

ORIGINAL ARTICLE

Open Access



# The combination operation of grouping and ensemble coding for structured biological motion crowds in working memory

Wei Chen<sup>1</sup>, Shujuan Ye<sup>1</sup>, Xin Yan<sup>1</sup> and Xiaowei Ding<sup>1\*</sup>

## Abstract

Massive studies have explored biological motion (BM) crowds processing for their remarkable social significance, primarily focused on uniformly distributed ones. However, real-world BM crowds often exhibit hierarchical structures rather than uniform arrangements. How such structured BM crowds are processed remains a subject of inquiry. This study investigates the representation of structured BM crowds in working memory (WM), recognizing the pivotal role WM plays in our social interactions involving BM. We propose the group-based ensemble hypothesis and test it through a member identification task. Participants were required to discern whether a presented BM belonged to a prior memory display of eight BM, each with distinct walking directions. Drawing on prominent Gestalt principles as organizational cues, we constructed structured groups within BM crowds by applying proximity and similarity cues in Experiments 1 and 2, respectively. In Experiment 3, we deliberately weakened the visibility of stimuli structures by increasing the similarity between subsets, probing the robustness of results. Consistently, our findings indicate that BM aligned with the mean direction of the subsets was more likely to be recognized as part of the memory stimuli. This suggests that WM inherently organizes structured BM crowds into separate ensembles based on organizational cues. In essence, our results illuminate the simultaneous operation of grouping and ensemble encoding mechanisms for BM crowds within WM.

**Keywords** Biological motion, Working memory, Grouping, Ensemble encoding, Gestalt principles

As social creatures, our ability to perceive and comprehend the motion of biological entities in crowds is of paramount importance. The social information available in crowds of biological motion (BM) not only benefits interpersonal interactions, but also provides clues about immediate predation risk and food resources (Galef & Giraldeau, 2001; Griffin, 2004). Consequently, crowds processing has intrigued vast investigations, including perceptions of emotions and intentions conveyed by BM crowds (e.g., Brunyé et al., 2014; Helbing et al., 2000;

Moussaïd et al., 2016), the working memory capacity dedicated to BM crowds (e.g., Ding et al., 2017), and crowd attention (e.g., Gallup et al., 2012; Ristic & Capozzi, 2022; Sweeny & Whitney, 2014), wherein the majority concerned uniformly distributed BM or BM crowds with high homogeneity. However, in reality, BM crowds frequently exhibit clustering structures rather than uniform arrangements. For instance, consider a basketball game where players inherently belong to two distinct teams, or a group seated on a lawn, often organized in pairs or trios rather than forming a unified entity. Despite these observations, the cognitive mechanisms underlying the processing of structured BM crowds remain insufficiently understood.

To address this gap, we embarked on an exploration of structured BM crowds processing by examining their

\*Correspondence:

Xiaowei Ding  
dingxw3@mail.sysu.edu.cn

<sup>1</sup> Present Address: Department of Psychology, Sun Yat-Sen University, Guangzhou, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

storage in working memory (WM), recognizing the crucial role of WM storage for BM in our daily social life. Numerous studies have examined the WM aspects related to BM, including BM's WM capacity (Shen et al., 2014; Smyth & Pendleton, 1989; Smyth et al., 1988; Wood, 2007, 2011), the binding between BM and other features (Ding et al., 2015; Lu et al., 2019; Wood, 2008, 2010), and the neural mechanisms underlying BM's representations in WM (Cai et al., 2018; Gao et al., 2015; Lu et al., 2016). However, these investigations have primarily centered on WM pertaining to individual BM or unstructured BM crowds, thereby failing to address the specificities of structured BM crowds. The present study aims to elucidate how structured BM crowds are stored in WM.

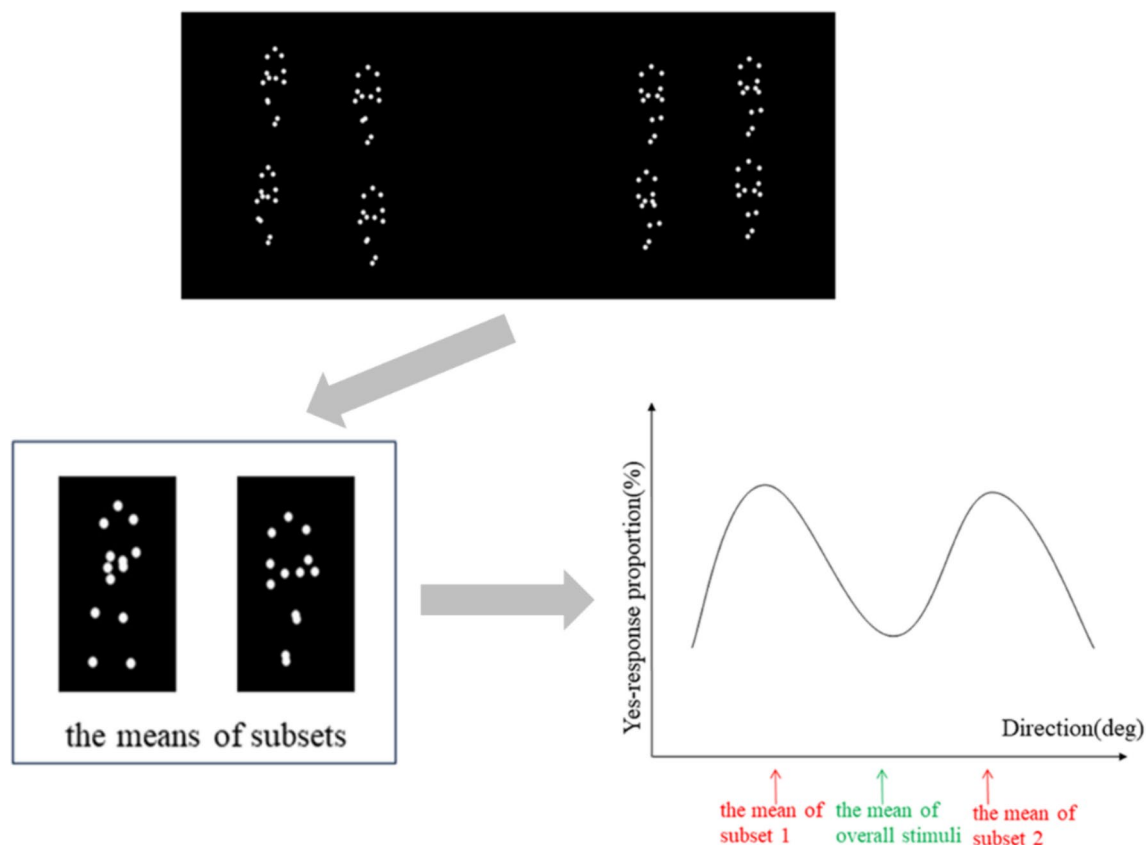
Ensemble coding may provide an efficient solution allowing WM to circumvent capacity limitations and simultaneously process substantial amount of information, whereby the statistical structure of BM crowds is rapidly and accurately extracted and maintained in WM (e.g., mean and variance; Alvarez, 2011; Michael et al., 2014; Whitney & Yamanashi Leib, 2018). This mechanism has undergone extensive exploration in relation to fundamental physical attributes (Ariely, 2001; Baek & Chong, 2020; Parkes et al., 2001; Webster et al., 2014), social traits (Elias et al., 2017; Lee & Chong, 2021; Marini et al., 2023; Sweeny et al., 2013), and even semantic categories (Khayat & Hochstein, 2019; Khayat et al., 2021) in perception. Within the realm of WM, this phenomenon also garners support from both behavioral findings (Brady & Alvarez, 2015; Schurgin & Brady, 2019; Son & Chong, 2023; Son et al., 2019; Utochkin & Brady, 2020) and neural evidence (Oh et al., 2019). Furthermore, this encoding process within WM seems to operate in an automatic fashion, as suggested by implicit member identification tasks. For instance, participants tasked with judging whether a probe belonged to a prior memory display showed a stronger inclination toward endorsing the probe as present in the memory display when it closely approximated the mean value of all stimuli (Khayat & Hochstein, 2018, 2019; Khayat et al., 2021; Oh et al., 2019). Drawing inspiration from these outcomes, which reaffirm the widespread and inherent nature of ensemble coding in WM, a practical and effective approach to representing structured BM crowds involves retaining their summary statistics, such as means.

However, considering the profound impact of the stimuli structure on WM representation, a seemingly more reasonable hypothesis is that structured BM crowds are stored within WM as discrete ensembles, guided by their organizational cues. Extensive research has consistently revealed that information stored in WM inherently follows a hierarchical organization based on its structural attributes. Within this hierarchy, Gestalt

grouping cues hold a pivotal role, encompassing factors like proximity, connectedness, and shared spatial region (e.g., Brady et al., 2011; Corbett, 2017; Gao et al., 2016a, 2016b; Peterson & Berryhill, 2013; Peterson et al., 2015; Son et al., 2019). From the computational level, ensemble representations, considered as pooling population responses, could naturally exhibit similarity-based clustering and segmentation effects (Im et al., 2021; Utochkin, 2015; Utochkin et al., 2024; Treue et al., 2000). Furthermore, within the domain of studies on memorizing BM, research has indicated that social relationships within BM can enhance WM performance, implying the extraction of crowd structure (Ding et al., 2017; Vestner et al., 2019, 2022). In light of these findings, we posit that the structural composition of BM crowds significantly influences their WM representations. Taken together, we propose a group-based ensemble hypothesis to tackle the storage of structured BM crowds in WM, which assumes that BM crowds undergo an automatic organization into distinct subsets and are maintained as separate ensembles within WM.

The direction of walking stands as a fundamental attribute of BM. To test the group-based ensemble hypothesis, the current study employed point light displays (PLDs) to depict walking, a method demonstrated to effectively probe the cognitive processing of human actions (e.g., Abernethy et al., 2001; Blake & Shiffrar, 2007; Johansson, 1973). Adopting the member identification task mentioned earlier (Oh et al., 2019), participants were tasked with determining whether a presented stimulus belonged to a previous memory display. This paradigm deliberately avoided explicit cues for participants to report the mean value of all stimuli, thereby offering a substantial advantage in implicitly detecting the automatic formation of ensemble representations. If a BM representation characterized by a specific walking direction is held in WM, it would be more prone to recognition as part of the memory stimuli. Thus, under the premise of the group-based ensemble hypothesis, the means of subsets would emerge as the most probable attributes to be perceived as components of the memory display (Fig. 1).

The underlying physical structure is a fundamental hallmark of group-based BM, typically adhering to Gestalt principles. Consequently, our study set out to manipulate two key Gestalt principles to emulate structured BM crowds: proximity (Experiment 1) and similarity (Experiments 2–3). In Experiment 1, eight BM exhibiting diverse directions were grouped based on their proximity, resulting in an equitable distribution on both sides of the screen. Experiments 2 and 3 positioned eight BM uniformly along an unseen circle centered on the screen. To achieve similarity-driven grouping, we positioned these stimuli in tight clusters around two distinct



**Fig. 1** Predictions of the group-based ensemble hypothesis. The group-based ensemble hypothesis predicts the highest proportion of 'yes' response to the BM featuring the mean direction of the subsets of memory stimuli

directions, with four stimuli aligned with each direction. The sole distinction between Experiments 2 and 3 lay in the composition of BM' walking directions. In Experiment 2, half of the stimuli walked to the left, while the remainder walked to the right, while in Experiment 3, all BM were directed either leftward or rightward. This strategic choice diminished the visibility of stimuli structures, enabling an assessment of result robustness.

### Experiment 1: WM representations for BM crowds structured by proximity

#### Methods

##### Participants

Twenty volunteers from Sun Yat-sen University participated in this experiment for payment. Two participants were subsequently excluded due to abnormal response patterns, characterized by unusually high proportions of 'yes' responses across all conditions. The remaining 18 participants (11 males and 7 females,  $M = 19.89$  years old) were all right-handed and reported normal or corrected-to-normal visual acuity.

The sample size was determined a priori based on PANGAEA (Westfall, 2016). Based on the results of previous studies (Oh et al., 2019; the responses of 'yes' at the mean orientation vs. the proportion combined across all memory orientations in the varied orientation condition of Experiment 1;  $t(19) = 3.74$ ,  $p = 0.001$ ), which used the similar design with ours, we calculated the effect size Cohen's  $d$  to be  $0.84 \left( \frac{t}{\sqrt{n}} \right)$  for the effect of paired  $t$ -tests in our experimental design. The suggested sample size was approximately 16 to obtain at least 95% power for the effect of paired  $t$ -test in the overall mean condition at a significance level of 0.05. Eighteen participants were recruited in Experiment 1 to ensure adequate power.

Before participation, all individuals provided signed informed consent. The study received approval from the Research Ethics Board of Sun Yat-sen University and was conducted in accordance with the approved guidelines.

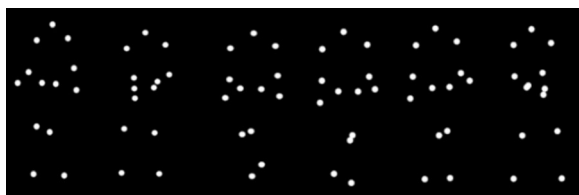
### Stimuli and apparatus

The experiment was run on a 27-inch LCD monitor, positioned at a viewing distance of 57 cm, with a resolution of  $2560 \times 1440$  pixels and a refresh rate of 60-Hz. The background was black (0, 0, and 0; RGB). The experiment was programmed using MATLAB (MathWorks, Natick, MA, USA) with Psychtoolbox extensions (Brainard, 1997; Pelli, 1997).

Point light displays (PLDs) were used to represent walking (Fig. 2). They were selected from the Motion Capture Database (<http://mocap.cs.cmu.edu>) built by the Graphics Lab at Carnegie Mellon University. This database offers a diverse array of PLDs, consisting of 13 points of light with 60 frames/s. The distribution of these 13 points was located in the following locations on the body: one on the head, two on the shoulders, two on the elbows, two on the wrists, two on the hips, two on the knees, and two on the ankles. We chose a sequence with walking movement from the database as our experimental stimulus. Every animation consisted of 30 distinct frames and was displayed in loop.

The memory display consisted of eight BM with different walking directions, four on the left side and four on the right side of the screen, constituting two subsets. The four BM of each subset were randomly distributed in a  $2 \times 2$  invisible square (each square:  $3.89^\circ \times 3.89^\circ$ ), positioned  $5.56^\circ$  to the left or the right of the center of the screen. Each BM randomly deviated  $0^\circ$  to  $0.56^\circ$  from the center of the square in which it was placed. The probe ( $2.23^\circ \times 2.23^\circ$ ) appeared in the center of the screen.

There were six levels of the mean directions of each subset ( $-40^\circ$ ,  $-30^\circ$ ,  $-20^\circ$ ,  $20^\circ$ ,  $30^\circ$  and  $40^\circ$ ). The mean directions of both subsets must be one positive and the other negative, and the combination of  $20^\circ$  and  $-20^\circ$  was excluded to avoid the two subsets being too similar, resulting in 8 combinations of subset means. For each combination of subset means, the left and right positions of each subset were counterbalanced. The four BM in each subset were derived from the subset's mean (plus  $-15^\circ$ ,  $-5^\circ$ ,  $5^\circ$ , and  $15^\circ$ ). Probe directions were determined by the mean (the mean of the left subset/the mean of the right subset) plus an angle ranging from  $-30^\circ$  to  $30^\circ$  in  $5^\circ$  increments.



**Fig. 2** Stimuli. From left to right, the walking directions were  $-40^\circ$ ,  $-30^\circ$ ,  $-20^\circ$ ,  $20^\circ$ ,  $30^\circ$ , and  $40^\circ$

### Experiment design and procedure

The experimental procedure is illustrated in Fig. 3. After a 500 ms fixation, the memory display composed of eight BM was presented for 1200 ms, which participants were required to all memorized. Then, a blank retention interval lasted for 1000 ms. Finally, a probe appeared in the center of screen. Participants should judge whether it was the member of the previous memory display (“Y” for “yes,” “N” for “no”). Responses should be completed within 2000 ms.

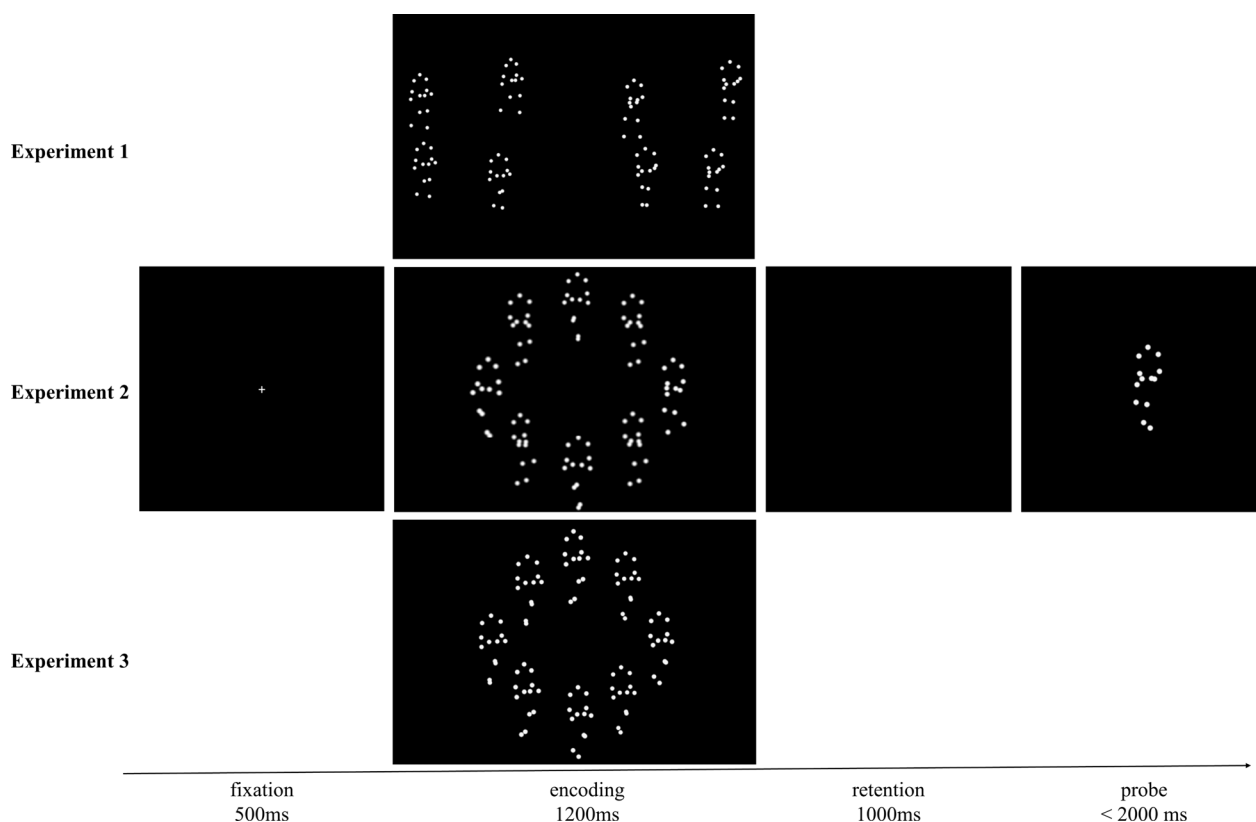
The experiment used a 2 (probe type: the left subset condition/the right subset condition)  $\times$  13 (probe offset:  $-30^\circ$  to  $30^\circ$  in  $5^\circ$  increments) within-subjects design. The first factor indicated the mean in reference to which the probe offset was calculated. Probe offset represents the orientation difference between the according mean direction and the probe. Each combined condition contained 16 trials, which were randomly divided into four blocks. Before formal trials, 16 practice trials were required to ensure that participants understood the procedure. The entire task took approximately 50 min.

### Data analysis

Only trials with reaction time (RT) longer than 200 ms and shorter than 2000 ms were included in further analysis. To detect whether the ensemble representations of subsets were represented, we merged trials in the left subset condition and the right subset condition and took the absolute values of probe offsets under these two conditions. The proportions of ‘yes’ responses were initially subjected to a one-factor seven-level (the absolute values of probe offset) repeated measures of variance (ANOVA). Then, we conducted separate comparisons between the proportion of ‘yes’ responses at each offset and that at the mean direction through paired *t*-tests to examine the extent to which the group-based ensembles were represented. Finally, the proportion of ‘yes’ responses at the mean direction was compared to the proportion combined across all memory directions using a paired *t*-test to directly test the group-based ensemble hypothesis.

### Results and discussion

5.6% of trials were removed for abnormal RTs. The ANOVA revealed a significant main effect of probe offset [ $F(1, 6) = 73.93$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.813$ ]. And a sharp tuning of ‘yes’ responses around the mean of subsets (probe offset =  $0^\circ$ ) appeared (Fig. 4a). Eleven out of eighteen participants had the highest proportion of ‘yes’ responses to the BM with a walking direction of the mean direction of subsets. These results suggest that the mean directions of subsets were maintained during the delay. Specifically, after correcting for multiple



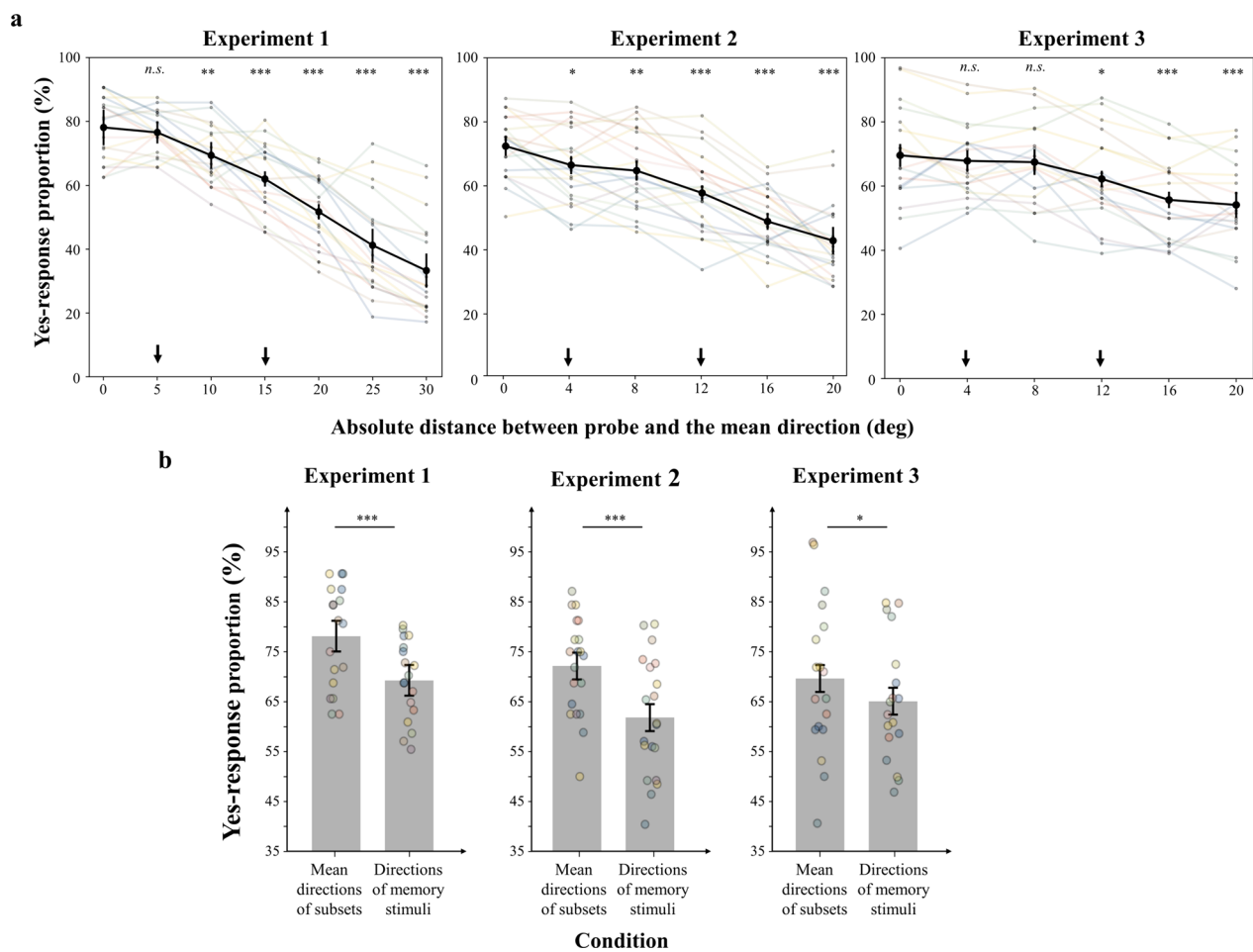
**Fig. 3** Procedure illustration in Experiments 1–3. Participants were required to remember the walking directions of all presented stimuli first. After a 1000-ms blank interval, a probe appeared. Participants should judge whether it was the member of the previous memory display

comparisons, the proportion of ‘yes’ responses was significantly higher when the probe offset was  $0^\circ$  compared to the range of  $10\text{--}30^\circ$  [ $5^\circ$ : 1.6%,  $t(17)=0.951$ ,  $p=0.355$ , Cohen’s  $d=0.22$ , 95%CI for mean difference= $[-4.3\%$ ,  $7.4\%$ ];  $10^\circ$ : 8.8%,  $t(17)=4.095$ ,  $p=0.002$ , Cohen’s  $d=0.97$ , 95%CI for mean difference= $[1.1\%$ ,  $16.4\%$ ];  $15^\circ$ : 16.1%,  $t(17)=5.650$ ,  $p<0.001$ , Cohen’s  $d=1.33$ , 95%CI for mean difference= $[6.0\%$ ,  $26.3\%$ ];  $20^\circ$ : 26.4%,  $t(17)=9.025$ ,  $p<0.001$ , Cohen’s  $d=2.13$ , 95%CI for mean difference= $[16.0\%$ ,  $36.9\%$ ];  $25^\circ$ : 36.9%,  $t(17)=8.009$ ,  $p<0.001$ , Cohen’s  $d=1.89$ , 95%CI for mean difference= $[20.5\%$ ,  $53.4\%$ ];  $30^\circ$ : 44.8%,  $t(17)=9.711$ ,  $p<0.001$ , Cohen’s  $d=2.29$ , 95%CI for mean difference= $[28.4\%$ ,  $61.3\%$ ] (Fig. 4a). Further paired- $t$  tests showed that the mean proportions of ‘yes’ responses at the mean directions of subsets were even higher than that for the directions of the memory display [ $5^\circ$  and  $15^\circ$ : 8.8%;  $t(17)=4.293$ ,  $p<0.001$ , Cohen’s  $d=1.01$ , 95%CI for mean difference= $[4.5\%$ ,  $13.2\%$ ] (Fig. 4b). Collectively, these results supported the group-based ensemble hypothesis.

However, a potential limitation exists that the proportion of ‘yes’ responses to the BM with a walking direction of the overall mean was higher compared to the subset

mean. To rule out this possibility, we further obtained the function representing the proportion of ‘yes’ response in relation to the probe offset relative to the subset mean. In Experiment 1, the probes with negative probe offsets, referenced to the positive subset mean, were closer to the global mean. Conversely, the probes with positive probe offsets, referenced to the negative subset mean, were closer to the global mean. Therefore, we first inverted the sign of the probe offsets determined relative to the negative subset mean. Consequently, the probes with negative offsets were consistently closer to the global mean, and the probes with positive offsets were farther, regardless of the subset mean used for reference. Then, we calculated the proportion of ‘yes’ response for each probe offset relative to the local mean. While the curves of the proportions varying with the offset display a slight bias toward the global mean (the negative direction), the proportion at the subset mean remains the highest (Fig. 5a). Moreover, we further conducted a paired- $t$  test to directly compare the proportion at the global mean with that at the local mean (Fig. 5b). The results revealed a significant lower proportion of ‘yes’ responses for the global mean compared to the local mean [ $-40.2\%$ ,  $t(17)=-6.882$ ,  $p<0.001$ , Cohen’s  $d=-1.62$ , 95%CI for





**Fig. 4** Results in Experiments 1–3. **a** The x-axis represents the absolute distance between probe orientation and the mean direction of subsets, and the y-axis shows the proportion of “yes” responses. The black dots represented group mean, with error bars indicating the within subject 95% confidence intervals. The smaller colorful dots and lines represented the results of each participant. The directions indicated by the arrows were the directions of the memory stimuli. ‘\*’ represented the significance of the data comparing to proportion at the 0° offset. n.s. > 0.05, \**p* < 0.05, \*\**p* < 0.005, \*\*\**p* < 0.001. **b** The proportion of ‘yes’ responses was plotted as a function of the type of the probe in Experiments 1–3 separately. The bars represented group mean, with error bars indicating the within subject 95% confidence intervals. The smaller colorful dots represented the results of each participant. \**p* < 0.05, \*\*\**p* < .001

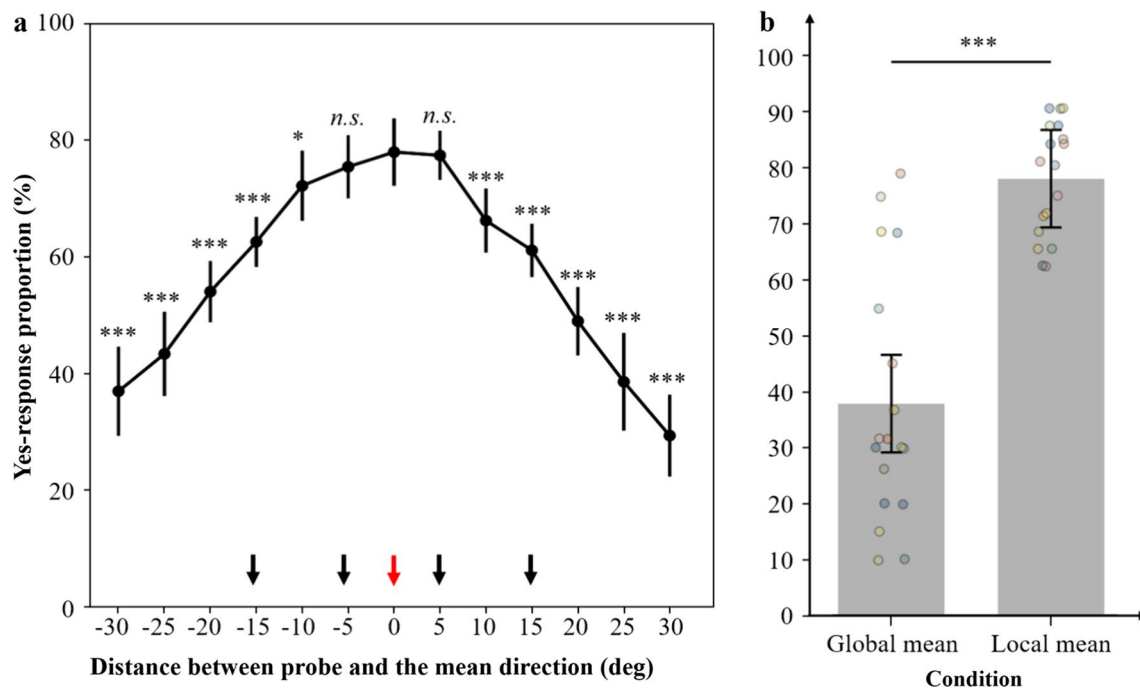
mean difference = [−52.5%, −27.9%]. Taken together, these findings provided additional evidence for that ensemble representations of the subsets were primarily maintained within the context of scenario structured by proximity.

### Experiment 2: WM representations for BM crowds structured by similarity

#### Methods

In Experiment 2, we advanced Experiment 1 by removing the spatial proximity cues to examine the sole effects of directional similarity. The sample size was determined in the same way as Experiment 1. A

separate group of 20 participants were recruited. The experimental procedure was similar to Experiment 1. The only two differences were that the memory stimuli were presented uniformly on an invisible circle (diameter = 8.89°) in the center of the screen (Fig. 3). To make the memory stimuli within each subset more similar to each other, their offsets from the means of the subsets were reduced to −12°, −4°, 4°, and 12°. Accordingly, the probe offsets were reduced to the range of −20–20° in 4° increments. Other experimental settings were the same as Experiment 1.



**Fig. 5** Results in Experiment 1. **a** The x-axis represents the distance between probe orientation and the mean direction of subsets, and the y-axis shows the proportion of ‘yes’ responses. The black dots represented group mean, with error bars indicating the within subject 95% confidence intervals. The directions indicated by the black arrows were the directions of the memory stimuli. The direction indicated by the red arrow was the directions of the subset mean. ‘\*’ represented the significance of the data comparing to proportion at the 0° offset. n.s. > 0.05, \* $p < 0.05$ , \*\*\* $p < 0.001$  (for detailed statistical results, see Table S1 in Appendix). **b** The proportion of ‘yes’ responses was plotted as a function of the type of the probe. The bars represented group mean, with error bars indicating the within subject 95% confidence intervals. \*\*\* $p < 0.001$

## Results and discussion

2.9% of trials were removed for abnormal RTs. The ANOVA revealed a significant main effect of probe offset [ $F(1, 5) = 57.45$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.751$ ]. And a sharp tuning of ‘yes’ responses around the mean directions of the subsets (probe offset = 0°) appeared (Fig. 4a). Thirteen out of twenty participants had the highest proportion of ‘yes’ responses to the BM with the walking direction of the mean directions of subsets. These results suggest that the mean directions of subsets were indeed represented during the delay. Specifically, after correcting for multiple comparisons, the proportion of ‘yes’ responses was significantly higher when the probe offset was 0° compared to other offsets [4°: 6.0%,  $t(19) = 3.015$ ,  $p = 0.021$ , Cohen’s  $d = 0.67$ , 95%CI for mean difference = [−0.7%, 12.6%]; 8°: 7.7%,  $t(19) = 3.966$ ,  $p = 0.003$ , Cohen’s  $d = 0.89$ , 95%CI for mean difference = [1.2%, 14.2%]; 12°: 14.7%,  $t(19) = 7.518$ ,  $p < 0.001$ , Cohen’s  $d = 1.68$ , 95%CI for mean difference = [8.1%, 21.3%]; 16°: 23.6%,  $t(19) = 11.651$ ,  $p < 0.001$ , Cohen’s  $d = 2.61$ , 95%CI for mean difference = [16.8%, 30.4%]; 20°: 29.7%,  $t(19) = 12.303$ ,  $p < 0.001$ , Cohen’s  $d = 2.75$ , 95%CI for mean difference = [21.6%, 37.8%]] (Fig. 4a). Further paired- $t$  tests showed that the average proportions of ‘yes’

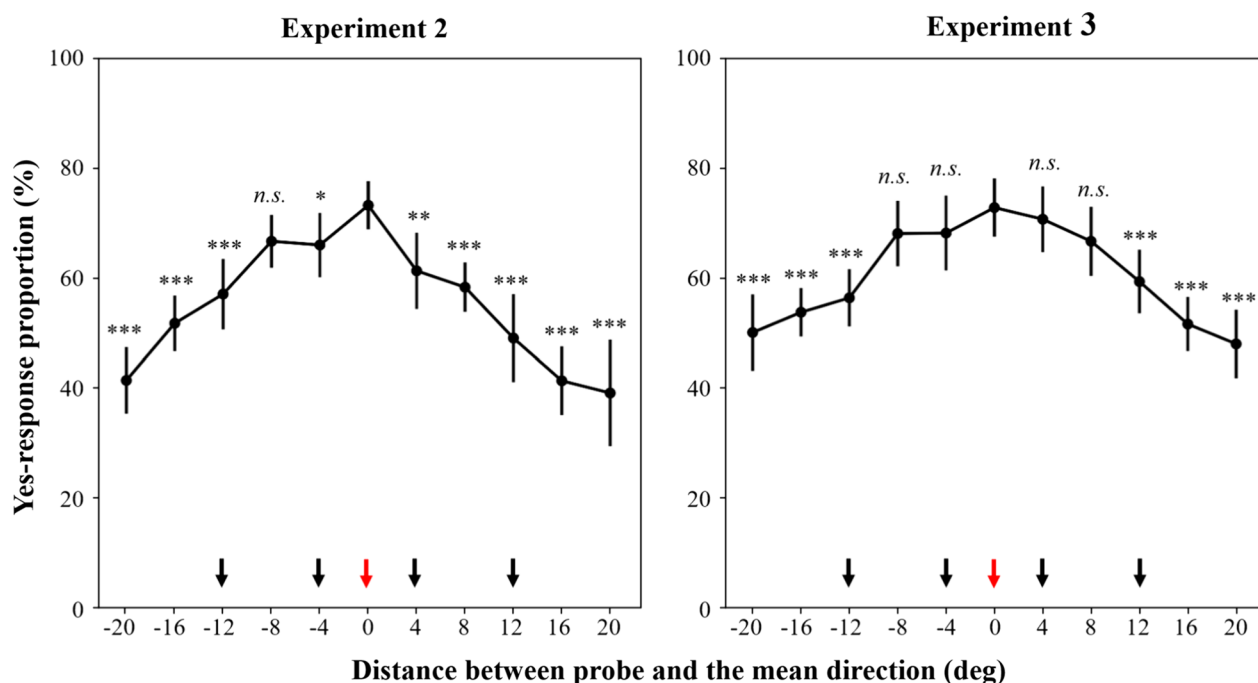
responses to the mean directions of subsets were even higher than that for all directions of the memory display [4° and 12°; 10.3%;  $t(19) = 5.67$ ,  $p < 0.001$ , Cohen’s  $d = 1.27$ , 95%CI for mean difference = [6.5%, 14.1%]] (Fig. 4b). Consistent with Experiment 1, the group-based ensemble hypothesis was supported.

Similar to Experiment 1, we also obtained the function representing the proportion of ‘presence’ response in relation to the probe offset relative to the subset mean. The curve also displayed a slight bias toward the global mean (the negative direction), but the proportion at the subset mean still remained the highest (Fig. 6). Due to the experimental settings, it is challenging to directly compare the proportions at the global mean with that at the local mean in Experiment 2. Nevertheless, this concern can be addressed in Experiment 3.

## Experiment 3: Reduction in the strength of similarity

### Methods

In Experiment 3, we ruled out the effects of directional consistency. The sample size was determined in the same way as Experiment 1. A separate group of 18 participants was recruited. The experimental procedure was similar



**Fig. 6** Results in Experiments 2 and 3. The x-axis represents the distance between probe orientation and the mean direction of subsets, and the y-axis shows the proportion of “yes” responses. The black dots represented group mean, with error bars indicating the within subject 95% confidence intervals. The directions indicated by the black arrows were the directions of the memory stimuli. The direction indicated by the red arrow was the directions of the subset mean. “\*” represented the significance of the data comparing to proportion at the 0° offset. n.s. > 0.05, \* $p < 0.05$ , \*\* $p < 0.005$ , \*\*\* $p < 0.001$  (for detailed statistical results see Table S2 and Table S3 in Appendix)

to Experiment 2. The differences were as follows. The means of the subsets own the same symbol, i.e., both were positive or both were negative, leading the strength of the similarity cue to be reduced. And there were only two combinations of subset means (20° and 65°, -20° and -65°). The other experimental settings were the same as Experiment 2 (Fig. 3).

### Results and discussion

1.4% of trials were removed for abnormal RTs. The ANOVA revealed a significant main effect of probe offset [ $F(1, 5) = 16.880$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.50$ ]. And a sharp tuning of ‘yes’ responses around the mean directions of subsets (probe offset = 0°) appeared (Fig. 4a). Nine out of eighteen participants had the highest proportion of ‘yes’ responses to the mean directions of subsets. Specifically, after correcting for multiple comparisons, the proportion of ‘yes’ responses was significantly higher when the probe offset was 0° compared to the range of 12–20° [4°: 1.7%,  $t(19) = 0.806$ ,  $p = 1.000$ , Cohen’s  $d = 0.19$ , 95%CI for mean difference = [-5.6%, 9.0%]; 8°: 2.1%,  $t(17) = 0.745$ ,  $p = 1.000$ , Cohen’s  $d = 0.18$ , 95%CI for mean difference = [-7.6%, 11.8%]; 12°: 7.3%,  $t(17) = 3.656$ ,  $p = 0.015$ , Cohen’s  $d = 0.86$ , 95%CI for mean difference = [0.5%, 14.2%]; 16°: 13.9%,

$t(17) = 6.716$ ,  $p < 0.001$ , Cohen’s  $d = 1.58$ , 95%CI for mean difference = [6.8%, 20.9%]; 20°: 15.4%,  $t(17) = 6.211$ ,  $p < 0.001$ , Cohen’s  $d = 1.46$ , 95%CI for mean difference = [7.0%, 23.9%]] (Fig. 4a). Further paired- $t$  tests showed that the average proportions of ‘yes’ responses to the mean directions of subsets were even higher than that for all directions of the memory display [4° and 12°; 4.5%,  $t(17) = 2.51$ ,  $p = 0.022$ , Cohen’s  $d = 0.59$ , 95%CI for mean difference = [0.7%, 8.3%]] (Fig. 4b). Consistent with Experiments 1 and 2, the group-based ensemble hypothesis was supported, although the line of proportion decreased less steep, showing a less apparent effect than that in Experiments 1 and 2.

Similarly, we further obtained the function representing the proportion of ‘presence’ response in relation to the probe offset relative to the subset mean. The only difference was that we inverted the sign of the probe offsets determined relative to the smaller subset mean (25° or -65°). The proportion at the subset mean remained the highest. Moreover, we further conducted a paired- $t$  test to directly compare the proportion at the global mean with that at the local mean (Fig. 6). Although the global mean was not directly probed in Experiment 3, the distance between the global and local means was fixed at 22.5°. Therefore, the point at -20° on the function representing



the proportion of 'presence' response in relation to the probe offset relative to the subset mean was consistently  $2.5^\circ$  away from the global mean. This was considered a reasonable approximation of the point representing the global mean. Consequently, we compared the proportion at  $-20^\circ$  and that at the local mean in Experiment 3. The results showed that the proportion of 'yes' response for the global mean was significantly lower compared to the local mean [ $-22.8\%$ ,  $t(17) = -4.586$ ,  $p < 0.001$ , Cohen's  $d = -1.08$ , 95%CI for mean difference =  $[-33.3\%, -12.3\%]$ , indicating that WM primarily maintained the ensemble representations of the subsets within the context of scenario structured by similarity.

### General discussion

The current study uncovered how BM crowds characterized by hierarchical structures were stored in WM. To achieve this, we employed the member identification task as an implicit measure to probe the specific representations maintained during the delay. In Experiment 1, where BM crowds were spatially partitioned into two separate units, participants exhibited a tendency to represent organized units as distinct ensembles in WM. These results provide support for the group-based ensemble hypothesis. Furthermore, Experiment 2 demonstrated that participants automatically maintained group-based ensemble representations even when multiple BM grouped by similarity were uniformly distributed. This result indicates directional similarity alone is sufficient to trigger the formation of subgroup-based ensemble representations in WM. Experiment 3 further corroborated the previous observations by reducing the strength of the similarity cue. After setting the directions of all memory stimuli to the same symbol, the results remained the same, indicating that group-based ensemble representations were still formed despite the unobvious similarity cue. Taken together, these findings provide compelling evidence for the existence of group-based ensemble representations in WM encoding of BM crowds.

In Experiment 1, we employed proximity as the organizational cue. Previous research comparing the roles of spatial and non-spatial grouping has highlighted the beneficial impact of proximity on subjects' statistical representation accuracy (Im & Chong, 2014). However, in our study, although no direct comparison between experiments was conducted, the ensemble coding effect stemming from proximity grouping did not exhibit a discernible difference from similarity in terms of effect size. It is possible that the greater within-group difference in Experiment 1 offset any potential benefits derived from proximity. But this difference may also be attributed to differences in task requirements. Participants in Im's

study were explicitly required to form ensemble representations of each group, while in our study ensemble representation was tested implicitly. It is conceivable that only when the task explicitly demands participants to more accurately estimate, the group mean would they actively leverage proximity to enhance their estimations.

In contrast, in Experiments 2 and 3, we employed similarity as the organizational cue. When similarity serves as the organizational cue, we found that, the distributional properties of each subgroup can significantly influence the grouping process. Recent findings by Ortego and Störmer (2024) demonstrate that participants' ensemble representations of one group tend to be biased toward the other group when the two groups overlap in feature space. In our study, we avoided overlap in the distribution of walking direction between two BM subsets, thus preserving the integrity of ensemble representations for each subset and mitigating any mutual influence between subsets. Moreover, researchers have proposed a series of factors affecting ensemble coding, including group differences in mean value and items' distance to mean (Im et al., 2021; Treue et al., 2000; Utochkin, 2015). A population-coding model of ensemble perception has been proposed for these various facets of ensemble perception (Utochkin et al., 2024). While these studies mainly focused on the grouping of simple features, our study extends these insights to BM to some extent. Comparing Experiments 3 versus 2, we notice that a smaller group differences in mean value resulted in a weaker bias to the group's mean, suggesting a less efficient ensemble coding due to diminished group disparity. However, it remains to be explored whether existing models accurately predict the grouping process of BM, warranting further investigation in this domain.

Although our results demonstrate that WM inherently organizes structured BM crowds into separate ensembles, the simultaneous maintenance of both ensemble representations for the two subsets in WM remains uncertain. Several studies have explored the feasibility of extracting and concurrently storing multiple ensemble representations in the context of ensemble perception, providing affirmative findings (Attarha & Moore, 2015; Attarha et al., 2014). However, this question persists within the framework of WM, which is characterized by limited capacity, particularly when dealing with complex features such as BM. This complexity is particularly pronounced given that while WM can typically store 3–4 simple features, the capacity for BM is confined to 2–3 instances (Gao et al., 2016a, 2016b; Wood, 2007). This disparity poses a challenge to maintaining numerous ensembles of BM within the WM framework. Future research could further investigate the simultaneous storage of ensemble representations for separate groups within WM.

Echoing previous studies which indicated that ensemble encoding was sensitive to grouping cues, our results provide direct evidence that WM efficiently groups information using Gestalt principles and simultaneously compresses this information as separate ensembles. Grouping and ensemble coding have both been extensively assumed as the strategies to alleviate capacity limitations (Alvarez, 2011; Ariely, 2001; Brady et al., 2009; Peterson et al., 2015; Xu & Chun, 2007). In light of this, the interaction between the two strategies has attracted intensive attention of studies, which confirmed the sensitivity of ensemble coding to grouping cues (Brady et al., 2011; Corbett, 2017; Lamer et al., 2018). For instance, researchers found that participants' judgements of summary statistics were less accurate when stimuli contained grouping structures (Lew & Vul, 2015; Marchant et al., 2013). Advancing these results, the current study directly reveals the simultaneous operating mechanisms of these two strategies in WM, i.e., WM automatically represents environmental regularities as subgroup-based ensemble representations. Besides, implicit ensemble tasks were adopted in the current study in which participants don't need directly report the mean of a stimuli set and were encouraged to memorize the individual stimulus, thus better verifying the robustness and automaticity of the pattern of subgroup-based ensemble representations in WM.

This study contributes to and extends research on ensemble coding, which has shown that BM can be automatically stored as ensembles in WM, in addition to simple features. There is substantial evidence that summary information, such as the mean, can be precisely and rapidly perceived from various types of features, including simple features (e.g., Alvarez, 2011; Michael et al., 2014) and complex social features (e.g., Elias et al., 2017; Florey et al., 2016; Yamanashi Leib et al., 2016), including the walking directions of BM (Sweeny et al., 2013). However, the question of whether ensembles can be maintained in WM has only been addressed for simple features (Brady & Alvarez, 2011; Brady & Alvarez, 2015; Brady et al., 2011; Schurgin & Brady, 2019). In this context, our findings fill this gap and provide further evidence for the generality of ensemble coding. Namely, in WM, a higher-level cognitive process, complex features such as BM can still be automatically represented as ensembles.

Our findings also present new evidence that representations in WM are not independent, but rather interconnected. Traditionally, WM has been conceived as a repository for storing limited and discrete items (Adam et al., 2017; Luck & Vogel, 1997; Zhang & Luck, 2008), often employing individual biological motion (BM) elements to explore WM capacity (Gao et al., 2016a, 2016b; Wood, 2007) and neural representations of BM (Cai

et al., 2018; Gao et al., 2015). Nevertheless, recent years have witnessed a growing body of evidence challenging this classical notion, with studies revealing interactions between WM representations (Bae & Luck, 2017; Czoschke et al., 2020; Lively et al., 2021; Utochkin & Brady, 2020). This shift in understanding is also mirrored in BM studies. For instance, Ding et al. (2017) observed that interactive BM was stored in WM as a unified chunk, yielding better memory performance compared to individual memorization. Consistent with these results, our study showed that participants inherently uphold ensemble representations, rather than isolated individual BM representations, within WM.

Although our previous discussion primarily focused the ensemble representations of BM within WM, we do not deny the existence of at least some memories for individual BM. In fact, drawing upon the hierarchical encoding theory, we posit that individual memories exist in the current study. Based on this theory, WM encodes both the "gist" of WM displays (ensemble statistics such as mean value) and information about specific items (Brady & Alvarez, 2011). Empirical evidence was largely derived from the delayed estimation task, wherein participants are tasked with reproducing a cued item from the memory display. The results of such studies have consistently revealed that the memories for individual items exhibit a bias toward the mean value, but displayed significantly lower errors compared to the chance level (Corbett, 2017; Duffy et al., 2010; Griffiths et al., 2018; Son et al., 2020). This suggests that WM retains information about individual items, and simultaneously, these individual representations are influenced by higher-order statistics (mean value) held in WM. Similarly, the member identification task employed in our current study also required the encoding of individual item memories. Therefore, we posit that some memories for individual items exist in the context of the current study.

Beyond the theoretical significance, this investigation provides valuable insights into the mechanisms and strategies employed by the human brain in processing and interpreting these social scenes, with practical implications across various domains. Firstly, understanding how individuals perceive and extract meaningful information from structured BM crowds can contribute to the development of effective crowd management strategies. For instance, our findings can inform the design of signage or visual cues that enhance safety in crowded environments such as airports or public gatherings. Additionally, our research has direct relevance to the domain of social interaction. The processing of structured BM crowds plays a crucial role in social contexts, where individuals rely on the interpretation of clustering structures to understand intentions

and emotions. By gaining a better understanding of how people perceive these structures within crowds, we can improve social interaction in various settings. Finally, our research contributes to the development of computational models or algorithms that simulate human visual processing, advancing computer vision applications.

Finally, this study has some limitations. While constructing structured BM crowds, we employed fundamental organizational cues of proximity and similarity, both rooted in basic physical attributes. However, one of the distinguishing aspects of BM lies in their inherent social nature. Consequently, in addition to the application of Gestalt principles, there exist other distinctive social organizational principles pertinent to BM, such as joint gaze (Corkum & Moore, 1995). Gaze cues represent ubiquitous social features that furnish a wealth of information during social interactions, particularly concerning attention and intention (Emery, 2000; Moll & Tomasello, 2007). Future research endeavors could consider incorporating social cues like joint gaze to formulate structured BM crowds. This avenue of investigation would provide an opportunity to explore whether higher-level cognitive processes influence the encoding and storage of these BM crowds within WM.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41235-024-00574-6>.

Supplementary material 1.

### Acknowledgements

#### Public Significance

Biological motion (BM) is a ubiquitous aspect of real-world scenarios, often manifesting in crowds with structured groups rather than uniform arrangements. In the cognitive processing of BM, working memory (WM) plays a pivotal role. Here we for the first time demonstrated that WM involuntarily encodes BM crowds as separate ensembles, even when the organizational cues are subtle. Our finding provides compelling evidence that WM efficiently groups BM crowds using Gestalt principles while concurrently compacting them into separate ensembles. This finding unveils the concurrent operational mechanism underlying the processes of grouping and ensemble coding for BM crowds within WM. Understanding how structured BM crowds are effectively processed within WM is crucial for our performance and adaptation in complex social interactions encountered in real-life scenarios. This investigation enables us to acquire insights regarding the mechanisms and strategies employed by the human brain in processing and interpreting these social scenes, which establishes the foundation for more effective social interaction technologies and training methods.

#### Funding

This work was supported by grants from the National Natural Science Foundation of China (32271103), Natural Science Foundation of Guangdong Province (2023A1515011762), Guangzhou Science and Technology Plan Project - Leading Elite Program (2024A04J3301), and the Fundamental Research Funds for the Central Universities, Major Project Cultivation and Emerging Interdisciplinary Cultivation Plan (24wkjc02) awarded to author Xiaowei Ding.

#### Availability of data and materials

Experimental data in an aggregated format and all materials are available on Open Science Framework <https://osf.io/ezk4v/>.

#### Code availability

The experiment was programmed using MATLAB (MathWorks, Natick, MA, USA) with Psychtoolbox extensions (Brainard, 1997; Pelli, 1997).

#### Declarations

#### Ethics approval and consent to participate

The study was approved by the Research Ethics Board of Sun Yat-sen University and performed according to the approved guidelines. For all participants, signed informed consent was obtained before the study.

#### Consent to publication

Consent to publish has been obtained from the participant.

#### Competing interests

The authors have no conflicts to disclose.

Received: 18 December 2023 Accepted: 4 July 2024

Published online: 10 July 2024

#### References

- Abernethy, B., Gill, D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. *Perception*, *30*, 233–252.
- Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Science*, *15*(3), 122–131.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.
- Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception & Psychophysics*, *77*(4), 1116–1131.
- Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of experimental psychology: Human Perception and Performance*, *40*(4), 1440–1449.
- Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, & Psychophysics*, *79*(8), 2376–2395.
- Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*, *82*, 63–79.
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.
- Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15), 6.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*, 487–502.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. *Journal of Vision*, *11*(5), 4.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Brunyé, T. T., Howe, J. L., & Mahoney, C. R. (2014). Seeing the crowd for the bomber: Spontaneous threat perception from static and randomly moving crowd simulations. *Journal of Experimental Psychology Applied*, *20*(4), 303–322.

- Cai, Y., Urgolites, Z., Wood, J., Chen, C., Li, S., Chen, A., & Xue, G. (2018). Distinct neural substrates for visual short-term memory of actions. *Human Brain Mapping, 39*(10), 4119–4133.
- Corbett, J. E. (2017). The whole warps the sum of its parts: Gestalt-defined-group mean size biases memory for individual objects. *Psychological Science, 28*(1), 12–22.
- Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Lawrence Erlbaum Associates Inc.
- Czoschke, S., Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2020). Visual objects interact differently during encoding and memory maintenance. *Attention, Perception, & Psychophysics, 82*(3), 1241–1257.
- Ding, X., Gao, Z., & Shen, M. (2017). Two equals one: Two human actions during social interaction are grouped as one unit in working memory. *Psychological Science, 28*(9), 1311–1320.
- Ding, X., Zhao, Y., Wu, F., Lu, X., Gao, Z., & Shen, M. (2015). Binding biological motion and visual features in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 850–865.
- Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review, 17*(2), 224–230.
- Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble perception of dynamic emotional groups. *Psychological Science, 28*(2), 193–203.
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604.
- Florey, J., Clifford, C. W., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports, 6*, 32210.
- Galef, B. G., & Giraldeau, L. A. (2001). Social influences on foraging in vertebrates: Causal mechanisms and adaptive functions. *Animal Behavior, 61*, 3–15.
- Gallup, A. C., Hale, J. J., Sumpter, D. J., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences, 109*(19), 7245–7250.
- Gao, Z., Bentin, S., & Shen, M. (2015). Rehearsing biological motion in working memory: An EEG study. *Journal of Cognitive Neuroscience, 27*(1), 198–209.
- Gao, Z., Gao, Q., Tang, N., Shui, R., & Shen, M. (2016a). Organization principles in visual working memory: Evidence from sequential stimulus display. *Cognition, 146*, 277–288.
- Gao, Z., Ye, T., Shen, M., & Perry, A. (2016b). Working memory capacity of biological movements predicts empathy traits. *Psychonomic Bulletin & Review, 23*(2), 468–475.
- Griffin, A. S. (2004). Social learning about predators: A review and prospectus. *Animal Learning & Behavior, 32*, 131–214.
- Griffiths, S., Rhodes, G., Jeffery, L., Palermo, R., & Neumann, M. F. (2018). The average facial expression of a crowd influences impressions of individual expressions. *Journal of Experimental Psychology: Human Perception and Performance, 44*(2), 311–319.
- Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature, 407*, 487–490.
- Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception, 43*(7), 663–676.
- Im, H. Y., Tiurina, N. A., & Utchkin, I. S. (2021). An explicit investigation of the roles that feature distributions play in rapid visual categorization. *Attention, Perception, & Psychophysics, 83*, 1050–1069.
- Johansson, G. (1973). Visual-perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211.
- Khayat, N., Fusi, S., & Hochstein, S. (2021). Perceiving ensemble statistics of novel image sets. *Attention, Perception, & Psychophysics, 83*, 1312–1328.
- Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision, 18*(9), 23.
- Khayat, N., & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception, & Psychophysics, 81*, 2850–2872.
- Lamer, S. A., Sweeny, T. D., Dyer, M. L., & Weisbuch, M. (2018). Rapid visual perception of interracial crowds: Racial category learning from emotional segregation. *Journal of Experimental Psychology: General, 147*(5), 683–701.
- Lee, J., & Chong, S. C. (2021). Quality of average representation can be enhanced by refined individual items. *Attention, Perception, & Psychophysics, 83*, 970–981.
- Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision, 15*(4), 10.
- Lively, Z., Robinson, M. M., & Benjamin, A. S. (2021). Memory fidelity reveals qualitative changes in interactions between items in visual working memory. *Psychological Science, 32*(9), 1426–1441.
- Lu, X., Huang, J., Yi, Y., Shen, M., Weng, X., & Gao, Z. (2016). Holding biological motion in working memory: An fMRI Study. *Frontiers in Human Neuroscience, 10*, 251.
- Lu, X., Ma, X., Zhao, Y., Gao, Z., & Shen, M. (2019). Retaining event files in working memory requires extra object-based attention than the constituent elements. *Quarterly Journal of Experimental Psychology, 72*(9), 2225–2239.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.
- Marchant, A., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142*, 251–258.
- Marini, F., Sutherland, C., Ostrovska, B., & Manassi, M. (2023). Three's a crowd: Fast ensemble perception of first impressions of trustworthiness. *Cognition, 239*, 105540.
- Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of National Academy of Sciences, 111*(21), 7873–7878.
- Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology, 43*(2), 309–317.
- Moussaïd, M., Kapadia, M., Thrash, T., Sumner, R. W., Gross, M., Helbing, D., & Hölscher, C. (2016). Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of the Royal Society Interface, 13*(122), 20160414.
- 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.
- Ortego, K., & Störmer, V. S. (2024). Similarity in feature space dictates the efficiency of attentional selection during ensemble processing.
- 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.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.
- Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review, 20*(6), 1282–1289.
- Peterson, D. J., Gözenman, F., Arciniega, H., & Berryhill, M. E. (2015). Contralateral delay activity tracks the influence of Gestalt grouping principles on active visual working memory representations. *Attention, Perception, & Psychophysics, 77*(7), 2270–2283.
- Ristic, J., & Capozzi, F. (2022). Mechanisms for individual, group-based and crowd-based attention to social information. *Nature Reviews Psychology, 1*, 721–732.
- Schurigin, M. W., & Brady, T. F. (2019). When “capacity” changes with set size: Ensemble representations support the detection of across-category changes in visual working memory. *Journal of Vision, 19*(5), 3.
- Shen, M., Gao, Z., Ding, X., Zhou, B., & Huang, X. (2014). Holding biological motion information in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1332–1345.
- Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology, 40*, 497–514.
- Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology, 41*, 235–250.
- Son, G., & Chong, S. C. (2023). Similarity-based clustering of multifeature objects in visual working memory. *Attention, Perception, & Psychophysics, 85*(2242), 2256.
- Son, G., Oh, B.-I., Kang, M.-S., & Chong, S. C. (2019). Similarity-based clusters are representational units of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(1), 46–59.
- Son, G., Oh, B. I., Kang, M. S., & Chong, S. C. (2020). Similarity-based clusters are representational units of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(1), 46–59.
- 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(2), 329–337.
- Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, 25(10), 1903–1913.
- Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion—physiology and psychophysics. *Nature Neuroscience*, 3(3), 270–276.
- Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision*, 15(4), 8–8.
- Utochkin, I. S., & Brady, T. F. (2020). Individual representations in visual working memory inherit ensemble properties. *Journal of Experimental Psychology: Human Perception and Performance*, 46(5), 458–473.
- Utochkin, I. S., Choi, J., & Chong, S. C. (2024). A population response model of ensemble perception. *Psychological Review*, 131(1), 36–57.
- Vestner, T., Flavell, J. C., Cook, R., & Tipper, S. P. (2022). Remembered together: Social interaction facilitates retrieval while reducing individuation of features within bound representations. *Quarterly Journal of Experimental Psychology*, 75(9), 1593–1602.
- Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.-A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*, 148(7), 1251–1268.
- Webster, J., Kay, P., & Webster, M. A. (2014). Perceiving the average hue of color arrays. *Journal of the Optical Society of America A*, 31(4), A283–A292.
- Westfall, J. (2016). PANGEA: Power analysis for general ANOVA designs. Available at <http://jakewestfall.org/publications/pangea.pdf>.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105–129.
- Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General*, 136(4), 639–652.
- Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition*, 108, 522–532.
- Wood, J. N. (2010). Visual working memory retains movement information within an allocentric reference frame. *Visual Cognition*, 18, 1464–1485.
- Wood, J. N. (2011). A core knowledge architecture of visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 357–381.
- Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, 104, 18766–18771.
- Yamanashi Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, 7, 13186.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.