AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2021, Vol. 47, No. 5, 648–661
https://doi.org/10.1037/xhp0000907

Tuning the Ensemble: Incidental Skewing of the Perceptual Average Through Memory-Driven Selection

Ryan S. Williams[1], Jay Pratt[1], Susanne Ferber[1], and Jonathan S. Cant[2]
[1] Department of Psychology, University of Toronto
[2] Department of Psychology, University of Toronto Scarborough

The process by which multiple items within an object grouping are rapidly summarized along a given visual dimension into a single mean value (i.e., perceptual averaging) has increasingly been shown to interact dynamically with visual working memory (VWM). Commonly, this interaction is studied with respect to the influence of perceptual averaging over VWM, but it is also the case that VWM can support perceptual averaging. Here, we argue that, in the presence of memory-matching elements, VWM exerts an obligatory influence over perceptual averaging even when it is detrimental to do so. Over four experiments, we tested our hypothesis by having individuals perform a mean orientation estimation task while concurrently maintaining a colored object in VWM. We anticipated that mean orientation reports would be attracted to the local mean of memory-matching items if such items are prioritized in perceptual average judgments. This was indeed the case as we observed a persistent bias in mean orientation judgments toward the subset mean of items matching the VWM item color, despite color being entirely irrelevant to the mean orientation task. Our results thus highlight a goal-invariant influence of VWM over perceptual averaging, which we attribute to amplification through memory-driven selection.

Public Significance Statement

Current understanding of the interaction between perceptual averaging and visual working memory (VWM) has largely centered on the influence of the former over the latter, with much less consideration of a potential bidirectional relationship between these two systems. In this study, we show that not only does VWM alter perceptual averaging judgments but that it also does so automatically (i.e., even when it is costly to do so). More broadly, this work provides confirmatory support to the idea that the amplification of items within object ensembles is set by an underlying priority map, which guides selective attention on the basis of physical salience, top-down goals, emotional valence, and reward history.

Keywords: ensemble statistics, visual working memory, feature-based attention, mean orientation judgments

Through the process of perceptual averaging, the visual system rapidly summarizes characteristics of object groupings into mean values that represent the group across the visual hierarchy, ranging from low-level visual features—for example, mean size (Ariely, 2001; Chong & Treisman, 2003), hue (Webster et al., 2014), and orientation (e.g., Alvarez & Oliva, 2009; Dakin & Watt, 1997)—to more complex visual properties, such as the mean facial expression of a crowd (e.g., Haberman et al., 2009; Haberman & Whitney, 2007, 2009). This process of perceptual averaging provides adaptive value in its own right, but there has also been a growing appreciation for how it may interact with and inform other cognitive systems, with a particular focus on visual working memory (VWM). Most notably, Brady and Alvarez (2011) demonstrated that when VWM arrays are segregable by a grouping feature (i.e., color), individuals' single-item memory reports incorporate higher-level information about the item's associated set. That is, when individuals were asked to report the size of memorized circles, size estimations were systematically biased in accordance with the mean size of a subset of items matching the probe item's color (Brady & Alvarez, 2011; see also Lowe et al., 2018). Corroborating this finding, subsequent works have since shown similar results on the basis of location (Lew & Vul, 2015) and other Gestalt grouping principles (Corbett, 2017). Furthermore, even

This article was published Online First March 15, 2021.

Ryan S. Williams https://orcid.org/0000-0003-1638-6799

Jonathan S. Cant https://orcid.org/0000-0002-8771-6671

This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants to Jay Pratt (RGPIN-2016-0635), Susanne Ferber (RGPIN-2018-05829), and Jonathan S. Cant (435647-13) and an NSERC Scholarship (CGS-D) to Ryan S. Williams. The authors thank Juanita Atton, Robert Newman, Richard Schmidt, and Nika Zahedi Neysiani for their assistance with data collection, and Andrew Clement for his thoughtful discussion.

Correspondence concerning this article should be addressed to Ryan S. Williams, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada. Email: ryanscott.williams@mail.utoronto.ca

when individual items are not readily separated into multiple clusters, information about the mean of the entire set has nonetheless been shown to influence VWM reports (e.g., Corbin & Crawford, 2018; Dubé et al., 2014; Griffiths et al., 2018; Sama et al., 2019, in press; Utochkin & Brady, 2020). Accounting for this pervasive influence of perceptual averaging on single-item VWM representations, two complementary functions have been hypothesized. First, by anchoring similar items to a single value (i.e., the subset mean), individuals can mitigate capacity limitations at the storage level (e.g., Alvarez, 2011; Son et al., 2020). Second, when faced with uncertainty, the set mean can be used at the decision level to approximate the identity of individual items (e.g., Dubé et al., 2014; Honig et al., 2020; Utochkin & Brady, 2020). As such, perceptual averaging can be viewed as a compensatory mechanism that supports VWM by reducing load and minimizing the effects of forgetting.

However, while the primary focus of works in this area to date has been concerned with the influence of perceptual averaging on the representation of individual items in VWM, there is good reason to believe that the inverse also holds true—namely, that VWM can influence perceptual averaging. For example, in much the same way that competition over shared cortical representation can shift the perception of an object's features toward (e.g., Teng & Kravitz, 2019) or away (e.g., Scocchia et al., 2013) from information of the same dimension in VWM, the perception of average motion is similarly shifted away from dimension-matching information held in VWM (Kang et al., 2011). Moreover, when items are segregable on a given dimension, individuals can use attentional templates stored in VWM (Berggren & Eimer, 2018; Bundesen, 1990; Carlisle & Woodman, 2011; Woodman & Arita, 2011) to improve the averaging of items possessing a task-relevant feature (Brand et al., 2012), despite previous evidence to the contrary (i.e., Chong & Treisman, 2005b). As this latter finding suggests, then, VWM can be used deliberately to prioritize the averaging of prespecified items through memory-driven selection. What is particularly interesting about this finding, though, is that this form of attentional guidance can also occur incidentally as simply maintaining a feature value in VWM can bias attention toward perceptually similar information in the visual field (e.g., Carlisle & Woodman, 2011; Hollingworth et al., 2013; Kumar et al., 2009; Olivers et al., 2006; Soto et al., 2005; Sun et al., 2015). This begs the question, then, does memory-driven selection automatically alter the contribution of individual items to perceptual averages even when it is costly to do so (i.e., when all items are equally relevant to one's task)?

Whether or not memory-driven selection can automatically bias the extraction of perceptual averages critically depends on whether the weighting of individual items is commensurate to the distribution of attention across the set, which indeed there is mounting evidence to suggest. Support in this regard was first provided by Chong and Treisman (2005a), who showed that broadly distributing attention across a set of items produces greater accuracy for mean-size recognition judgements relative to conditions where attention is localized to specific items or locations at encoding. Extending from this finding, de Fockert and Marchant (2008) further showed that individually selected items (in this case, for a concurrent localization task) produce an undue influence on perceptual averaging, with more recent work confirming that this effect is at least partially owed to increased weighting of selected items at the time of encoding (Choi & Chong, 2020). Similarly, using a multiple-object tracking procedure with faces, Chen and Zhou (2018) found that individuals were more likely to endorse the mean identity of the tracked faces as being a member of the tracked set compared to the mean identity of the untracked faces or the set as a whole, suggesting that even the implicit influence of perceptual averaging on single-item recognition is biased in favor of attended items.

Still, although the above findings provide strong support to suggest that perceptual averaging is biased in accordance with the allocation of attention across set members, the described works largely rely on tasks that explicitly require specific items be attended over others (though to a lesser extent with respect to Chong & Treisman, 2005a), whereas to establish that such effects occur automatically, such a pattern of results ideally would be observable in the absence of direct instruction. On this matter, Kanaya et al. (2018) recently demonstrated that the effect of attentional weighting on perceptual averaging is observable for more basic tasks. Using a procedure where individuals simply had to compare the mean size or temporal frequency of circular disks to a test disk, participants were shown to reliably overestimate the mean size and temporal frequency of the disks, particularly when the set size was large or interstimulus variability was high. Accounting for this finding, Kanaya et al. (2018) put forth an amplification hypothesis of perceptual averaging, stating that physically salient items (in their case, the largest- or highest-frequency items) are more heavily weighted than less salient items in the determination of such summary statistics. More specifically, this hypothesis rests on the idea that perceptual averaging occurs via the sampling of just a subset of items (e.g., Allik et al., 2013; Myczek & Simons, 2008) approximately equal to the square root of all items (e.g., Dakin, 2001; Whitney & Yamanashi Leib, 2018), rather than through exhaustive sampling of all items in a set, as others have argued (e.g., Ariely, 2001; Chong & Treisman, 2005b; Chong et al., 2008). Physically salient items, which attract attention, are argued to be preferentially included in the sample with proportionally greater representation as the set size is increased, with set variability working to increase the relative saliency of individual items and thus strengthen the amplification effect.

More recently, Iakovlev and Utochkin (2020) extended this work and offered further clarity with respect to whether this attentional amplification is based solely on a subset of physically salient items. Much like the study by Kanaya et al. (2018), physical salience was defined according to the relative size of set members; however, the items themselves were oriented arrows, to which participants were asked to determine the mean orientation. By making the biasing dimension independent of the probed dimension, the authors were able to manipulate the mean of the largest items to be oriented counterclockwise from the global mean, clockwise from the global mean, or equal to the global mean. Additionally, rather than relying on forced-choice judgments, participants were asked to estimate the mean through a continuous-report procedure. These methodological differences allowed for greater confidence in the effect of physical salience on perceptual averaging in that differences in the estimation of the mean could be directly mapped to changes in the assignment of the large arrows. Overall, the observed results were consistent with the amplification hypothesis;

estimations were biased toward the mean of the largest set items and more or less variable depending on the range of orientations used for the large set items. Of note, though, the effects were less than what would be expected had individuals relied solely on the most salient items in the displays, highlighting that while salient items received greater weighting, less salient items nonetheless contributed to the perceived perceptual average.

Last, while both Kanaya et al. (2018) and Iakovlev and Utochkin (2020) demonstrated that physically salient items are automatically favored in the determination of perceptual averages, it is worth noting that the amplification effect is not limited to such stimuli. For instance, Dodgson and Raymond (2020) showed that individuals overvalue the contribution of items possessing a feature (i.e., color) previously associated with a reward when making mean size estimations. Similarly, when making judgments of mean expression, individuals' reports are biased toward emotionally salient faces (Goldenberg et al., 2020). Taken together, it can be argued that the weighting of items contributing to perceptual averages is set by an underlying priority map, which distributes attentional resources on the basis of bottom-up salience, top-down goals, reward history, and emotional valence, akin to what is observed in the case of visual search (e.g., Awh et al., 2012; Wolfe, 2007).

Returning, then, to the matter of whether memory-driven selection automatically influences perceptual averaging, we can confidently assert that if VWM does indeed bias attention toward feature-matching elements, then this will have a measurable impact on the extraction of summary statistics. We thus sought to demonstrate this in the present study. For our general method, we used a modified version of a dual-task paradigm typically used to study memory-driven selection (e.g., Kumar et al., 2009; Olivers et al., 2006; Soto et al., 2005; Sun et al., 2015) but replaced the visual search task that is commonly employed with a mean orientation estimation task that incorporated design elements used by Iakovlev and Utochkin (2020) and Dodgson and Raymond (2020). To start each trial, individuals were asked to memorize a colored object to be tested through change detection at the end of the trial. Between the time of study and test, individuals were shown displays of oriented bars containing two intermixed subsets of elements grouped by color and were then asked to report the mean orientation of all presented items. Critically, the color of the item maintained in VWM could match one of the two subsets (oriented either counterclockwise or clockwise from the global mean of the entire set). Subscribing to the amplification hypothesis and other related works, our primary hypothesis was that items matching a color maintained in VWM would receive greater levels of attention and thus be overweighted in estimations of the perceptual average. As such, we predicted that estimation errors would be biased in a counterclockwise direction when the local mean of elements sharing the color of the item held in VWM was oriented counterclockwise from the global mean. Similarly, errors would be biased in a clockwise direction when the local mean of such color-matching elements was clockwise from the global mean.

Experiment 1

The purpose of Experiment 1 was to directly address our primary research question: Does memory-driven selection obligatorily bias the extraction of perceptual averages? As described above, this entailed having participants perform mean orientation estimations while simultaneously maintaining a colored object in VWM that could match the color of a subset of items. We predicted that, through the implicit biasing of feature-based attention, global mean orientation estimates would be attracted toward the local mean of a subset of items matching the color of the object held in VWM.

Method

Participants

Twenty-five undergraduates (19 female; M = 20.8 years, SD = 1.9) from the University of Toronto participated in this experiment for course credit. We determined through an a priori power analysis that this sample size would yield a .95 probability of detecting a difference between clockwise and counterclockwise conditions at an alpha level of .05 if a true effect were present. We did so using G*Power (Faul et al., 2009) and assuming an effect size of d_z = .76, which corresponds to the average effect size in the study by de Fockert and Marchant (2008) when contrasting mean size recognition accuracy for targets presented alongside foils congruent with an attended set member versus targets presented alongside foils incongruent with an attended set member. Notably, this sample size was approximately equal to that used by Iakovlev and Utochkin (2020), who employed a similar orientation estimation task to the one used in the current study. We use this sample size for all proceeding experiments and additionally include Bayesian and cross-experiment analyses where key conclusions depend on acceptance of the null hypothesis to ensure that our findings are not due to insufficient power. All participants provided written informed consent and reported to have normal color vision and normal or corrected-to-normal visual acuity.

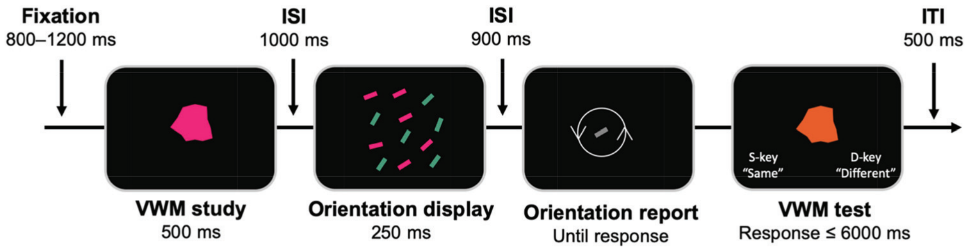
Stimuli, Apparatus, and Procedure

Participants were seated 57 cm from a 19-in. CRT monitor (resolution: 1,280 × 1,024 pixels; refresh rate: 60 Hz). A mounted chin rest was used to stabilize head position. The experiment was carried out using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in MATLAB. All stimuli were displayed against a black background.

A schematic of the task is shown in Figure 1. Each trial began with fixation on a white central cross (.5° × .5°) for 800–1,200 ms. A colored, irregularly shaped 2D object was then presented centrally for 500 ms. Based on stimuli used by Cohen and Singh (2007), these objects were constructed by defining 12 evenly spaced angles from 1–360° and then introducing a random jitter between −5° and 5° to each. These angles were then used to create the object's vertices at randomly selected distances subtending 1.1° to 2.2° from center. Six colors were used for these objects (i.e., orange, gold, teal, blue, violet, and magenta), chosen from evenly spaced locations around a 360° color wheel (centered at L* = 54, a* = 22, b* = 11). Participants were asked to memorize both the form and color of the presented object while it was on the screen.

After studying the object, participants were shown a central fixation cross for 1,000 ms. An orientation display of 12 colored bars (item size: 1.5° × .5°) was then shown for 250 ms within an imaginary 17° × 17° square, with all bars separated by a minimum

Figure 1
Schematic of the Task Used in Experiment 1



Note. Participants studied the color and form of an irregular 2D object. A display of 12 oriented bars was then shown, consisting of two subsets oriented counterclockwise and clockwise from the global mean orientation, respectively. Depending on the condition, the color of one of these subsets could match the color of the studied object. Participants then reported the mean orientation of the bars by rotating a gray bar presented at the center of the screen. Last, participants judged whether a newly presented object matched the object studied at the start of the trial. Shown is an example of a trial with a clockwise feature-matching subset in the orientation display and a color change at visual working memory (VWM) test. ISI = interstimulus interval; ITI = intertrial interval. See the online article for the color version of this figure.

center-to-center distance of 2.2°. This display consisted of two equally numbered subsets of bars, the colors of which were defined by high-contrast color pairs from the six colors used in the VWM task (i.e., orange/blue, gold/violet, and teal/magenta). The locations of all items were randomly assigned within the predefined spatial constraints. In each case, the mean of one subset was oriented 15° counterclockwise to the global mean orientation of the entire set, while the other was oriented 15° clockwise to the global mean. The orientations of individual items within each subset of bars varied by a standard deviation of 8–12°. For one third of the trials, the color of the studied item from the VWM task matched the color of the counterclockwise-oriented subset (CCW condition); for another third of trials, the studied item matched the color of the clockwise-oriented subset (CW condition); and for the remaining third of trials, neither of the subsets matched the color of the VWM study item (control condition). Within each display condition, each of the six colors served as the color of the VWM study item, CCW subset, and CW subset an equal number of times, and each color pair used for the orientation displays occurred with equal frequency across conditions. Further, six predetermined global means ranging from 15°–165° in intervals of 30° were used for the orientation displays, rather than allowing the overall mean to vary randomly from 1–180°, to ensure that the contents of the displays would be matched across conditions.

The orientation display was followed by 900 ms of central fixation. A gray bar then appeared at the center of the screen at a random orientation. Using the mouse cursor, participants rotated the bar to estimate the mean orientation of the orientation display and then clicked the left mouse button to lock in their answer. Participants were instructed to be as accurate as possible and to base their judgments on all of the bars that were present in the display. There was no time limit on the response.

Immediately following the mean orientation judgment, memory for the studied item was tested. Participants were shown another 2D object and were asked to indicate if it was identical to or different from the object studied at the start of the trial by making an "s" (same) or "d" (different) key press, respectively. For different trials, either the color or the form could differ from the studied object (but never both at the same time), with each occurring with equal frequency. When the color of the object changed, one of the remaining colors was randomly selected for the object, with the exception of the studied color's high-contrast counterpart, which was never chosen. When the form of the object changed, a new object was generated following the same procedure used to create the study items. Participants were allotted up to 6 s to respond to the test item, with accuracy being emphasized over speed. Once a response was registered or the time limit was reached, the trial ended and was followed by a 500-ms blank intertrial interval. The task consisted of six blocks separated by short breaks, with each block comprised of 36 trials containing an equal number of trials from each of the three conditions. Prior to starting the task, participants were given 18 practice trials (six trials from each condition) to familiarize them with the VWM and mean orientation estimation tasks.

Analysis

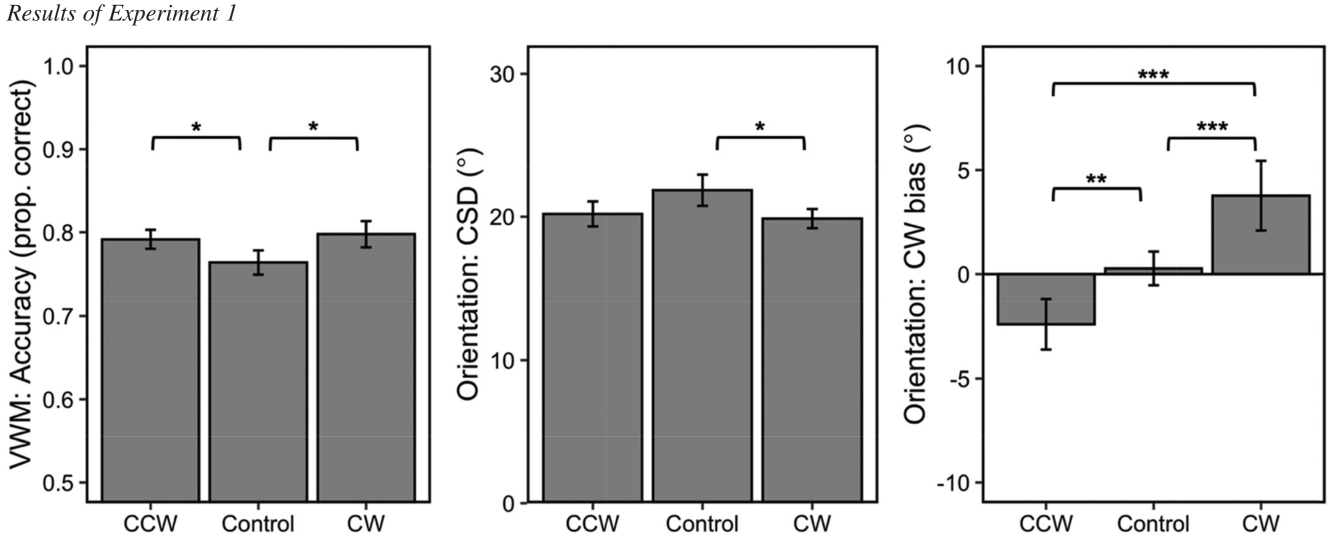
Accuracy was used to gauge performance in the VWM task and was measured as the proportion of correct responses to the test item. Analysis of mean orientation judgments was limited to trials in which participants correctly responded to the VWM test item. This was to ensure that participants had an accurate representation of the study idem in VWM while attending the orientation display. We analyzed response errors, measured as the difference between the participant's estimation of the global mean and the actual mean, with errors ranging from −90° to 90°. For each participant, we scaled the errors to a 360° circular space and estimated circular standard deviation (CSD) and clockwise/counterclockwise bias (i.e., central tendency) parameters of the error distributions corresponding to each of the three conditions (see Bays et al., 2009; code available at http://www.bayslab.com). Last, we employed within-subject one-way analyses of variance (ANOVAs) to compare performance across the three display types (i.e., CW, CCW, control) for our measures of VWM accuracy, CSD, and bias. Moreover, following a similar analysis used by Iakovlev and Utochkin (2020), we compared the average difference on the bias parameter between our control and memory-matching item displays (reversing the signs of the difference scores for the CCW condition) to the value that would be expected if individuals were to solely base their mean orientation judgments on the memory-matching subset of items (i.e., 15°). For better interpretability, we report the estimated parameters in the original 180° orientation space. Effect sizes are provided alongside the results of each ANOVA using partial eta squared (η_p^2), as well as Cohen's d_z and d_s for follow-up within-subject and between-subjects comparisons, respectively (Lakens, 2013). Data for all experiments are openly available online (https://osf.io/u45h7/).

Results and Discussion

Means for the CW, CCW, and control conditions for each of our dependent measures (i.e., VWM accuracy, CSD, and bias) are presented in Figure 2. As is shown, performance on the VWM task significantly differed by display type, $F(2, 48) = 4.25$, $p = .020$, $\eta_p^2 = .15$. Specifically, the proportion of correct responses was lower in the control condition relative to the CCW condition, $t(24) = 2.55$, $p = .017$, $d_z = .51$, and the CW condition, $t(24) = 2.39$, $p = .025$, $d_z = .48$, while VWM performance was equivalent for CCW and CW conditions, $t(24) = .54$, $p = .596$, $d_z = .11$. With respect to the mean orientation task, there was a significant effect of display type on CSD, $F(2, 48) = 3.62$, $p = .034$, $\eta_p^2 = .13$. Further inspection of this effect revealed that there was greater variance in the error distribution for the control condition relative to the CW condition, $t(24) = 2.46$, $p = .022$, $d_z = .49$. There was no significant difference in variance for the control condition relative to the CCW condition, $t(24) = 1.74$, $p = .094$, $d_z = .35$, nor was there a difference between the CCW and CW conditions, $t(24) = .58$, $p = .570$, $d_z = .12$. Critically, there was a significant effect of display type on the bias parameter, $F(2, 48) = 14.80$, $p = < .001$, $\eta_p^2 = .38$. Relative to the control condition, both the CCW, $t(24) = 4.37$, $p < .001$, $d_z = .87$, and CW, $t(24) = 2.92$, $p = .007$, $d_z = .58$, conditions exhibited bias in the direction of the memory-matching subset mean (i.e., counterclockwise and clockwise bias, respectively), and the difference between these two conditions was also significant, $t(24) = 4.29$, $p < .001$, $d_z = .86$. Last, while these results do suggest that estimates of mean orientation for the entire set were attracted toward the local mean of individual items corresponding to the color of the item held in VWM, the average magnitude of the memory-matching item bias relative to the control condition ($M = 3.08°$) was significantly smaller than what would be expected had such judgments been based solely on the memory-matching subset, $t(24) = 16.60$, $p < .001$, $d_z = 3.32$.

Overall, these results provide strong evidence to support our prediction that memory-driven selection can alter perceptual averaging independent of intention. Considering that the guidance of attention toward memory-matching elements is the likely source of this bias, these results align well with the amplification hypothesis of perceptual averaging (Kanaya et al., 2018), which argues that perceptual averaging occurs through a process of nonrandom sampling that prioritizes high-salience items. Moreover, with respect to the fate of the less salient (i.e., the nonmatching items), our results are consistent with Iakovlev and Utochkin (2020) in that we too demonstrate that while salient items overcontribute to estimations of the mean, such judgments are not based solely on these items. Indeed, while we cannot directly speak to whether individuals employed exhaustive sampling of all presented items (e.g., Ariely, 2001; Chong & Treisman, 2005b; Chong et al., 2008) or instead relied on just a partial sample of the items (e.g., Allik et al., 2013; Dakin, 2001; Myczek & Simons, 2008; Whitney & Yamanashi Leib, 2018), it is worth noting that to achieve the observed bias of approximately 3° using the latter strategy, on average, individuals would have needed to sample memory-matching items over nonmatching items at a rate of 3 to 2 (i.e., sampling three items from a distribution centered at 15° and two items from a distribution centered at −15° would yield an average value of 3°).

Figure 2
Results of Experiment 1



Note. $N = 25$. Dependent measures include accuracy for the visual working memory (VWM) task (left), as well as circular standard deviation (CSD; center) and bias (right) parameters of the error distributions for estimations of mean orientation. Error bars represent 95% confidence intervals for within-subject designs (Cousineau, 2005). CCW = counterclockwise condition; CW = clockwise condition.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Experiment 2

In Experiment 1, we demonstrated that estimations of mean orientation are biased by the presence of elements that shared a common feature with an item maintained in VWM, supporting the notion that memory-driven selection exerts an obligatory influence over perceptual averaging. Still, it is nonetheless possible that our findings could be attributable to perceptual priming rather than top-down alterations of an underlying priority map by the contents of VWM. In other words, simply being exposed to the color of one of the subsets prior to the presentation of the orientation display may be sufficient to alter one's perceptual sensitivity to such items when subsequently encountered (Belopolsky et al., 2010; Maljkovic & Nakayama, 1994; Wiggs & Martin, 1998). To test this account, we conducted a second experiment in which the color of the studied item was never tested and thus did not need to be maintained in VWM. If we continued to observe an estimation bias in favor of items matching the color of the studied item, then this would indicate that our initial finding is better accounted for by effects related to one's perceptual history rather than the explicit maintenance of a feature in VWM.

Method

Participants

Twenty-six undergraduates from the University of Toronto participated in this experiment for course credit; however, one participant was excluded because of poor performance on the VWM task (i.e., overall accuracy $<$ 3 standard deviations below the group mean). The final sample thus consisted of 25 participants (18 female; M = 19.9 years, SD = 1.5). All participants provided written informed consent and reported to have normal color vision and normal or corrected-to-normal visual acuity.

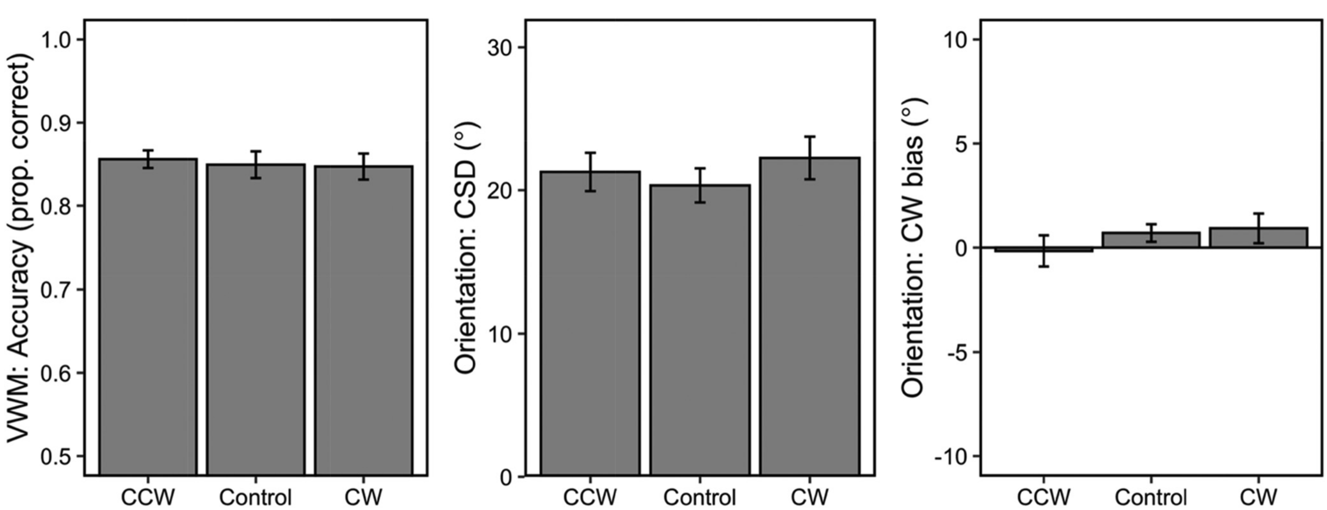
Stimuli and Procedure

The task used in Experiment 2 was identical to that of Experiment 1, with one exception. That is, for the VWM component of the task, the form, but not the color, of the studied object could change at the time of the test display. Thus, while the color of the studied object still matched the color of the CCW subset on one third of the trials and the CW subset on another third of the trials, it was no longer necessary to remember the color of the studied object. Participants were explicitly informed that only the form of the object could change at test.

Results and Discussion

Means for all three display types (i.e., the CW, CCW, and control conditions) are illustrated in Figure 3 for each of our dependent measures. In contrast to the results of Experiment 1, there was no difference in accuracy for the VWM task across display types, $F(2, 48) = .46$, $p = .634$, $\eta_p^2 = .02$. Further, the error distributions for the orientation task did not differ across condition for either the CSD parameter, $F(2, 48) = 1.30$, $p = .282$, $\eta_p^2 = .05$, or the bias parameter, $F(2, 48) = 2.04$, $p = .141$, $\eta_p^2 = .08$. However, because our conclusions here rest on the acceptance of the null hypothesis, we additionally examined the Bayes factor (BF) associated with the effect of display type on our bias parameter. From this analysis, we found that support for the null was positive but inconclusive ($BF_{01} = 1.78$). We thus performed a cross-experiment ANOVA on the bias parameter, which included a between-subjects factor of experiment along with our data from Experiment 1. In doing so, we found an overall effect of display type, $F(2, 96) = 16.32$, $p < .001$, $\eta_p^2 = .25$, that critically interacted with the factor of experiment, $F(2, 96) = 8.21$, $p < .001$, $\eta_p^2 = .15$. Follow-up comparisons showed that the bias toward the memory-matching subset was significantly larger in Experiment 1 relative to the current experiment for both the CCW condition, $t(48) = 2.63$, $p =$

Figure 3
Results of Experiment 2



Note. N = 25. Dependent measures include accuracy for the visual working memory (VWM) task (left), as well as circular standard deviation (CSD; center) and bias (right) parameters of the error distributions for estimations of mean orientation. Error bars represent 95% confidence intervals for within-subject designs (Cousineau, 2005). CCW = counterclockwise condition; CW = clockwise condition.

.012, d_s = .74, and the CW condition, $t(48) = 2.69$, $p = .010$, d_s = .76, while there was no difference for the control condition, $t(48) = .69$, $p = .495$, $d_s = .19$.

From these results, we can conclude that the observed effect of the VWM task on mean orientation estimations in Experiment 1 was not due to perceptual priming since this effect is significantly attenuated, if not completely abolished, when color did not need to be actively stored in VWM, but participants are nevertheless exposed to the color of one of the subsets prior to encountering the orientation display. These results thus strengthen our conclusion that the observed misestimations in the mean orientation task in Experiment 1 were driven by the inadvertent prioritization of items following memory-driven selection. From a broader perspective, the absence of a memory-matching bias in the current experiment also aligns well with a proposal that the guidance of attention critically depends on the nature of the maintained feature representation. That is, while VWM is thought to guide attention when feature values are actively maintained for immediate use, features that are only encoded incidentally, or not required for immediate use, are thought to be maintained only in an accessory state that does not interact with attentional selection (see Olivers et al., 2011).

Experiment 3

Although the results of Experiment 2 demonstrate that the observed bias on estimates of mean orientation in Experiment 1 was not the result of perceptual priming, it is important to acknowledge that factors other than the automatic guidance of attention toward memory-matching items may have contributed to our finding. Most notably, it is possible that the observed bias simply reflects a strategic attempt to improve performance on the VWM task itself. That is, by strategically devoting more attention to memory-matching elements, individuals would be better able to detect a change in the item actively maintained in VWM when tested at the end of the trial. Indeed, this explanation is consistent with the higher accuracy observed for the CCW and CW conditions relative to the control condition in the VWM task of Experiment 1. We thus sought to examine this possibility in Experiment 3 by manipulating the duration of orientation displays to be either brief (i.e., 150 ms) or long (i.e., 500 ms). In doing so, we reasoned that voluntary attentional allocation would be more difficult for brief displays given that strategic attentional allocation is less viable for short presentation durations. In the case of VWM tasks, for example, even if it is known that one of two items is more likely to be probed at test than the other, individuals are unable to strategically prioritize the encoding of the likely test item when the exposure duration is less than 200 ms (but can do so for longer durations; Bays et al., 2011). With respect to perceptual averaging, manipulations of display duration have similarly been effective for understanding effects related to processing time at encoding (e.g., Li et al., 2016; Whiting & Oriet, 2011). Indeed, Goldenberg et al. (2020, Study 2) recently used this method to clarify the time course of attentional amplification in the extraction of mean facial expressions. Here, it was shown that while both positive and negative faces are amplified in estimations of mean expression, longer display durations are required for the bias toward positive faces to match that for negative faces. Considering that negative faces are believed to capture attention more readily than positive faces (e.g., Eastwood et al., 2001; Hansen & Hansen, 1988), this seems to suggest that effects of involuntary attention on perceptual averaging are present early on, whereas effects related to the deliberate allocation of attention may occur at a later stage. As such, if the bias toward memory matching that we observed in Experiment 1 was due to a strategic attempt to improve accuracy on the VWM task, then we would expect that the observed bias would be larger for long displays versus short displays.

Method

Participants

Twenty-seven undergraduates from the University of Toronto participated in this experiment for course credit. After excluding two participants for having accuracy less than 3 standard deviations below the group mean on the VWM task, the final sample consisted of 25 participants (18 female; M = 19.9 years, SD = 1.5). All participants provided written informed consent and reported to have normal color vision and normal or corrected-to-normal visual acuity.

Stimuli and Procedure

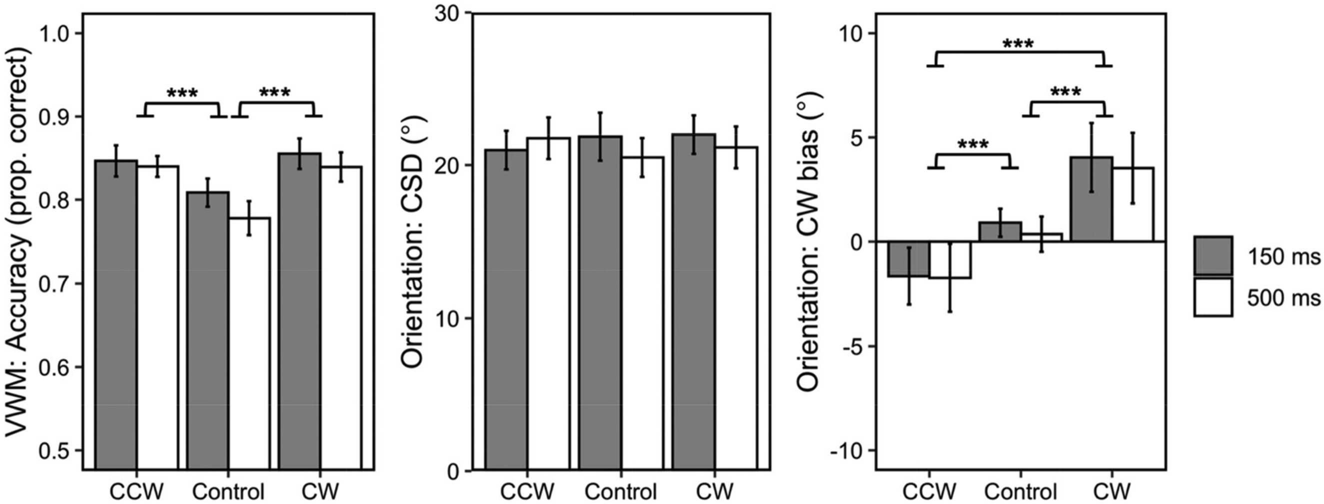
The experimental task used in Experiment 3 was similar to that used in Experiment 1; however, rather than using a constant presentation duration of 250 ms for ensemble displays, such displays were now presented for either 150 ms or 500 ms, randomly intermixed throughout each block. Within each of our conditions, trials were balanced according to study item color and ensemble display duration. Again, six predetermined orientation means were used, which were equally present for each display type and duration. The task consisted of 12 blocks (30 trials each; 360 trials total), with each consisting of an equal number of trials from each display type and display duration. Participants were given 12 practice trials prior to beginning the experiment.

Results and Discussion

Means for the CW, CCW, and control conditions for each of our dependent measures are displayed in Figure 4. Data for each dependent measure were analyzed using a 2 (Display Duration: 150 ms vs. 500 ms) × 3 (Display Type: CCW vs. CW vs. control) repeated-measures ANOVA. Regarding VWM accuracy, a significant main effect of display duration was observed, $F(1, 24) = 6.65$, $p = .016$, $\eta_p^2 = .011$, with reduced accuracy for items tested following longer display durations. As observed in Experiment 1, we again found a main effect of display type on VWM performance, $F(2, 48) = 13.36$, $p < .001$, $\eta_p^2 = .089$, with worse performance in the control condition relative to the CCW, $t(24) = 4.29$, $p < .001$, $d_z = .86$, and CW conditions, $t(24) = 4.10$, $p < .001$, $d_z = .82$, but no difference between CCW and CW conditions, $t(24) = .39$, p = .697, $d_z = .08$. The display type by duration interaction was not significant, $F(2, 48) = .91$, $p = .410$, $\eta_p^2 < .01$.

For orientation estimations, no significant results were observed in the analysis of CSD, main effect of display duration: $F(1, 24) = .63$, $p = .436$, $\eta_p^2 < .01$; main effect of display type: $F(2, 48) = .13$, $p = .882$, $\eta_p^2 < .01$; interaction: $F(2, 48)$ = 1.17, $p = .320$, $\eta_p^2 < .01$. Analysis of our bias parameter, however, revealed a significant main effect of display type,

Figure 4
Results of Experiment 3



Note. $N = 25$. Dependent measures include accuracy for the visual working memory (VWM) task (left), as well as circular standard deviation (CSD; center) and bias (right) parameters of the error distributions for estimations of mean orientation. Gray bars depict trials containing short orientation display durations (i.e., 150 ms), and white bars depict trials containing long orientation display durations (i.e., 500 ms). Error bars represent 95% confidence intervals for within-subject designs (Cousineau, 2005). CCW = counterclockwise condition; CW = clockwise condition.
*** $p < .001$.

$F(2, 48) = 12.34$, $p < .001$, $\eta_p^2 = .26$. As was the case in Experiment 1, mean orientation judgments were rotated more counterclockwise in the CCW condition relative to the control condition, $t(24) = 3.83$, $p < .001$, $d_z = .77$, and more clockwise in the CW condition relative to the control condition, $t(24) = 4.52$, $p < .001$, $d_z = .90$. The difference between the CCW and CW conditions was also significant, $t(24) = 5.10$, $p < .001$, $d_z = 1.02$. The main effect of display duration did not significantly affect response bias, $F(1, 24) = 2.07$, $p = .164$, $\eta_p^2 < .01$, and importantly, the effect of display type did not interact with display duration, $F(2, 48) = .18$, $p = .837$, $\eta_p^2 < .01$, with substantial evidence observed for the null hypothesis ($BF_{01} = 8.62$).

Overall, these results further support the notion that the bias exerted by VWM on perceptual averaging occurs automatically, as opposed to being explained by a purposeful cognitive strategy meant to improve accuracy on the VWM task. Still, while these results do well to minimize the probability that the memory-matching item bias is driven by a deliberate strategy to improve VWM performance, it is not insignificant that we again see an effect of display type on VWM with color now being relevant to the VWM task (unlike in Experiment 2). This may simply reflect that participants benefited from the repeated exposure of the VWM color in the memory-matching displays. Alternatively, participants may have guessed strategically during the VWM task by endorsing a change if the color of the test item was not present in the orientation display. Indeed, this strategy would have been more disadvantageous following the presentation of the control displays as the memory item never matched the color of the items in these displays (either at the time of study or test) and thus may account for why VWM accuracy was reduced following control displays.

Experiment 4

The findings in Experiment 3 again demonstrated that reports of mean orientation were biased toward the color of elements that matched the color of a single item actively held in VWM. Importantly, the magnitude of this bias was invariant to changes in display duration (i.e., short vs. long presentations), supporting the idea that VWM automatically biases attention toward perceptually similar objects in the environment, which are then prioritized in the determination of perceptual averages. Still, while the results of Experiment 3 do work to minimize the likelihood that our results are explained by a strategic attempt to improve accuracy on the VWM task, they alone cannot do so definitively. As such, to add greater confidence to our conclusions, in Experiment 4, we sought to further rule out the involvement of strategic attempts to improve VWM accuracy. To accomplish this, we made it such that the color of the studied item only changed on rare occasions at test. Thus, while individuals were still required to maintain the item's color in VWM, strategically allocating more attention to memory-matching elements would provide little advantage in the VWM task.

Method

Participants

Twenty-five undergraduates from the University of Toronto participated in this experiment for course credit (22 female; $M = 19.8$ years, $SD = 1.4$). All participants provided written informed consent and reported to have normal color vision and normal or corrected-to-normal visual acuity.

Stimuli and Procedure

The procedure used for Experiment 4 was identical to Experiment 1 with one important exception. As was the case in each of our previous experiments, a feature of the studied object changed at the time of test on half of all trials. However, in the current experiment, "change" trials were designed such that the shape of the object changed on 83.3% of these trials (i.e., 41.7% of all trials), whereas the color changed on only 16.7% of these trials (i.e., 8.3% of all trials). In other words, the color of the VWM item had to be actively maintained in VWM (and thus should still bias feature-based attention), but because the probability of a color change was so low, there would be much less motivation to strategically direct more attention to one colored subset over the other. Prior to starting the task, participants were given 36 practice trials, in which the color of the studied object changed at test on three trials (once for each display condition).

Results and Discussion

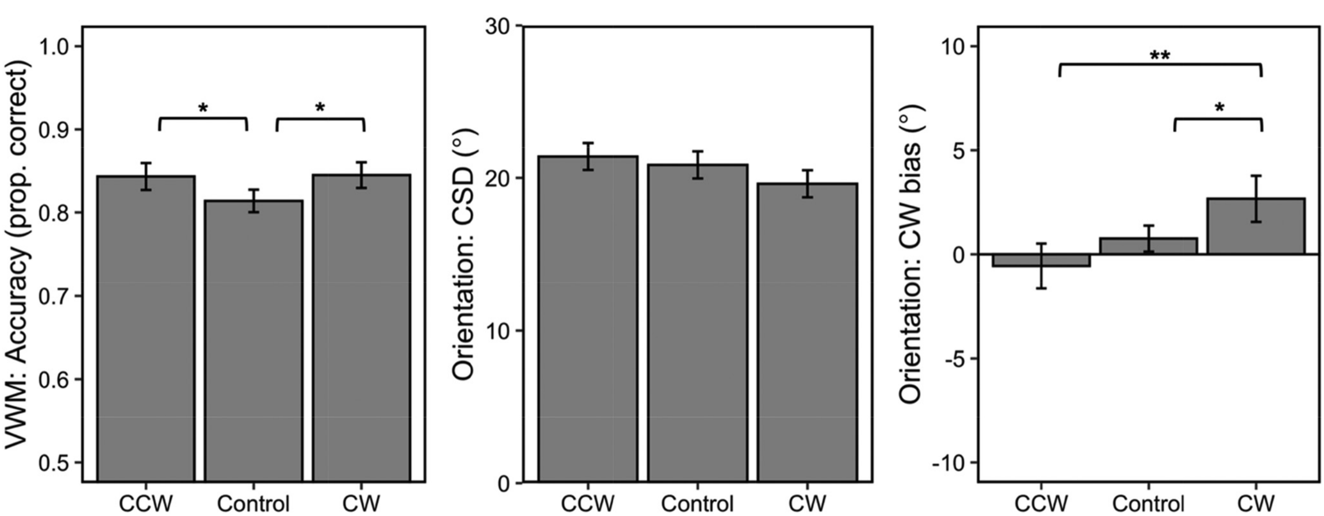
Means for all three display conditions for each of our dependent measures are depicted in Figure 5. Performance on the VWM task varied by display type, $F(2, 48) = 3.38$, $p = .042$, $\eta_p^2 < .12$. In line with the results of Experiments 1 and 3, accuracy was lower in the control condition relative to the CCW, $t(24) = 2.25$, $p = .034$, $d_z = .45$, and CW conditions, $t(24) = 2.48$, $p = .020$, $d_z = .50$, with the CCW and CW conditions not differing significantly from one another, $t(24) = .08$, $p = .934$, $d_z = .02$.

With regard to judgments of mean orientation, while the main effect of display type was not significant for the CSD parameter, $F(2, 48) = 2.74$, $p = .075$, $\eta_p^2 = .10$, a significant main effect was observed for the bias parameter, $F(2, 48) = 7.32$, $p = .002$, $\eta_p^2 = .234$. Post hoc investigation revealed that, relative to the control condition, estimation errors were biased in a clockwise direction for the CW condition, $t(24) = 2.60$, $p = .016$, $d_z = .52$. While the difference between the CCW and control conditions did not reach significance, $t(24) = 1.89$, $p = .070$, $d_z = .38$, there was a significant difference between the CCW and CW conditions, $t(24) = 3.03$, $p = .006$, $d_z = 1.02$, with mean orientation errors being more clockwise in the CW condition. Taken together, these results are quite similar to those reported in Experiments 1 and 3, despite the marginally significant difference between the CCW and control conditions. Importantly, the fact that mean orientation errors continued to exhibit a bias related to the maintenance of a color value in VWM when the probability of a color change was rare further rules out the possibility that our results are due to a strategic attempt to improve accuracy in the VWM task (in line with the results from Experiment 3). Thus, the most likely cause of our results is an automatic bias to select visual information matching the contents of VWM, which is then subsequently overvalued in the integration process.

General Discussion

In the current study, we examined whether VWM exerts an obligatory influence over perceptual averaging when a subset of items possesses a memory-matching feature. To do so, we had individuals perform mean orientation judgments while concurrently maintaining a colored shape in VWM, which could match the color of a subset of items in the orientation displays. Supporting our hypothesis, in Experiment 1, we found global estimations of mean orientation to be attracted toward the local subset mean of items possessing a color common to the VWM item, despite color being an irrelevant, secondary feature in the orientation task. Further, we were able to rule out the possibility that this finding was due to more bottom-up driven processes (namely perceptual priming) since the effect was eliminated in Experiment 2 where individuals saw the color of one subset prior to the presentation of the ensemble display but did not need to store it in VWM. Moreover, our findings cannot be reduced to a strategic attempt to improve VWM task

Figure 5
Results of Experiment 4



Note. $N = 25$. Dependent measures include accuracy for the visual working memory (VWM) task (left), as well as circular standard deviation (CSD; center) and bias (right) parameters of the error distributions for estimations of mean orientation. Error bars represent 95% confidence intervals for within-subject designs (Cousineau, 2005). CCW = counterclockwise condition; CW = clockwise condition.
* $p < .05$. ** $p < .01$.

performance since, in Experiment 3, we showed the biasing effect of VWM to be insensitive to display duration, whereas one would expect the effect to increase with processing time if individuals were strategically (rather than incidentally) devoting more attention to color-matching elements. Further, even when the probability of a color change was low, and thus attending color should not have been prioritized for the VWM task, we continued to observe the biasing effect in Experiment 4. Taken together, across four experiments, these results clearly demonstrate that VWM works to automatically prioritize memory-matching elements in perceptual averaging.

As we have noted, much of the previous works concerned with the interaction between perceptual averaging and VWM have approached the matter with respect to understanding the influence of perceptual averaging over VWM, with such works converging on the conclusion that perceptual averaging complements VWM by reducing capacity limitations and minimizing the effects of forgetting (Alvarez, 2011; Brady & Alvarez, 2011; Corbett, 2017; Corbin & Crawford, 2018; Dubé et al., 2014; Griffiths et al., 2018; Lew & Vul, 2015; Lowe et al., 2018; Son et al., 2020; Utochkin & Brady, 2020). Again, while much less work has considered the inverse relationship—namely, the effect of VWM on perceptual averaging—evidence of template-based filtering (i.e., Brand et al., 2012) suggests that VWM can similarly facilitate perceptual averaging. We confirm this filtering role of VWM in that we demonstrate that the maintenance of a visual feature in VWM leads to a significant bias in perceptual average judgments toward memory-matching items. More importantly, though, we highlight that this influence of VWM on perceptual averaging can occur incidentally and thus has an obstructive effect on the extraction of summary statistics, particularly when memory-matching items and nonmatching items are of equal relevance. The question, then, is how does the active maintenance of a visual feature from one dimension (i.e., color) interfere with the perceptual averaging of an entirely different feature dimension (i.e., orientation)?

At the conceptual level, we draw on the amplification hypothesis, which states that physically salient items are more heavily weighted than less salient items in the determination of ensemble summary statistics (Kanaya et al., 2018), to account for the memory-matching item bias. Again, this hypothesis takes the position that ensemble coding is achieved through the integration of a subset of items (e.g., Dakin, 2001; Im & Halberda, 2013; Whitney & Yamanashi Leib, 2018) and argues that physically salient items have a higher probability of being included within this subset than less salient items. While we did not systematically vary physical salience in the current study, the match between the color of the item held in VWM and the color of the oriented bars encountered in the orientation displays effectively had the same consequence. That is, because attentional selection is biased toward elements that bear perceptual similarity to information actively represented in VWM (e.g., Bundesen, 1990; Desimone & Duncan, 1995; Carlisle & Woodman, 2011; Hollingworth et al., 2013; Kumar et al., 2009; Olivers et al., 2006; Soto et al., 2005; Sun et al., 2015; Wolfe, 1994), we can confidently assume that the allocation of attentional resources was greater for the feature-matching items than the nonmatching items. It is this unequal distribution of attention following memory-guided selection that we attribute to the bias toward memory-matching items.

Where we diverge from the amplification hypothesis, however, has to do with pinpointing the exact mechanism by which the unequal distribution of attention gives rise to the memory-matching bias. In large part, this is because the amplification hypothesis is somewhat ambiguous with respect to whether the sample used to derive perceptual averages is comprised entirely of the most salient items or just proportionally more of these items. In this regard, much like Iakovlev and Utochkin (2020), when we compared the observed bias to that which would be expected had individuals only sampled memory-matching items, we found the observed bias to be significantly smaller than the expected bias, indicating that nonmatching items were included in the mean orientation judgments. As such, we ultimately arrive at the same conclusion as Iakovlev and Utochkin (2020) in that we come to three possible routes by which memory-driven selection may bias perceptual averaging. First, in line with the amplification hypothesis and partial sampling theories of perceptual averaging more generally (e.g., Allik et al. 2013; Dakin, 2001; Myczek & Simons, 2008; Whitney & Yamanashi Leib, 2018), it is possible that through memory-guided selection, memory-matching items more freely gain access to a privileged sample of items to which perceptual average calculations are based. For example, to achieve the bias of $\sim 3°$ observed in Experiments 1 and 3 (where color was most relevant to the VWM task), from a partial sampling perspective, this would imply that memory-matching items were sampled over nonmatching items at a rate of 3 to 2. Alternatively, the observed bias may instead be accounted for by an exhaustive sampling account of perceptual averaging (e.g., Ariely, 2001; Chong & Treisman, 2005b; Chong et al., 2008). That is, it is possible that all items were included in perceptual average calculations but differed in their weighting, as determined by an attentional scaling parameter. Last, a hybrid model that takes into account both partial and exhaustive sampling may also explain the current findings. For example, an exhaustive sampling mechanism may gather summary information about the group as a whole (independent of the focus of attention), while a partial sampling mechanism collects information about the items with the greatest priority (i.e., memory-matching items). These two mechanisms could then jointly contribute to perceptual average judgments, with the magnitude of the bias reflecting the weighting given to the two sampling mechanisms.

It is worth noting, however, that all three of these proposed accounts stand at odds with a recent distributed attention model (Baek & Chong, 2020a), which suggests that the role of attention in perceptual averaging is better conceptualized as a zoom lens (i.e., all items receive equal level of attention) rather than as a spotlight (i.e., some items receive more attentional weighting than others). From this perspective, there should be no bias toward memory-matching items as all items should equally contribute to the mean. To account for this discrepancy, we argue that Baek and Chong (2020a) mischaracterized the spotlight model of attention because they treat the selection of individual items as a random process. This ignores Kanaya et al.'s (2018) proposal that nonrandom sampling underlies perceptual averaging but, more broadly, is inconsistent with what is known about the guidance of attention. Namely, the allocation of focal attention is believed to follow an underlying priority map (e.g., Awh et al., 2012; Fecteau & Munoz, 2006; Stemmann & Freiwald, 2019; Wolfe, 2007), which orients attention in accordance with one's top-down goals (Bacon & Egeth, 1994; Folk et al., 1992; Soto et al., 2005), the physical salience of items (Corbetta & Shulman, 2002; Theeuwes, 1992; Yantis, 1993), learned value associations (Anderson, 2016; Anderson et al., 2011; Theeuwes & Belopolsky, 2012), and the emotional valence or potential threat of stimuli (e.g., Eastwood et al., 2001; Hansen & Hansen, 1988; Koster et al., 2004; Schmidt et al., 2015). It is this underlying priority map that we believe sets the weighting of individual items within perceptual averaging by selectively guiding attention to items that receive the greatest levels of activation. Indeed, while our results can only speak directly to the influence of memory-guided selection on perceptual averaging, as mentioned previously, all of these described factors have been shown to bias perceptual averaging (e.g., Chen & Zhou, 2018; Choi & Chong, 2020; de Fockert & Marchant, 2008; Dodgson & Raymond, 2020; Goldenberg et al., 2020; Iakovlev & Utochkin, 2020; Kanaya et al., 2018).

Still, while our results stand at odds with the distributed attention model of perceptual averaging in its pure form, Baek and Chong (2020b) did recognize the existence of a focal attention mechanism that they argued serves object recognition independent of the distributed attention mechanism. This distinction is particularly intriguing when considered alongside a hybrid account of attentional amplification in that it may serve to bridge the disparity between our findings and the predictions of the distributed attention model. That is, it is possible that these two mechanisms do in fact extract information independent of one another but either become integrated over time or, at the response stage, jointly contribute to perceptual average judgments. Offering some support to this idea, it is important to note that when Dodgson and Raymond (2020) presented stimulus masks immediately following the presentation of mean size displays (Experiment 2), the bias toward items matching a previously rewarded color was eliminated. As such, masking may have worked to selectively disrupt the representation of the focally attended item, leaving accessible only the distributed representation at the time of response. We believe this to be an interesting possibility that should be pursued further by future works.

It should additionally be acknowledged that while both our results and conclusions align strongly with those of Iakovlev and Utochkin (2020), there was one marked point of difference. Specifically, in addition to showing that mean orientation errors are biased toward salient items, Iakovlev and Utochkin (2020) also found the variance of such errors to be sensitive to the variance of salient items. That is, when the orientations of salient stimuli spanned a narrow range, mean orientation reports were less variable relative to when the orientations of salient stimuli spanned a wide range. From this perspective, it might have been expected that, in the present study, mean orientation judgments would have been less variable in the memory-matching display conditions compared to the control display conditions. Yet we found only one contrast where CSD was reduced for memory-matching displays relative to the control displays (see Experiment 1). We suggest that this difference has to do with the range of orientations used across the two studies as the means of our colored subsets were oriented $\pm 15°$ from the global mean, whereas the stimuli used by Iakovlev and Utochkin (2020) could range from $-30°$ to $30°$ relative to the global mean. As such, had we used a wider range of orientation values, our task may have been better able to detect such stimulus-based effects on response variance.

Furthermore, while the current work clearly demonstrates an influence of memory-driven selection on perceptual averaging, it is worth noting that, for our experimental design, several items matched the VWM item in the orientation displays, whereas in the traditional dual-task procedure used to examine memory-driven selection, only a single, salient distractor typically matches the VWM item color (e.g., Kumar et al., 2009; Olivers et al., 2006; Soto et al., 2005; Sun et al., 2015). This is a particularly noteworthy distinction when it is considered that studies examining the effect of visual outliers (i.e., singletons) on perceptual averaging tend to find that such outliers are filtered out or strongly devalued (e.g., Cant & Xu, 2020; Epstein et al., 2020; Haberman & Whitney, 2010). This might suggest that the bias we observed may depend on whether the attended items are perceived as belonging to the group since half of the items in the orientation displays matched the color of the VWM item in the current study, rather than a single, unique item. A worthwhile future direction would thus be to explore how the results of the current study are affected by the proportion of matching elements within the orientation displays. That is, if only a single item matches the color of the VWM item, would one continue to observe a bias of memory-guided selection, or would this effect be corrected by an outlier filtering mechanism?

Conclusion

Across four experiments, we show that the active maintenance of a feature value in VWM exerts an obligatory influence over perceptual averaging such that, if a portion of items possess the maintained feature value, perceptual average judgments are systematically attracted toward the mean of these items. We attribute the source of this bias to an underlying priority map, which works to guide attention toward memory-matching items while also taking into account a host of other factors (including additional top-down goals of the observer and the physical salience, learned value associations and emotional valence of the stimuli). How such attentional allocation ultimately leads to the memory-matching bias, however, remains an open topic of investigation as partial sampling models, exhaustive sampling models, and hybrid sampling models all present plausible routes to attentional amplification. It should thus be the aim of future works to definitively test these three models against one another.

References

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39. https://doi.org/10.1016/j.visres.2013.02.018

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

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7345–7350. https://doi.org/10.1073/pnas.0808981106

Anderson, B. A. (2016). The attention habit: How reward learning shapes attentional selection. *Annals of the New York Academy of Sciences*, *1369*(1), 24–39. https://doi.org/10.1111/nyas.12957

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the*

United States of America, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. https://doi.org/10.3758/bf03205306

Baek, J., & Chong, S. C. (2020a). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*, *82*(1), 63–79. https://doi.org/10.3758/s13414-019-01827-z

Baek, J., & Chong, S. C. (2020b). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic Bulletin & Review*, *27*(4), 602–606. https://doi.org/10.3758/s13423-020-01718-7

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), Article7. https://doi.org/10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(6), 1–15. https://doi.org/10.1167/11.10.6

Belopolsky, A. V., Schreij, D., & Theeuwes, J. (2010). What is top-down about contingent capture? *Attention, Perception, & Psychophysics*, *72*(2), 326–341. https://doi.org/10.3758/APP.72.2.326

Berggren, N., & Eimer, M. (2018). Visual working memory load disrupts template-guided attentional selection during visual search. *Journal of Cognitive Neuroscience*, *30*(12), 1902–1915. https://doi.org/10.1162/jocn_a_01324

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. https://doi.org/10.1177/0956797610397956

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

Brand, J., Oriet, C., & Tottenham, L. S. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(1), 63–69. https://doi.org/10.1037/a0026950

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547. https://doi.org/10.1037/0033-295x.97.4.523

Cant, J. S., & Xu, Y. (2020). One bad apple spoils the whole bushel: The neural basis of outlier processing. *NeuroImage*, *211*, Article 116629. https://doi.org/10.1016/j.neuroimage.2020.116629

Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta Psychologica*, *137*(2), 217–225. https://doi.org/10.1016/j.actpsy.2010.06.012

Chen, B., & Zhou, G. (2018). Attentional modulation of hierarchical ensemble coding for the identities of moving faces. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(10), 1542–1556. https://doi.org/10.1037/xhp0000549

Choi, Y. M., & Chong, S. C. (2020). Effects of selective attention on mean-size computation: Weighted averaging and perceptual enlargement. *Psychological Science*, *31*(10), 1261–1271. https://doi.org/10.1177/0956797620943834

Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327–1334. https://doi.org/10.3758/pp.70.7.1327

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/s0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/bf03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

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. https://doi.org/10.1177/0956797616671524

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Corbin, J. C., & Crawford, L. E. (2018). Biased by the group: Memory for an emotional expression biases towards the ensemble. *Collabra: Psychology*, *4*(1), Article 33. https://doi.org/10.1525/collabra.186

Cohen, E. H., & Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research*, *47*(22), 2825–2840. https://doi.org/10.1016/j.visres.2007.06.021

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A: Optics, Image Science & Vision*, *18*(5), 1016–1026. https://doi.org/10.1364/josaa.18.001016

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/s0042-6989(97)00133-8

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

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

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

Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research*, *96*, 8 –16. https://doi.org/10.1016/j.visres.2013.12.016

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & Psychophysics*, *63*(6), 1004–1013. https://doi.org/10.3758/bf03194519

Epstein, M., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. (2020). The outlier paradox: The role of iterative ensemble coding in discounting outliers *Journal of Experimental Psychology: Human Perception and Performance*, *46*(11), 1267–1279. https://doi.org/10.1037/xhp0000857

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Goldenberg, A., Weisz, E., Sweeny, T., Cikara, M., & Gross, J. (2020). *The crowd emotion amplification effect*. https://doi.org/10.31219/osf.io/cn6qy

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. https://doi.org/10.1037/xhp0000446

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

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), 751–753. https://doi.org/10.1016/j.cub.2007.06.039

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

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825–1838. https://doi.org/10.3758/APP.72.7.1825

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*(6), 917–924. https://doi.org/10.1037/0022-3514.54.6.917

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological Science*, *24*(5), 790–796. https://doi.org/10.1177/0956797612459767

Honig, M., Ma, W. J., & Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8391–8397. https://doi.org/10.1073/pnas.1918143117

Iakovlev, A. U., & Utochkin, I. S. (2020). Roles of saliency and set size in ensemble averaging. *Attention, Perception, & Psychophysics*. Advance online publication. https://doi.org/10.3758/s13414-020-02089-w

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*(2), 278–286. https://doi.org/10.3758/s13414-012-0399-4

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

Kang, M. S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, *18*(5), 860–869. https://doi.org/10.3758/s13423-011-0126-5

Koster, E. H., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does imminent threat capture and hold attention? *Emotion*, *4*(3), 312–317. https://doi.org/10.1037/1528-3542.4.3.312

Kumar, S., Soto, D., & Humphreys, G. W. (2009). Electrophysiological evidence for attentional guidance by the contents of working memory. *European Journal of Neuroscience*, *30*(2), 307–317. https://doi.org/10.1111/j.1460-9568.2009.06805.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

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), 10–14. https://doi.org/10.1167/15.4.10

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*, Article 1332. https://doi.org/10.3389/fpsyg.2016.01332

Lowe, M. X., Stevenson, R. A., Barense, M. D., Cant, J. S., & Ferber, S. (2018). Relating the perception of visual ensemble statistics to individual levels of autistic traits. *Attention, Perception, & Psychophysics*, *80*(7), 1667–1674. https://doi.org/10.3758/s13414-018-1580-1

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. https://doi.org/10.3758/bf03209251

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788. https://doi.org/10.3758/pp.70.5.772

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Sama, M. A., Nestor, A., & Cant, J. S. (2019). Independence of viewpoint and identity in face ensemble processing. *Journal of Vision*, *19*(5), 2–17. https://doi.org/10.1167/19.5.2

Sama, M. A., Srikanthan, D., Nestor, A., & Cant, J. S. (in press). Global and local interference effects in ensemble encoding are best explained by interactions between summary representations of the mean and range. *Attention, Perception, & Psychophysics*.

Schmidt, L. J., Belopolsky, A. V., & Theeuwes, J. (2015). Attentional capture by signals of threat. *Cognition and Emotion*, *29*(4), 687–694. https://doi.org/10.1080/02699931.2014.924484

Scocchia, L., Cicchini, G. M., & Triesch, J. (2013). What's "up"? Working memory contents can bias orientation processing. *Vision Research*, *78*, 46–55. https://doi.org/10.1016/j.visres.2012.12.003

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*, 46–59. https://doi.org/10.1037/xlm0000722

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248–261. https://doi.org/10.1037/0096-1523.31.2.248

Stemmann, H., & Freiwald, W. A. (2019). Evidence for an attentional priority map in inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(47), 23797–23805. https://doi.org/10.1073/pnas.1821866116

Sun, S. Z., Shen, J., Shaw, M., Cant, J. S., & Ferber, S. (2015). Automatic capture of attention by conceptually generated working memory templates. *Attention, Perception, & Psychophysics*, *77*(6), 1841–1847. https://doi.org/10.3758/s13414-015-0918-1

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*(8), 827–836. https://doi.org/10.1038/s41562-019-0640-4

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. https://doi.org/10.3758/bf03211656

Theeuwes, J., & Belopolsky, A. V. (2012). Reward grabs the eye: Oculomotor capture by rewarding stimuli. *Vision Research*, *74*, 80–85. https://doi.org/10.1016/j.visres.2012.07.024

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. https://doi.org/10.1037/xhp0000727

Webster, J., Kay, P., & Webster, M. A. (2014). Perceiving the average hue of color arrays. *Journal of the Optical Society of America A: Optics, Image*

Science & Vision, *31*(4), A283–A292. https://doi.org/10.1364/JOSAA.31.00A283

Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, *18*(3), 484–489. https://doi.org/10.3758/s13423-011-0071-3

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

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189193.003.0008

Woodman, G. F., & Arita, J. T. (2011). Direct electrophysiological measurement of attentional templates in visual working memory. *Psychological Science*, *22*(2), 212–215. https://doi.org/10.1177/0956797610395395

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science*, *2*(5), 156–161. https://doi.org/10.1111/1467-8721.ep10768973

Received September 1, 2020
Revision received December 17, 2020
Accepted January 16, 2021 ■