration and feature binding provides indirect support. For example, Liu et al. (2017) studied brain activity during integrated pattern perception and found

that during integration of discrete stimuli in the visual field, the posterior intraparietal sulcus (IPS) in the dorsal visual pathway was rapidly activated (within 100 ms) and modulated activity in early visual areas (EVAs) through feedback. In feature binding tasks, researchers also found feedback connections from V5 to V2 (Zhang et al., 2014).

6.2 Functions of Different Visual Processing Pathways in Ensemble Perception

First, regarding whether ensemble perception relies on ventral or dorsal pathways, existing research evidence is relatively scarce. Whitney et al. (2014) argue that ensemble perception is not mediated by a single brain region or pathway. For example, low-level visual features like orientation depend on the occipital lobe, object motion or position information depends on the dorsal pathway, while high-level stimuli like faces or shapes depend more on the ventral temporal pathway. However, while empirical research partially supports the multi-pathway view, it reveals different roles for different brain regions: ensemble perception of facial emotion relies more on the dorsal visual pathway (Im et al., 2017, 2021; Zhao, Shen et al., 2023), while representation of orientation ensembles is not limited to the occipital lobe but extends to parietal cortex (Tark et al., 2021). Similarly, the roles of magnocellular (M) and parvocellular (P) pathways—closely related to dorsal and ventral pathways—in ensemble perception remain controversial (Im et al., 2021; Lee & Chong, 2021). Evidence supporting magnocellular involvement comes from Im et al.' s (2021) MEG study, which observed rapid activation of the dorsal pathway around 68 ms after ensemble face stimulus presentation, with temporal characteristics highly consistent with magnocellular rapid transmission properties. However, a behavioral study using a flicker adaptation paradigm reached the opposite conclusion: suppressing magnocellular activity not only failed to impair ensemble perception precision but actually improved it, leading the authors to argue that ensemble perception relies more on fine information provided by parvocellular pathways (Lee & Chong, 2021). Future research needs to examine whether dorsal and ventral pathways and magnocellular and parvocellular pathways play different roles and have different importance in ensemble perception.

Furthermore, whether V1 can process statistical information remains controversial. V1 neurons have extremely small receptive fields that often cannot simultaneously cover multiple items in a set, leading to the belief that ensemble perception occurs in higher visual areas beyond V1. However, direct evidence supporting this view is limited: Joo et al. (2009) found that neural computation of visual statistical information occurs after binocular fusion and binocular suppression, i.e., not earlier than V1. In contrast, evidence supporting V1 or earlier stage involvement comes from multiple reports. On one hand, from early animal physiology studies (Bonin et al., 2006) to observations of patients with extrastriate and high-level visual cortex damage (Norman et al., 2015), variance representation appears to occur in V1. On the other hand, recent behavioral

studies found that the brain can represent and extract mean and variance information for circle sizes before binocular information fusion, suggesting that ensemble perception may even originate in subcortical structures before V1 (Zhao, Zeng et al., 2023; Zeng et al., 2024).

Importantly, existing neuroimaging studies have inconsistent conclusions about V1. Tark et al. (2021) used fMRI and IEM analysis and found that although occipital V1/V3 regions showed significant ensemble orientation selectivity when average orientation was an irrelevant feature in individual tasks, considering V1 receptive field size limitations, the authors ultimately concluded that V1 does not directly participate in ensemble perception, emphasizing instead the role of V3. In contrast, Gong et al. (2025) used high temporal resolution MEG combined with IEM and source reconstruction analysis to show that V1, V2, and V3 all contributed to average orientation coding. They revealed delayed activation of V1 from a temporal dimension, proposing that V1 may encode ensemble information through two mechanisms: first, relying on horizontal connections within early visual areas to promote interactions between individual items; second, using iterative or recursive processes of recurrent signals to optimize weight distribution within ensembles for precise ensemble perception.

From behavioral to neural evidence, it is difficult to draw definitive conclusions about human neural mechanisms of ensemble perception. Differences in research results likely stem from variations in stimulus features themselves. Future research addressing this issue should also distinguish potential effects of different stimulus levels while further investigating whether average and variance representations have different neural mechanisms.

6.3 Domain-General or Domain-Specific Mechanisms

A core and unresolved question is whether the visual system relies on domain-general neural mechanisms or multiple domain-specific mechanisms when processing statistical summary information across different levels (e.g., orientation, facial identity), different visual features within the same level (e.g., color, orientation), or different statistical indices (e.g., mean, variance). Current research on this issue mainly focuses on the behavioral level, with many conflicting conclusions (Chang, Cha, McGugin et al., 2024; Haberman, Brady et al., 2015).

First, early individual-differences-based studies found no correlation between tasks across different visual levels (i.e., between-domain, high vs. low), supporting specific mechanisms (Haberman, Brady et al., 2015). However, recent studies using latent variable models found that a single latent variable could explain performance across levels, supporting a domain-general hypothesis (Chang, Cha & Gauthier, 2024). Cross-domain variance adaptation aftereffect studies also suggest that variance representation may involve domain-general mechanisms (Maule & Franklin, 2020). Second, within the same visual level (within-domain), low-level color and brightness (Rajendran et al., 2021) and size and orientation (Yoruk & Boduroglu, 2020) may involve specific mechanisms, but other stud-

ies found shared mechanisms between size and orientation (Kacin et al., 2021; Yang et al., 2018). High-level features (e.g., facial identity and emotion) may be dissociated (Haberman & Whitney, 2009; Kwon & Chong, 2023), but studies of non-face complex objects found evidence of domain-general ability (Chang & Gauthier, 2022). Finally, regarding different statistical indices such as mean and variance, some studies reveal independent processing (Norman et al., 2015; Michael et al., 2015), such as no significant correlation between mean and variance estimation for size and orientation (Khvostov & Utochkin, 2019; Yang et al., 2018). However, more studies reveal interactive influences and shared mechanisms (Chang, Cha & Gauthier, 2024; Hansmann-Roth et al., 2021; Jeong & Chong, 2020; Michael et al., 2014; Tong et al., 2015). For example, in two-alternative forced-choice tasks with larger sample sizes, participants' judgments of mean size and variance size for dot sets showed significant correlation (Cha et al., 2022), and perceptual adaptation effects for color variance depended on color mean (Maule & Franklin, 2020).

Conflicting behavioral results highlight the limitations of inferring underlying mechanisms from behavioral measures alone. Behavioral correlations cannot distinguish early visual feature extraction from later statistical processing in temporal course, nor can they differentiate shared neural computational processes from other potential confounding factors. Therefore, to truly resolve the "specificity"versus "generality"debate in ensemble perception coding mechanisms, future research urgently needs to go beyond behavioral correlation analysis and directly employ neuroscientific techniques.

In fact, neural evidence reveals that magnocellular-related dorsal pathways may play a key role in ensemble perception, with significant parietal activation found for both grating and face stimuli (Tark et al., 2021; Im et al., 2017, 2021). We therefore propose that in the early stage of ensemble coding, there exists a domain-general mechanism dependent on magnocellular pathways, followed by domain-specific mechanisms dependent on parvocellular systems and various brain regions (see the "Coarse-Fine-Refine" model below). This model integrates the debate between domain-general and domain-specific mechanisms. Future research can use fMRI-based MVPA combined with Representational Similarity Analysis (RSA) to directly compare whether activation patterns for different stimulus types (high/low level), different features, or different statistical indices (mean/variance) overlap or show similar neural representations, thereby determining whether their neural bases are shared or dissociated. The high temporal resolution of EEG/MEG can help reveal similarities and differences in processing time courses across different task types, more clearly elucidating differences and commonalities in neural mechanisms of ensemble perception across dimensions and providing more solid evidence for understanding how the visual system efficiently processes complex information.

6.4 Neural Development and Experience Effects

Ensemble perception is a relatively automatic psychological process, leading us to naturally expect genetic influences to outweigh experiential ones. Unfortunately, existing limited research cannot resolve this issue.

First, in typical neural development, ensemble perception ability gradually improves with age. Studies show that although 4–5-year-old children can represent average size of object sets, their representation efficiency is lower than adults' , suggesting that ensemble perception ability is a gradually maturing process (Sweeny et al., 2015). Karaminis et al. (2018) observed similar developmental trends in ensemble perception of facial emotion, with children performing worse than adults in both average emotion discrimination and single-face emotion discrimination. These findings collectively indicate that as the visual system and related cognitive functions mature, individuals' ability to extract and utilize statistical summary information is enhanced.

Second, cultural background and individual experience also shape ensemble perception. Zhao et al. (2024) examined ensemble perception of facial emotion in cross-cultural contexts and found significant "other-ethnicity effects," where participants perceived emotions of their own ethnicity more accurately but tended to overestimate the emotional weight of other ethnicities in mixed-race groups. This finding resonates with Peng et al.' s (2021) observation that British participants showed stronger average representation tendencies for their own gender faces (while Chinese participants did not), collectively indicating that cultural experience and social group identity may modulate ensemble perception mechanisms for high-level social stimuli like faces. Additionally, reward value may not affect averaging itself but only the late conscious representation stage of ensemble perception (Dodgson & Raymond, 2020). Similarly, specific long-term experiences and related clinical disorders may affect ensemble perception. Chang et al.' s (2025) ERP study on individuals with Internet Gaming Disorder (IGD) found that IGD patients showed attentional bias toward negative emotions in early automatic processing stages and deficits in later cognitive regulation and interference inhibition. This suggests that long-term specific behavior patterns (e.g., excessive gaming) may alter ensemble perception of social emotional information.

Beyond typical developmental changes, individual differences in ensemble perception and its manifestations in specific neurodevelopmental disorders have begun to receive attention, such as in individuals with Autism Spectrum Disorder (ASD) and Attention-Deficit/Hyperactivity Disorder (ADHD). Karaminis et al. (2018) found that ASD children (typically considered to have difficulties in social information processing) did not differ significantly in task performance from typically developing children, but computational modeling indicated that their ensemble perception ability correlated with nonverbal reasoning ability, suggesting ASD children may adopt different cognitive strategies for ensemble perception. For ADHD individuals, Yuan et al.' s (2025) large-sample rERPs

study found that ADHD patients' attention deficit symptoms correlated significantly with reduced neural responses in both early global and late local information processing, suggesting that core cognitive deficits in ADHD may involve multiple levels of statistical summary information formation.

Thus, ensemble perception ability is not static, and its formation and development are not entirely determined by innate factors. Neural system maturation and individual experience also significantly influence it. Previous research reveals possible interactions between atypical development and environmental factors in shaping ensemble perception ability and provides new perspectives for understanding individual differences and potential interventions. Future research should adopt longitudinal designs to track developmental trajectories and deeply explore specific neural mechanisms by which different experiential factors affect ensemble perception, for example, by combining neuroimaging techniques to examine how experience reshapes neural circuits for ensemble perception.

6.5 Whether Neural Activity in Ensemble Perception Equates to Statistical Summary Representation

Although most researchers use the two concepts interchangeably—for example, Whitney' s team argues that when observing an ensemble, people can automatically extract statistical summary information, and that automated coding of ensemble information equates to statistical summary representation (Whitney et al., 2014; Whitney & Yamanashi Leib, 2018)—this equivalence is reasonable in behavioral research but problematic in neuroscience research. The reason is that when the brain processes ensemble information, increased member number leads to accumulated neural signals, i.e., enhanced activity levels. For example, studies found that N170 amplitude increased with set size (Puce et al., 2013), and even in primary visual cortex, C1 amplitude evoked by multiple items in a set could be a linear superposition of its member items (Chen et al., 2016). This neural activity enhancement due to quantity increase is not the neural mechanism of statistical summary representation, which depends on global integration of ensemble content.

However, most current research on neural mechanisms of ensemble perception infers mechanisms by comparing differences between ensemble and individual item processing (Im et al., 2017, 2021; Roberts et al., 2019). These neural activity differences include not only the brain mechanisms of statistical summary representation but also neural activity differences caused by quantity differences, making psychological processes impure. Therefore, future research needs to be clearly aware that ensemble perception and statistical summary representation may differ in neural mechanisms.

6.6 Coarse-Fine-Refine Model

Based on existing research, we propose that ensemble perception follows a "Coarse-Fine-Refine" multi-stage process (see Figure 2 [Figure 2: see original

paper]): The "Coarse" stage relies on dorsal visual pathways driven by magnocellular pathways, representing a widely distributed, cross-domain domain-general mechanism that can rapidly extract scene or ensemble "gist" to form coarse summary representations or predictions under limited attentional resources. The subsequent "Fine" stage relies on ventral visual pathways driven by parvocellular pathways, representing relatively refined processing specific to particular stimulus features and showing domain specificity. Finally, the coarse representation or prediction is fed back to stimulus-specific brain regions, and the initial summary impression is calibrated through iterative feedforward-feedback loops to achieve final precise ensemble perception.

Figure 2. Schematic diagram of the "Coarse-Fine-Refine" model of visual ensemble perception (using facial emotion stimuli as an example): (1) The "Coarse" stage processes low-frequency information to rapidly form summary representations; (2) The "Fine" stage processes high-frequency information to form relatively refined representations; (3) The "Refine" stage forms final precise representations through iterative feedforward-feedback loops.

Inferences about the model are as follows:

First, the formation of precise ensemble perception involves multiple stages, a view proposed by integrating series of neural evidence on ensemble perception temporal dynamics. Existing studies show that ensemble information can be extracted by the brain at very early post-stimulus stages and differs neurally from individual/member stimulus processing (Epstein & Emmanouil, 2021; Im et al., 2021; Roberts et al., 2019). Moreover, early EEG component emergence reflects the rapid and even partially automatic processing characteristics of ensemble perception (Ji et al., 2018; Ji et al., 2024; Epstein & Emmanouil, 2021; Chen & Ji, 2025). However, studies using IEM to decode EEG/MEG signals reveal that although early neural signals exist, stable and precise ensemble perception only correlates significantly with behavioral performance at later stages, a process that may involve more complex processing such as recursion and iteration (Yashiro et al., 2024; Gong et al., 2025). These combined evidence suggest that ensemble perception is a multi-stage process from rapid, coarse initial processing to late, precise deep processing.

Second, coarse and fine information in ensemble perception are transmitted via fast magnocellular and slow parvocellular pathways, respectively, a reasonable speculation based on previous theories. Previous research on efficient recognition of objects and natural scenes proposed a Coarse-to-Fine view, suggesting that the visual system first processes low spatial frequency coarse information via magnocellular pathways and projects it to high-level cortex (e.g., frontoparietal cortex) to form initial "gist" perception or predictions (Bar, 2003, 2004; Bar et al., 2006; Kveraga et al., 2007). Subsequently, parvocellular pathways transmit high spatial frequency information for detailed processing (Kveraga et al., 2007; Kandel et al., 2021; Kauffmann et al., 2014). Although ensemble information is not objects or natural scenes, it can be understood as special scenes composed of multiple objects. The physiological characteristics of magnocel-

lular pathways—high temporal resolution and low spatial resolution—facilitate rapid, pre-attentive processing of ensemble information, forming coarse statistical summary signals in dorsal pathways (e.g., intraparietal sulcus). Existing neural evidence has revealed important roles of both magnocellular and parvocellular pathways in ensemble perception (Im et al., 2017, 2021; Lee & Chong, 2021; see Section 6.2), and the Coarse-to-Fine view 恰好 unifies these two opposing pieces of evidence.

Third, coarse representations dependent on magnocellular pathways are domain-general, while fine representations dependent on parvocellular pathways are domain-specific, an inference synthesizing existing evidence. Studies show that high-level emotional face ensembles can activate ventral pathways (Im et al., 2017), while low-level grating orientation ensembles activate occipital cortex (Tark et al., 2021), seemingly supporting domain-specific hypotheses (Haberman, Brady et al., 2015; Whitney et al., 2014). However, what cannot be ignored is that both high-level emotional face ensembles and low-level line orientation ensembles can cause significant activation in frontoparietal regions (Im et al., 2017, 2021; Tark et al., 2021; Gong et al., 2025; Michael et al., 2015), which cannot be explained by single domain-general or domain-specific mechanisms. Combined with extensive research revealing that magnocellular pathways play rapid coarse representation roles across various visual stimulus types (Bar, 2003, 2004; Bar et al., 2006; Kveraga et al., 2007), and that slower parvocellular pathways undertake detailed analysis and identification functions (Kveraga et al., 2007; Kandel et al., 2021; Kauffmann et al., 2014), we have reason to speculate that ensemble perception may involve an early, domain-general neural basis in the "Coarse" stage, followed by differential representations in different brain regions for specific stimulus features at different levels (specific processing) that refine "Fine" representations. This view integrates the debate over whether ensemble perception relies on domain-general or multiple domain-specific mechanisms (see Section 6.3).

Finally, coarse representations or predictions formed by frontoparietal networks can act top-down on feedforward signals representing details in stimulus-specific brain regions (Gilbert & Sigman, 2007; Shipp, 2016), calibrating initial summary impressions through feedforward-feedback loops. This view is also a theoretical inference based on predictive coding theory and existing neural evidence. When task demands require higher precision or initial perception mismatches predictions, high-level brain regions such as frontal cortex intervene via feedback connections. For example, through feedback connections, attentional resources are focused on one or more key items in the ensemble (e.g., extreme values or outliers), causing items to be weighted differently in final perception (Choi & Chong, 2020; De Fockert & Marchant, 2008). Additionally, outlier suppression is another manifestation of feedback calibration. Studies show that the visual system can robustly compute averages after excluding outliers (De Gardelle & Summerfield, 2011), a process considered relatively slow and iterative, highly dependent on feedback mechanisms (Epstein et al., 2020). Through iterative feedforward-feedback processing, the visual system can not only integrate en-

semble information but also flexibly adjust representations to ultimately achieve precise and robust ensemble perception.

In summary, the "Coarse-Fine-Refine" model is developed based on predictive coding theory and Coarse-to-Fine theoretical perspectives (Gilbert & Sigman, 2007; He et al., 2012; Shipp, 2016; Bar, 2003, 2004; Bar et al., 2006; Kveraga et al., 2007), tailored to the characteristics of ensemble perception and the latest research evidence. It differs from major theoretical viewpoints in the ensemble perception field. First, the "Coarse-Fine-Refine" model includes both feedforward and feedback processes, while Signal Pooling Hypothesis and population response model emphasize direct feedforward effects. Signal Pooling Hypothesis is a descriptive hypothesis based on behavioral evidence proposing the possibility of neural signal pooling along visual hierarchy (Haberman & Whitney, 2012; Whitney et al., 2014). Population response model provides more specific computational pathways for hierarchical pooling by simulating neuronal population tuning curves (Utochkin et al., 2024; Iakovlev & Utochkin, 2023). The "Fine" process in "Coarse-Fine-Refine" likely achieves hierarchical pooling. In fact, feedback connections from top to bottom are possible (Utochkin et al., 2024), and pure feedforward hierarchical models are insufficient to explain nonlinear phenomena like outlier suppression that require iterative processing (Gong et al., 2025; Wang et al., 2023). Therefore, after fine processing, the feedforward-feedback iterative "Refine" process begins. Second, the "Coarse-Fine-Refine" model also differs in emphasis from Reverse Hierarchy Theory. RHT advocates that conscious summary information forms during feedforward stages, with feedback allocating attentional resources to local details to achieve local member representation (Hochstein & Ahissar, 2002; Hochstein et al., 2015). Although this theory also emphasizes feedback, its role is to achieve local representation, following a "global before local" process. In contrast, our model corresponds to the "global" representation stage in RHT, where the formation of statistical information itself undergoes multiple dynamic stages.

The "Coarse-Fine-Refine" hypothesis integrates multi-faceted evidence on ensemble perception regarding temporal dynamics, neural pathways, and computational mechanisms, providing a unified perspective for understanding how the visual system balances efficiency and precision. Although this provides a richer understanding of how the brain efficiently processes visual ensemble information, it still requires further empirical testing.

In conclusion, exploration of neural mechanisms underlying visual ensemble perception has made significant progress but remains at a critical stage of in-depth development. Future research needs to integrate multidisciplinary perspectives and methods from cognitive psychology, neuroscience, and computational modeling, focusing on elucidating precise neural pathways and temporal courses, synergistic roles of feedforward and feedback processes, domain-generality and specificity of information coding, influences of development and experience, and clarifying differences between concepts. These efforts will greatly deepen our understanding of how the visual system "simplifies complexity" to achieve efficient

abstract representation from complex information.

References

Chen, Z., & Ji, L. (2025). Automatic processing of multiple facial emotion variability: Evidence from visual mismatch components. *Acta Psychologica Sinica*, *57*(9), 1553–1571. https://journal.psych.ac.cn/xlxb/CN/10.3724/SP.J.1041.2025.1553

Hao, S., Ye, Q., & He, W. (2023). Holistic coding of crowd facial emotion and its influencing factors. *Psychological Science*, *46*(01), 50–56. https://jps.ecnu.edu.cn/CN/Y2023/V46/I1/50

Li, Q., & Chen, W. (2022). Differences in ensemble representation across stimuli and attributes. *Chinese Science Bulletin*, *67*(21), 2463–2472. https://doi.org/10.1360/TB-2021-1068

Liu, D. (2021). *Cognitive and neural mechanisms of ensemble representation' s modulatory effect on perceptual decision-making* [Doctoral dissertation]. University of Chinese Academy of Sciences, Beijing.

Tian, X., Hou, W., Ou, Y., Yi, B., Chen, W., & Shang, J. (2021). Average representation mechanism based on composite average stimuli: Evidence from average facial attractiveness. *Acta Psychologica Sinica*, *53*(7), 714–728. https://doi.org/10.3724/sp.J.1041.2021.00714

Tong, K., Tang, W., Chen, W., & Fu, X. (2015). Content and mechanisms of statistical summary representation. *Advances in Psychological Science*, *23*(10), 1723–1731. https://doi.org/10.3724/sp.J.1042.2015.01723

Zhao, B., He, J., Liu, Y., Yang, S., Wang, Z., Zhang, Q., & Bai, X. (2024). Effects of salient stimuli on the numerical effect of holistic perceptual ensembles. *Psychology and Behavior Research*, *22*(2), 212. https://psybeh.tjnu.edu.cn/CN/Y2024/V22/I2/212

Allard, R., Ramanoël, S., Silvestre, D., & Arleo, A. (2021). Variance-dependent neural activity in an unvoluntary averaging task. *Attention, Perception, & Psychophysics*, *83*(3), 1094–1105. https://doi.org/10.3758/s13414-020-

Allik, J., Toom, M., Naar, R., & Raidvee, A. (2022). How are local orientation signals pooled? *Attention, Perception, & Psychophysics*, *84*(3), 981–991. https://doi.org/10.3758/s13414-022-02456-9

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Andrew H. Chwe, J., & Freeman, J. B. (2024). Trustworthiness of crowds is gleaned in half a second. *Social Psychological and Personality Science*, *15*(3), 351–359. https://doi.org/10.1177/19485506231164703

Ariely, D. (2001). Seeing sets: Representation by Statistical Properties. *Psychological Science*, *12*, 157–162. https://doi.org/10.1111/1467-9280.00327

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. https://doi.org/10.1162/089892903321662976

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), https://doi.org/10.1038/nrn1476

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ···& Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, & Psychophysics*, *71*(3), 655–665. https://doi.org/10.3758/APP.71.3.655

Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *Journal of Neuroscience*, *26*(23), 6346–6353. https://doi.org/10.1523/JNEUROSCI.0284-06.2006

Campana, F., Rebollo, I., Urai, A., Wyart, V., & Tallon-Baudry, C. (2016). Conscious vision proceeds from global to local content in goal-directed tasks and spontaneous vision. *Journal of Neuroscience*, *36*(19), 5200–5213. https://doi.org/10.1523/JNEUROSCI.3619-15.2016

Cha, O., Blake, R., & Gauthier, I. (2022). Contribution of a common ability in average and variability judgments. *Psychonomic Bulletin & Review*, *29*(1), 108–115. https://doi.org/10.3758/s13423-021-01982-1

Chaney, W., Fischer, J., & Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. *Frontiers in Integrative Neuroscience*, *8*, 73. https://doi.org/10.3389/fnint.2014.00073

Chang, Q., Hao, B., Fan, C., Luo, W., & He, W. (2025). Ensemble coding of crowd facial emotion in Internet gaming disorder under the emotional interference condition: An ERP study. *Journal of Behavioral Addictions*, *14*(2), 817–830. https://doi.org/10.1556/2006.2025.00027

Chang, T. Y., & Gauthier, I. (2022). Domain-general ability underlies complex object ensemble processing. *Journal of Experimental Psychology: General*, *151*(4), 966. https://doi.org/10.1037/xge0001110

Chang, T. Y., Cha, O., & Gauthier, I. (2024). A general ability for judging simple and complex ensembles. *Journal of Experimental Psychology: General*, *153*(6), 1517. https://doi.org/10.1037/xge0001582

Chang, T. Y., Cha, O., McGugin, R., Tomarken, A., & Gauthier, I. (2024). How general is ensemble perception? *Psychological Research*, *88*(3), 695–708. https://doi.org/10.1007/s00426-023-01883-z

Chen, J., Yu, Q., Zhu, Z., Peng, Y., & Fang, F. (2016). Spatial summation revealed in the earliest visual evoked component C1 and the effect

of attention on its linearity. *Journal of Neurophysiology*, *115*(1), 500–509. https://doi.org/10.1152/jn.00044.2015

Choi, J., & Chong, S. C. (2020). Contextual cueing in target absent trials by distractor–distractor associations. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(12), https://doi.org/10.1037/xhp0000867

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

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

Davis, E. E., Matthews, C. M., & Mondloch, C. J. (2021). Ensemble coding of facial identity is not refined by experience: Evidence from other-race and inverted faces. *British Journal of Psychology*, *112*(1), 265–281. https://doi.org/10.1111/bjop.12457

De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789–794. https://doi.org/10.3758/PP.70.5.789

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722. https://doi.org/10.1080/17470210902811249

De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346. https://doi.org/10.1073/pnas.1104517108

Dodgson, D. B., & Raymond, J. E. (2020). Value associations bias ensemble perception. *Attention, Perception, & Psychophysics*, *82*(1), 109–117. https://doi.org/10.3758/s13414-019-01744-1

Epstein, M. L., & Emmanouil, T. A. (2021). Ensemble Statistics Can Be Available before Individual Item Properties: Electroencephalography Evidence Using the Oddball Paradigm. *Journal of Cognitive Neuroscience*, *33*(6), 1056–1068. https://doi.org/10.1162/jocn_a_{01704}

Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The outlier paradox: The role of iterative ensemble coding in discounting outliers. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(11), 1267. https://doi.org/10.1037/xhp0000857

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, *106*(3), 1389–1398. https://doi.org/10.1152/jn.00904.2010

Florey, J., Clifford, C. W., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Sci Rep*, *6*, 32210. https://doi.org/10.1038/srep32210

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*(5), 677–696. https://doi.org/10.1016/j.neuron.2007.05.019

Gong, X., He, T., Wang, Q., Lu, J., & Fang, F. (2025). Time Course of Orientation Ensemble Representation in the Human Brain. *Journal of Neuroscience*, *45*(7). https://doi.org/10.1523/JNEUROSCI.1688-23.2024

Haberman, J., & Suresh, S. (2021). Ensemble size judgments account for size constancy. *Attention, Perception, & Psychophysics*, *83*(3), 925–933. https://doi.org/10.3758/s13414-020-02144-6

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1–1. https://doi.org/10.1167/9.11.1

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, *15*(4), 16. https://doi.org/10.1167/15.4.16

Hansmann-Roth, S., Kristjánsson, Á., Whitney, D., & Chetverikov, A. (2021). Dissociating implicit and explicit ensemble representations reveals the limits of visual perception and the richness of behavior. *Scientific Reports*, *11*(1), 3899. https://doi.org/10.1038/s41598-021-83358-y

He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high-and low-level visual aftereffects by perceptual grouping. *Current Biology*, *22*(11), 1040–1045. https://doi.org/10.1016/j.cub.2012.04.026

Hochstein, S., & Ahissar, M. (2002). View from the Top: Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*, *36*, 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision*, *15*(4), 7. https://doi.org/10.1167/15.4.7

Iakovlev, A. U., & Utochkin, I. S. (2023). Ensemble averaging: What can we learn from skewed feature distributions? *Journal of Vision*, *23*(1), 5–5. https://doi.org/10.1167/jov.23.1.5

Im, H. Y., Albohn, D. N., Steiner, T. G., Cushing, C. A., Adams, R. B., Jr., & Kveraga, K. (2017). Differential hemispheric and visual stream contributions to ensemble coding of crowd emotion. *Nature Human Behaviour*, *1*, 828–842. https://doi.org/10.1038/s41562-017-0225-z

Im, H. Y., Cushing, C. A., 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. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(4), 776–792. https://doi.org/10.3758/s13415-021-00880-2

Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research*, https://doi.org/10.1016/j.visres.2020.01.002

Ji, L., Chen, Z., Zeng, X., Sun, B., & Fu, S. (2024). Automatic processing of unattended mean emotion: Evidence from visual mismatch responses. *Neuropsychologia*, https://doi.org/10.1016/j.neuropsychologia.2024.108963

Ji, L., Rossi, V., & Pourtois, G. (2018). Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs. *Neuropsychologia*, https://doi.org/10.1016/j.neuropsychologia.2018.01.022

Jia, J., Wang, T., Chen, S., Ding, N., & Fang, F. (2022). Ensemble size perception: Its neural signature and the role of global interaction between individual items. *Neuropsychologia*, https://doi.org/10.1016/j.neuropsychologia.2022.108290

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, *9*(9), 7–12. https://doi.org/10.1167/9.9.7

Kacin, M., Gauthier, I., & Cha, O. (2021). Ensemble coding of average length and average orientation are correlated. *Vision Research*, *187*, 94–101. https://doi.org/10.1016/j.visres.2021.04.010

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20172770. https://doi.org/10.1098/rspb.2017.2770

Kandel, E. R., Koester, J. D., Mack, S., & Siegelbaum, S. A. (Eds.). (2021). *Principles of neural science* (6th ed.). McGraw-Hill.

Karaminis, T., Neil, L., Manning, C., Turi, M., Fiorentini, C., Burr, D., & Pellicano, E. (2018). Reprint of "Investigating ensemble perception of emotions in autistic and typical children and adolescents" . *Developmental Cognitive Neuroscience*, *29*, 97–107. https://doi.org/10.1016/j.dcn.2018.02.003

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 37. https://doi.org/10.3389/fnint.2014.00037

Khvostov, V. A., & Utochkin, I. S. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision*, *19*(9), 3. https://doi.org/10.1167/19.9.3

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145–168. https://doi.org/10.1016/j.bandc.2007.06.007

Kwon, D., & Chong, S. C. (2023). The relationship between ensemble representations of facial information. *Vision Research*, *203*, 108156. https://doi.org/10.1016/j.visres.2022.108156

Lee, J., & Chong, S. C. (2021). Quality of average representation can be enhanced by refined individual items. *Attention, Perception, & Psychophysics*, *83*(3), 970–981. https://doi.org/10.3758/s13414-020-02139-3

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of Individual Items during Ensemble Coding of Facial Expressions. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2016.01332

Liu, L., Wang, F., Zhou, K., Ding, N., & Luo, H. (2017). Perceptual integration rapidly activates dorsal visual pathway to guide local processing in early visual areas. *PLoS Biology*, *15*(11), e2003646. https://doi.org/10.1371/journal.pbio.2003646

Liu, R., Ye, Q., Hao, S., Li, Y., Shen, L., & He, W. (2023). The relationship between ensemble coding and individual representation in crowd facial emotion. *Biological Psychology*, https://doi.org/10.1016/j.biopsycho.2023.108593

Lukashevich, A., Sigurdardottir, H. M., Kudriavtsev, N., & Utochkin, I. (2025). The role of attention in basic ensemble statistics processing. *Neuropsychologia*, https://doi.org/10.1016/j.neuropsychologia.2025.109086

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average perception. *Acta Psychologica*, *142*(2), https://doi.org/10.1016/j.actpsy.2012.11.002

Marini, F., Sutherland, C. A., Ostrovska, B., & Manassi, M. (2023). Three' s a crowd: Fast ensemble perception of first impressions of trustworthiness. *Cognition*, *239*, 105540. https://doi.org/10.1016/j.cognition.2023.105540

Maule, J., & Franklin, A. (2020). Adaptation to variance generalizes across visual domains. *Journal of Experimental Psychology: General*, *149*(4), 662–675. https://doi.org/10.1037/xge0000678

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences*, *111*(21), 7873–7878. https://doi.org/10.1073/pnas.1308674111

Michael, E., de Gardelle, V., Nevado-Holgado, A., & Summerfield, C. (2015). Unreliable evidence: 2 sources of uncertainty during perceptual choice. *Cerebral Cortex*, *25*(4), 937–947. https://doi.org/10.1093/cercor/bht287

Nemrodov, D., Ling, S., Nudnou, I., Roberts, T., Cant, J. S., Lee, A. C. H., & Nestor, A. (2020). A multivariate investigation of visual word, face, and ensemble processing: Perspectives from EEG-based decoding and feature selection. *Psychophysiology*, *57*(3), e13511. https://doi.org/10.1111/psyp.13511

Nguyen, T. T. N., Vuong, Q. C., Mather, G., & Thornton, I. M. (2021). Ensemble coding of crowd speed using biological motion. *Attention, Perception, & Psychophysics*, *83*(3), 1014–1035. https://doi.org/10.3758/s13414-

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, *15*(4), 3. https://doi.org/10.1167/15.4.3

Oh, B. I., Kim, Y. J., & Kang, M. S. (2019). Ensemble representations reveal distinct neural coding of visual working memory. *Nature Communications*, *10*(1), 5665. https://doi.org/10.1038/s41467-019-13592-6

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744. https://doi.org/10.1038/89532

Pavlovskaya, M., Soroker, N., Bonneh, Y. S., & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience*, *27*(7), https://doi.org/10.1162/jocn_a_{00791}

Peng, S., Liu, C. H., & Hu, P. (2021). Effects of subjective similarity and culture on ensemble perception of faces. *Attention, Perception, & Psychophysics*, *83*(3), 1070–1079. https://doi.org/10.3758/s13414-020-02133-9

Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology*, *114*(5), 766–785. https://doi.org/10.1037/pspi0000120

Puce, A., McNeely, M. E., Berrebi, M. E., Thompson, J. C., Hardee, J., & Brefczynski-Lewis, J. (2013). Multiple faces elicit augmented neural activity. *Frontiers in Human Neuroscience*, https://doi.org/10.3389/fnhum.2013.00282

Rajendran, S., Maule, J., Franklin, A., & Webster, M. A. (2021). Ensemble coding of color and luminance contrast. *Attention, Perception, & Psychophysics*, *83*(3), 911–924. https://doi.org/10.3758/s13414-020-02136-6

Roberts, T., Cant, J. S., & Nestor, A. (2019). Elucidating the Neural Representation and the Processing Dynamics of Face Ensembles. *The Journal of Neuroscience*, *39*(39), 7737–7747. https://doi.org/10.1523/JNEUROSCI.0471-

Robinson, M. M., & Brady, T. F. (2023). A quantitative model of ensemble perception as summed activation in feature space. *Nature Human Behaviour*, *7*(10), 1638–1651. https://doi.org/10.1038/s41562-023-01602-z

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*(8), 363–370. https://doi.org/10.1016/j.tics.2004.06.003

Sama, M. A., Nestor, A., & Cant, J. S. (2024). The neural dynamics of face ensemble and central face processing. *Journal of Neuroscience*, *44*(7), e1027232023. https://doi.org/10.1523/JNEUROSCI.1027-23.2023

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, *7*, 1792. https://doi.org/10.3389/fpsyg.2016.01792

Sun, P., Chu, V., & Sperling, G. (2021). Multiple concurrent centroid judgments imply multiple within-group salience maps. *Attention, Perception, & Psychophysics*, *83*(3), 934–955. https://doi.org/10.3758/s13414-020-02197-7

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 329–337. https://doi.org/10.1037/a0028712

Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4-5-year-old children. *Developmental Science*, *18*(4), 556–568. https://doi.org/10.1111/desc.12239

Tark, K. J., Kang, M. S., Chong, S. C., & Shim, W. M. (2021). Neural representations of ensemble coding in the occipital and parietal cortices. *Neuroimage*, *245*, 118680. https://doi.org/10.1016/j.neuroimage.2021.118680

Tiurina, N. A., Markov, Y. A., Whitney, D., & Pascucci, D. (2024). The functional role of spatial anisotropies in ensemble perception. *BMC Biology*, *22*(1), 28. https://doi.org/10.1186/s12915-024-01822-3

Tokita, M., Ueda, S., & Ishiguchi, A. (2016). Evidence for a Global Sampling Process in Extraction of Summary Statistics of Item Sizes in a Set. *Frontiers in Psychology*, *7*, 711. https://doi.org/10.3389/fpsyg.2016.00711

Tong, K., Ji, L., Chen, W., & Fu, X. (2015). Unstable mean context causes sensitivity loss and biased estimation of variability. *Journal of Vision*, *15*(4), 15. https://doi.org/10.1167/15.4.15

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, *146*, 7–18. https://doi.org/10.1016/j.actpsy.2013.11.012

Utochkin, I. S., Choi, J., & Chong, S. C. (2024). A population response model of ensemble perception. *Psychological Review*, *131*(1), 36–57. https://doi.org/10.1037/rev0000426

Wang, T., Zhao, Y., & Jia, J. (2023). Nonadditive integration of visual information in ensemble processing. *Iscience*, *26*(10). https://doi.org/10.1016/j.isci.2023.107988.

Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition*, *152*, 78–86. https://doi.org/10.1016/j.cognition.2016.01.010

Wardle, S. G., Bex, P. J., Cass, J., & Alais, D. (2012). Stereoacuity in the periphery is limited by internal noise. *Journal of Vision*, *12*(6), 12–12. https://doi.org/10.1167/12.6.12

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*, *69*, 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: multiple levels of summary statistical perception. In J. S. Werner, L. M. Chalupa, & M. E. Burns (Eds.), *The new visual neurosciences* (pp. 695–710). MIT Press.

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4), 11. https://doi.org/10.1167/15.4.11

Yamanashi Leib, A., Landau, A. N., Baek, Y., Chong, S. C., & Robertson, L. (2012). Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. *Frontiers in Human Neuroscience*, *6*, 267. https://doi.org/10.3389/fnhum.2012.00267

Yang, Y., Tokita, M., & Ishiguchi, A. (2018). Is there a common summary statistical process for representing the mean and variance? A study using illustrations of familiar items. *i-Perception*, *9*(1), 2041669517747297. https://doi.org/10.1177/2041669517747297

Yashiro, R., Sawayama, M., & Amano, K. (2024). Decoding time-resolved neural representations of orientation ensemble perception. *Frontiers in Neuroscience*, *18*, 1387393. https://doi.org/10.3389/fnins.2024.1387393

Ying, H. (2022). Attention modulates the ensemble coding of facial expressions. *Perception*, *51*(4), 276–285. https://doi.org/10.1177/03010066221079686

Yoruk, H., & Boduroglu, A. (2020). Feature-specificity in visual statistical summary processing. *Attention, Perception, & Psychophysics*, *82*(2), 852–864. https://doi.org/10.3758/s13414-019-01942-x

Yuan, J., Pan, H., Sun, Y., Wang, Y., & Jia, J. (2025). Neural responses to global and local visual information processing provide neural signatures of ADHD symptoms. *International Journal of Psychophysiology*, 112582. https://doi.org/10.1016/j.ijpsycho.2025.112582

Zeng, T., Zhao, Y., Cao, B., & Jia, J. (2024). Perception of visual variance is mediated by subcortical mechanisms. *Brain and Cognition*, *175*, 106131. https://doi.org/10.1016/j.bandc.2024.106131

Zhang, X., Qiu, J., Zhang, Y., Han, S., & Fang, F. (2014). Misbinding of color and motion in human visual cortex. *Current Biology*, *24*(12), 1354–1360. https://doi.org/10.1016/j.cub.2014.04.045

Zhao, D., Shen, X., Li, S., & He, W. (2023). The Impact of Spatial Frequency on the Perception of Crowd Emotion: An fMRI Study. *Brain Sciences*, *13*(12), 1699. https://doi.org/10.3390/brainsci13121699

Zhao, Y., Zeng, T., Wang, T., Fang, F., Pan, Y., & Jia, J. (2023). Subcortical encoding of summary statistics in humans. *Cognition*, *234*, 105384. https://doi.org/10.1016/j.cognition.2023.105384

Zhao, Z., Yaoma, K., Wu, Y., Burns, E., Sun, M., & Ying, H. (2024). Other ethnicity effects in ensemble coding of facial expressions. *Attention, Perception, & Psychophysics*, *86*(7), 2412–2423. https://doi.org/10.3758/s13414-

Note: Figure translations are in progress. See original paper for figures.