Attention, Perception, & Psychophysics (2024) 86:776–798
https://doi.org/10.3758/s13414-024-02850-5

What is the basis of ensemble subset selection?

Vladislav A. Khvostov[1,2] · Aleksei U. Iakovlev[1] · Jeremy M. Wolfe[3,4] · Igor S. Utochkin[5]

Accepted: 23 January 2024 / Published online: 13 February 2024

Abstract

The visual system can rapidly calculate the ensemble statistics of a set of objects; for example, people can easily estimate an average size of apples on a tree. To accomplish this, it is not always useful to summarize all the visual information. If there are various types of objects, the visual system should select a relevant subset: only apples, not leaves and branches. Here, we ask what kind of visual information makes a "good" ensemble that can be selectively attended to provide an accurate summary estimate. We tested three candidate representations: basic features, preattentive object files, and full-fledged bound objects. In four experiments, we presented a target and several distractors' sets of differently colored objects. We found that conditions where a target ensemble had at least one unique color (basic feature) provided ensemble averaging performance comparable to the baseline displays without distractors. When the target subset was defined as a conjunction of two colors or color-shape partly shared with distractors (so that they could be differentiated only as preattentive object files), subset averaging was also possible but less accurate than in the baseline and feature conditions. Finally, performance was very poor when the target subset was defined by an exact feature relationship, such as in the spatial conjunction of two colors (spatially bound object). Overall, these results suggest that distinguishable features and, to a lesser degree, preattentive object files can serve as the representational basis of ensemble selection, while bound objects cannot.

Keywords Ensemble summary statistics · Ensemble selection · Subset selection · Preattentive object file · Binding problem

Introduction

Visual input can deliver information about hundreds of different objects at any moment, a number that far exceeds the processing capacity of attention and working memory (Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988). However, while the visual system cannot fully process all its input, neither does it ignore that input. Some processing occurs everywhere in the visual field. Selective processing is much more limited (Wolfe et al., 2011). What is the nature of this non-selective processing that allows the visual system to process incoming information beyond the capacity limits on selective processing? One popular idea builds on the redundancy of the visual input, i.e., objects' characteristics usually do not vary randomly but form groups of similar items, like leaves on a tree or regions of similar features, like a textured carpet. These regularities make it possible for the visual system to deal with structured groups of similar objects, so-called "ensembles" (Alvarez, 2011; Cohen et al., 2016). This may avoid problems that would arise if hundreds of separate objects or locations needed to be processed individually.

Statement of Significance People can rapidly summarize multiple objects to form ensemble statistics; for example, assessing the average size of apples on a tree. To accomplish this, the visual system cannot just summarize information about all objects – it should select a relevant subset: apples but not leaves. We asked what kind of information makes a "good" ensemble that can be selectively attended and can provide an accurate summary estimate. We showed that people can calculate an ensemble summary for a subset defined by a basic feature, a feature conjunction (though, less accurately), but not by an exact feature relationship like "green to the left of red".

✉ Vladislav A. Khvostov
vkhvostov@hi.is; khvostov.vladislav@gmail.com

[1] Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland

[2] HSE University, Moscow, Russia

[3] Visual Attention Laboratory, Brigham and Women's Hospital, Boston, MA, USA

[4] Harvard Medical School, Boston, MA, USA

[5] Institute for Mind and Biology, University of Chicago, Chicago, IL, USA

What does it mean to say that the visual system can treat multiple objects as a whole and represent them in terms of their ensemble summary statistics (Ariely, 2001; Chong & Treisman, 2003, 2005a, b)? For instance, it has been shown that people can roughly and rapidly estimate the number of objects (Burr & Ross, 2008; Chong & Evans, 2011; Halberda et al., 2006). They can also calculate some summary statistics over features of a large set of items (Whitney & Yamanashi Leib, 2018). Those statistics include the central tendency of a feature across a group of objects, i.e., mean (Ariely, 2001; Chong & Treisman, 2003, 2005a) and measures of variability in an objects' feature, i.e., variance or range (Haberman et al., 2015; Khvostov & Utochkin, 2019; Morgan et al., 2008; Solomon et al., 2011; Suárez-Pinilla et al., 2018). These ensemble summaries can act as perceptual stimuli in their own right, for example, producing adaptation aftereffects (Burr & Ross, 2008; Corbett et al., 2012; Norman et al., 2015). Ensemble statistics are generated after very brief exposure to the stimulus (as fast as 50–200 ms) (Chong & Treisman, 2003; Whiting & Oriet, 2011). Moreover, ensemble statistics can be computed even if observers have limited or no conscious access to individual objects (Alvarez & Oliva, 2008; Ariely, 2001; Corbett & Oriet, 2011; Parkes et al., 2001). These computations are not very demanding of attentional resources (Alvarez & Oliva, 2008; Bauer, 2009; Epstein & Emmanouil, 2017; but see Jackson-Nielsen et al., 2017).

Many visual dimensions can be compressed into a statistical summary; basic features, certainly: size (Ariely, 2001; Chong & Treisman, 2003, 2005a, b), orientation (Alvarez & Oliva, 2009; Dakin & Watt, 1997), color (Gardelle & Summerfield, 2011; Maule & Franklin, 2015), speed of motion (Emmanouil & Treisman, 2008; Watamaniuk & Duchon, 1992), but also, perhaps, higher-order properties like emotional expressions (Haberman & Whitney, 2007), as well as others (Florey et al., 2016; Sweeny & Whitney, 2014). Recent studies show that these calculations generate more than just a single number. The visual system is capable of representing the whole distribution of objects' features (Chetverikov et al., 2016, 2017a, b; Kim & Chong, 2020). The ability to extract the mean and variance of an ensemble would be of little use in the real world unless that ensemble could be a subset of all the stimuli in the field. For instance, imagine that you are watching a soccer game and want to know which team is taller to estimate the relative chances to score during a corner. In this case, it would be useless to estimate a simple average height over all players. You would need to split the players into groups and calculate this summary separately for each subset. Our interest in this paper is in the properties that can be used to create subsets for ensemble calculations.

It is known that it is possible to extract summary statistics from a subset of items even when they are intermixed with numerous items from other subsets (Chong & Treisman, 2005b; Drew et al., 2010; Emmanouil & Treisman, 2008; Halberda et al., 2006; Im & Chong, 2014; Poltoratski & Xu, 2013; Sun et al., 2016, 2018; Utochkin & Vostrikov, 2017). Thus, it is possible to assess the average size of the blue circles in a display that mixes circles of multiple colors (Chong & Triesman, 2005b). Chong and Treisman (2005a) suggested that subset selection can be based on features like those that can guide attention in the visual search for a single object (e.g., color, size, orientation; Wolfe & Horowitz, 2017). We even know that the subset does not need to be completely uniform to be selected. Research shows that if one group of items forms a peak in a basic feature distribution, distinct from the remaining items, the visual system can rather easily segment these two ensembles from one another (e.g., the first group can consist of heterogeneously reddish objects, while the second group is greenish). In contrast, if the distribution of features over all objects has only one smooth peak, subset selection becomes a much harder task. For example, it would be hard to estimate the properties of reddish and orangish objects separate from yellowish and greenish objects that were part of the same, continuous color distribution (Im et al., 2021; Khvostov et al., 2021; Utochkin, 2015; Utochkin et al., 2018; Utochkin & Yurevich, 2016).

However, less is known about how various objects intermixed over space are categorized as being members of one or another ensemble. Note that almost all studies investigating ensemble subset selection have used basic features (mostly vivid, highly discriminable, uniform colors) to define the different subsets. There is no doubt that simple and salient basic features can efficiently guide the global selection for ensemble processing. But can the visual system use anything more complex than a single feature for the ensemble selection? The main question of this paper is the following: *what is the representational basis of ensemble subset selection?* What kind of distinguishing information about various kinds of objects makes a "good" ensemble that can be selectively attended and not confused with other ensembles?

To our knowledge, such questions have not been settled in the ensemble literature, but the question of how deeply the visual system processes objects before the involvement of focused attention has been well addressed in parallel literature on visual attention. This question gained much popularity since the visual search paradigm was introduced (Treisman & Gelade, 1980). In search tasks, observers are typically presented with an array of objects and are asked to determine whether this array contains a target (an odd-one-out object or an object with predefined features) or not (only distractors are present in a display). Based on this literature, one can come up with three "candidates" for the representational basis of a subset selection stage in ensemble perception. We order these from the "shallowest" characteristic to the "deepest" one. This study aimed to evaluate these three candidates.

The first candidate is a "*basic feature.*" As mentioned above, there is a lot of evidence that observers can selectively attend to all objects having a common, distinctive feature value. For example, observers easily select all blue objects and compute their mean size independently from the mean size of green objects (Chong & Treisman, 2005b). Influential theories such as Feature Integration Theory (Treisman & Gelade, 1980) and the Boolean map theory of attention (Huang & Pashler, 2007) claim that this is the only information that is available for global preattentive segmentation of a display (that is, before focused attention deploys locally to individual items). The indirect support of the "feature-based" view also comes from the field of feature-based attention (Maunsell & Treue, 2006). For example, it was shown that paying attention to a stimulus feature facilitates the processing of other stimuli sharing the same feature (Saenz et al., 2002, 2003). The "feature-based" view of ensemble selection predicts that *fast and accurate ensemble selection is possible only if the target group of objects has a feature value distinct from all distractors.*

The second candidate for being the representational basis of ensemble selection is the so-called *preattentive object file*. This term assumes that the visual system can extract more elaborate information during the pre-attentive stage than simply the locations of all feature values (Wolfe & Bennett, 1997). It means that before selective attention comes to a place and binds all object features to an "object file" (Kahneman & Treisman, 1984), the visual system creates preattentive object files: shapeless bundles of basic features, which represent a "list" of all basic features the objects contain without knowing how exactly these features are combined. Guided Search theory (Wolfe, 2021) can be seen as endorsing a version of this idea. Guidance to two or more basic features could in principle create a group of items with all the correct features. A preattentive object file hypothesis of ensemble selection would predict: *as long as the target subset has a distinct preattentive object file (i.e., the list of features), the fast ensemble selection is possible.*

The third candidate is the full-fledged *bound object*. This possibility would mean that the visual system works with bound representations even before focused attention arrives. This is a version of a classic "late selection" position (Deutsch & Deutsch, 1963; Norman, 1968). In this view, the ensemble section could be based on object identity (e.g., "rabbit" or "red-blue conjunction") even if those sets were not defined by a unique feature or a unique preattentive object file. The prediction based on the bound-object account would be*: fast subset selection is possible as long as the items in the set have a distinct identity.*

Our study

In our study, we tested these predictions about the basis of ensemble subset selection by varying the nature of the similarity between targets and distractors. In Experiments 1 and 2, observers performed an orientation averaging task where a target set of objects (bicolor triangles of different orientations) were spatially intermixed with distracting sets of objects and shared with them none, one, or two colors (the more colors are shared between targets and distractors – the greater their similarity – the deeper level of representation is needed to select the target subset). We tested the feature candidate in conditions where the target set shared no colors with distractors (the target set had a unique color). The "preattentive object file" as a candidate was tested in a condition where the target set shared one color with distractors (the target set had a unique conjunction of colors). The bound-object hypothesis was tested in a condition where the target set shared both colors with distractors but had an opposite spatial combination (the target set had a unique spatial conjunction of colors – e.g., target: red on the left, blue on the right; distractors: blue on the left, red on the right). We compared the ability to extract the average of the target group in each of these conditions to a baseline condition where only the target subset was presented. Based on the results, we can conclude that, as previously shown, fast ensemble selection can be readily based on a unique basic feature. Weaker, but reliable ensemble selection is possible for items defined by a unique combination of two features (our preattentive object files). The subset selection based on spatial relations between otherwise identical combinations of features proved to be impossible. In Experiment 3, we confirm our conclusions about the "preattentive object file" candidate testing the condition where a target subset was defined by a conjunction of two dimensions. In Experiment 4 using similar conditions, we show that these conclusions can be generalized from orientation averaging to size averaging in an adjustment task.

Experiment 1

Method

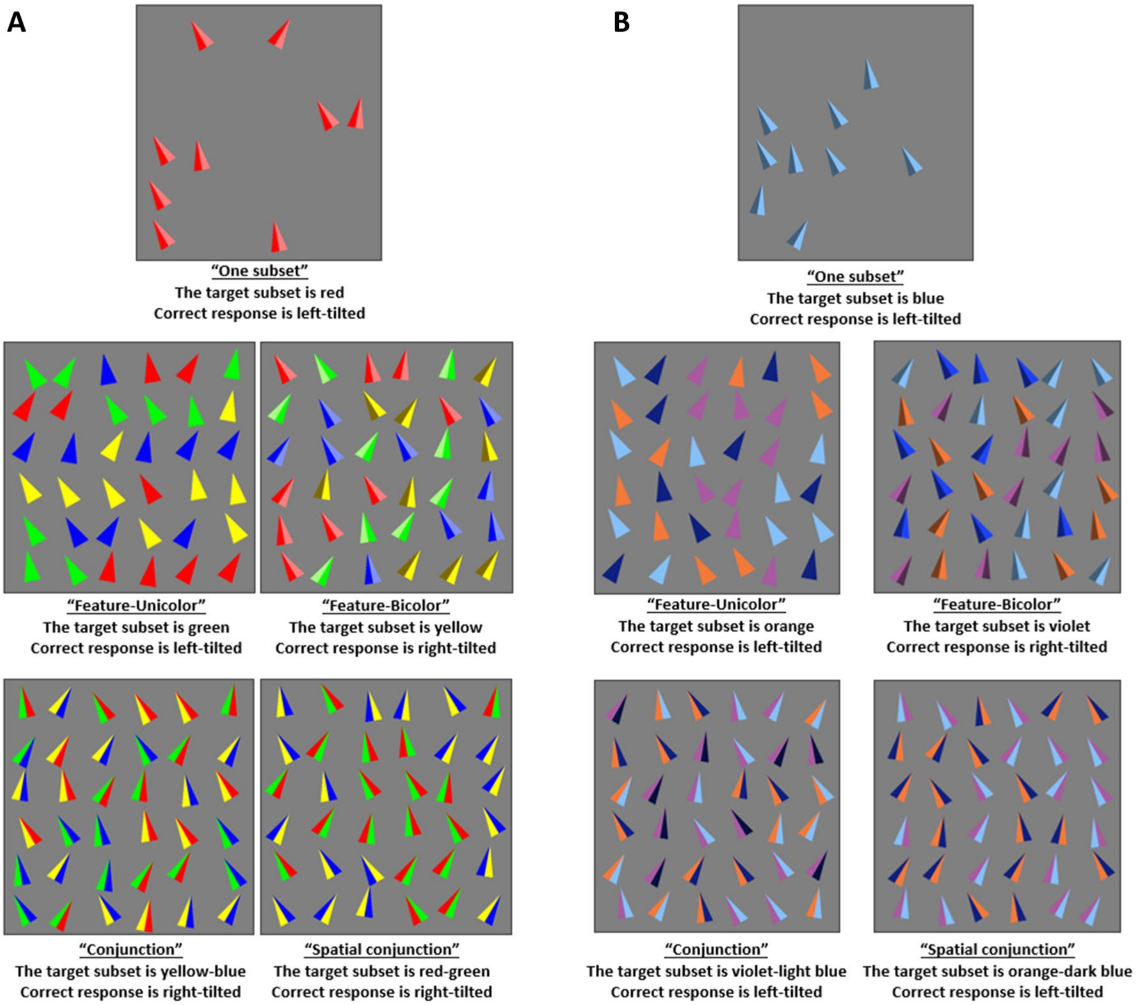
Participants and power analysis

We used G-power software 3.0.10 (Faul et al., 2007) to determine a sample size for this experiment such that a medium effect size η^2 = .06 with α = .05 and power (1-β) = .8 could be found, if present, using a one-way repeated-measures ANOVA with five conditions. This yielded an estimated sample size of 26 participants. Considering possible technical problems and poor performance in some observers, we planned to recruit a sample of 30 observers.

Thirty undergraduate students at the HSE University (Moscow, Russia) participated in Experiment 1 for extra course credits (27 females and three males, mean age = 19 years, SD = 1.47 years). All participants reported having normal or corrected-to-normal vision without color-vision deficiency and provided informed consent electronically. The results of two participants were excluded from the analysis because they committed more than 35% errors. This exclusion criterion, as well as procedures and data analysis (for all experiments in this paper) were preregistered at https://osf.io/2qmca/registrations (see Open Practices Statement). All procedures performed in a study were in accordance with the Declaration of Helsinki.

Stimuli

Stimuli, illustrated in Fig. 1A, were developed using PsychoPy 3 software (Peirce et al., 2019). The experiment was run online using Pavlovia (https://pavlovia.org) on participants' personal computers. A 720 px × 720 px square at the center of the screen was used as the "working" field for presenting stimuli; the remaining screen space remained grey. This square was divided into 6 × 6 = 36 cells by an imaginary grid (each cell side was 120 pixels). Each cell was used as the location for a triangle. Within the cell, the position of a triangle was randomly jittered within a 15-pixel range in both horizontal and vertical directions.



"One subset"
The target subset is red
Correct response is left-tilted

"Feature-Unicolor"
The target subset is green
Correct response is left-tilted

"Feature-Bicolor"
The target subset is yellow
Correct response is right-tilted

"Conjunction"
The target subset is yellow-blue
Correct response is right-tilted

"Spatial conjunction"
The target subset is red-green
Correct response is right-tilted

"One subset"
The target subset is blue
Correct response is left-tilted

"Feature-Unicolor"
The target subset is orange
Correct response is left-tilted

"Feature-Bicolor"
The target subset is violet
Correct response is right-tilted

"Conjunction"
The target subset is violet-light blue
Correct response is left-tilted

"Spatial conjunction"
The target subset is orange-dark blue
Correct response is left-tilted

Fig. 1 (**A**) The display examples of five experimental conditions in Experiment 1. (**B**) The colorblind-friendly version of Fig. 1A (for illustrative purposes only). Correspondence between four original colors and their substitutions on this picture: red – violet, yellow – orange, green - light blue, blue – dark blue. Note that participants in the actual experiment were screened for colorblindness

The stimuli were bicolor or unicolor isosceles triangles (width – 50 pixels, height – 94 pixels) of different orientations. In each trial, we presented four subsets of nine triangles in each (36 objects in total). All subsets had different colors/combinations of colors and contained four different orientations of triangles (-30˚, -10˚, 10˚, and 30˚) in different proportions. Two sets contained five 30˚ triangles, two 10˚ triangles, one a -10˚ triangle, and one a -30˚ triangle (right-tilted average orientation: 14.4˚). The other two sets contained the opposite proportion of orientations: one 30˚ triangle, one 10˚ triangle, two -10˚ triangles, and five -30˚ triangles (left-tilted average orientation: -14.4˚). The subsets were spatially intermixed. Within each block, triangles from one subset were targets and three other subsets were distractors (all consistent throughout the block). In different blocks, a target subset was defined by a different unique attribute (see Fig. 1A).

(1) Feature – Unicolor. Here, subsets were defined by a uniform color. For example, a target subset consisted of red triangles, and distractors subsets consisted of green, blue, and yellow triangle; so that the target subset shared no colors with distractors.

(2) Feature – Bicolor. This condition was similar to condition (1) in the sense that target subsets were defined by features not shared with any of the distractors. However, whereas each triangle in condition (1) was colored uniformly, triangles in the present condition were bicolor, which was accomplished by dividing each triangle into two halves vertically. For example, target triangles could be dark-red and light-red, whereas distractors could be dark-green and light-green, dark-blue and light-blue, light-yellow, and dark-yellow. This condition was introduced to have a version of a feature condition but with the same spatial complexity as in the Conjunction and the Spatial conjunction conditions (which are described below). In addition, the comparison between the Feature-Unicolor and the Feature-Bicolor conditions lets us control for the use of "bicolorness" as a potential cue for orientation averaging (e.g., because bicolor triangles have a vertical boundary between two colored halves that unicolor triangles do not have).

(3) Conjunction of two colors. Here, a target subset shared each of its colors with one of the distractor subsets, so that no color was unique to the target subset. For example, a target subset consisted of red-green triangles whereas distractor subsets consisted of yellow-GREEN, RED-blue, and yellow-blue.

(4) Spatial conjunction of color. Here, the target definition includes the relative position of colors within an object. For example, a target subset could be red-green whereas distractor subsets could be green-red, yellow-blue, and blue-yellow. Here, targets have the same set of two colors as one subset of distractors but in a different spatial arrangement (mirror-reversed). Importantly, the target subset and its mirror-reverse counterpart always had opposite average tilts. For example, if the red-green target subset were on average tilted to the right, then the green-red distractor subset was on average tilted to the left. This made it impossible to judge the average orientation of the target subset based on picking any items just having the relevant colors. Instead, it required distinguishing between the two spatial arrangements of these colors.

(5) Targets alone. This condition was used as a baseline to measure the accuracy of orientation averaging when observers do not need to filter out distractors so that all errors come only from the averaging itself. In this condition, only a target subset of triangles and no distractors were presented. The presented triangles were colored the same way as in the Feature-Bicolor condition.

Procedure, design, and data analysis

The experiment consisted of five blocks, each dedicated to a single stimulus condition described in the previous section. The color or color combination of a target subset was set randomly for each block and remained consistent during the entire block. Each trial started with the presentation of a fixation point for 500 ms (see Fig. 2) followed by a

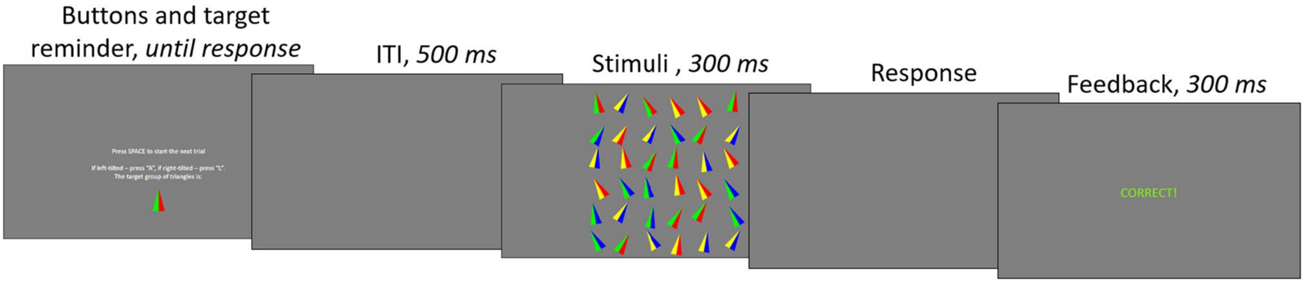

Buttons and target reminder, *until response*
Press SPACE to start the next trial
If left-tilted – press "A", if right-tilted – press "L".
The target group of triangles is:
ITI, *500 ms*
Stimuli , *300 ms*
Response
Feedback, *300 ms*
CORRECT!

Fig. 2 The time course of a typical trial in Experiment 1

brief presentation of a sample set of triangles for 300 ms. Observers were asked to report whether the target subset had a left-tilted or right-tilted average orientation (two-alternative forced-choice; 2AFC) by pressing an "L" or an "A" button on a keyboard, respectively. After the button press, the observers received feedback (300 ms) informing them whether the response had been correct or not. The feedback screen was followed by an intertrial interval when observers were shown a vertically oriented triangle from the target subset as a reminder. The next trial started upon the pressing of the spacebar so that participants could progress at a comfortable pace and rest whenever they wanted.

At the beginning of each block, the participants performed a practice session consisting of 26 trials for familiarization with the task and stimuli in the block. The practice session was immediately followed by an experimental session. Each experimental session consisted of 120 trials (600 trials in total).

The design of Experiment 1 was within-subject. Each participant completed all five conditions/blocks (Targets alone, Feature-Unicolor, Feature-Bicolor, Conjunction, Spatial conjunction) in random order. The dependent variable was the percent of correct responses within each condition.

Results

As shown in Fig. 3, there was a clear effect of the way the target and distractor subsets are defined on the ability to estimate the average orientation of the target subset. Repeated-measures ANOVA showed a strong main effect of the experimental condition on the proportion of correct responses (F[2.6, 70.11] = 133.98, $p < .001$, $\eta^2 = .73$; Greenhouse–Geisser correction was applied to the degrees of freedom). Pairwise post hoc comparisons showed equal performance in the Feature-Unicolor and the Feature-Bicolor conditions ($t(27) = 0.5$, $p = .62$, Bonferroni-corrected $\alpha = .005$, Cohen's $d = 0.09$). All other conditions differed significantly from each other: the Targets alone yielded the best accuracy (M = 91% correct); the Feature-Unicolor and the Feature-Bicolor were slightly less accurate (M = 84% correct); then the Conjunction condition followed (M = 74% correct); and, finally the Spatial conjunction caused the worst performance (M = 56% correct); pairwise ts(27) > 5.68, ps < .001, Bonferroni-corrected $\alpha = .005$, Cohen's ds > 1.07. Note that observers performed above chance (50% correct) in all conditions, even in the Spatial conjunction one (ts(27) > 3.34, ps < .003, Bonferroni-corrected $\alpha = .01$, Cohen's ds > .63).

Discussion

First, we should note that the absence of a difference between the Feature-Unicolor and the Feature-Bicolor conditions showed that the "bicolorness" of our stimuli and the number of colors in a display (four or eight) do not influence the accuracy of determining the mean orientation. Therefore,

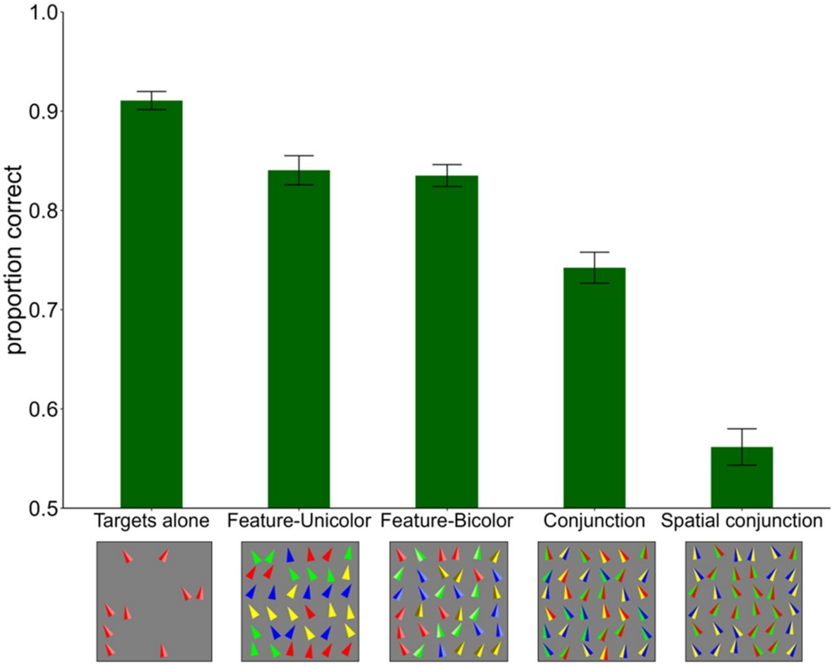


Fig. 3 The proportion of correct responses in different conditions of Experiment 1. Error bars denote the SEM, with between-subject variance removed following Cousineau's (2005) method. The lower point on y-axis represents the "random guess" performance level (0.5)

we show that even relatively complex, two-part objects can be quite successfully selected for ensemble processing if they have unique features. The absence of the difference in these conditions allows us to rule out some of the low-level accounts such as display heterogeneity that might be used to explain the observed reduction in performance in the conjunction and spatial conjunction conditions (e.g., see Lleras et al., 2019; Wang et al., 2017).

Secondly, our results showed that observers could select the target objects to some degree in almost all conditions except for the Spatial conjunction condition, as their performance was well above chance in those conditions. The baseline condition (Targets alone) represents performance under ideal circumstances – no possible confusion between targets and distractors (also, potential crowding effects would be smaller due to the reduced number of objects in the Targets alone display). Performance in this condition was very high (91% correct) but still not perfect, suggesting that subset selection was not the only source of errors in this task. Some portion of the errors could come from lapses (e.g., accidental mixing up of the response keys) or imperfect estimates of the ensemble mean orientation. Since other conditions differ from the baseline by the need to select a subset and they differ from each other only by the number of features shared between targets and distractors, we can explain the differences in performance by the difficulty of selection.

The performance in the two conditions where the target subset was defined by a unique feature (Feature-Unicolor and Feature-Bicolor) was rather high (~ 84%: only 7% lower than in the baseline). Thus, we conclude that, though the presence of differently colored distractors interferes a bit with the calculation of mean orientation in the target subset, still, observers filtered out most of the distractors and selected the target subset successfully. This implies that objects defined by unique features, not shared with other objects, can be used as a basis for ensemble selection.

In contrast with these "feature-based" conditions, performance in the Spatial conjunction condition (56% correct) was much worse than the baseline (although slightly better than chance), which suggests that it is much harder or practically impossible to select a target subset based on the spatial conjunction of two colors. A clever observer might beat chance by

1) Picking one item
2) Deciding if it is a target
3) If it is a target, guess that orientation is the ensemble orientation (correct seven out of nine times)
4) If it is not a target, guess the opposite.

While possible, this requires a lot of work in 300 ms. Some such strategy might, however, explain the modest, above chance behavior. We will address this "subsampling" strategy more carefully in Experiment 2. For now, we conclude that it is unlikely that ensemble selection was used as the basis for performance in the Spatial Conjunction condition.

At 74% correct, performance in the Conjunction condition fell in-between the relatively easy, feature-based conditions and the very hard Spatial conjunction condition. We cannot unequivocally decide between an ensemble account and a sampling account for performance in this condition. It could be that observers used a sampling strategy similar to that we proposed for the Spatial conjunction condition, but, in the easier Conjunction condition, they were able to better sample a proper item more often. Alternatively, observers might be selecting a large target subset, but that selection might be imperfect. Either some distractors were erroneously included in the subset, or a smaller number of target items were selected (which would decrease the accuracy of the calculation of the mean). We address these two alternatives in Experiment 2. For now, we can conclude that it is possible that a preattentive object file can be used to some degree to select relevant items for computing subset ensemble statistics.

Experiment 2

A recurring question in the ensemble literature asks if ensemble representations are built on global processing of all items in the relevant subset or if performance can be based on a smaller sample from that set. There are multiple claims that the level of accuracy, seen in the data, can be achieved if observers efficiently sampled only a small handful of random items (e.g., Allik et al., 2013; Gorea et al., 2014; Myczek & Simons, 2008; Solomon, 2010). This could allow ensemble summary statistics to be computed using mechanisms whose capacity would not exceed those of focused attention and/or working memory. Although the most extreme versions of this idea (e.g., that the average feature is computed exclusively from one to three samples with no contribution from other items) have been shown to be wrong (Chong et al., 2008; Utochkin & Tiurina, 2014), the debate about the capacity of ensemble processing still lives on; for example, perhaps the sample size grows with the square root of set size (Whitney & Yamanashi Leib, 2018) or perhaps different weights are placed on attended and unattended sets of items (Iakovlev & Utochkin, 2021; Kanaya et al., 2018).

Extreme versions of sampling theories should be considered as offering an alternative explanation for the results in Experiment 1. Instead of selecting multiple target objects and averaging them, observers could just find a random target item and respond based on its orientation (performing a version of a visual search task where they look for at least one out of many possible targets). This strategy can provide a reasonable performance level (at least above chance). In Experiment 2, we compare subset selection to sampling by analyzing the performance in displays where we showed the whole target distribution of orientations and in displays where the target subset contained nine copies of a single orientation. The single orientation was randomly sampled from the whole distribution. If observers just sample one target object, average performance should not differ between these two conditions. If observers select the whole target subset or at least some of its members greater than one, then they should benefit from presenting the whole, heterogeneous distribution and the performance in the condition with the whole distribution should be higher.

Method

Participants and power analysis

We planned to recruit the same number of participants (N = 26) as in Experiment 1. We calculated the power for this experiment based on the pilot study (N = 7) with the same experimental design (the data can be found at https://osf.io/2qmca/) as the main Experiment 2. We evaluated the expected mean differences, SD, and other parameters for the effect of interest (the interaction between factors). Based on these parameters, our power analysis (GLIMMPSE program: https://glimmpse.samplesizeshop.org/) asserts that 26 observers give us the power (1-β) > .99 to detect an interaction similar to the results of the pilot experiment (with α = .05). Given concerns about possible technical problems, poor performance in some observers, and some variations between the pilot and the main experiment, we increased the planned sample size to 31 observers.

Thirty one observers (23 females and eight males, mean age = 23.83 years, SD = 7.27 years) were recruited via Prolific (www.prolific.ac; Palan & Schitter, 2018; Peer et al., 2017) and run using Pavlovia (https://pavlovia.org). They gave informed consent and were paid £5 per hour (the experiment lasted approximately 40 min). All participants reported normal or corrected-to-normal vision without color-vision deficiency. The results of three participants were excluded from the analysis because they committed more than 35% errors.

Stimuli

Software and stimuli were the same as in Experiment 1 in terms of their number, coloring, orientation distributions, and layout. The critical changes concerned presenting a target subset. We presented either the whole target distribution (the "Whole-distribution" condition), as described in Experiment 1, or nine copies of one random value, drawn from this distribution of orientations (the "One-value" condition). Thus, displays in the Whole-distribution condition were identical to those in Experiment 1. Displays in the One-value condition were identical to the Whole-distribution condition in terms of distractors, but all nine triangles of the target distribution had the same orientational value sampled from the whole distribution (see Fig. 4). As in Experiment 1, the whole orientation distribution included four different orientations (-30°, -10°, 10°, and 30°) in different proportions. The sampling of a random orientation from that distribution to make nine copies for the One-value condition followed these proportions across trials. For example, if the whole distribution was left-tilted, as described in Experiment 1, then the fractions of trials of a corresponding One-value condition would be as follows: in 5/9 trials the target triangles will be -30°, and in 2/9 trials, all triangles will be -10°, in 1/9 trials, all triangles will be 10°, and in 1/9 trials, all triangles will be 30° (see Fig. 4B). If observers were sampling only one random target triangle and gave the response based only on it, we expect no difference between the Whole-distribution and the One-value conditions, because the latter condition provides the same amount of information (averaged over all trials) as the sampling one random target in the former one. But if observers select more than one target triangle and average them, the performance should be better for the Whole-distribution condition. This prediction is based on the fact that the Whole-distribution always contains items predominantly tilted to the real average direction, whereas 2/9 trials of the One-value condition are completely misleading because the sampled value has an opposite tilt compared to the implied average of the distribution from which this value has been sampled. The use of these 2/9 trials is required to model performance as if it were based on selecting one random target item – in 2/9 cases, the observer would select the wrong item without knowing that it is incorrect (see Fig. 4B).

As in Experiment 1, we manipulated the attributes that defined target and distractor subsets. We had the same list of conditions except that we dropped the Feature-Unicolor condition, as it was shown to provide the same performance as its bicolor analog. Therefore, we had the following conditions for this experiment: Feature-Bicolor (a target subset was defined by a unique color), Conjunction (by a unique conjunction of two colors), Spatial conjunction (by a unique conjunction of position × color), and Targets alone (baseline condition when one a target subset is present).

Procedure, design, and data analysis

The procedure was the same as in Experiment 1. Each participant completed four blocks (Feature-Bicolor, Conjunction, Spatial conjunction, and Targets alone) in random

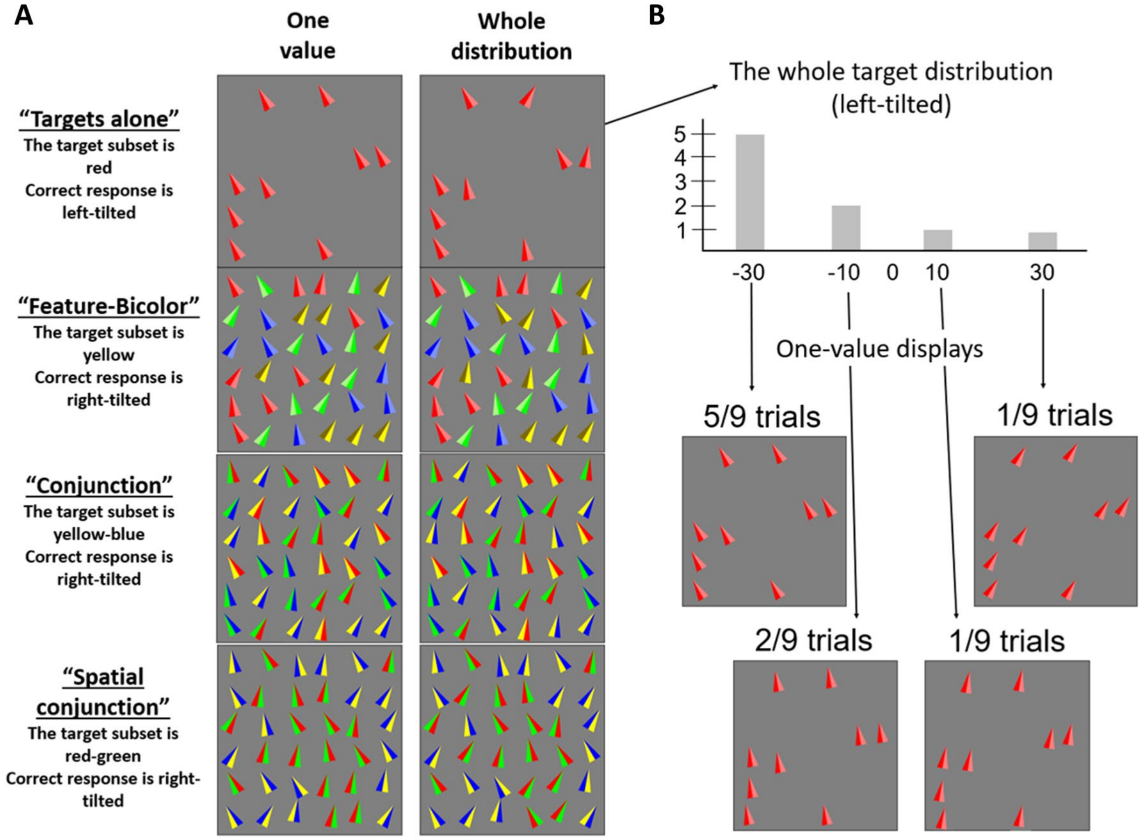

A
One value
Whole distribution
"Targets alone"
The target subset is red
Correct response is left-tilted
"Feature-Bicolor"
The target subset is yellow
Correct response is right-tilted
"Conjunction"
The target subset is yellow-blue
Correct response is right-tilted
"Spatial conjunction"
The target subset is red-green
Correct response is right-tilted
B
The whole target distribution (left-tilted)
5
4
3
2
1
-30
-10
0
10
30
One-value displays
5/9 trials
1/9 trials
2/9 trials
1/9 trials

Fig. 4 (**A**) The display examples of the Whole-distribution and the One-value displays for the Targets alone, Feature-Bicolor, Conjunction, and Spatial conjunction conditions in Experiment 2. (**B**) All possible types of the One-value displays for the Targets-alone condition. The picture depicts the left-tilted distribution case – the right-tilted distribution trials were constructed symmetrically

order. The Whole-distribution and the One-value trials of each condition were intermixed within each block. At the beginning of each block, the participants performed a practice session consisting of 36 trials for familiarization with the task and stimuli in the block. Each experimental session consisted of 216 trials (864 trials in total).

In Experiment 2, we used a 2 (Target distribution: Whole-distribution vs. One-value) × 4 (Target attributes: Feature-Bicolor, Conjunction, Spatial conjunction, and Targets alone) within-subject design. The feedback to the observer about the correctness of the response in a trial is based on the true average orientation of target triangles present on the screen. However, the primary measure for data analysis purposes was based on the percentage of the "implied" correct responses. The "implied" correct response is based on the average of the distribution from which individual orientations were sampled in each trial. The "implied" average always matches the screen-based average in the Whole-distribution trials, so there would be no mismatch between feedback to the observer and the implied correct response. However, in 2/9 of the One-value trials, there is a mismatch. In those trials, the nine triangles, displayed in the target subset, could be tilted left even though they were sampled from a right-tilted distribution. If an observer responds "left-tilted," that would be an incorrect "implied" response for data analysis purposes, but the observer would receive feedback informing them that this was a correct response (which, of course, it was, given what was on the screen). The One-value data show what performance would look like if observers were sampling only one random item out of nine targets. Based on the whole distribution, they should pick items with a "wrong" tilt in 2/9 cases.

Results

The main results of Experiment 2 are depicted in Fig. 5. We found a strong main effect of the target-defining attribute on accuracy relative to the implied average orientation ($F[2.33,62.86] = 200.61$, $p < .001$, $\eta^2_p = .881$; Greenhouse–Geisser correction was applied to the degrees of freedom). All pairwise differences were significant: Targets alone > Feature-Bicolor > Conjunction > Spatial conjunction (all ts(27) > 3.2, ps < .004, Bonferroni-corrected $\alpha = .008$, Cohen's ds > 0.606). The main effect of the distribution of orientations in the target subset was also found to be significant (F[1,27] = 66.16, p < .001, η^2_p = .71). Over all conditions, observers were more accurate in the Whole-distribution than in the One-value condition (t(27) = 8.13, p < .001, Cohen's d = 1.54). This is of modest interest. Of more importance is the difference between the One-value and the Whole-distribution displays for each of the subset-defining, target attributes. The difference varies with target attribute, as seen in the significant target attribute × distribution interaction (F[3,81] = 33.231, p < .001, η^2_p = .552). Looking at the individual conditions, in the Targets-alone condition, there was a clear advantage of the Whole-distribution condition over the One-value (t(27) = 9.69, p < .001, Cohen's d = 1.83). The same is true in the Feature-Bicolor condition (t(27) = 8.67, p < .001, Cohen's d = 1.63). A significant, if smaller benefit is still found in the Conjunction condition (t(27) = 2.78, p = .009, Cohen's d = 0.53). On the other hand, there is a small, but significant effect in the *opposite* direction for the Spatial conjunction condition. Performance with the One-value subset was better than with the Whole-distribution (t(27) = 3.68, p = .001, Bonferroni-corrected α = .012, Cohen's d = 0.7). In this experiment, observers performed above chance in all combination of conditions (ts(27) > 4.3, ps < .001, Bonferroni-corrected α = .01, Cohen's ds > .55), except for the Whole-distribution × the Spatial conjunction (t(27) = 0.82, p = .42, Bonferroni-corrected α = .006, Cohen's d = 0.15). That is, in the Spatial conjunction condition, observers actually did worse with the Whole-distribution than they could have done if they had selected one item and guessed on the basis of that item.

Discussion

As we would expect, we observed a substantial increase in performance within the Whole-distribution compared to the One-value condition for the Targets alone (10% difference) and the Feature-Bicolor (7%). This difference comes from the fact that the percent of the implied correct responses in the One-value condition could not raise higher than 78% (even for an ideal observer) because only in 7/9 (~78%) trials all nine target triangles had the "correct" orientation (e.g., the left-tilted orientation was sampled from the left-tilted distribution). In the remaining 2/9 trials, even an ideal observer could not give an implied correct response because all nine target triangles had a "wrong" tilt compared to the implied correct. Overall, it means that observers take advantage of availability the whole target distribution (rather than one random value) and use this additional information for a more accurate calculation of the target average orientation. We can conclude that in the Targets alone and the Feature-Bicolor, observers use more than one target object to make their decision about the subset orientation. This result is consistent with the calculation of ensemble statistics. Based on the similar results of the Targets-alone and the Feature-Bicolor condition, we can also confirm our conclusion from Experiment 1: a unique feature can be used as a basis of ensemble selection.

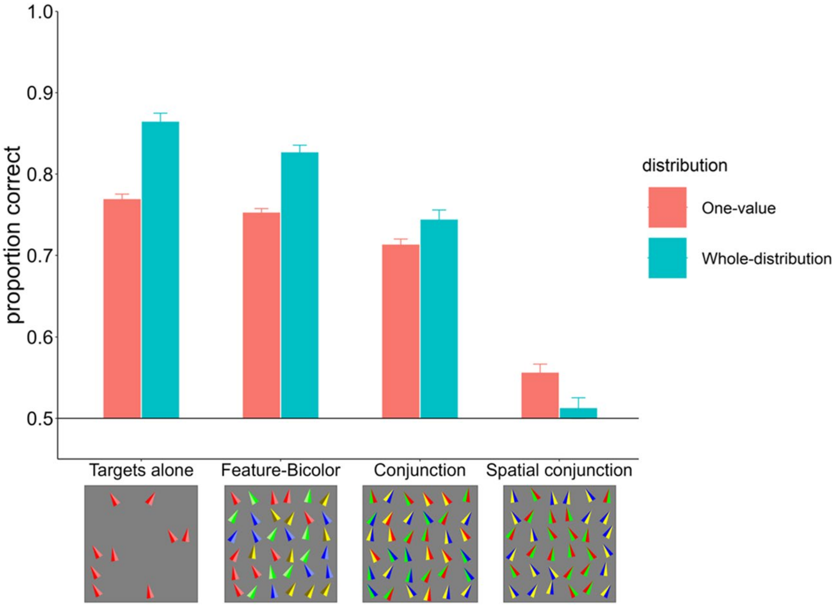


Fig. 5 The percentage of correct responses relative to the "implied" average orientation as a function of the target-defining attribute target distribution (Experiment 2). Note that correct responses in the One-value condition were calculated based on the whole target distribution, not based on what was presented on the screen. Error bars denote the SEM, with between-subject variance removed following Cousineau's (2005) method

The opposite results can be seen for the Spatial conjunction condition: the Whole-distribution condition produces worse performance than the One-value (4% disadvantage). We consider this to be strong evidence that observers in the Spatial conjunction condition did not base their decisions on a subset of even a few relevant items. Their very poor performance looks more like they based their decision on one item, and sometimes even that one item was not a member of the correct subset. Observers could not reliably distinguish between the target subset and the distractors having the same colors but with the opposite mean tilt (e.g., if the target subset of red-green was left-tilted on average, then the distractor subset of green-red would be right-tilted on average). For the Whole distribution condition, this means that selecting a random object with red and green colors gives them only a 50% chance to respond correctly because exactly half of the 18 objects with red and green colors were tilted to the correct direction (note that observers indeed performed at chance in this condition). The One-value displays within the Spatial conjunction condition produce better than chance performance. There are several possible paths to this modest performance. For example, there were 11 out of 18 objects with red and green colors tilted in the correct direction (all nine target objects + two distractors) in 7/9 of the trials (when the sampled orientation matched the correct distributional response). In this case, the strategy of sampling one random object with red and green colors should provide slightly better than chance (and the Whole-distribution) performance because the opposite situation (when all target objects tilted to the wrong direction) happened only in 2/9 trials. The straight-forward calculation would be ((7/9)*(11/18)+(2/9)*(7/18)) = 56%, exactly the result for the One-value version of our Spatial conjunction condition. The important conclusion is that there is no evidence that observers could select a subset, based on the spatial conjunction of colors – performance in both the Whole-distribution and One-value conditions can be explained by the strategy of selecting of one item with two correct colors.

The Conjunction condition provides results that can support intermediate conclusions. Clearly, the results are different from those for the Spatial conjunction condition. Like the Targets-alone and the Feature-Bicolor conditions, there is a significant advantage of the Whole-distribution over the One-value condition, although it is rather small (3%). These results show that observers do not sample just one target object for a response. They use some extra information.

In the discussion of Experiment 1, we put forward two different explanations of the Conjunction condition results. (1) Observers sample one random target object and base their responses solely on that orientation. (2) They select a larger target subset but not an ideal target subset. It could be a noisy subset, including non-target items. Alternatively, it could be a subset greater than one, but not much greater than one. The current data cannot distinguish between these hypotheses. Experiment 3 uses other preattentive object file stimuli in an effort to clarify this issue.

Experiment 3

The Conjunction condition of Experiments 1 and 2 might not have been ideal for subset selection and calculation of ensemble statistics because we used a color × color conjunction. Conjunctions of two values from the same dimension (color × color or orientation × orientation) are known to lead to inefficient visual search (Wolfe et al., 1990; but see Wolfe et al., 1994, for an exception of part-whole color relations) and it is reasonable to imagine that such conjunctions would not support easy subset selection. Perhaps a feature bundle or preattentive object file would support more effective subset formation and ensemble calculation if the conjunction was between two different feature dimensions. Buetti et al. (2019) has argued that, during visual search, many objects can be processed and compared along two feature dimensions (e.g., color and shape) in parallel in space and time. Thus, in Experiment 3, we tested color × shape conjunctions along with the relevant simple feature control conditions for our ensemble selection task. As in Experiment 2, we wanted to show directly that observers were indeed selecting a subset of target items to perform the task and not basing their response on just one random target item. Thus, we used the methods from Experiment 2 and compared the performance in a condition where observers were shown the whole target distribution of orientations with a condition where they were shown nine copies of one randomly sampled item from the whole distribution.

Method

Participants and power analysis

The power calculation was identical to Experiment 2. Thirty observers (nine females and 21 males; mean age = 28 years, SD = 9.1 years) were recruited via Prolific (www.prolific.ac; Palan & Schitter, 2018; Peer et al., 2017) and run using Pavlovia (https://pavlovia.org). They gave informed consent and were paid £5 per hour (the experiment lasted approximately 45 minutes). All participants reported normal or corrected-to-normal vision without color-vision deficiency. The results of one participant were excluded from the analysis because they committed more than 35% errors (preregistered exclusion criterion).

Stimuli

Software and stimuli were the same as in Experiment 2 in terms of their number, orientation distributions, and layout. As in Experiment 2, we used the same manipulation of the composition of the target distribution. We presented either the whole target distribution (the Whole-distribution condition) or nine copies of one random orientation value from this target distribution (the One-value condition). The critical changes in Experiment 3 involved stimulus identities. In different blocks, a target subset was defined by a different unique attribute (see Fig. 6).

(1) Shape. Subsets were defined by shape and designed to be highly discriminable. There were four different shapes in the experiment: a bar, an oval with three holes in it (looks like a pod), a mirrored S-shaped figure, and an oval with jagged edges (nicknamed “kiki”; Ramachandran & Hubbard, 2001). Each of these shapes fit in a rectangle 31 px × 93 px. All stimuli had the same color – blue. For example, a target subset could consist of blue bars among distractors subsets composed of blue pods, S-shaped figures, and kikis.

(2) Color. Subsets were defined by color. All stimuli had the same shape – a bar. For example, a target subset could consist of red bars among distractors subsets composed of green, blue, and yellow bars.

(3) Conjunction of two dimensions (Conjunction-Dimensions). Here, a target subset shared each of its features with one of the distractor subsets, so that no feature was unique to the target subset. We used only two forms (bars and kikis) and two colors (red and blue) for this condition. For example, a target subset could consist of red kikis among distractor subsets composed of blue kikis, red bars, and blue bars.

The experiment had two further control conditions.

(4) Conjunction of two colors (Conjunction-Colors). This condition was the same as the Conjunction condition from Experiments 1 and 2. We wanted to measure the performance in this condition for a new set of observers to directly compare it with the Conjunction-Dimension condition (3). All stimuli were bicolor triangles (width – 40 pixels, height – 94 pixels). A target subset shared each of its colors with one of the distractor subsets so that no color was unique to the target subset. For example, a target subset could consist of red-green triangles among distractor subsets composed of yellow-green, red-blue, and yellow-blue.

(5) Targets alone. As previously, this condition was used as a baseline to measure the accuracy of orientation averaging when observers did not need to filter out distractors. Any errors would come from the averaging itself. In this condition, only a target subset of objects and no distractors were presented. We used only stimuli from the Conjunction-Dimension condition (3): either red kikis, blue kikis, red bars, or blue bars.

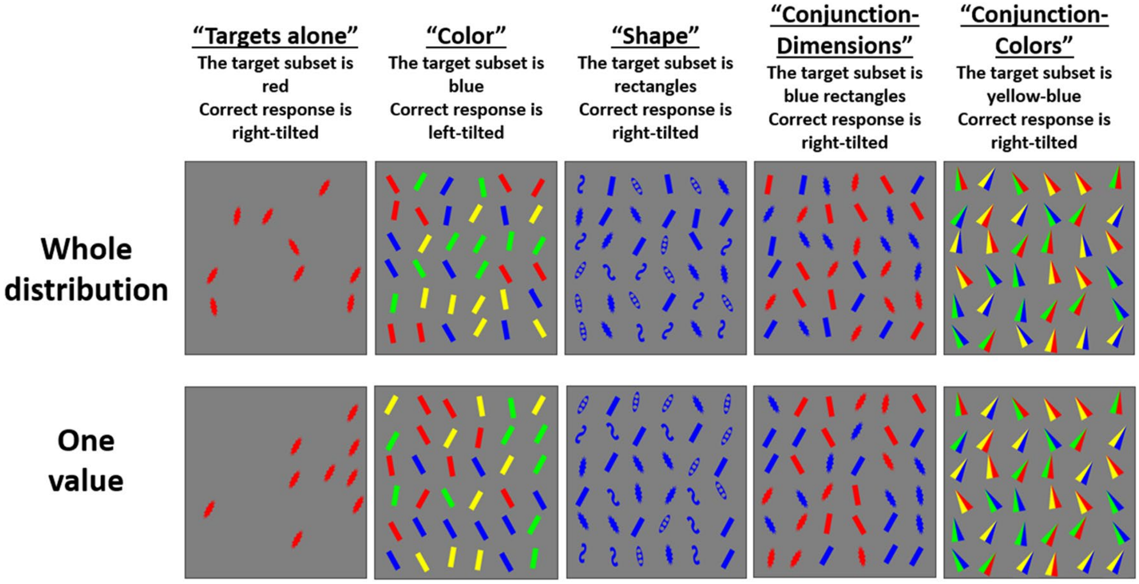

“Targets alone”
The target subset is red
Correct response is right-tilted
“Color”
The target subset is blue
Correct response is left-tilted
“Shape”
The target subset is rectangles
Correct response is right-tilted
“Conjunction-Dimensions”
The target subset is blue rectangles
Correct response is right-tilted
“Conjunction-Colors”
The target subset is yellow-blue
Correct response is right-tilted
Whole distribution
One value

Fig. 6 The display examples of the Whole-distribution and the One-value displays for the Targets alone, Color, Shape, Conjunction-Dimensions, and Conjunction-Colors conditions in Experiment 3

Procedure, design, and data analysis

The procedure was the same as in Experiment 2. Each participant completed five blocks (Color, Shape, Conjunction-Dimensions, Conjunction-Colors, and Targets alone) in random order. The Whole-distribution and the One-value trials of each condition were intermixed within each block. At the beginning of each block, the participants performed a practice session consisting of 36 trials for familiarization with the task and stimuli. Each experimental session consisted of 216 trials (1,080 trials in total).

In Experiment 3, we used a 2 (Target distribution: Whole-distribution vs. One-value) × 5 (Target attributes: Color, Shape, Conjunction-Dimensions, Conjunction-Colors, and Targets alone) within-subject design. As in Experiment 2, the feedback about the correctness of the response in a trial was given based on the real average orientation of target triangles presented on the screen. However, the primary measure was the percentage of the "implied" correct responses (the true average of a distribution from which individual orientations were sampled in each trial).

Results

We found a strong main effect of the distribution of orientations in the target subset ($F[1,28] = 85.48$, $p < .001$, $\eta^2_p = .759$). Over all conditions, observers were more accurate in the Whole-distribution than in the One-value condition ($t(28) = 9.25$, $p < .001$, Cohen's $d = 1.72$). The main effect of the target-defining attribute was also significant ($F[4,112] = 22.67$, $p < .001$, $\eta^2_p = .447$). The most important results concern the difference between the One-value and the Whole-distribution displays for each of the subset-defining, target attributes (see Fig. 7). The difference varies with target attribute, as seen in the significant interaction between target attribute and distribution factors ($F[4,112] = 14.849$, $p < .001$, $\eta^2_p = .347$). Looking at the individual conditions, in the Targets alone, Color, Shape, and Conjunction-Dimension conditions, there was a clear advantage of the Whole-distribution condition over the One-value ($ts(28) > 4.29$, $ps < .001$, Cohen's $ds > .79$). A small benefit is also present in the Conjunction-Colors condition but it does not survive after the correction for multiple comparisons ($t(28) = 2.26$, $p = .03$, Bonferroni-corrected $\alpha = .01$, Cohen's $d = 0.42$).

To compare these advantages between different conditions of the target-defining attribute, we calculated the difference between the percentage of correct responses in the Whole-distribution and the One-Value for each observer. Pairwise comparisons confirmed the impression of Fig. 7: this difference smoothly decreases while moving from conditions on the left part of the figure to the right. There is no clear boundary between big and small differences. Statistically, the Target Alone condition produced a larger difference between Whole-distribution and One-Value version than any other condition (all $ts(28) > 3.27$, $p < .004$, Cohen's $d > 0.6$) except for the Color condition ($t(28) = 0.858$, $p = .398$, Bonferroni-corrected $\alpha = .01$, Cohen's $d = 0.159$). The Color result was not statistically different from Shape ($t(28) = 2.26$, $p = .03$, Bonferroni-corrected $\alpha = .01$, Cohen's $d = 0.42$), but it was higher than both of the Conjunction conditions ($ts(28) > 5.09$, $p < .001$, Cohen's

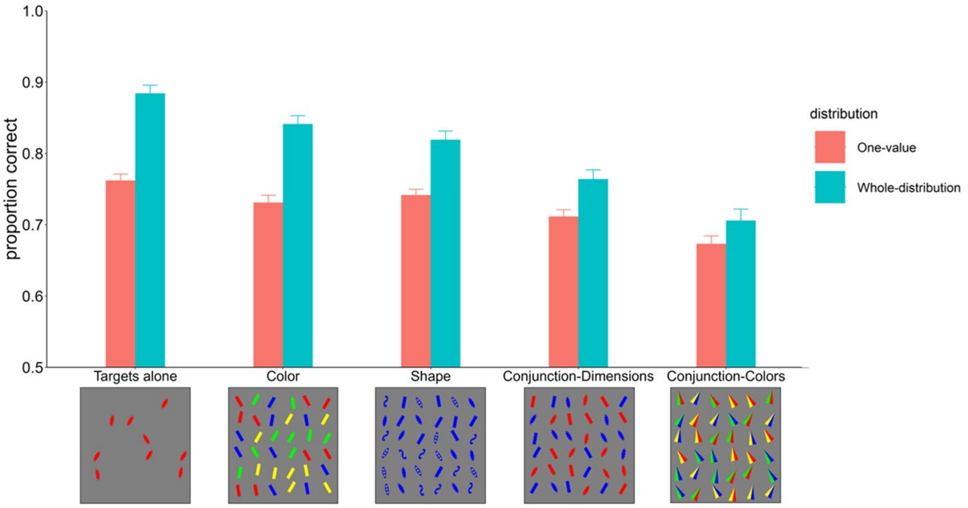


Fig. 7 The percentage of correct responses relative to the "implied" average orientation as a function of the target-defining attribute target distribution (Experiment 3). Note that correct responses in the One-value condition were calculated based on the whole target distribution, not based on what was presented on the screen. Error bars denote the SEM, with between-subject variance removed following Cousineau's (2005) method

$d > 0.946$). Shape was not statistically different from Conjunction-Dimensions ($t(28) = 2.18$, $p = .037$, Bonferroni-corrected $\alpha = .005$, Cohen's $d = 0.407$), but was greater than Conjunction-Colors ($t(28) = 3.07$, $p < .005$, Cohen's $d = 0.57$); the Conjunction-Dimensions was not statistically different than Conjunction-Colors ($t(27) = 1.35$, $p = .187$, Cohen's $d = 0.251$).

Discussion

As in the previous experiments, the baseline condition without distractors (Targets alone) provided the best overall performance and the greatest advantage of the availability of the whole target distribution (12%). The Color condition provided almost the same advantage (11%) as in the baseline, which replicates our result from previous experiments: observers can easily select a subset of targets based on a distinct color. Our new condition (Shape) which tested another basic feature also showed good results but a little worse than the Targets alone baseline (7% advantage). Overall, we can conclude that observers take advantage of the availability of more than one member of the target distribution in each trial of the Target alone, Color, and Shape conditions and use this additional information for a more accurate calculation of the target average orientation. This result is consistent with the creation of an ensemble representation. Overall, we can also confirm our conclusion from Experiments 1 and 2: a unique feature (both color and shape) can be used as a basis of ensemble selection.

Two other conditions (Conjunction-Colors and Conjunction-Dimensions) tested the hypothesis that preattentive object files can be the basis of ensemble selection. The Conjunction of two colors showed the same advantage as in Experiment 2 (3%), but due to greater variability this time, this difference did not reach significance. This increase in variability might be explained if some observers "gave up" in this condition because it was the hardest one in this experiment, in contrast to Experiment 2, where observers had to perform the even harder Spatial conjunction condition. Based on the results, we cannot conclude that observers used more than one target item defined by conjunction of two colors to perform the task in this experiment.

The conjunction of two dimensions (shape and color) was the condition of primary interest in this experiment. Based on the visual search literature, we expected this condition to provide more reliable results compared to the Conjunction of two colors. We see a clear, if modest advantage in presenting the whole target distribution (5%) in this condition. Though it is a bit smaller than in the baseline and the Color conditions, it is statistically indistinguishable from that of the Shape condition. Based on this, we can conclude that observers used more than one conjunctively defined object to judge an average orientation of a target subset. This means that a preattentive object file could be used as a basis of ensemble selection.

An important conclusion of Experiment 3 is that we cannot draw a strict boundary between cases where observers can select a subset of target items and where it is impossible. Unlike Experiment 2, where the results of the bound object condition (Spatial conjunction) showed the inability to select a target subset, there no large drop in the performance among conditions of Experiment 3. Both the proportion of correct responses and the advantage of the Whole-distribution over the One-value condition slowly decreases moving from left to right along the x-axis of Fig. 7: from the baseline through the feature condition to the conjunction conditions.

Experiment 4

Experiments 1–3 involve estimates of average subset orientation. To test the generalizability of the results, Experiment 4 uses estimation of the average size of items. Many ensemble perception studies have been done using size estimation (for a review see: Whitney & Yamanashi Leib, 2018). Experiment 4 uses the method of adjustment, instead of the 2AFC method of Experiments 1–3. This provides finer estimates of potential errors that may arise due to the noise in sampling and averaging. It is the error distribution that is usually analyzed when using the method of adjustment, rather than binary accuracy as used in Experiments 1–3. Our critical predictions in Experiment 4 are framed in terms of the mean shift of this error distribution relative to the actual mean size of a target subset. If observers can select the subset and extract a size estimate, the average estimated size should be close to the actual average of the subset. If there are any difficulties in selecting a target subset, the estimated mean size should be biased toward the mean size of the whole set (target subset and other distractors put together). In Experiment 4, subsets were defined as in Experiments 1 and 2. We compared subset size estimates against two baselines (also measured directly): ceiling performance, defined by size estimates on the subset in the absence of any distractors, and floor performance, defined by the estimated size of the whole display.

Method

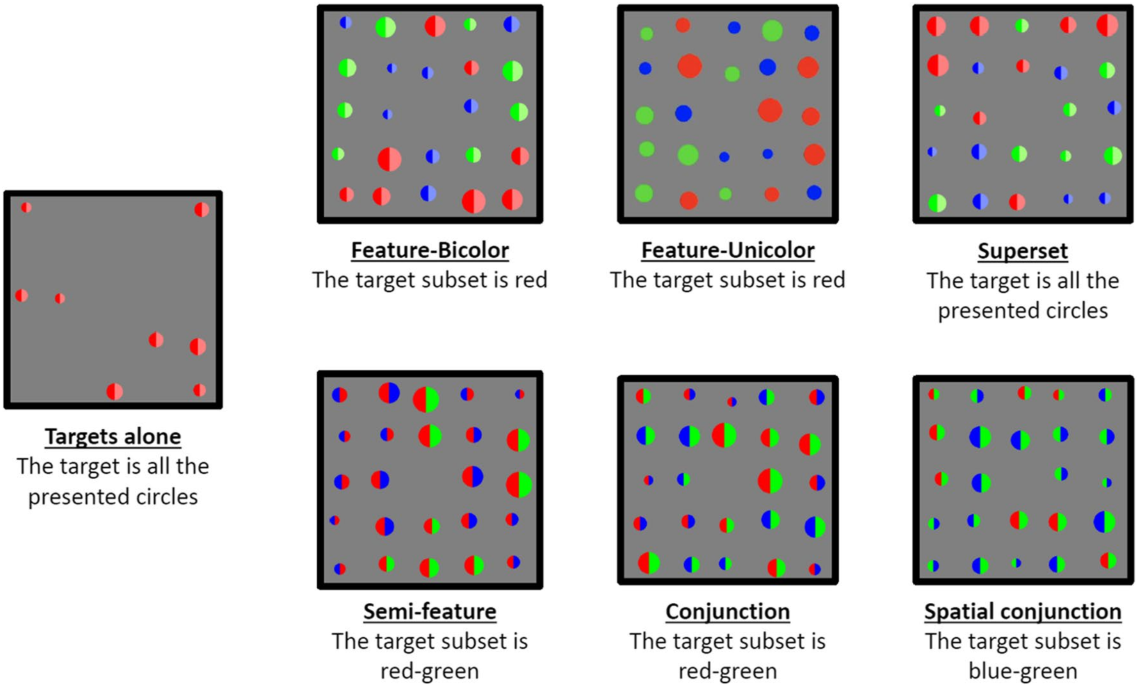
Participants

One hundred and seventeen undergraduate students at the HSE University (99 females and 18 males; mean age = 20.1 years, SD = 1.5 years) were recruited for participation in Experiment 4 for course credits. All participants reported having normal or corrected-to-normal vision and no neurological problems or color deficiency. At the beginning of the experiment, they gave informed consent online. The number of trials and participants (N = 100) were estimated using "power contour estimator" (https://shiny.york.ac.uk/powercontours/), which has been developed by Baker and colleagues (Baker et al., 2021) and pre-registered. Parameters for "power contour estimator" have been taken from two conditions with the smallest differences in a bias of a pilot experiment using a similar design but not including the full list of conditions (mean difference = 0.05, within-subject SD = 0.26, between-subject SD = 0.16). The data of 16 participants were excluded from the analysis because the accuracy of their responses or the reaction time did not meet our preregistered criteria of inclusion into the analysis (see "Data analysis and design" section for a description). Therefore, the data of 101 participants were analyzed.

Stimuli

The experiment was run online via the Pavlovia platform. Stimuli were presented on a gray background within the central part of a screen subtending 600 × 600 px; the rest of the screen was not used during the experiment. This central part of the screen was subdivided into an imaginary grid with 5 × 5 = 25 cells. Each cell side was 120 × 120 px. Each cell, except the central one, contained one item (a circle) from a sample set. Each item was placed in the cell center with a random jitter within 15 px along horizontal and vertical dimensions. The central cell of the grid never contained any circle and was only used to display a target item at the beginning of a trial and a response item after a sample set presentation.

A sample set in each trial consisted of 24 or eight circles depending on the condition. The circles had different sizes and colors. All circles were subdivided into three subsets based on their colors. One color subset was used as targets and two other subsets were used as distractors subsets (eight circles per subset). Each subset had its own mean size. The average size of one subset was larger than the grand average of all three subsets, the average size of a second subset was smaller than the grand average, and the average size of a third subset matched the grand average. The target subset could have either the largest or the smallest mean size. The grand average was randomly selected from a uniform distribution ranging from 40 to 50 px in diameter in each trial. To generate the mean sizes of subsets and sizes of individual circles in proportion to the initially chosen grand average, we used a psychophysical scale of perceived sizes where the perceived size is a power function of the physical area with an exponent of 0.76 (Teghtsoonian, 1965). The average size of the largest subset was randomly chosen from an interval of 120–140% of the grand average; the average size of the smallest subset was randomly chosen from an interval of 80–60% of the grand average. Sizes of individual circles were 64%, 88%, 112%, and 136% of the subset size average. These individual sizes were presented in two exemplars within each subset, thus yielding eight items with a uniform



Feature-Bicolor
The target subset is red

Feature-Unicolor
The target subset is red

Superset
The target is all the presented circles

Targets alone
The target is all the presented circles

Semi-feature
The target subset is red-green

Conjunction
The target subset is red-green

Spatial conjunction
The target subset is blue-green

Fig. 8 The display examples of seven experimental conditions in Experiment 4

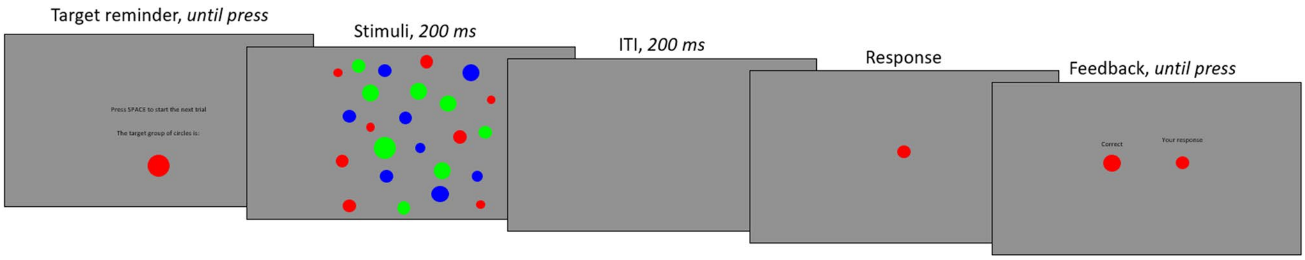

Target reminder, *until press*
Press SPACE to start the next trial
The target group of circles is:
Stimuli, *200 ms*
ITI, *200 ms*
Response
Feedback, *until press*
Correct
Your response

Fig. 9 The time course of a typical trial of Experiment 4. Each trial started with a precue of the target subset followed by a sample set for 200 ms. Participants were then asked to adjust the size of a probe circle to match the mean size of the target subset of circles. After response confirmation, feedback was shown

size distribution. The actual mean size was never presented as a subset member.

There were seven conditions each reflecting how target and distractor attributes were defined (Fig. 8). Four of them were the same as in Experiment 1: (1) Feature-Unicolor (uniformly colored circles, each subset is defined by a unique color); (2) Feature-Bicolor (each circle is divided by two halves colored in a light and dark versions of the same hue, hues are unique to subsets); (3) Conjunction of two colors (targets have two colors, each color is shared by either of the distractor subsets); and (4) Spatial conjunction of position and color (one of the distractor subset shares both colors with the target but in a different spatial arrangement). In condition (5) "Semi-feature", the target subset and both distractor subsets had one color in common, while another color was unique for the target subset. For example, the target set could be red-green circles while the distractors sets consisted of red-blue and blue-red circles. Thus, green would be a unique target feature. In this condition, we were able to test whether the presence of a unique feature supports feature-based selection when a common color potentially supports grouping between targets and distractors. The other two conditions were baselines. Condition (6), "Targets Alone," where only the target subset of bicolor circles was presented, served to measure the best possible averaging performance without distractors ("ceiling" performance). Condition (7), "Superset," was identical to the Feature-Bicolor in terms of stimuli used but the task was to adjust to the mean size of all circles regardless of their colors. It was used to measure "floor" performance, that is, how observers would average sizes if they were "subset-blind." It is important to measure performance in these two baseline conditions, rather than simply using the physical subset and superset mean sizes, because the baselines themselves can be biased. It has been previously shown that observers tend to give more weight to larger items in averaging tasks, which results in average estimates biased towards the larger items. This bias is referred to as an *amplification effect* (Iakovlev & Utochkin, 2021; Kanaya et al., 2018). Therefore, properly estimated baselines should take into account these amplification biases. In turn, any biases that we can potentially observe in the various subset selection conditions should be recalibrated given these biased baselines to draw correct conclusions about the selection processes.

Procedure

As in Experiments 1–3, each of the seven conditions described in the previous section was presented in a separate block of trials. The order of blocks was randomized across observers. The colors of a target subset were assigned randomly at the beginning of each block and kept consistent throughout the entire block. Each trial began with a reminder presentation of a single, 40 pix diameter circle from the target subset at the center of the screen (see Fig. 9). It was visible until participants pressed the spacebar. Next, the critical display of circles was presented for 200 ms, followed by a blank screen for 200 ms, and then an adjustable probe circle was presented at the screen center. The probe circle had the same color(s) as the target subset, or it was black in the Superset condition when all presented circles had to be averaged. The initial size of the probe was randomly chosen from the interval between 15 and 85 px in diameter. Participants could increase or decrease the size of the probe by holding the left mouse button and moving the mouse up or down. The participants were asked to adjust the test circle size to match the average size of all presented circles or to the size of the target subset, depending on the condition. When the answer was confirmed by pressing the spacebar, the participants were given feedback: two circles appeared on the screen next to each other, one having the just adjusted size and another having the correct mean size of the target set along the perceived size scale (Teghtsoonian, 1965).

At the beginning of the experiment, participants completed a block of ten practice trials using the stimuli and task from the Superset condition. These trials were intended to get observers familiar with the averaging task and the adjustment procedure. In addition, two first trials at the beginning of each block were considered as practice trials.

Data analysis and design

Practice trials were excluded from the analysis. As per preregistration, we excluded from analysis individual trials if the reaction time was below 300 ms or the adjustment error was greater than 3 standard deviations (SDs) from the overall error distribution obtained from a given participant. We excluded all data from a participant if their average reaction time was below 300 ms in any block or if their average adjustment error was greater than 3 SDs of all participants. The excluded participants were replaced by new participants to reach the preregistered sample size.

The experiment had a 7 (Target-defining attribute: Feature-Unicolor, Feature-Bicolor, Semi-feature, Conjunction, Spatial conjunction, Targets alone, and Superset) × 2 (mean size of the target subset: largest and smallest relative to the grand average of the superset) within-subject design. Overall, each participant was exposed to 22 trials per condition (308 trials in total). The experiment took approximately 30 min.

The primary behavioral measure in this experiment (in all conditions, except Superset – see below) was the adjustment error calculated as follows: *Error = (Reported mean size – Correct mean size) / Correct mean size.* This normalized error was used as a measure of a systematic deviation from the true mean of a target set, or bias. It was calculated for each participant and each condition separately. When the bias is equal to zero, it can be interpreted as an unbiased mean size estimation; when the bias is positive or negative, it means an over- or underestimation.

To correctly estimate the amount of bias specifically associated with subset selection, we used two reference points corresponding to "floor" and "ceiling" performance. To estimate the "ceiling" performance, we used the mean bias obtained in the Targets-alone condition (no selection required). To calculate bias corresponding to the "floor" performance, we asked participants to estimate the mean size of all the presented circles, but we calculated the bias relative to the mean size of the smallest and the largest subsets using the following formula *(Reported superset mean size – Subset mean size) / Subset mean size.* In each trial of superset condition, the bias was calculated relative to both large and small subset means; both these biases were included into the analysis as separate measurements. Note that unlike the formula for other conditions, here we used the term "*Subset mean size*" instead of "*Correct mean size*" because the bias was calculated relative to the subset mean size the participants were not asked about. As a result, calculated biases in the Superset conditions modeled trials where participants completely failed the subset selection and report the mean size of all presented circles instead of the mean size of a target subset.

For the main part of the data analysis, we collapsed the biases from the smallest and the largest target subset conditions. The main parameter of interest was the bias toward the grand average as a result of the subset selection difficulties. Without the collapse of the data, these difficulties caused a positive bias (i.e., an overestimation) when the target set had the smallest mean size; they led to a negative bias (i.e., an underestimation) when the target set had the largest mean size. Therefore, we reversed the sign of the errors (biases) in trials with the largest mean size of a target subset. As a result, positive and negative biases in a new scale corresponded to the bias toward the grand average and away from it (Fig. 10).

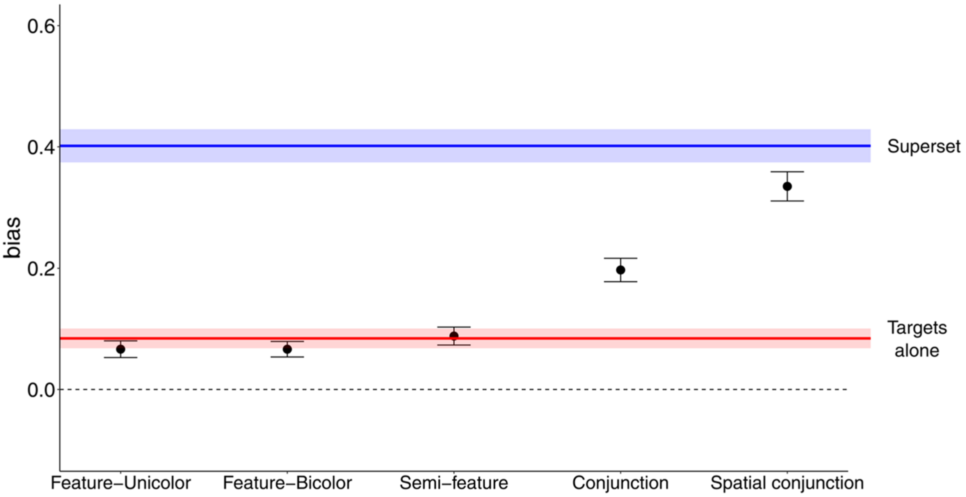


Fig. 10 The bias as a function of the unique attribute. The lower boundary indicates the baseline bias for the perfect selection (i.e., Targets-alone condition). The upper boundary indicates the maximum bias caused by the complete failure of target subset selection. Error bars denote the SEM, with between-subject variance removed following Cousineau's (2005) method

Results

Repeated-measures ANOVA showed a strong main effect of the target-defining attribute on bias (F(6,600) = 378.7.6, $p < 0.001$, $\eta^2_p = .67$). The post hoc comparisons showed that participants' responses were significantly biased in all the conditions compared to zero (ts > 4.86, ps < .001, Bonferroni-corrected $\alpha = .01$, Cohen's ds > 0.34). The biases in the Feature conditions (i.e., Feature-Unicolor, Feature-Bicolor and Semi-feature) were statistically indistinguishable from ones in the Targets-alone condition (ts < 1.51, ps > .13, Cohen's ds < 0.09). The biases in the other conditions statistically differ from each other (ts > 6.30 , ps < .001 , Bonferroni-corrected $\alpha = .017$, Cohen's ds > 0.20); the smallest bias was in the Feature and Targets-alone conditions (M = 0.08), there was a larger bias in the Conjunction condition (M = 0.2), then in the Spatial conjunction condition (M = 0.34), and the largest bias was in the Superset condition (M = 0.4), though that bias is not an error, per se. It is the definition of the upper bound to which we compare the other conditions.

We also performed analyses of biases in the smallest and the largest target subset conditions separately. The results were similar to the main analysis (see Online Supplementary Materials for details).

Discussion

Although we used a different method of measuring ensemble averaging in Experiment 4, our results show the same pattern as in Experiment 1. First, we observed that when the target subset had a unique feature not shared with any of the distractors, observers' estimates were similar to the Targets alone baseline. That is, their ability to pick relevant items was not strongly impaired by the presence of the distractors. A new condition not tested in the previous two experiments was the Semi-feature: a target subset with one unique feature, and another feature shared with the distractors. Since performance in the Semi-feature trials was the same as in the Targets Alone as well as in both Feature-Unicolor conditions, we conclude that a unique feature can be a good basis for ensemble selection even if this feature is just a part of the objects making an ensemble.

On the other hand, when the target subset was defined as a specific spatial conjunction of colors that were also presented in one of the distractor subsets with a different mean size, participants showed performance very close to what they showed in the Superset, floor condition. That can be compared to near-chance performance in the same condition of Experiments 1 and 2. These results suggest that the ability to correctly select a subset or even a few items in this condition is severely limited.

Finally, as in the other Experiments, we found a somewhat intermediate result in the Conjunction condition, where the pairing of colors in the target subset was unique but where each target color was also present in half of the distractor items. This intermediate pattern suggests that ensemble selection and processing based on conjunction is harder than that based on features, but, unlike spatial conjunctions, some selection is still possible, even with only a 200-ms presentation.

Since we measured baseline performances in the Targets-alone conditions, we also were able to capture a systematic overestimation bias that we interpret as an amplification effect (Kanaya et al., 2018). Indeed, even when the target subset is presented alone and no distractors interfere with sampling only relevant items, observers overestimate the mean.

General discussion

This study aimed to answer questions about the representational basis of ensemble selection. We tested three candidates for this role: basic features, preattentive object files, and bound objects. In four experiments, we found that conditions where a target ensemble was defined by a feature not shared with any of the distractors provided a rather good ensemble averaging performance (both for orientation and size), i.e., comparable to the baseline without distractors. When the target subset was defined by a conjunction of features (either from one or two dimensions) partly shared with distractors, some subset averaging was also possible, as the performance was more accurate than that predicted based on subset-blind sampling and averaging of all items on a display including distractors. Yet, averaging of subsets defined by a preattentive object file was less accurate than in the baseline and in the feature conditions. Finally, performance was very poor when the target subset was distinguished by an exact spatial conjunction of features. Overall, these results suggest that basic features support ensemble selection and calculation. For preattentive object files – here defined as conjunctions of two features, but without specific knowledge of their spatial relationship – some selection is possible. Responses may be based on a selection of an imperfect subset of the display or, perhaps, simply a small subset. In any case, we can reject the hypothesis that observers are basing their responses on the properties of a single item. When the subset is defined by the spatial relationship of features (a "bound object"), selection fails.

One could propose that the differences between conditions reflect a difference in "crowding." Perhaps the spatial conjunction conditions produce more crowding and, thus, poor subset selection. This would not contradict the idea that spatial conjunctions do not produce good subsets. It offers a hypothetical account that could be the basis for future research.

Experiments 2 and 3 directly tested the possibility that the results could be explained if observers selected only one random item from the subset and based their response on its orientation alone. The results showed that observers performed better than what would be expected from this one-item strategy in all conditions, except for the Spatial Conjunction condition. This condition produced performance near-chance, even worse than what would be expected from sampling just one relevant item per trial. The finding that the observers were better than with one-item sampling in most of the conditions does not rule out the possibility that observers were sampling a small number of items, rather than the entire subset (Whitney & Yamanashi Leib, 2018). Based on the performance in our task, it is hard to distinguish between genuinely parallel processing of all relevant objects as an ensemble (e.g. Chong et al., 2008) and sampling strategies of different sorts (e.g., sampling the square root of the number of objects, Dakin, 2001; Kanaya et al., 2018; Whitney & Yamanashi Leib, 2018). Although the distinction between these two modes of processing is central to the discussion of the mechanism of ensemble representation, this is not the focus of our study. Here, our primary interest is in defining the attributes that can make an impression of a distinct ensemble out of heterogeneous visual "stuff." The explanations for the observed pattern of results will be quite different depending on the view of sampling capacity. If the computational mechanism of ensemble averaging is parallel and exhaustive, then any decrement in performance should be attributed to an increased probability of confusion between items from target and distractor subsets. The more difficult the task of distinguishing between the target and distractor objects is, the greater is the chance of occasionally including some of the distractors into a selected group. These intrusions would corrupt an estimated summary of that group which would otherwise be fully processed. In contrast, the sampling account would explain the deviations from the correct answer by proposing that the number of sampled objects is smaller in the conditions with harder distinction. With a smaller effective subset, the sample is less representative of the true target subset, leading to the decrease in performance. Another convenient framework complimentary to sampling account is a "functional visual field" (FVF) account that proposes that different portions of the visual display can be processed under different stimulus condition (Hulleman & Olivers, 2017; Motter & Simoni, 2008; Sanders, 1963; Young & Hulleman, 2013). The FVF account would hold that only one item would fall into the FVF in the hardest condition. All the items in the display would be eligible for subset processing in the easiest conditions. In the intermediate conditions, the FVF would not cover all the items in the display, making ceiling performance impossible. Note that the two abovementioned scenarios (the including distractors into a target subset and decreasing the number of sampling objects) are not mutually exclusive and can be combined depending on circumstances. That is, the difficulty of target-distractor discrimination might cause both greater confusion and limited sampling. Moreover, some conditions, like the Feature conditions, might represent parallel ensemble calculations over the entire subset, while results for other conditions, like the Conjunction conditions, might reflect the sampling of a few items (e.g., Baek & Chong, 2020; Treisman, 2006).

Our study can shed light on some questions related to preattentive vision, namely, which information is available for the visual system in a non-selective mode, without the need for focused attention to individual objects (Wolfe et al., 2011). Our results suggest that the visual system can quickly select and statistically summarize a group or ensemble defined by a basic feature, which is not that surprising, as such a possibility has been previously documented (e.g., Chong & Treisman, 2005b; Halberda et al., 2006; Im & Chong, 2014; Sun et al., 2016). This fact reaffirms the predictions made by such models as Feature Integration Theory (Treisman & Gelade, 1980), Boolean map theory of visual attention (Huang & Pashler, 2007), and Guided Search theory (Wolfe, 2021).

Also, we found some version of this ability for ensembles defined by specific preattentive object file attributes (feature bundles, Wolfe & Bennett, 1997). That is, not only do observers know that there are some green-red things present among green-blue and blue-red things in a scene, but they can also appreciate that these green-red things are larger or smaller on average than other things or that these green-red things are tilted to a particular direction. This is an impressive result considering the minimal training received by participants and the fact that the calculations of average size and orientation were shown to be quite challenging (e.g., compared to the centroid estimation – Inverso et al., 2016; Rodriguez-Cintron et al., 2019). Unlike the first result, this can be seen as a confirmation of a unique prediction of the Guided Search theory (Wolfe, 2021). The theory suggests that the visual system has access to preattentive object file information via the priority map – a theoretical and, perhaps, physiological structure summing parallel inputs from multiple feature maps (each feature map sends a signal about the presence of a target feature in each location). This architecture allows observers to have a rough idea of where the objects with the target characteristics are and to select them for later processing. However, the judgments made in the conjunction conditions are much less accurate than in the basic feature conditions. Observers may be sampling fewer items or including some of the wrong items in the sample. Nevertheless, observers are able to extract some ensemble information in a relatively short stimulus duration and with quite irregular spatial organization (target subset is spatially intermixed with distractors).

Finally, when the target subset is defined by a specific spatial combination of features (as in the Spatial conjunction condition), we found that selection of such an ensemble is not possible. Consistent with many influential theories (e.g., Logan, 1994), fine encoding of exact feature conjunctions taking spatial relations into account (such as distinguishing between red-green and green-red objects) require the slow serial deployment of focused attention for feature binding (Treisman & Gelade, 1980; Wolfe et al., 2011; Wolfe & Bennett, 1997). Therefore, it is no surprise that observers did not discriminate target subsets in this condition and judged the mean based on a superset (Experiment 4) or guessed it at near chance level (Experiments 1 and 2).

The question about the representational basis of ensemble selection is closely related to the question of how deeply the visual system processes objects before focused attention comes into play. This question has been broadly studied in visual attention literature using the visual search paradigm in particular. While visual search ultimately requires detection, recognition, and localization of individual objects, ensemble perception is more about the global summarized (gist) impression of a large collection of objects. Yet, our results converge with those from the visual search. Of particular interest, evidence for the availability of preattentive object file characteristics for the formation of a distinct ensemble representation can be related to the fact that these characteristics can be subsequently used to guide the search of individual objects.

Conclusions

Our study advances our understanding of what attributes of visual items can be effectively used by the visual system for the selection of a group of similar items among other items and for ensemble processing of this group. The results showed that observers can select an ensemble of objects, labeled with a distinct feature. It may be possible to assemble an ensemble of objects with a specific conjunction of features but without information about their exact relationships. This supports the idea that, even before attention arrives at an object, the visual system has global access to quite elaborate representations of the large set of objects in the visual field. This information includes not only the distribution of separate features, but also what kinds of preattentive object files these features roughly form, and ensemble summary statistics of separate groups.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13414-024-02850-5.

Open Practices Statement All experiments reported in this article have been preregistered online before data collection. The filled preregistration forms as well as experimental data, stimuli, and code can be accessed at: https://osf.io/2qmca.

Funding Experiments 1 and 2 were supported by funding from the Russian Science Foundation under grant 18-18-00334. Experiment 3 is supported by the Basic Research Program of the National Research University Higher School of Economics. Experiment 4 was supported by funding from the US National Institutes of Health (EY01700, CA207490) and the US National Science Foundation (1848783, 2146617).

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. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. *Psychological Science, 19*(4), 392–398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

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, 106*(18), 7345–7350. https://doi.org/10.1073/pnas.0808981106

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

Baek, J., & Chong, S. C. (2020). 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

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods, 26*(3), 295–314. https://doi.org/10.1037/met0000337

Bauer, B. (2009). Does Stevens's Power Law for Brightness Extend to Perceptual Brightness Averaging? *The Psychological Record, 59*(2), 171–185. https://doi.org/10.1007/BF03395657

Buetti, S., Xu, J., & Lleras, A. (2019). Predicting how color and shape combine in the human visual system to direct attention. *Scientific Reports, 9*(1), 20258. https://doi.org/10.1038/s41598-019-56238-9

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology: CB, 18*(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition, 153*, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017a). Chapter 5 - Learning features in a complex and changing environment: A distribution-based framework for visual attention and vision in general. In C. J. Howard (Ed.), *Progress in Brain Research* (236th ed., pp. 97–120). Elsevier. https://doi.org/10.1016/bs.pbr.2017.07.001

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017b). Representing Color Ensembles. *Psychological Science, 28*(10), 1510–1517. https://doi.org/10.1177/0956797617713787

Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count vs estimate). *WIREs Cognitive Science, 2*(6), 634–638. https://doi.org/10.1002/wcs.136

Chong, S. C., Joo, S. J., Emmmanouil, 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

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences, 20*(5), 324–335. https://doi.org/10.1016/j.tics.2016.03.006

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138*(2), 289–301. https://doi.org/10.1016/j.actpsy.2011.08.002

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition, 20*(2), 211–231. https://doi.org/10.1080/13506285.2012.657261

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A, 18*(5), 1016. 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

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80–90. https://doi.org/10.1037/h0039515

Drew, S. A., Chubb, C. F., & Sperling, G. (2010). Precise attention filters for Weber contrast derived from centroid estimations. *Journal of Vision, 10*(10), 20–20. https://doi.org/10.1167/10.10.20

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70*(6), 946–954. https://doi.org/10.3758/PP.70.6.946

Epstein, M. L., & Emmanouil, T. A. (2017). Ensemble coding remains accurate under object and spatial visual working memory load. *Attention, Perception, & Psychophysics, 79*(7), 2088–2097. https://doi.org/10.3758/s13414-017-1353-2

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports, 6*(1), 32210. https://doi.org/10.1038/srep32210

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13341–13346.

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision, 14*(9), 22–22. https://doi.org/10.1167/14.9.22

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

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science, 17*(7), 572–576. https://doi.org/10.1111/j.1467-9280.2006.01746.x

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*(3), 599–631. https://doi.org/10.1037/0033-295X.114.3.599

Hulleman, J., & Olivers, C. N. L. (2017). The impending demise of the item in visual search. *The Behavioral and Brain Sciences, 40*, e132. https://doi.org/10.1017/S0140525X15002794

Iakovlev, A. U., & Utochkin, I. S. (2021). Roles of saliency and set size in ensemble averaging. *Attention, Perception, & Psychophysics, 83*(3), 1251–1262. https://doi.org/10.3758/s13414-020-02089-w

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception, 43*(7), 663–676. https://doi.org/10.1068/p7719

Im, H. Y., Tiurina, N. A., & Utochkin, I. S. (2021). An explicit investigation of the roles that feature distributions play in rapid visual categorization. *Attention, Perception, & Psychophysics, 83*(3), 1050–1069. https://doi.org/10.3758/s13414-020-02046-7

Inverso, M., Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). Evidence against global attention filters selective for absolute bar-orientation in human vision. *Attention, Perception, & Psychophysics, 78*(1), 293–308. https://doi.org/10.3758/s13414-015-1005-3

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition, 48*, 149–160. https://doi.org/10.1016/j.concog.2016.11.007

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. A. Davies (Eds.), *Varieties of attention.* New York: Academic Press.

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

Khvostov, V. A., Lukashevich, A. O., & Utochkin, I. S. (2021). Spatially intermixed objects of different categories are parsed automatically. *Scientific Reports, 11*(1), 377. https://doi.org/10.1038/s41598-020-79828-4

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

Kim, M., & Chong, S. C. (2020). The visual system does not compute a single mean but summarizes a distribution. *Journal of Experimental Psychology: Human Perception and Performance, 46*(9), 1013–1028. https://doi.org/10.1037/xhp0000804

Lleras, A., Wang, Z., Madison, A., & Buetti, S. (2019). Predicting Search Performance in Heterogeneous Scenes: Quantifying the Impact of Homogeneity Effects in Efficient Search. *Collabra: Psychology, 5*(1), 2. https://doi.org/10.1525/collabra.151

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 20*(5), 1015–1036. https://doi.org/10.1037/0096-1523.20.5.1015

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281. https://doi.org/10.1038/36846

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision, 15*(4), 6. https://doi.org/10.1167/15.4.6

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A "dipper" function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11), 9–9. https://doi.org/10.1167/8.11.9

Motter, B. C., & Simoni, D. A. (2008). Changes in the functional visual field during search with and without eye movements. *Vision Research, 48*(22), 2382–2393. https://doi.org/10.1016/j.visres.2008.07.020

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

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*, 522–536. https://doi.org/10.1037/h0026699

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

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

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

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision, 13*(8), 6–6. https://doi.org/10.1167/13.8.6

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197. https://doi.org/10.1163/156856888X00122

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3–34.

Rodriguez-Cintron, L. M., Wright, C. E., Chubb, C., & Sperling, G. (2019). How can observers use perceived size? Centroid versus mean-size judgments. *Journal of Vision, 19*(3), 3. https://doi.org/10.1167/19.3.3

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631–632. https://doi.org/10.1038/nn876

Sàenz, M., Buracâs, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research, 43*(6), 629–637. https://doi.org/10.1016/S0042-6989(02)00595-3

Sanders, A. F. (1963). *The selective process in the functional visual field*. Assen, NL.: Van Gorcum.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10*(14), 19–19. https://doi.org/10.1167/10.14.19

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12), 13–13. https://doi.org/10.1167/11.12.13

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision, 18*(7), 4. https://doi.org/10.1167/18.7.4

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). Human attention filters for single colors. *Proceedings of the National Academy of Sciences, 113*(43), E6712–E6720. https://doi.org/10.1073/pnas.1614062113

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2018). High-capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of the National Academy of Sciences, 115*(52), E12153–E12162. https://doi.org/10.1073/pnas.1814657115

Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science, 25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Teghtsoonian, M. (1965). The Judgment of Size. *The American Journal of Psychology, 78*(3), 392. https://doi.org/10.2307/1420573

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision, 15*(4), 8. https://doi.org/10.1167/15.4.8

Utochkin, I. S., Khvostov, V. A., & Stakina, Y. M. (2018). Continuous to discrete: Ensemble-based segmentation in the perception of multiple feature conjunctions. *Cognition, 179*, 178–191. https://doi.org/10.1016/j.cognition.2018.06.016

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

Utochkin, I. S., & Vostrikov, K. O. (2017). The numerosity and mean size of multiple objects are perceived independently and in parallel. *PLOS ONE, 12*(9), e0185452. https://doi.org/10.1371/journal.pone.0185452

Utochkin, I. S., & Yurevich, M. A. (2016). Similarity and heterogeneity effects in visual search are mediated by "segmentability." *Journal of Experimental Psychology: Human Perception and Performance, 42*(7), 995–1007. https://doi.org/10.1037/xhp0000203

Wang, Z., Buetti, S., & Lleras, A. (2017). Predicting Search Performance in Heterogeneous Visual Search Scenes with Real-World Objects. *Collabra: Psychology, 3*(1), 6. https://doi.org/10.1525/collabra.53

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32*(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036-I

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

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review, 28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive Object Files: Shapeless Bundles of Basic Features. *Vision Research, 37*(1), 25–43. https://doi.org/10.1016/S0042-6989(96)00111-3

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 3. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Võ, M.L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences, 15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R., & Cave, K. R. (1990). Limitations on the parallel guidance of visual search: Color × Color and Orientation × Orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 879–892. https://doi.org/10.1037/0096-1523.16.4.879

Wolfs, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of park-whole information in visual search tasks.

Perception & Psychophysics, 55(5), 537–550. https://doi.org/10.3758/BF03205311

Young, A. H., & Hulleman, J. (2013). Eye movements reveal how task difficulty moulds visual search. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 168–190. https://doi.org/10.1037/a0028679

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