Mechanisms of Visual Statistical Summary Representation

Authors: Zhang Xiuling, Zhang Fan, Ge Mingxiao, Li Sijia, Jiang Yi, Zhang Xiuling, Zhang Fan, Jiang Yi

Date: 2022-12-03T00:00:00+00:00

Abstract

Although sensory registration is capacity-unlimited, this does not imply that the visual system is exempt from efficiently processing the vast amounts of visual information encountered in daily life. Statistical summary representation constitutes an efficient information processing mechanism whereby, when presented with a set comprising multiple visual stimuli, the visual system can rapidly extract statistical summary information such as the average properties of the set. Statistical summary representation is an important psychological process for understanding neural computation. Numerous studies have currently uncovered certain cognitive mechanisms underlying this process, including automatic processing mechanisms and mechanisms of specificity or generality across different levels or features. However, only a limited number of studies have examined its neural mechanisms, and future research should employ neuroscientific techniques to directly investigate the neural mechanisms of statistical summary representation.

Full Text

Mechanism of Visual Statistical Summary Representations

Xiuling Zhang[1], **Fan Zhang**[1], Mingxiao Ge[1], Sijia Li[1], Yi Jiang[2]* [1]School of Psychology, Northeast Normal University, Changchun 130024, China [2]State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

Abstract: Despite the unlimited capacity of sensory registration, this does not mean that the visual system does not need to efficiently process the massive amount of visual information we face daily. Statistical summary representation is an efficient information processing mechanism whereby, when presented with

a set composed of multiple visual stimuli, the visual system can rapidly extract statistical summary information such as the average properties of the set. Statistical summary representation is an important psychological process for understanding neural computation. Currently, numerous studies have revealed certain cognitive mechanisms underlying this process, such as automatic processing mechanisms and mechanisms of domain specificity or generality across different levels or features. However, only a small number of studies have focused on its neural mechanisms, and future research requires more investigations using neuroscience techniques to directly examine the neural mechanisms of statistical summary representation.

Keywords: statistical summary representation; ensemble coding; ensemble processing; neural mechanism

Although sensory registration has unlimited capacity, this does not mean that the visual system can ignore processing efficiency issues and leave everything to attention and working memory. In fact, our visual system is not only capable of perceptual organization but can also extract core critical information from ensembles. When presented with a set containing a group of similar stimuli, the visual system can rapidly and accurately extract the overall statistical properties of the set (rather than individual stimulus properties), forming statistical summary representations (Ariely, 2001; Haberman & Whitney, 2007). Both statistical summary representations and perceptual organization reflect the efficiency of visual system processing.

Statistical summary representation is an important psychological process for studying neural computation. Building upon the encoding of individual stimuli, neurons form statistical summary representations of sets composed of multiple stimuli. How this neural computation process works and what its characteristics are have attracted significant attention from researchers. This review covers research on the automatic processing mechanisms of statistical summary representations, as well as debates regarding their domain-specific and domain-general mechanisms, with a focus on introducing recent studies that have employed cognitive neuroscience methods to investigate their neural mechanisms. Based on this foundation, we provide a summary and outlook to offer ideas and directions for future research on the neural mechanisms of statistical summary representations.

Central tendency (such as mean) and dispersion tendency (such as variance) are two important metrics of visual statistical summary representations (Haberman, Lee, et al., 2015; Norman et al., 2015). For a stimulus set, both central tendency representation and dispersion tendency representation appear indispensable for a relatively complete representation (Tong et al., 2015). Mean representation is an important indicator for describing the central tendency of stimulus sets.

In his seminal study on statistical summary representations of circle size, Ariely (2001) found that mean representation was rapid, accurate, and unaffected by set size (i.e., the number of members), yet participants could not accurately

represent individual items within the set, suggesting that ensemble mean representation is privileged over member representation. In addition to stimulus size, researchers have found mean representations across a wide range of stimuli from low-level to high-level features, including stimulus size (Haberman & Suresh, 2021), spatial location (Sun et al., 2021), depth (Wardle et al., 2012), orientation (Parkes et al., 2001), motion direction (Watamaniuk & Duchon, 1992), color and contrast (Rajendran et al., 2021), brightness (Bauer, 2009), and higher-level features such as facial emotion (Haberman, Harp, et al., 2009), facial gender (Haberman & Whitney, 2007), facial identity (de Fockert & Wolfenstein, 2009; Davis et al., 2021), gaze direction (Florey et al., 2016), and biological motion (Sweeny et al., 2013; Nguyen et al., 2021).

Variance representation is an important indicator for describing the dispersion tendency of stimulus sets. Haberman, Lee, and Whitney (2015) were the first to examine statistical summary representations of facial emotion from the perspective of variance. Using an adjustment-matching method, they asked participants to adjust the variance of one set of emotional faces to match the variance of another set of faces with different emotions, using the average adjustment error as the metric. The results showed that participants could reliably represent the variance of emotional face sets, and this ability was unaffected by set size. Subsequent studies have found that our visual system can compute variance across various visual features, such as color (Maule & Franklin, 2020; Ward et al., 2016), brightness (Tong et al., 2015), size (Tokita et al., 2016), and facial gender and race (Phillips et al., 2018), demonstrating that the visual system can effectively represent the dispersion tendency of ensembles.

2. Holistic and Automatic Processing of Visual Statistical Summary Representations

Tong et al. (2015) reviewed two processing modes of statistical summary representations: holistic parallel distributed processing (unaffected by set size or the number of members) and individual-based sampling processing. Holistic processing differs from individual processing mechanisms in that participants can accurately represent the ensemble mean but cannot accurately represent individual items within the set (Ariely, 2001; Whitney & Yamanashi Leib, 2018). Holistic processing is automatic (“fast, task-irrelevant stimuli can also form statistical representations, unaffected by attention allocation, and capable of detecting global changes even when local changes cannot be detected”) (Tong et al., 2015). Therefore, this review will not elaborate further on many studies concerning processing time and attentional resources, focusing instead on new neuroscientific evidence and evidence from unconscious processing studies.

Automatic processing has the following characteristics: fast, unconscious, unaffected by attentional resources (distributed attention or parallel processing, can still be processed when attentional resources are limited), unintentional, and uncontrollable. It should be noted that very few cognitive processes and behaviors satisfy all automatic features, and even driving activities cannot meet all the

above criteria. Therefore, when understanding and distinguishing automatic from non-automatic processes, it is unnecessary to adopt a dogmatic all-or-none standard.

Fortunately, a small number of neuroscience studies have provided supporting evidence for the automatic processing of statistical summary representations. First, regarding the influence of attention, an ERP study investigated whether mean representation of multiple facial expressions requires attention. Using a cueing paradigm, participants were asked to judge whether the average expression of a face set or the expression of a single face was positive or negative under valid and invalid cue conditions. The results showed that participants could still extract the average expression under invalid cue conditions, without observing a significant N2pc component (which reflects spatial selective attention to target stimuli), indicating that participants could extract the average expression under limited attentional conditions. Interestingly, under valid cue conditions, there was no significant difference in the SPCN component between ensemble and individual tasks (which typically increases with the number of items in visual short-term memory until reaching memory capacity), suggesting that multiple faces in a set may be compressed into a "single" object stored in visual short-term memory (Ji et al., 2018).

Second, regarding processing speed or time course, an ERP study used an oddball paradigm to explore the temporal dynamics of ensemble representation composed of multiple lines. The researchers distinguished between oddballs based on an individual item within the set and oddballs based on the entire set, finding that the P3b latency evoked by ensemble-based oddballs was significantly earlier than that for individual-based oddballs. Multivariate pattern analysis (MVPA) also revealed that neural signals could differentiate between standard and deviant stimuli under ensemble conditions earlier than under individual conditions (classification began at 102 ms). This indicates that ensemble perception can occur rapidly and that representation of the whole precedes representation of individual object properties (Epstein & Emmanouil, 2021), consistent with behavioral research by Li et al. (2016) which found that mean representation had already begun by 50 ms and was better than member representation. It should be noted that although ensemble processing occurs earlier than individual processing, some behavioral studies have revealed that statistical summary representations occur at relatively later stages, at least later than size illusions, perceptual constancy using depth cues, and binocular fusion. Specifically, mean size judgments are influenced by the Ebbinghaus illusion, suggesting that mean size may be computed after perceptual size (Im & Chong, 2009); under depth cues from binocular disparity, statistical summary representations are based on objects' real size rather than retinal size, meaning that mean representation occurs on the basis of perceptual constancy (Tiurina & Utochkin, 2019); studies on binocular rivalry have also found that statistical summary representations are formed after binocular fusion (Joo et al., 2009).

The impact of unconscious processing on statistical summary representations re-

mains controversial. Using the crowding paradigm from unconscious processing research (Crowding, where discrimination of a target object in the peripheral vision becomes difficult when other objects are presented around it), researchers found that although participants could not report the orientation of individual gratings, they could reliably estimate the mean orientation of grating ensembles, suggesting that orientation information in primary visual cortex, while not yet reaching consciousness, had already been averaged under crowding (Parkes et al., 2001). Similarly, mean representations of location and facial emotion can overcome the crowding bottleneck (Fischer & Whitney, 2011; Greenwood et al., 2009). However, for mean size of circles, unconscious individual members cannot contribute to mean representation in either crowding or binocular rivalry paradigms (Banno & Saiki, 2012; Joo et al., 2009).

3.1 Encoding Mechanisms of Statistical Summary Representations Across Different Levels and Features

Whether there exist separate neural codes or a general neural code across different visual levels or features—that is, whether the neural mechanisms of visual statistical summary representations are domain-specific or domain-general (or universal, general)—remains a key question. Whitney et al. (2014) proposed that visual mean representations may be computed at multiple levels in the visual system. Some ensembles, such as mean brightness, color, and orientation, may be generated in early cortical or even subcortical areas, while mean facial identity may be represented in later stages of the ventral pathway.

Haberman et al. (2015) first examined the domain-specificity of encoding for visual stimuli at different levels using behavioral experiments. They employed an individual differences approach, having observers extract mean features from different stimulus sets. In a series of experiments (facial identity and grating orientation, facial emotion and dot color, triangle orientation and triangle color, grating orientation and dot color, grating orientation and triangle orientation, facial identity and facial emotion), the researchers calculated the correlation of errors between participants' mean representations for two types of stimuli to investigate whether statistical summary representations across different domains share a common mechanism or multiple mechanisms. The results showed that individuals' mean representation performance was uncorrelated between high- and low-level domains, suggesting the existence of multiple representation mechanisms specific to different high-low level domains—that is, statistical summary representations between high- and low-level domains are domain-specific rather than mediated by a single, general mechanism spanning different levels. This finding has been supported by subsequent studies. Peng et al. (2019) found that priming of interdependent self-construal enhanced mean representation of facial identity but had no effect on mean representation of dot size. Sama et al. (2019), using face ensemble stimuli and a series of behavioral experiments, revealed the independence of mean representations for high-level features (facial identity) and low-level features (facial viewpoint).

Notably, although Haberman et al. (2015) found correlations in mean representation performance between different features within high- and low-level domains respectively, they did not conclude whether mean representations of different features within the same level operate via a general mechanism or separate mechanisms. They acknowledged two possibilities: first, that statistical summary representations within the same level domain may indeed share a general mechanism, or second, that correlations between mean representations of different features may result from error. However, some other studies have drawn conclusions in favor of domain-general mechanisms based on correlational results. In fact, regarding whether domain-general mechanisms exist for different features within high- and low-level domains, there are indeed some controversies among different empirical studies.

Regarding research within low-level domains, most studies suggest that statistical summary representations of different features are domain-specific. Rajendran et al. (2021), in their study of mean representations for color and brightness, found differences between color and brightness mean representations: color mean representation requires indirect inference, whereas brightness can be directly statistically estimated, possibly reflecting the specificity of two distinct summary representation mechanisms. Attarha and Moore (2015), using gratings as stimuli, found that participants could simultaneously represent both size and orientation of gratings, suggesting that mean representations of size and orientation are not limited by processing capacity and implying that the mechanisms underlying statistical summary representations of size and orientation are independent. Yörük and Boduroglu (2020) demonstrated that mean representations of line length and orientation were uncorrelated, possessing independent mechanisms. In contrast, Kacin et al. (2021), attempting to replicate Yörük and Boduroglu' s study, reached the opposite conclusion, arguing for a common, domain-general mechanism across different features. By improving stimulus parameter settings and task designs, they found correlations between mean representations of line length and orientation. Similarly, Yang et al. (2018), using an individual differences approach, found significant correlations in accuracy between mean representations of strawberry size and lollipop orientation, leading the researchers to propose that size and orientation summary representations share a common mechanism. Their view is that if performance on mean representations for two features shows strong correlation, this indicates that these two features' mean representations share a common mechanism.

Domain-specificity in statistical summary representations also exists between different features within high-level domains. Haberman and Ulrich (2019) found certain differences in precision between mean representations of facial identity and facial expression. Studies on face inversion have also revealed a dissociation between mean identity and mean emotion: Haberman and Whitney (2009) found that inversion impaired the precision of mean emotion representation, whereas other studies found that inversion did not affect mean representation of face identity sets (Sun & Choo, 2020; Davis et al., 2021). The differences between mean identity and mean emotion may be based on fundamental differ-

ences between identity and emotion themselves. Haxby et al. (2000), in their model of "the distributed human neural system for face perception," mentioned differences in processing and encoding between facial identity and facial expression, where the fusiform gyrus primarily analyzes invariant facial features and plays an important role in identity recognition, while the superior temporal sulcus mainly analyzes changeable facial features and plays an important role in perceiving facial expression, eye gaze direction, and lip movements.

Currently, very few studies have investigated whether variance representations across different domains are domain-specific. Maule and Franklin (2020) proposed that variance representation may be domain-general. Using a variance adaptation aftereffect paradigm, they found that adaptation to a high-variance hue stimulus ensemble caused participants to underestimate the variance of subsequently presented orientation stimuli—this being the first evidence for cross-domain adaptation aftereffects in visual statistical summary representations. This result implies that visual variance representation is encoded by a domain-general mechanism.

In summary, these behavioral studies show uncontroversially that mean representations exhibit domain-specific mechanisms between high- and low-level domains, while whether variance representations do so remains to be explored. What remains controversial is whether domain-specific mechanisms exist for different features within the same level domain. These controversies suggest that we need not only further detailed and systematic behavioral investigations but also, more importantly, direct examination of the neural encoding mechanisms underlying visual statistical summary representations across different levels and different features within the same level—work that would be significantly pioneering.

3.2 Encoding Mechanisms Between Different Statistical Indices of Statistical Summary Representations—Mean and Variance Representations

Studies on adaptation aftereffects for mean and variance representations have revealed the existence of specific neural codes in the brain responsible for statistical summary representations (Corbett et al., 2012; Norman et al., 2015). Whether the processing mechanisms for these two metrics are identical is an important question in investigating the mechanisms of visual statistical summary representations.

Some studies have found that mean and variance representations are independent of each other. For both orientation and color, adaptation aftereffects to variance are unaffected by changes in mean values (Maule & Franklin, 2020; Norman et al., 2015). In dual-task paradigms, these two ensemble tasks of mean and variance representation do not interfere with each other, also suggesting that they can be estimated independently (Khvostov & Utochkin, 2019). On the other hand, no significant correlation exists between performance on mean

discrimination tasks and variance discrimination tasks (Yang et al., 2018).

However, other evidence challenges the independence of mean and variance representations. In perceptual priming studies, participants could complete mean representation tasks for targets well as long as variability matched (i.e., variability of prime and target stimuli matched in the same stimulus dimension), indicating that variability processing influences mean representation (Michael et al., 2014). Conversely, mean representation can also influence variance representation (Tong et al., 2015). Similarly, in adaptation aftereffect studies with orientation stimuli, adaptation to variance was found to affect mean discrimination, while adaptation to mean affected subsequent variance discrimination (Jeong & Chong, 2020).

4. Possible Neural Mechanisms of Visual Statistical Summary Representations

Visual statistical summary representation is a process that efficiently processes complex visual input, and investigating its neural mechanisms helps us understand how the visual system computes statistical information such as the mean.

Haberman and Whitney (2012) and Whitney et al. (2014) elaborated on theoretical proposals regarding the neural pathways and representation processes through which the brain performs statistical summary representations, suggesting that visual mean representations may be computed at multiple hierarchical levels in the visual system. For example, mean brightness, color, and orientation may be represented in early cortical or even subcortical regions, high-level shapes and faces may be represented in the ventral pathway, while mean motion and location may be represented in the dorsal pathway. The mean representation process may be related to signal pooling. For instance, when viewing a set of lines or gratings with different orientations, orientation-selective neurons (possibly in V1) are activated by visual stimuli, and activity from some or all orientation-selective neurons is pooled together to ultimately generate a holistic percept.

Corbett et al. (2012) used adaptation aftereffects to study mean size representation of dot ensembles, revealing specific neural codes for statistical summary representations in the brain. Participants adapted to dot sets of different sizes and then judged the size of test dots. After adapting to a set with a large mean size, test dots were perceived as smaller, and vice versa, producing a perceptual aftereffect. The authors proposed that mean size appears to be explicitly represented as a feature dimension in the brain. However, where this representation is located remains uncertain, with possibilities ranging from primary visual cortex to lateral occipital cortex neuron populations that encode object size. Behavioral studies have revealed that neural computation of visual statistical information occurs after binocular fusion and binocular suppression, that is, no earlier than primary visual cortex.

Specifically, participants can judge the mean size of circle ensembles distributed

across both eyes, and if some items are suppressed through binocular rivalry, the mean representation of set size is impaired (Joo et al., 2009).

However, directly investigating the neural mechanisms of statistical summary representations is not straightforward. One major reason is the difficulty in selecting a control condition that matches the statistical summary representation of ensembles. If single stimuli are used as controls for comparison with ensembles, the stimulus materials are mismatched, potentially causing differences in evoked neural activity from the outset. Indeed, the research path for statistical summary representations has been tortuous. Initially, researchers inferred possible neural mechanisms underlying statistical summary representations through behavioral experiments. Later, some researchers examined the neural mechanisms of ensemble coding, which can be divided into two categories: one is object ensemble representation that does not contain statistical summary information (similar to texture or spatial layout), and the other is object ensemble representation that contains statistical summary information. Such studies often compare ensembles composed of multiple stimuli with single stimuli, but because the number of stimuli is mismatched, differences between the two cannot reflect pure neural mechanisms of statistical summary representation.

4.1 Spatial Distribution and Outliers in Ensemble Representation

Cant and Xu's research group has conducted a series of studies on the function of the parahippocampal place area (PPA) in ensemble representation. They found that when object identity remains consistent within an ensemble (Cant & Xu, 2012), neural adaptation occurs in the PPA even when density changes (Cant & Xu, 2015). However, when an ensemble contains two types of objects and their relative density changes (Cant & Xu, 2015), or when the ensemble contains outliers of different objects (Cant & Xu, 2020), the degree of adaptation is affected.

Cant and Xu (2012) used an fMRI adaptation paradigm to examine brain activation when participants processed object ensembles. They found that when the same object ensemble with identical texture statistics was repeatedly presented, even as different images (with different local features), the anterior-medial ventral visual cortex, including the parahippocampal place area (PPA), showed adaptive changes. The PPA is a brain region that processes scenes and textures, suggesting that object ensemble representation may employ neural coding similar to texture statistics representation.

In another study, they manipulated the spatial distance between individuals in an ensemble and their relative proportions to investigate the effects of absolute density and relative density on ensemble representation. They found that when the density of the same object ensemble changed, the PPA also showed adaptation, with adaptation magnitude comparable to when density remained unchanged. However, when an ensemble contained two types of objects and the proportion between them was changed (altering the relative density of each

object type), neural adaptation in the PPA decreased. This demonstrates that the PPA' s processing of object ensembles relies more on high-level visual information such as object proportion rather than low-level visual information like spacing or spatial frequency (Cant & Xu, 2015).

Using the same fMRI adaptation paradigm to study the representation of outliers in the human brain, experiments first presented a homogeneous ensemble containing one object type (e.g., strawberries), then presented either the same homogeneous ensemble (strawberries) or a heterogeneous ensemble containing mostly the same objects with a few different objects as outliers (strawberries plus a few watermelons). Participants judged whether the two ensembles were identical. The results showed that different outliers reduced the adaptation effect in the PPA, a scene-selective region. Interestingly, when a strawberry ensemble was first presented, followed by a watermelon ensemble or mostly watermelons with a few strawberries, and participants judged whether the two ensembles were different, matching outliers enhanced the PPA adaptation effect. This suggests that the PPA and related brain regions may mark outliers during visual perception and weight them in subsequent behavioral decision-making (Cant & Xu, 2020).

Cant and Xu' s research found that the parahippocampal place area adapts to the same ensemble with different spatial layouts, but the amount of adaptation changes when the objects in the ensemble are altered. Specifically, the hippocampal representation of ensembles does not depend on statistical properties like spatial layout. We speculate that the PPA' s representation of ensembles may depend on concepts, so when relative density or outliers affect concepts, hippocampal adaptation also changes.

4.2 Ensemble Coding in Extracting Visual Statistical Summary Representations

Ensemble coding refers to the encoding of ensemble stimuli by the visual system. We primarily focus on the following points: there exist specific neural mechanisms in the brain for holistically processing ensemble properties (Jia et al., 2022; Roberts et al., 2019), and ensemble coding and individual coding are dissociated in the dorsal and ventral pathways (Im et al., 2017). The neural mechanisms of grating ensembles are related to the occipital and parietal lobes, possibly involving signal pooling (Tark et al., 2021) and interactions among ensemble members (Jia et al., 2022).

To investigate whether specialized neural mechanisms for processing holistic ensemble perception exist in the brain and how the brain computes holistic ensemble perception, researchers used frequency-tagged EEG (SSVEP) to track brain activity related to holistic representation of periodically changing circle ensemble sizes. Neural responses to holistic ensemble size were detected at occipitoparietal electrodes. They then used temporal response functions (TRF) to separate neural responses to individual member sizes and interactions among

individuals (including global interaction and local interaction), finding that only global interaction directly contributed to holistic size perception. These findings indicate the existence of specific neural mechanisms for holistically processing ensemble size (Jia et al., 2022).

To examine perceptual differences between ensemble coding and individual coding of emotional faces, researchers used fMRI technology to explore the dissociation of activated brain regions. They found that the intraparietal sulcus and superior frontal gyrus in the dorsal pathway participated in holistic perception of emotional faces, while the fusiform cortex in the ventral pathway participated in individual perception of emotional faces. This study also found that holistic coding of emotional faces shows right-hemisphere lateralization advantages (Im et al., 2017). In 2021, they used MEG technology to validate these findings, confirming that the dorsal pathway participates in holistic processing of face ensembles while the ventral pathway identifies and discriminates individual faces. Importantly, MEG revealed that the dorsal pathway can perform very rapid holistic coding of face ensembles. They proposed that the dorsal pathway may rely on fast magnocellular pathway input to form holistic representations of crowd emotion (Im et al., 2021).

Researchers used EEG to explore the neural basis of facial identity statistical summary representations. In experiments presenting participants with face ensembles or single faces, the P1 amplitude was smaller and N1 and N2 latencies were shorter in the ensemble condition. Using multivariate pattern analysis (MVPA) with linear support vector machines (SVM) (incorporating time points and electrode sites into the model) for neural decoding, they found that neural signals could not only discriminate between different single faces but also discriminate between face ensembles with different average identities. Interestingly, when two ensembles had the same average identity, neural signals could not discriminate between them even when the individual members differed (marginally significant) (Roberts et al., 2019). In another article by these researchers (which actually used face ensemble data from Roberts et al., 2019), multivariate feature selection based on linear SVM and recursive feature elimination similarly found that neural signals (in both time and frequency domains) could discriminate between different face ensembles (Nemrodov et al., 2020).

Using fMRI and inverted encoding models (IEM), researchers constructed selective neural responses to mean orientation and individual member orientation in grating ensembles. They found that although BOLD signals did not differ significantly between ensemble and individual tasks, the neural responses constructed by IEM differed. Under task-relevant conditions, selective neural responses to both mean orientation (in ensemble tasks) and individual orientation (in individual tasks) existed in the occipital and frontoparietal lobes. Notably, under task-relevant conditions, selective responses to mean orientation were not significant in V1 but were significant in V2 and V3, with important linear increases from V1 to V2 to V3. These results suggest that pooled signals at multiple levels of the visual system form neural representations of holistic perception (Tark

et al., 2021).

Researchers examined whether working memory representations are structured like hierarchical representations in the visual system. Using line orientation ensemble stimuli and IEM to decode EEG signals, results supported the structured representation hypothesis. Static coding observed in frontocentral regions could represent both simple features and abstract ensemble averages, correlating with behavioral measures, while dynamic and static coding observed in occipitoparietal regions was modulated by top-down task demands (Oh et al., 2019).

Currently, only a few studies have preliminarily explored the neural coding of variance. Researchers found that adaptation to variance depends on retinal coordinates (dependent on fixation position) rather than spatiotopic coordinates. Furthermore, in a brain-damaged patient with only the left hemisphere V1 intact, variance representation accuracy showed no significant difference compared to normal participants. This patient had bilateral ventromedial occipitotemporal cortex damage, right hemisphere primary visual cortex damage, and partial damage to bilateral V2, V3, and V4, suggesting that variance representation may occur at early stages of the visual system, such as primary visual cortex (Norman et al., 2015). In summary, the brain regions involved in variance neural coding and its temporal origin require extensive further research.

5. Research Outlook

We propose that two statistical summary representation mechanisms coexist in the brain: a hierarchical pooling mechanism (signal pooling) and a neural ensemble coding mechanism. The former represents statistical properties of low-level visual stimuli, while the latter represents statistical properties of high-level visual stimuli through simultaneous activation of neurons that encode individual members. Whitney et al. (2014) theoretical model suggests that mean representation processes may involve signal pooling. For example, when viewing a set of lines or gratings with different orientations, orientation-selective neurons are activated, and the activity of multiple orientation-selective neurons is pooled together to ultimately generate holistic ensemble perception. This theory is supported by empirical evidence showing significant linear increases in selective responses to mean orientation from primary visual cortex to extrastriate cortex (Tark et al., 2021). This theory appears to well explain statistical summary representations of low-level visual stimuli, as visual cortex indeed contains simple neurons and higher-level complex neurons that receive and pool information from simple neurons. However, for high-level visual stimuli such as faces, forming a mean representation would imply the existence of higher-level neurons in the brain that receive input from multiple faces. This would require not only a considerable number of neurons representing individual faces but also an even larger number of neurons for statistical summary representation. The hierarchical pooling theory, like the Grandmother Cells theory, would similarly fail to solve the efficiency problem. In this case, statistical representation may rely on neural ensemble coding, where simultaneous activation of multiple faces forms

a statistical summary representation of the ensemble. Therefore, for statistical summary representations, the brain may contain not only hierarchical pooling mechanisms but also ensemble coding mechanisms based on simultaneous activation of member-representing neurons.

Although the preceding review has provided preliminary understanding of statistical summary representation mechanisms, many unresolved or unexplored issues remain for future research to investigate thoroughly.

5.1 Systematic Investigation of Neural Mechanisms Using Neuroscience Techniques

First, regarding whether ensemble representation relies on the ventral or dorsal pathway, the very limited existing studies have not reached consistent conclusions (Im et al., 2017; Im et al., 2021; Cant & Xu, 2020). There are also inconsistencies regarding whether the magnocellular system participates in statistical summary representations (Im et al., 2021; Lee & Chong, 2021). Future research needs to examine whether the ventral and dorsal pathways, as well as magnocellular and parvocellular systems, play different roles and have different importance in statistical summary representation processes. Additionally, we need to investigate whether top-down feedback signals and interactive signals within the same level play essential roles in statistical summary representations. To clarify these issues, we must also consider the modulatory effects of stimulus level and task demands.

Second, we need to examine the cortical origins and temporal processing dynamics of statistical summary representations. Since receptive fields in primary visual cortex are very small and often cannot cover multiple stimuli in an ensemble, it seems reasonable to believe that statistical summary representations occur after primary visual cortex. However, this view lacks direct neuroscientific evidence. Behavioral evidence suggests that mean size representation of circles occurs no earlier than primary visual cortex (Joo et al., 2009). Research on variability suggests that statistical summary representations may occur earlier. A study of a patient with damage to extrastriate and higher visual areas found that orientation variance representation was not impaired, suggesting that orientation variance representation may rely on primary visual cortex (Norman et al., 2015). Animal studies have found that the cat' s lateral geniculate nucleus (LGN) can respond specifically to the standard deviation of brightness, indicating that early visual processing pathways can process variability information (Bonin et al., 2006). From animal experiments to single-patient brain damage studies to behavioral research, it is difficult to draw definitive conclusions about the cortical origins of human statistical summary representation mechanisms. Future research addressing this issue also needs to distinguish potential effects of different stimulus levels while answering whether mean and variability representations have different neural mechanisms.

Finally, we need more direct investigation of domain-specific or domain-general

mechanisms for statistical summary representations across different levels and different metrics (central tendency vs. variability)—whether they are mediated by specific brain regions or share a common processing pathway, and whether the temporal dynamics across different levels and metrics are consistent. Revealing these important issues cannot remain at the behavioral level as described previously (Haberman, Brady & Alvarez, 2015). Future research must employ clever experimental designs and various brain imaging techniques to directly address these questions.

5.2 Are Statistical Summary Representations Innate or Experience-Dependent?

Statistical summary representation is a psychological process with relatively high automaticity, naturally leading to the expectation that genetic influences outweigh experiential influences. Unfortunately, the few studies on cultural differences cannot resolve this issue. Statistical summary representations show no significant differences between own-race and other-race faces. However, British participants show stronger averaging tendencies for faces of their own gender, while Chinese participants do not show this tendency, possibly due to differences in holistic processing biases (Peng et al., 2021). A study on statistical summary representations of size in 4-5-year-old children found that although children' s visual functions were not yet fully developed at this age (such as selective attention, spatiotemporal attentional resolution, and visual working memory capacity), they could already perform summary representations of object ensembles, identifying mean size rather than individual items in the set. Of course, compared to adults, children' s representation efficiency was lower, because as age increases and visual functions gradually mature, the capacity for statistical summary representation also improves (Sweeny et al., 2015). Additionally, reward value may not affect averaging itself but only affects high-level conscious representations of statistical summary representations (Dodgson & Raymond, 2020).

5.3 Are Neural Mechanisms of Ensemble Coding and Texture Representation Equivalent to Those of Statistical Summary Representation?

Many studies investigating neural mechanisms have inferred the neural mechanisms of statistical summary representations by comparing differences between ensembles and individuals. This approach is problematic. Differences in neural activity between ensembles and individuals may include brain mechanisms of statistical representation but cannot exclude differences caused by quantitative disparities—that is, the psychological process is not pure. Indeed, the number of items affects neural activity: as ensemble size increases, N170 amplitude increases (Puce et al., 2013), and even in primary visual cortex, C1 amplitude evoked by multiple items in an ensemble can be a linear sum of its member items (Chen et al., 2016).

Compared to statistical summary representation, texture representation seems to lack a neural computation process. Future research needs to employ clever methods to isolate neural activity that is purely related to statistical summary representation.

References

[1] Tong, K., Tang, W., Chen, W., & Fu, X. (2015). Statistical summary representation: Contents and mechanisms. *Advances in Psychological Science*, 23(10), 1723-1731. https://doi.org/10.3724/sp.J.1042.2015.01723 [2] Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157-162. https://doi.org/10.1111/1467-9280.00327 [3] Attarha, M., & Moore, C. M. (2015). The perceptual processing capacity of summary statistics between and within feature dimensions. *Journal of Vision*, 15(4), 9. https://doi.org/10.1167/15.4.9 [4] Banno, H., & Saiki, J. (2012). Calculation of the mean circle size does not circumvent the bottleneck of crowding. *Journal of Vision*, 12(11). https://doi.org/10.1167/12.11.13 [5] Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *The Journal of Neuroscience*, 26(23), https://doi.org/10.1523/JNEUROSCI.0284-06.2006 [6] Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience*, 32(22), https://doi.org/10.1523/JNEUROSCI.3325-11.2012 [7] Cant, J. S., & Xu, Y. (2015). The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. *Cerebral Cortex*, 25(11), 4226-4239. https://doi.org/10.1093/cercor/bhu145 [8] Cant, J. S., & Xu, Y. (2020). One bad apple spoils the whole bushel: The neural basis of outlier processing. *Neuroimage*, https://doi.org/10.1016/j.neuroimage.2020.116629 [9] 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 [10] Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of visual adaptation to size. *Cognition*, 20(2). https://doi.org/10.1080/13506285.2012.657261 [11] 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 [12] de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, 62(9). https://doi.org/10.1080/17470210902811249 [13] 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- [14] 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}

[15] Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, 106(3). https://doi.org/10.1152/jn.00904.2010 [16] Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31). https://doi.org/https://doi.org/10.1073/pnas.0901352106 [17] Haberman, J., & Whitney, D. (2007a). 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 [18] Haberman, J., & Whitney, D. (2007b). Supplemental data: Rapid extraction of mean emotion and gender from sets of faces. [19] 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 [20] 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 [21] Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. S. Werner, L. M. Chalupa & M. E. Burns (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339-349). MIT Press. [22] 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 [23] 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 [24] Haberman, J. M., & Ulrich, L. (2019). Precise ensemble face representation given incomplete visual input. *i-Perception*, 10(1). https://doi.org/10.1177/2041669518819014 [25] 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- [26] Haxby, J. V., Hoffman, E., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6). https://doi.org/10.1016/S1364-6613(00)01482-0 [27] Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics*, 71(2). https://doi.org/10.3758/APP.71.2.375 [28] 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- [29] 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 [30] Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research*, 167, 46-53. https://doi.org/10.1016/j.visres.2020.01.002 [31] Jia, J., Wang, T., Chen, S.,

Ding, N., & Fang, F. (2022). Ensemble size perception: Its neural signature and the role of global interaction over individual items. *Neuropsychologia*, 173, 108290. https://doi.org/10.1016/j.neuropsychologia.2022.108290 [32] Ji, L., Rossi, V., & Pourtois, G. (2018). Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs. *Neuropsychologia*, 111, 92-102. https://doi.org/10.1016/j.neuropsychologia.2018.01.022 [33] 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 1-12. https://doi.org/10.1167/9.9.7 [34] Kacin, M., Gauthier, I., & Cha, O. (2021). Ensemble coding of average length and average orientation are correlated. *Vision Research*. https://doi.org/10.1016/j.visres.2021.04.010 [35] 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 [36] 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 [37] 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*, 7, 1332. https://doi.org/10.3389/fpsyg.2016.01332 [38] Maule, J., & Franklin, A. (2020). Adaptation to variance generalizes across visual domains. *Journal of Experimental Psychology: General*, 149(4). https://doi.org/10.1037/xge0000678 [39] Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences of the United States of America*, 111(21), 7873-7878. https://doi.org/10.1073/pnas.1308674111 [40] 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 [41] 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/s13414-020-02163-3 [42] 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 [43] 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 [44] 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/https://doi.org/10.1038/89532 [45] Peng, S., Zhang, L., Xu, R., Liu, C. H., Chen, W., & Hu, P. (2019). Self-construal priming modulates ensemble perception of multiple-face identities. *Frontiers in Psychology*, 10, 1096. https://doi.org/10.3389/fpsyg.2019.01096 [46] 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 [47] 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 [48] 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*, 7, 282. https://doi.org/10.3389/fnhum.2013.00282 [49] Rajendran, S., Maule, J., Franklin, A., & Webster, M. A. (2021). Ensemble coding of color and luminance contrast. *Attention, Perception, & Psychophysics*, 83(3). https://doi.org/10.3758/s13414-020-02136-6 [50] 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.0471-19.2019 [51] Sama, M. A., Nestor, A., & Cant, J. S. (2019). Independence of viewpoint and identity in face ensemble processing. *Journal of Vision*, 19(5), 2. https://doi.org/10.1167/19.5.2 [52] Sun, J., & Chong, S. C. (2020). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology: General*, 149(3), 550-563. https://doi.org/10.1037/xge0000667 [53] 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 [54] 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 [55] Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4-5-year-old children. *Developmental Science*, 18(4). https://doi.org/10.1111/desc.12239 [56] 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 [57] Tiurina, N. A., & Utochkin, I. S. (2019). Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging. *Journal of Experimental Psychology: General*, 148(4), 728-738. https://doi.org/10.1037/xge0000485 [58] 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 [59] 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 [60] 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 [61] 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). https://doi.org/10.1167/12.6.12 [62] Watamaniuk, S. N. J., & Duchon, A.

(1992). The human visual system averages speed information. *Vision Research*, 32(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036- [63] 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. [64] Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105-129. https://doi.org/10.1146/annurev-psych-010416-044232 [65] 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 [66] Ying, H., & Xu, H. (2017). Adaptation reveals that facial expression averaging occurs during rapid serial presentation. *Journal of Vision*, 17(1). https://doi.org/10.1167/17.1.15 [67] Yoruk, H., & Boduroglu, A. (2020). Feature-specificity in visual statistical summary processing. *Attention, Perception, & Psychophysics*, 82(2). https://doi.org/10.3758/s13414-019-01942-x

Corresponding Authors: Xiuling Zhang: zhangxl556@nenu.edu.cn Yi Jiang: yijiang@psych.ac.cn Fan Zhang: yffs9762@163.com

Author Contribution Statement: Xiuling Zhang: Responsible for writing and revising the manuscript, reviewing and finalizing the paper, literature organization, and submission support and coordination Yi Jiang: Responsible for supervision and review of the paper Fan Zhang: Responsible for initial draft writing, content organization and supplementation, and literature collection and organization Mingxiao Ge: Responsible for supplementary literature resources and paper content Sijia Li: Responsible for supplementary literature resources and paper content, and reference organization

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