PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES
NEUROSCIENCE

Deriving the number of salience maps an observer has from the number and quality of concurrent centroid judgments

Lingyu Gan[a,1], Peng Sun[b], and George Sperling[a,c,1]

Contributed by George Sperling; received February 1, 2023; accepted March 31, 2023; reviewed by Misha Pavel and John Reynolds

[C. Koch, S. Ullman, *Hum. Neurobiol.* 4, 219–227 (1985)] proposed a 2D topographical salience map that took feature-map outputs as its input and represented the importance "saliency" of the feature inputs at each location as a real number. The computation on the map, "winner-take-all," was used to predict action priority. We propose that the same or a similar map is used to compute centroid judgments, the center of a cloud of diverse items. [P. Sun, V. Chu, G. Sperling, *Atten. Percept. Psychophys.* 83, 934–955 (2021)] demonstrated that following a 250-msec exposure of a 24-dot array of 3 intermixed colors, subjects could accurately report the centroid of each dot color, thereby indicating that these subjects had at least three salience maps. Here, we use a postcue, partial-report paradigm to determine how many more salience maps subjects might have. In 11 experiments, subjects viewed 0.3-s flashes of 28 to 32 item arrays composed of M, M = 3,...,8, different features followed by a cue to mouse-click the centroid of items of just the post-cued feature. Ideal detector response analyses show that subjects utilized at least 12 to 17 stimulus items. By determining whether a subject's performance in (M-1)-feature experiments could/could-not predict performance in M-feature experiments, we conclude that one subject has at least 7 and the other two have at least five salience maps. A computational model shows that the primary performance-limiting factors are channel capacity for representing so many concurrently presented groups of items and working-memory capacity for so many computed centroids.

neural networks | salience maps | perceptual grouping | statistical summary judgments | centroid judgments

Significance

In order to determine what in the visual field to process first, Koch and Ullman proposed the brain uses a topographical map that combines a wide range of inputs to record at each spatial location just one value—salience (likely importance). The location with the highest salience is processed first. Here, we demonstrate that the same or a similar salience map is used for centroid judgments, the judged center of a cloud of items. Following a brief flash of five different item types, all our observers were able to make five accurate concurrent centroid judgments. This indicates that our observers each have at least five concurrently available salience maps, an unexpected visual information processing capacity of the brain.

Human visual system can extract summary statistical information from groups of similar items in a brief glance (1, 2). Such ensemble statistics are interesting because they result from brain mechanisms that can quickly distill a large amount of sensory information for subsequent cognitive processes that have much lower capacity. Prior studies have shown that humans can form statistical representations, for example, of size (3), spatial location (4), orientation (5), motion (6), brightness (7), hue (8), numerosity (9) and facial expression (10).

Humans are also capable of selectively forming statistics out of spatially intermingled visual stimuli defined by different features. Of particular relevance here are centroids, the mean location of a group of items. Drew et al. (4) report that in a display of interleaved light and dark dots, subjects can selectively compute different centroids for the light and for the dark dots. Sun et al. (11) report that subjects could efficiently compute the centroid of three target dots of a pre-cued color even when the target dots were intermixed with 21 dots of 7 other equiluminant distracter colors.

Salience Maps and Centroid Judgments. To compute the centroid of a set of items of a particular feature, the target items need to be separated from distracter items, and spatial information of the segregated target items is required. A topographical map that inherently contains the required spatial information of the target items would meet the requirement for computing a centroid of the target items. Because the brain is finite, and because there is an infinity of features and feature combinations that could define target items, it is clear that the feature locations must be represented in a topographic map that is indifferent to feature identities. This kind of topographic map was first proposed by Koch and Ullman (12) to represent the overall saliency of items (Fig. 1). Their saliency map has inputs from many different feature maps and summarizes the inputs at each x, y location as a single real number, saliency, which is represented in the map. The output of their saliency map was fed into a winner-take-all network that computed the location with the highest saliency. The output of the winner-take-all network could be used to

Author affiliations: [a]Department of Cognitive Sciences, University of California, Irvine, CA 92697; [b]Digital Imaging Solutions, Canon USA, Irvine, CA 92618; and [c]Department of Neurobiology and Behavior, University of California, Irvine, CA 92697

Author contributions: L.G., P.S., and G.S. designed research; L.G. performed research; L.G. and G.S. analyzed data; and L.G. and G.S. wrote the paper.

Reviewers: M.P., Northeastern University; and J.R., Salk Institute for Biological Studies.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: lingyug1@uci.edu or sperling@uci.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2301707120/-/DCSupplemental.

Published May 15, 2023.

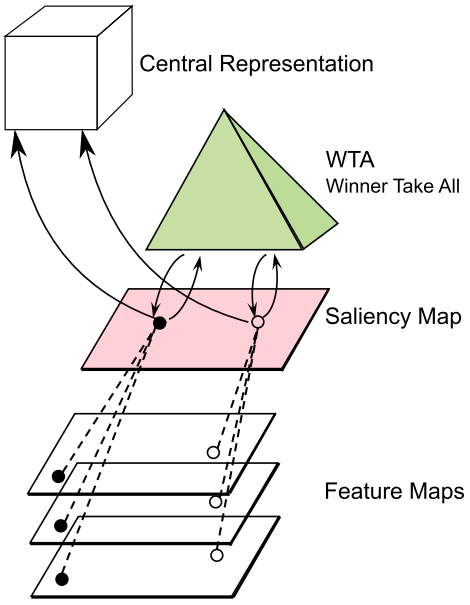

Fig. 1. Koch and Ullman's (12) original representation of a saliency processing system. (Colors have been added).

predict priority in visual search and other visual processes. Koch and Ullman (12) is a conceptual model. Many subsequent papers clarify, quantify, and further elaborate the original concept, e.g., refs. 13–15, and 16. A recent Google Scholar search yielded over one million citations to salience maps, most of the relevant citations were for priority.

Sun et al. (17) proposed a similar topographic salience map to compute not a winner-take-all but rather a centroid. Their conception was that the input to a salience map is flexible, rather than hardwired and depends on the task. Sun et al. (17) reported that after a 250-msec view of a display of 24 dots, 8 dots per color, subjects were able to report three centroids, one for each color, with approximately equal accuracy. An ideal detector that knows the exact location of each stimulus item and that computes a centroid perfectly would require 13.0 of 24 stimulus items to match the average subject's performance. The number of stimulus items required by such an ideal detector is the lower bound of the number of stimulus items processed by subjects. The aim of the present study is to extend these results to determine whether more than three centroids can be computed concurrently and to determine how many perceptually processed items are incorporated into these centroids. But first, even to draw the conclusion that subjects can concurrently compute three centroids, two possible confounds needed to be excluded.

(1) Visual information might be available for much longer than the brief exposure to enable computation of a centroid long after the stimulus exposure was terminated. Indeed, visual stimulus continues to be visible for some time after the offset of the stimulus (18). However, a poststimulus mask (19) can strictly control the duration for which the stimulus is visually available. In ref. 17, for 3 of 5 subjects, there was no poststimulus mask following the stimulus display. Therefore, for these three subjects, visual information is available longer than the duration of the stimulus display; two subjects viewed stimuli followed by masks. No statistical difference was found between the mask data and no-mask data suggesting that visual information available in any mask-susceptible visual memory after the offset of the stimulus display did not significantly benefit the centroid computation.

(2) Suppose subjects computed only a random two of three centroids on each exposure and optimally guessed the third centroid (subsampling). If subjects were to adopt this subsampling strategy, the quality of the subject's attention filters—a measure that describes the weights that subjects give to different features—when three centroids are required would be worse than the quality of the subject's attention filters when only one centroid is required (subjects know which centroid to report ahead of the stimulus presentation). The subsampling confound is excluded by the observation that the quality of the subject's attention filters when subjects reported three centroids was as good as the quality of the subject's attention filters when subjects reported only one centroid.

The facts 1) that subjects process at least 13.0 stimulus items (implying more than 4 dots utilized times 3 centroids), 2) that there is no significant impairment in subject's attention filters when reporting three versus one centroid, and 3) that there are no significant differences in error magnitude between the three different concurrent centroid judgments demonstrate that subjects are indeed able to compute three concurrent centroids. Because a salience map is feature-independent, and a single salience map cannot discriminate colors, Sun et al. (17) concluded that three salience maps are required to concurrently compute three centroids.

Procedure Outline. The question that we investigate here is: "How many salience maps do subjects have?" Each centroid computation requires its own salience map, so we seek to determine the maximum number of centroids that subjects can compute concurrently. Sun et al. (17) required their subjects to report three centroids following a brief exposure. The problem with asking subjects to report more than three centroids is that the process of reporting the first few centroids strongly interferes with the memory of the still unreported centroids. Therefore, we used a partial-report paradigm (18), a postcue procedure that requires only one report on each trial (Fig. 2). On a post-cued trial, subjects viewed briefly flashed 300 ms arrays containing M, M = 3, 4, 5, 6, 7, or 8, spatially interleaved features. The number of different dot colors was the main features variable. When the number of colors was 5 and greater, colors became difficult to discriminate so different black shape items were introduced. M denotes the number of items in a stimulus consisting of different dot colors, or different shapes, or a mixture of colored dots and black shapes. A cue indicating a randomly selected target feature was presented immediately after a 50-ms blank interval and a 100-ms mask that followed the stimulus exposure. Subjects' task was to compute only the centroid of items defined by cued feature and to mouse-click its location on the same screen that displayed the stimuli.

How Do post-cued Trials in the Current Centroid Paradigm Enable Us to Determine How Many Salience Maps a Subject Has?. The brief stimulus exposure and the postcue masking stimulus require all the possibly-to-be-reported centroids to be computed during the brief period in which stimulus information is visually available. On a post-cued trial, subjects do not know which centroid to report until the postcue appears 150 msec after the onset of the poststimulus mask. At that point, there is no longer any significant amount of visual information available in any visual memory that is susceptible to postexposure masking,

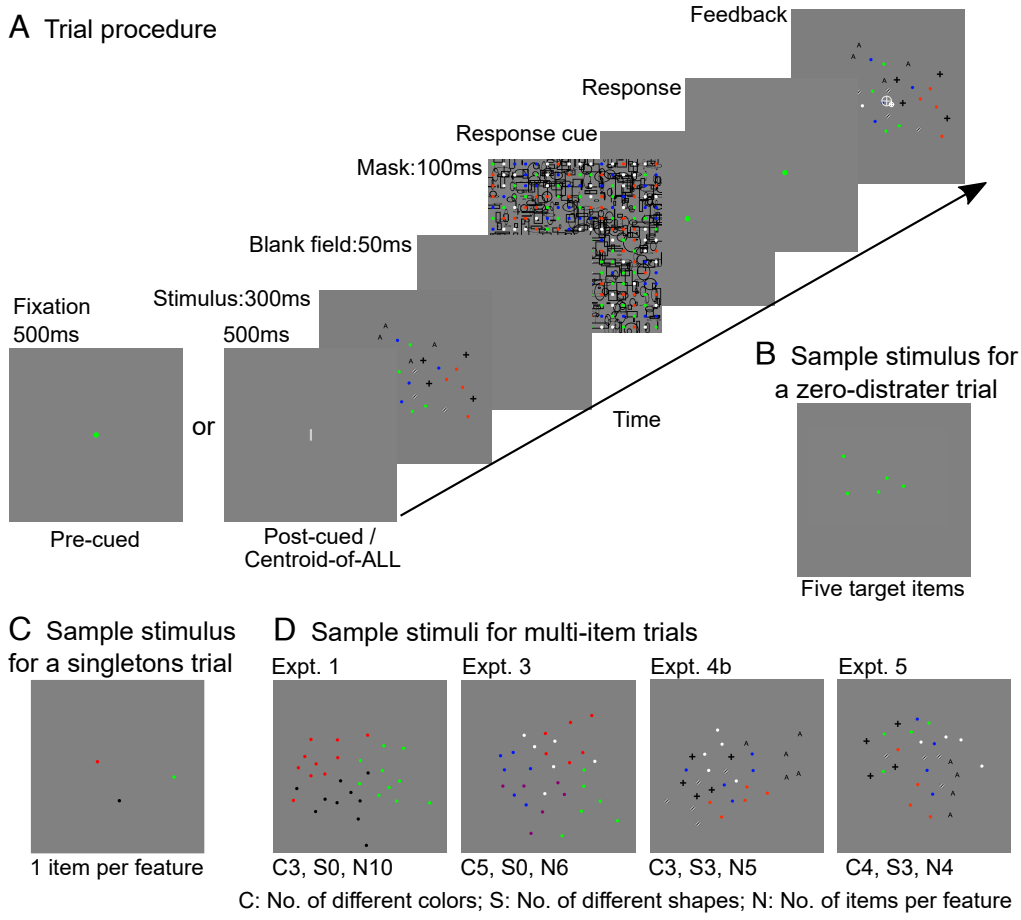

Fig. 2. Experimental procedures and sample stimuli. (*A*) Every trial begins with a 500-ms blank field with a fixation point, followed by a 300-ms stimulus, a 50-ms blank field, a 100-ms mask, a blank field with an example target stimulus that the subject moves to the judged centroid location, and finally, a feedback display. Feedback shows the stimulus, the centroid of the target feature as the large plus sign inscribed in a light gray open circle, and the subject's response as the smaller plus sign inscribed in a white circle. For pre-cued trials, the fixation point is a target stimulus. For post-cued trials, the fixation point is a white bar, and the cue indicating the target feature is presented 50 ms after the poststimulus mask. Centroid-of-ALL trials have the same procedure as the post-cued trials, except that the task is to judge the centroid of ALL stimulus items. (*B*) Sample stimulus for a zero-distracter trial. The procedure for zero-distracter trials is the same as for postcue trials, except that only target items are presented on the screen. (*C*) Sample stimulus for a singleton trial. For each singleton trial, the stimulus display consists of one item from each feature. The singleton items are located at centroids of the corresponding full stimuli from which they are derived. (*D*) Sample stimuli as displayed on an 800 × 800 pixel screen. "C" in the title under the stimulus display represents the number of different colors of dots in the stimulus display, "S" is the number of different shapes, and "N" is the number of items in each feature class. For example, "C3, S3, N5" means that this stimulus contains dots of 3 different colors, 3 different shapes, and the number of items of each feature is 5. There are (3+3) * 5 = 30 items in this display.*

i.e., by the time the subject receives the postcue, there is no information available in "visual persistence" (18) or "iconic memory" (20). Memory that survives a strong postexposure mask is commonly called "working memory." Sun et al. (21) found that, following a brief display of 26 dots, less than two dots were available in working memory whereas, in the same display, at least 15 dots were incorporated into centroid computations. Therefore, all centroid computations have to be completed on the basis of information that fails to survive the postexposure mask, and only computed centroids, not individual stimulus items, are retained in working memory. To reiterate, on a post-cued trial, all the centroids that a subject computes are computed on the basis of visual information that is available only during a 300-stimulus exposure and a 50-msec blank period before the strong postexposure masking field.

Eleven Centroid Experiments (Summary). The current study used a variant of the centroid paradigm that was originally developed by Drew et al. (4) and was considerably enhanced by Sun et al. (11). Subjects viewed briefly flashed arrays containing 3, 4, 5, 6, 7, or 8 spatially interleaved features. Fig. 2 and *SI Appendix*, S1 show sample experimental stimuli. An initial set of 3 experiments was conducted with 24 items per stimulus. Subjects performed so well in these experiments that, to avoid ceiling effects, the total number of items in each of the main experimental stimuli was increased to 30 ± 2, the largest number that allowed us to maintain the 56-pixel center-to-center spacing of items (required for the shape stimuli) within the display boundaries. The subjects' task in all experiments was to move a cursor to the judged centroid location of the cued-feature items (targets) and to mouse-click.

There were four primary trial types (pre-cued vs. post-cued trials X multi-item stimuli versus singletons) and several variations.

(1) pre-cued multi-item trials: The cue indicating the target feature whose centroid was to be reported was presented 500 msec before the stimulus.
(2) Post-cued multi-item trials: The cue was presented immediately after a 50-msec blank interval and a 100-msec mask that followed the stimulus exposure.

SI Appendix, Fig. S1 displays sample stimuli for each experiment.

(3, 4) pre-cued and post-cued singleton trials: For every type of multi-item trial, there was matched singleton trial. A singleton stimulus consisted of just one item of each feature type. Each singleton item was located at the centroid of its feature in the corresponding full stimulus. The task was reporting the location of the cued singleton item. Singleton trials control for difficulties in computing centroids.

(5) Zero-distracter trials: There was no cue, distracters were eliminated and only target dots or target shapes were presented. Zero-distracter trials control for attention-filter imperfections in extracting target items.

(6) Centroid-of-ALL trials: Stimulus items were the same as the pre-cued and post-cued trials, but subjects were instructed to judge the centroid of ALL displayed items. Centroid-of-ALL trials estimate how many items can be processed when the problems associated with separating the items into separate groups are eliminated.

The stimuli of the six kinds of trials and the experimental procedure are depicted in Fig 2.

Results

Overview. In every experiment, the basic datum from each trial is the centroid response location, represented as response error distance. To gain more insight into the processes by which subjects arrive at their responses, three other statistics are computed as further explained below: A subject's attention filter (the weights of target and distracter items in the filter output) is estimated directly from the locations of a subject's responses. The number of stimulus items that survive to be incorporated into a subject's centroid computation is estimated by an ideal detector that processes only a sufficient fraction of the stimulus items to match the accuracy of the subject's centroid judgments. Two ideal detectors are considered: One assumes the subject's estimated attention filter, and the other assumes a perfect attention filter that admits all target items and rejects all others.

Mean Error Magnitude. The most basic measurement of a subject's performance is the response error—the difference between the true centroid location of the target feature and the subject's mouse-click response. Although this difference is a 2D quantity, there were no significant differences between x and y components; therefore, they are treated equally. Error magnitude is the Euclidean distance error measured in units of display pixels.

All the data of all the experiments are exhibited in Fig. 3 *A* and *B*, except the implicit data that are used to compute attention filter weights. Fig. 3*A* depicts the response accuracy of multi-item trials in terms of mean response error magnitude averaged over the M different target conditions. The x-axis represents the 11 experiments, and the y-axis represents the response error magnitude in pixels. Response accuracy is best for zero-distracter and centroid-of-ALL trials, slightly worse for precue trials (in 29 of 31 instances), and very significantly worse for post-cued trials. For zero-distracter and centroid-of-ALL trials, subjects' response accuracy was independent of experimental conditions. For pre-cued trials, subjects' response accuracy was independent of experimental conditions with one exception, S2, Expt. C0S4N6, which reflects this particular subject's inability to discriminate four shapes. One-way ANOVA tests confirmed that there were no significant differences among experiments for pre-cued ($F(10,20) = 1.385$, $P = 0.256$), zero-distracter ($F(10,20) = 0.721$, $P = 0.696$) and centroid-of-ALL trials ($F(10,20) = 1.125$, $P = 0.392$). For post-cued trials, subjects' response accuracy dropped significantly as the number of attended groups increased ($F(10,20) = 14.979$, $P = 0.000$).

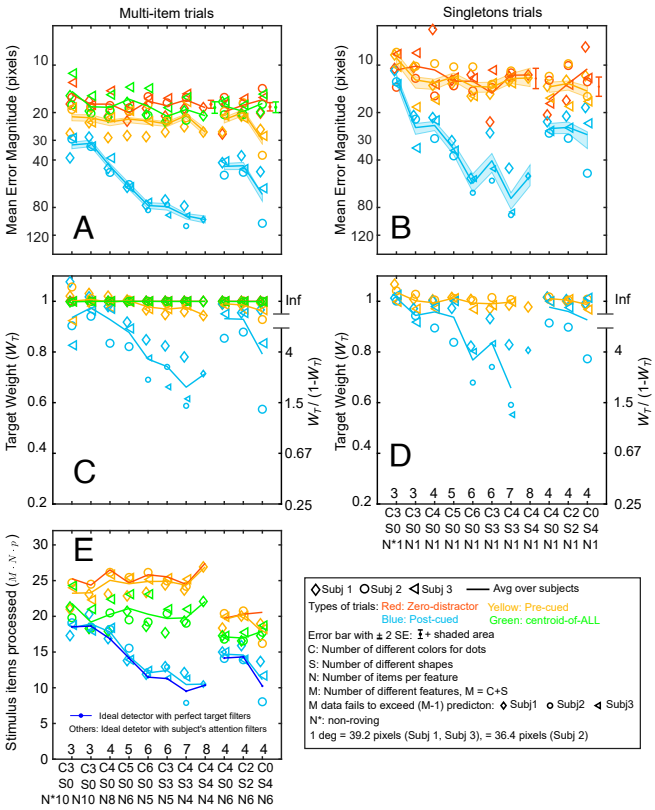

Fig. 3. Results of 11 multicentroid judgment experiments for three subjects. In all five panels, the x-axis represents the 11 experiments with their stimulus parameters as coded in Fig. 2. The number above each bottom x-axis tick represents *M*, the number of different features the subject must attend. In each of the 11 experiments, the symbols (triangle, diamond, circle) represent a particular subject's data. Small-size data points are from an *M*-feature experiment that failed to exceed the prediction from an available corresponding (*M*-1)-feature experiment. The 8 leftmost data points are from experiments with 30 stimulus items, the 3 rightmost data points are from experiments with 24 stimulus items. Colored lines represent the mean data of three subjects. (*A* and *B*) Response accuracy in terms of mean error magnitude in pixels—the distance between a subject's centroid judgment and the true centroid. The shaded areas around mean post-cued and mean pre-cued data curves represent the confidence intervals ± 2SE (Standard Errors) for each computed mean. Error bars (mean± 2SE) to the right of C4S4N4 apply to the first 8 experiments; error bars on the far right apply to the three rightmost experiments. (*C* and *D*) Filter weight for target features (*Left* ordinate). The *Right* ordinate is a ratio, the average filter weight for target features (W_T) divided by average filter weight for distracters ($1 - W_T$). For centroid-of-ALL trials, there are no distracters, target weight is defined as 1. (*E*) $M \cdot N \cdot p$, where *M* is the number of different features, *N* is the number of items of each feature, and *p* is the probability of each item surviving an initial decimation process that produces the observed mean error magnitude. $M \cdot N \cdot p$ is the minimum number of stimulus items that an ideal observer with the same attention filter as the subject needs to process perfectly in order to match the subject's performance (*Upper* four curves); the dark blue bottom curves represent the slightly smaller number of items required by an ideal detector with a perfect target filter.†

Fig. 3*B* depicts the mean response error magnitude of singleton trials averaged over the *M* different target conditions. The mean error magnitude of singleton trials is remarkably similar to the mean error magnitude of multi-item trials for pre-cued, post-cued, and zero-distracter trial types in all 11 experiments. The variability of multi-item data points is smaller than for

† For the nonroving C3S0N*10 experiment, the centroid SD for each feature was 56.6 ± 2.6. Details in *Materials and Methods*.

singleton data because there were 2.5 times more multi-item trials. Response accuracy for zero-distracter and pre-cued trials is independent of experimental conditions, whereas response accuracy for post-cued trials decreases as M increases.

Number of Surviving Target Items. When a subject's mean centroid error is 20 pixels, does this represent good or bad performance? We measure the quality of a subject's performance in terms of the minimum number of stimulus items that an ideal detector, which knows the exact location of every item and that computes a perfect centroid, needs in order to match a subject's performance.

Fig. 4*A* shows the ideal detector model. The ideal detector has only one source of error, an initial decimation process during which each stimulus item has only a probability p, $0 \leqslant p \leqslant 1$, of surviving. The ideal detector considered here has a perfect target filter that passes only the N target items and discards all distracters. The ideal detector knows the exact location of each of the $N \cdot p$ admitted target items and computes a perfect centroid. A Monte Carlo method is used to estimate the mean error magnitude of the ideal detector as a function of p for each of the 11 experiments (curves in Fig. 4*B*). $N \cdot \hat{p}$ is the minimum number of target items that the ideal detector needs to match its mean error to a subject's mean error.

The curves in Fig. 4*B* depict the error magnitude (abscissa) as a function of the number of surviving target items (ordinate) for the ideal detector for each of the 11 experiments. Because the ideal detector has a perfect filter, its performance depends only on N, the number of target items in the stimulus, and on p, their survival probability. Five curves are shown for the five values of N plus a sixth curve for the one nonroving condition C3S0N*10. Data averaged over three subjects for the 3 kinds of trials in 11 experiments are shown as points on the ideal detector curves. All subsequent references to the number of surviving target items are based on these 33 numbers.

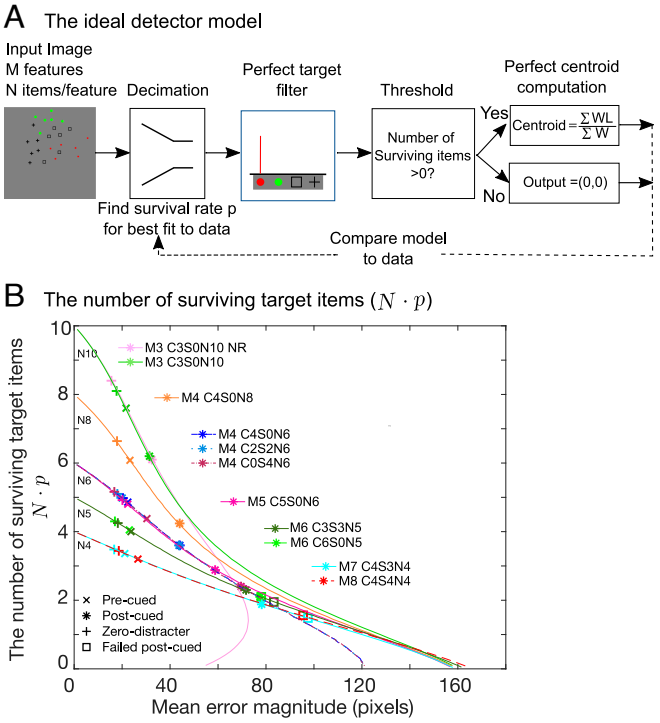

Fig. 4. The ideal detector model and its expected centroid error as a function of the number of surviving target items. (*A*) The ideal detector model. The input is a stimulus image. In this example, it is a C2S2N6 stimulus (2 colors, 2 shapes, 6 dots for each feature, 24 dots total). The input image first goes through a decimation process during which each stimulus item has a probability p of surviving. Then, a perfect target filter removes all nontarget items and passes only surviving target items. If no target item has survived the decimation process, the output is (0,0), the center of the screen. Otherwise, a perfect centroid computation is carried out on the surviving target items. A Monte Carlo method is used to estimate the mean error magnitude of the ideal detector as a function of p (the curves in Fig. 4*B*). The number of surviving target items is the product p*N of the survival rate, p, and the number of target items, N (N = 6 in this example). Fig. 4*A* is adapted from Fig. 5 in ref. 17. (*B*) Mean error magnitude (abscissa) as a function of the number of surviving target items (ordinate) for 3 kinds of trials in 11 experiments. The data points show $N \cdot p$, the least number of target items that must be processed by an ideal detector with a perfect target filter to match a subject's performance. The colored lines represent the mean error magnitude of the ideal detector as a function of the number of target items being processed by the ideal detector for each of the 11 experiments using 8 different stimulus configurations. Data points indicated by "x," "∗," and "+" represent the mean error magnitude averaged over subjects' data from an *M*-feature experiment in which performance was better than could be predicted from the corresponding (*M-1*)-feature experiment. Squares represent postcue data from an M-feature experiment that failed to exceed the prediction from the corresponding (*M-1*)-feature experiment.

Target Weight. Response error magnitude informs us about how far a subject's response is away from the correct response, but error magnitude is not informative about the nature of the response computation. In the centroid paradigm, in addition to error magnitude, the particular distribution of errors allows us to estimate a subject's attention filter for each experiment. Attention filters, first used by ref. 4 and refined by ref. 11, describe the weight for each target and each distracter feature that best predicts the judged centroids. Attention filter weights are estimated separately for each experiment, condition, and subject to minimize the sum of the squared distances between responses predicted purely by attention filters and observed responses (details in *SI Appendix*).

For each subject, in each experiment, the *Left* ordinate in Fig. 3*C* depicts the target weight W_T of multi-item trials averaged over the M features. In zero-distracter trials, only the target feature was displayed, therefore the target weight for zero-distracter trials is always 1. In centroid-of-ALL trials, subjects were instructed to give equal weight to each feature. As no significant weight differences were observed for various different targets in centroid-of-ALL trials, in Fig. 3*C* target weight is plotted as 1. For pre-cued trials, for all three subjects, target weights are nearly 1.0, with one exception, Subj 2, Expt. C0S4N6 where Subject 2 found the 4 shapes difficult to discriminate. post-cued trials are much more difficult than precue trials, and target weight drops significantly as M increases. The *Right* ordinate in Fig. 3*C* depicts the ratio between the aggregate weights of the target features and the aggregate weight of all the distracter features, $W_T/(1 - W_T)$. The ratio of the average target item filter weight to the average distracter item filter weight is larger by a factor of M-1, i.e., it is $W_T(M - 1)/(1 - W_T)$; this ratio is a better indicator of filter selectivity, it is not illustrated here.

Fig. 3*D* depicts the target weight in singleton trials' attention filters averaged over M features for each subject in each experiment. Although the target weights of singleton trials are superficially similar to the target weights of multi-item trials, a more detailed analysis (*SI Appendix*, Figs. S4 and S5) shows that the error distributions differ. Wide multicentroid error distributions reflect filter error, whereas distributions for singletons are mixtures of narrow and flat distributions indicating memory errors.

Number of Stimulus Items Processed. The number of surviving target items is the absolute smallest number of stimulus items that an ideal detector with a perfect attention filter needs to process

in order to match a subject's performance (Fig. 3*E*, dark blue curves). While that is a useful characterization of the absolute quality of performance, we here begin to derive a description of how subjects arrive at their performances rather than merely characterizing the quality of their performances. The first step is replacing the perfect attention filter with the attention filter that best characterizes a subject's performance—the subject's *achieved attention filter*. Because a subject's achieved attention filter is rarely perfect, an ideal detector using a subject's achieved attention filter requires more items to match that subject's performance than a detector with a perfect filter. For a stimulus containing $M \cdot N$ stimulus items, the number of stimulus items processed by a detector with an achieved attention filter is expressed as $M \cdot N \cdot \widehat{p}$. Here, $\widehat{p}$ is the probability of survival that equates the mean error magnitude of an ideal detector with a subject's attention filter to that subject's measured error magnitude.

Fig. 3*E* depicts the number of stimulus items processed by an ideal detector with a different attention filter for each of 3 subjects, 4 kinds of trials, and 11 experiments, averaged only over target types. For comparison, the mean number of stimulus items processed $M \cdot N \cdot \widehat{p}$ computed using a perfect filter is shown only for post-cued trials (the dark blue curve). (The other perfect vs. actual filter comparisons are shown in Fig. 5*A*.) For all 11 experiments, zero-distracter trials have the highest number of stimulus items processed; pre-cued trials are very similar, and for both kinds of trials there is no significant drop in performance as the number of features to be processed increases. In contrast, post-cued trials have the lowest number of surviving target items, and post-cued performance declines significantly as the number of features increases.

Two things are worthy of notice. First, the mean number of stimulus items processed by an ideal detector with a perfect target filter (the dark blue curve) closely follows the mean number of stimulus items processed by an ideal detector with the subject's attention filter (the light blue curve). The great similarity between these two curves indicates that subjects' attention filters are not a major source of error. Second, the number of stimulus items processed for the pre-cued trial type is only slightly worse than the corresponding zero-distracter trials. This indicates that, in pre-cued trials, when subjects need to form only a single attention filter, this single attention filter virtually eliminates distracters.

Number of Stimulus Items Processed Corrected for Motor Error. We are interested here in the costs of computing and remembering multiple centroids. However, even the report of the position of a single stimulus item, the easiest possible condition, has a significant error that we call "motor error". We estimate motor error variance and subtract it from the overall error variance to enable us to better estimate perceptual and memory errors that are of greater interest. Here, we report surprising properties of zero-distracter and pre-cued centroid trials that are revealed by subtracting motor error.

We estimate each subject's motor error from zero-distracter singleton trials in which only one stimulus item is shown on the screen, and the task is to report the location of this item. The error variance for such trials includes both motor error (positioning the cursor) and memory noise (remembering the target location) and possibly other components. Each subject's motor error for a particular stimulus type is computed as the overall error variance of 50 zero-distracter singleton trials. Motor error variance is then subtracted from the overall error variance in each experiment to produce a set of data that is "corrected for motor error" for that type of stimulus. The number of stimulus items processed by an ideal detector is then recomputed in two slightly different ways based on the data corrected for singleton error: 1) assuming the subject's attention filter and 2) assuming a perfect target filter.

Fig. 5 *A* and *B* depict the ideal detector's estimates of the total number of stimulus items processed, both corrected and uncorrected for motor error, averaged over subjects. These estimates are shown for the 4 types of trials in each of the 11 experiments using the subject's attention filters and also using perfect attention filters.

Fig. 5*B* shows that the numbers of stimulus items processed for zero-distracter, pre-cued, and centroid-of-ALL trials after correction for motor error are close to the data points representing the total number of stimulus items. The average difference between the zero-distracter subject's filter curve and the number of stimulus items is 1.37 items, the pre-cued difference is 1.47

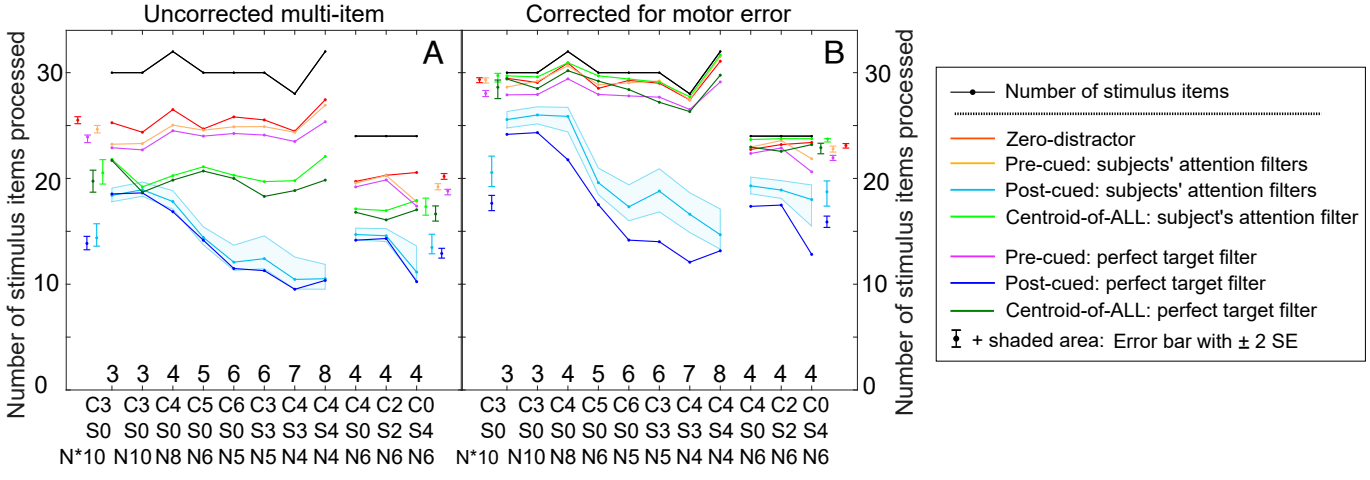

Fig. 5. Number of stimulus items processed, averages of three subjects, in four types of multi-item trials in each of 11 experiments, uncorrected and corrected for motor error. In (*A*) and (*B*), the shaded area around the post-cued curves represents the confidence interval for ± 2 standard errors (SE) for each experiment. Error bars (*mean*± 2SE) at the extreme *Left* of Fig. 4 apply to the 8 experiments with 28 to 32 items; error bars on the far *Right* apply to the 24-item experiments. (*A*) The mean number of stimulus items processed by an ideal detector with each individual subject's filter and with perfect target filters for pre-cued, post-cued, zero-distracter, and centroid-of-ALL trials in 11 experiments. The black line represents the performance ceiling, the total number of stimulus items in each experiment. (*B*) The number of stimulus items processed assuming subject's attention filters and also perfect target filters, all corrected for motor error—i.e., the variance of zero-distracter singleton trials (motor error) was subtracted from mean error variance in each experiment and condition.

items, and the centroid-of-ALL difference is 1.01 items. These small differences indicate an extraordinary processing capacity: For zero-distracter, pre-cued, and centroid-of-ALL multi-item trials, subjects utilized essentially all the 30 ± 2 target items displayed on the screen to compute their centroid response, and that the centroid error observed in data that include motor error is due to the subjects' inability to perfectly output the result of their centroid computations.

Discussion

How Many Centroids Can Observers Compute Concurrently?

Prior studies. (9, 22, 23) attempted to measure the number of features that can be processed concurrently in ensemble perception. The task in all three studies was estimating the number of items composed of a particular cued feature. The criterion for all three studies was the minimum number of item features for which performance was reliably different between the pre-cued and the post-cued conditions. Poltoratski and Xu (22) and Luo and Zhao (23) reported 2 and Halberda et al.(9) reported 3 concurrently processed numerosity estimates.

Figs. 3 and 5 illustrate pre-cued performance that is almost identical to zero-distracter performance. By definition, in zero-distracter trials, there is only one centroid available to be computed, the target centroid. The almost identical performance between pre-cued and zero-distracter trials indicates that, in pre-cued trials, subjects compute only the pre-cued target centroid, i.e., subjects have very good attention filters. Therefore, all that one can learn from the difference between pre-cued and post-cued performance is that more than one feature is processed in the post-cued trials. That post-cued trials have a significantly greater error than pre-cued trials (the measure of capacity in prior studies, e.g., ref. 9, 22, 23) only tells us that there may be some cost in processing more than one group of items, not that processing a particular number of groups is possible or impossible. To estimate the number of features that can be processed concurrently, we determine the extent to which performance on M-1 feature trials can predict performance on M feature trials. When performance on M feature trials exceeds the optimal extrapolation of performance on M-1 feature trials it means at least some additional capacity is displayed on M feature trials, versus that the available capacity was merely divided more finely. This capacity analysis is developed in the next section.

Subsampling analysis: Predicting target weight and errors in M-feature experiments from (M-1) and (M-2)-feature experiments. How many centroids can a subject compute simultaneously? To estimate the number of centroids that a subject can compute simultaneously, we used the data of (M-1)- and (M-2)-feature experiments to predict a subject's performance in the M-feature experiment. Specifically, subjects are assumed to have only M-1 or M-2 attention filters. This results in two kinds of M-feature trials: a) $(M-1)/M$ or $(M-2)/M$ "available-filter" trials in which subjects report a feature using the same filters as in the corresponding M-1 and M-2 feature trial, and b) $1-(M-1)/M$ or $1-(M-2)/M$ "unavailable-filter" trials in which subjects must report a feature they cannot process. Available-filter trials were analyzed as if the stimulus items for which the subjects did not have a filter were invisible. i.e., these trials are basically similar to trials in experiments with M-1 or M-2 features, and these stimuli are processed similarly in terms of target weight. On the $1-(M-1)/M$ or $1-(M-2)/M$ "unavailable-filter" trials on which subjects are presented with a feature they cannot process, they click the center of the screen. These strategies are statistically optimal for stimuli that contain items that cannot be processed.

The three possible outcomes of the subsampling analysis. When the subject's actual performance with M centroids is better than the predicted performance from M-1 or M-2 centroids, it means that subjects must have computed M centroids. When the subject's actual performance is equivalent to the predicted performance, it is ambiguous whether a) subjects could have computed just M-1 or M-2 centroids and still have produced the observed performance on M-feature trials or b) have computed M centroids but just not better than the M-1 prediction. When the subject's actual performance is worse than the predicted performance, it implies that the subject used a suboptimal centroid-computation strategy.

The review of ref. 17 in the introduction section pointed out that subjects were able to concurrently compute three centroids. Therefore, the stimuli in the present experiments which were designed to find the limits of performance contained 3 to 8 centroids. Fig. 6 displays the data and the predictions of the subsampling analysis.

The results of the subsampling analysis: The number of subject–computed centroids. For multi-item post-cued trials, Fig. 6 *A*, *C*, and *E* depict observed and predicted target weights; Fig. 6 *B*, *D*, and *F* depict observed and predicted response errors individually for three subjects when (M-1)- and (M-2)- feature experiments were used to generate predictions for M-feature experiments. Predictions were based on Efron's (24) method of stimulus resampling (200 runs); the error bars were so small they are omitted in the graph. The target weight subsampling predictions are similar to but slightly more discriminating than the response accuracy predictions, so we concentrate on target weights.

For Subject 1 (Fig. 6*A*), the predicted target weights are worse than the observed weight from $M4$ to $M7$ and are equal to $M8$. If either target weight or response accuracy indicates that a subject must have computed M centroids and the other measure is ambiguous (i.e., does not contradict the first measure), that is sufficient to demonstrate the success of an M-centroid computation. (That an M-1 prediction equals performance does not mean that M-1 was the basis of performance, only that it could have been.) This logic implies that Subject 1 is able to establish at least 7 concurrent attention filters and to compute 7 concurrent centroids because the attention filters are derived from successful centroid performance. For response errors (Fig. 6*B*), Subject 1's M-data are better than predicted from M-1 data for M = 4, 5; equivalent for M = 6, 7, and worse for M = 8. Based only on response error, Subject 1 can compute at least 5 concurrent centroids, for M = 6, 7 the response-error data are ambiguous. The failure of Subject 1 to match the (M-1)-prediction for M = 8 means that attempting to compute 8 centroids leads to worse performance than could have been achieved by consciously ignoring one of the eight presented features. Based on attention filter accuracy and not contradicted by the response error analysis, Subject 1 was able to compute 7 but not 8 concurrent centroids.

For Subjects 2 and 3, the target weight (Figs. 6 *C*, *E*) and response error (Figs. 6 *D*, *F*) predictions. Following similar reasoning as for Subject 1, we conclude that Subjects 2 and 3 can create 5 concurrent attention filters and compute 5 centroids concurrently following a brief stimulus exposure. As centroids are computed on data that are represented in salience maps, we conclude that our three subjects have at least 7,5,5 salience maps. These numbers of salience maps are significantly greater than the 3 salience maps observed by Sun et al. (17)—the maximum possible in their procedure.

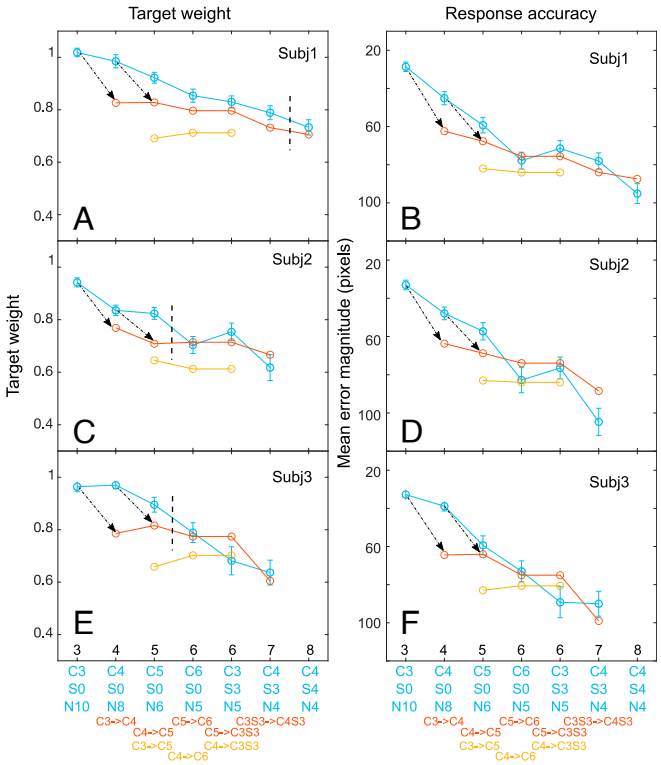

Fig. 6. A subsampling analysis of multi-item post-cued trials to determine the number of centroids that a subject computes. To demonstrate *M* computed centroids, subjects' location errors and/or target-filter weights in *M*-feature experiments must exceed predictions from (*M*-1)- or (*M*-2)-feature experiments. For all six panels, the black integers above the horizontal axis represent *M*, the number of different stimulus features. Underneath the figures, *C*, *S*, and *N* indicate the number of stimulus Colors, Shapes, and target items. In panels *A*, *C*, *E*, the ordinate is the estimated weight of the target feature; in panels *B*, *D*, *F*, the ordinate is the observed mean error in judged centroid location. Blue dots indicate a subject's mean data; the error bars indicate ± 2 standard deviations of the mean. Red symbols indicate the result when an (*M*-1)-feature experiment is used to predict performance in an *M*-feature experiment. To illustrate, the black arrows from blue *M*3 data to red *M*4 predictions and from *M*4 to *M*5 illustrate the source and target of two of the 4 or 5 predictions in each panel. The figure is too crowded to show the other arrows. The yellow symbols indicate predictions of an (*M*-2)-feature experiment. When a prediction of an (*M*-1)- or (*M*-2)- feature experiment equals or exceeds observed performance on an *M*-feature experiment, it means that a subject potentially could have processed only *M*-1 or *M*-2 features and have achieved the observed *M*-feature performance. It does not exclude the possibility that the subject can process *M* features, it just means that this particular prediction is ambiguous with respect to the number of features actually processed. Experiments on the *Left* of the vertical lines in the first column were the experiments in which subjects' performance in *M*-feature experiments exceeds the predictions from corresponding *M*-1 feature experiments.

Process Models of Pre-cued and Post-cued Centroid Trials.

Model of a post-cued centroid trial. Fig. 7 *A* and *B* are elaborations of the model in Figure 6 of reference (17), they are based on the data herein. Fig. 7*A* illustrates how the retinal image of the stimulus display in a post-cued trial is transformed into a response. Fig. 7*B* does the same for pre-cued trials. The concern here is identifying possible sources of error. Error sources in Fig. 7*A* are (1a) failure to encode stimulus items prior to the point at which the target items have to be segregated (encoding failure), (1b) the number of stimulus items that survive to be grouped; 1a and 1b are not distinguished in the subsequent error analysis and treated together as multichannel loss. 2) the precision of attention filters (% inclusion), 3) error in the computation of centroids, 4) working memory (the limited capacity to record centroids and the precision of the recording as determined by memory noise), and 5) "motor" error in producing the response. The blue areas in Fig. 7 *A* and *B* illustrate these five error sources, which are estimated in a subsequent section.

Model of a pre-cued centroid trial. pre-cued trials begin with a small fixation stimulus (the precue) that is a sample of the items whose centroid is to be reported. As in post-cued trials, it is followed by the stimulus display, a poststimulus mask, a reminder screen with a sample of items whose centroid is to be reported, and finally, the response screen. In the pre-cued model, prior availability of the precue is assumed to enable selective filtering that is so effective that only the cued items are encoded and there is no encoding noise (no perceptual loss). As illustrated in Fig. 5 *A* and *B*, the centroid accuracy of a pre-cued subset is nearly as accurate as the response to zero-distracter trials. The remarkable performance of our practiced subjects indicates that, although pre-cued centroid tasks involve some processes similar to post-cued centroid task processes, pre-cued centroids are completely different because they involve encoding only the pre-cued items.

Estimating the Sources of Error in Multicentroid Processing.

Five error sources defined. Based on the models in Fig. 7, the data from multicentroid and singleton trials enable the derivation of five component errors for each experiment that are assumed to add independently to produce the observed response errors: *1*, loss in each channel that conveys information about a particular class of stimuli from the input all the way to the memory component but not including *2*, attention filter imperfections and *3*, centroid computation errors, each of which is computed separately; *4*, memory error in storing and recalling the computed centroids; and *5*, motor error in producing a cursor response to represent the recalled memory content.

Distance and item error estimates. In the prior sections, we considered two ways of measuring error: a) the pixel distance of the response from the target, and b) the number of stimulus items an ideal detector must discard (items lost) in order to match a subject's performance level. To produce total pixel error squared for Experiments with *M* different features, E_M^2, the squared component errors of five assumed-to-be-independent stages *j* add independently. The sequence in which these stages occur is irrelevant.

$$\text{Post-cued error: } E_M^2 = \sum_{j=1,5} E_{j,M}^2. \qquad [1]$$

Unfortunately, there is no equivalent item theory to derive total item error *I* from independent component item errors. In an item theory, the items lost at each stage depend on the number and quality of items delivered by the prior stage (interdependence). However, to aid intuition about how well the stimulus information is utilized, we derive a "number of items lost" that would produce the observed "distance error' (and vice versa) after three processing levels: perceptual, memory, and motor.

To estimate the five component error sources, only the 12/19 conditions in which the analysis of Fig. 6 indicates that a subject did not use a subsampling strategy are included. Two different conditions (C6S0N5 and C3S3N5) in which there are six different item types are combined. Parameter estimates of mean-square pixel-distance errors and of items-lost errors are made individually for each subject, the averages of these are reported in the text and Fig. 8.

Motor error E_5 is the pixel-distance error observed in trials in which only one item is presented and the subject's task is to mouse-click the location of that item (Fig. 3*B*, Zero-distracter

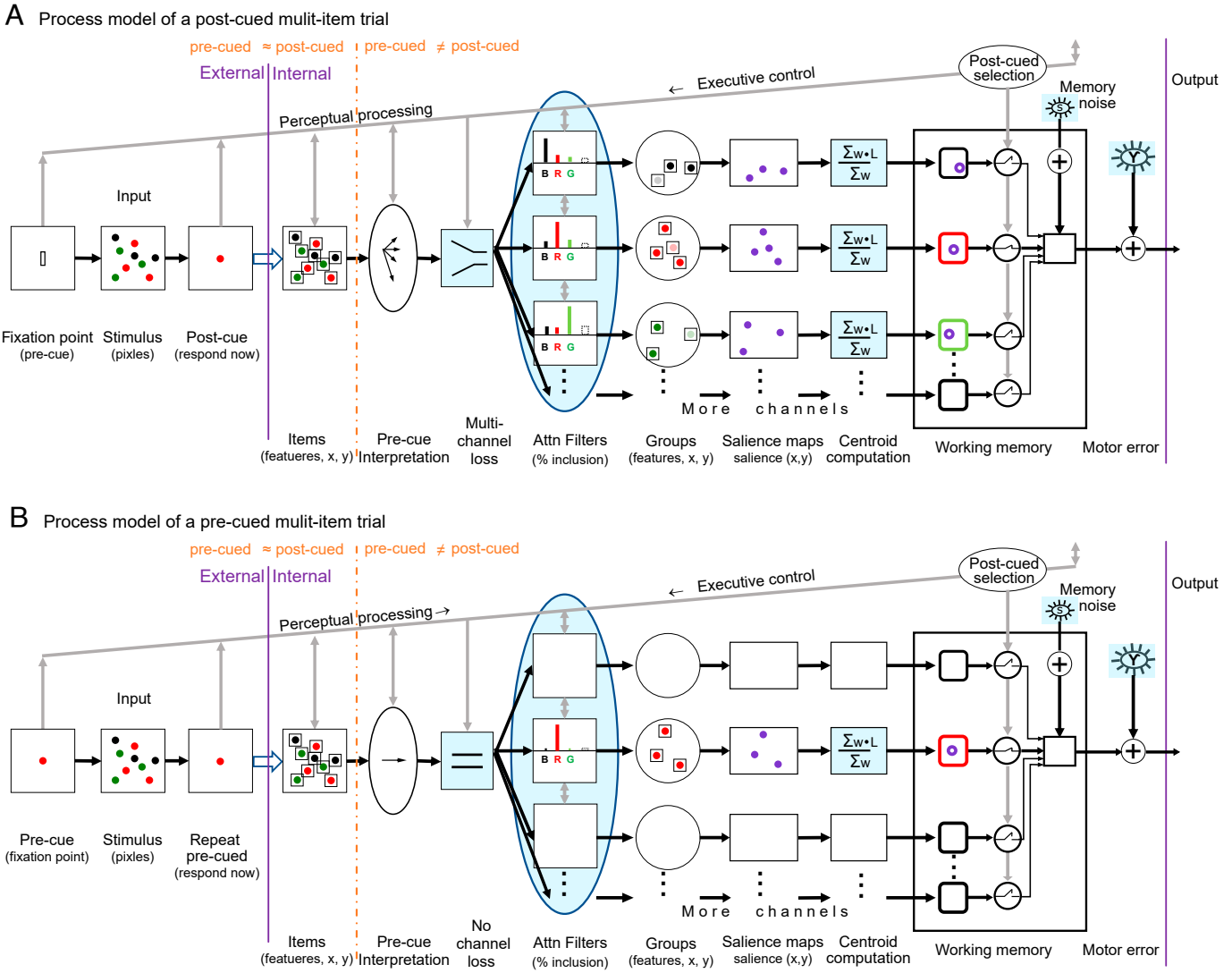

Fig. 7. Flow chart of the essential components of computational models of post-cued and pre-cued multi-item trials. (*A*) A computational model of a post-cued multi-item trial in which a stimulus array (9 dots of 3 colors in this example) is transformed into a single response, the centroid of the dots whose color matches the postcue. The first three components represent the stimulus sequence. The rectangular fixation bar indicates a post-cued trial. The stimulus image is represented as pixels, early visual processing transforms it into an array of items (indicated as small squares) each of which consists of a feature and a location. Interpretation of the precue fixation point indicates that multiple attention filters are needed. Not all stimulus items can be fully processed, some of the multichannel loss of stimulus information is represented by a decimation process. Every stimulus item has the same probability p, $0 \leqslant p \leqslant 1$, of survival. Three attention filters assign a weight to each surviving stimulus item according to its feature (color, in this example). The filter output (weighted items) constitutes a group. The content of each group—the weight and spatial location of each group item—is passed to a salience map. A centroid process computes a centroid location based on the weights and locations of the items in the salience map. In this path, feature information is associated with items only by an executive control process that binds each centroid with its filter label. The labeled centroids are transferred into working memory and perturbed by memory noise. Response output is triggered by a postcue that indicates which centroid is to be reported. A mouse-click response is made with additive motor error. Blue color in a component indicates it is a source of information loss in the five-error-source model. (*B*) Pre-cued multicentroid model. The components and processes of the precue computational model are the same as the postcue computational model except for 1) Pre-cued trials begin with a precue indicating which feature-centroid to report, therefore only one attention filter is needed, and it assigns nearly all weight to items whose feature matches the precue. 2) Only one channel is used, and only one centroid is computed. The single centroid stored in working memory is the response. 3) Because the precue defines the feature of a small group of to-be-attended items, there is very little channel loss (only cued items are processed to the level of entering into a group) and little memory loss (only one centroid needs to be recorded).

Singleton trials. There are insufficient trials to discriminate motor error in different conditions, so they are all combined in these estimates: $E_5 = 14.8$, 14.6, and 13.4 pixels for the three subjects. For reference, the diameter of an individual stimulus dot is 15 pixels (0.39 deg for subjects 1 and 3; 0.43 deg for subject 2). Motor error squared for 10 conditions is shown Fig. 8 *E* and *F*; the motor error item loss is in Fig. 8 *I* and *J*. The motor error proportion of the total error variance is much smaller than in the item analysis for two reasons: 1) Squaring distance errors give a much greater relative weight to larger versus smaller errors than does the item analysis. 2) The item analysis asks "if there were no motor error how much better would the total item score have been." The raw item score indicates that at least 12 stimulus items have been processed in every post-cued multicentroid condition. To further improve this already good accuracy to the level it would be without motor error requires processing more additional items than intuition suggests (*SI Appendix*, Fig. S3). This discrepancy between two different measurements (distance, item) illustrates a fundamental measurement principle: The relative importance of various

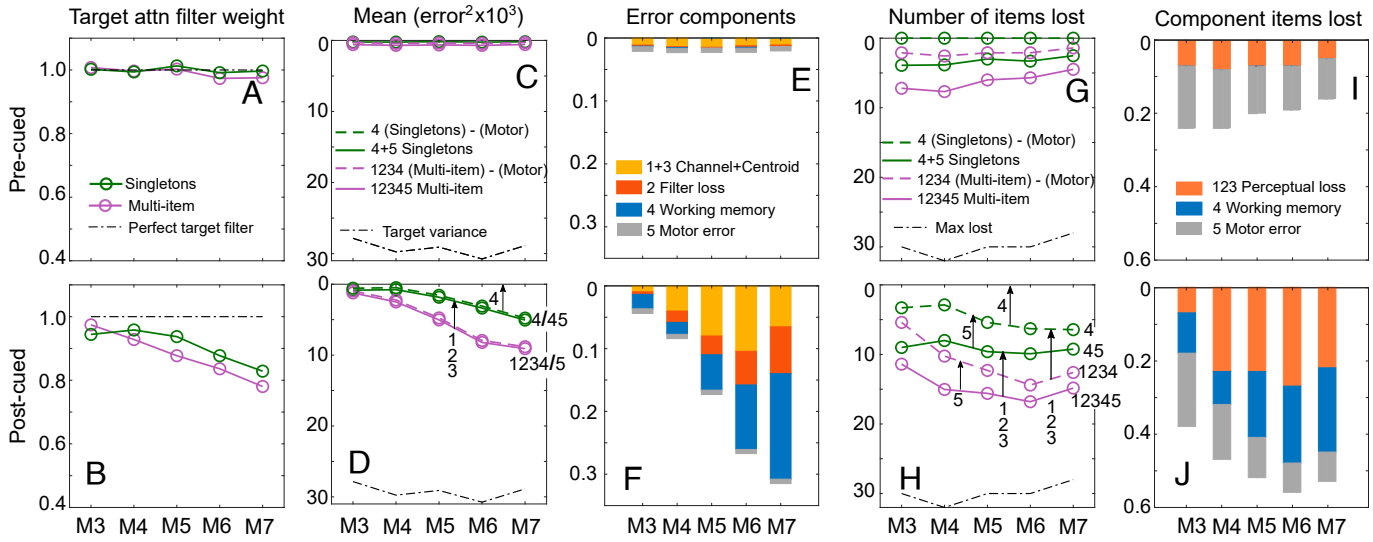

Fig. 8. Estimates of five additive component error sources in multicentroid computations: Perceptual errors include 1) channel loss 2) attention-filter error, 3) centroid computation error; postperceptual errors include 4) storing and recalling the computed centroid (working memory, and 5) motor error in producing a cursor response to represent the memory content. For all ten panels, the x-axis is the number of different stimulus features. The *Upper* row represents pre-cued trials; the bottom row, post-cued trials. Data are only from subjects and multi-item conditions that exceeded subsampling predictions (Fig. 6). (*A* and *B*) Attention filter weight of targets in multi-item and singleton trials. (*C* and *D*) Mean squared error in units of *display_pixels*^{2} for multi-item and singleton trials with and without correcting for motor error (tiny differences are not visible). The numbers under the vertical arrows represent the component(s) measured by the arrows except motor error (5) which is too small to display clearly. (*E* and *F*) Error component variances as a fraction of the target location variance. (*G* and *H*) The number of stimulus items lost for singleton and multi-item trials, with and without correcting for motor error. (*I* and *J*) Component item loss as a fraction of total stimulus items.

sources of error depends on the particular choices of how error is measured.

Centroid computation errors $E^2_{3,M}$ are estimated from trials in which only 4 to 10 target items and zero distracters are presented (Fig. 3*A*, multi-item zero-distracters trials). The only other significant source of error in these trials is motor error which is subtracted from the response error to yield the squared centroid computation error $E^2_{3,M}$. Our experienced subjects were so accurate in estimating single centroids that these centroid errors are too small to represent on the scales used for the other errors in Fig. 8. Centroid computation errors are probably greater when multiple centroids have to be computed, but we have not found an efficient way to estimate them here.

Attention filter discrimination of targets from nontargets is indexed by the fraction of the total filter weight of 1.0 that is assigned to the target filter. The pre-cued attention filter weights in Fig. 8*A* represent virtually perfect target discrimination. For post-cued trials (Fig. 8*B*), filter weights for (multi-item, singletons) decrease from (0.97, 0.94) to (0.78, 0.83) as the number of target types M increases. The distance errors $E^2_{2,M}$ that would be produced by these attention filters are shown Fig. 8 *E* and *F*.

Memory loss $E^2_{4,M}$ in Experiment *M* is the error difference between singletons and motor error averaged over the *M* target conditions. The only error sources for post-cued singletons are remembering the feature and location of the to-be-cued single item until the postcue appears, and then moving the cursor to the remembered location of cued item:

$$\begin{aligned} E^2_{4,M} &= \text{Singleton error}^2_M - \text{Motor error}^2 \\ &= (E^2_{4,M} + E^2_5) - E^2_5 \\ I_{4,M} &= (I_{4,M} + I_{5,M}) - I_{5,M}. \end{aligned} \quad [2]$$

In both the distance error and item error analysis, memory loss increases with the number of locations to be remembered.

Channel loss $E^2_{1,M}$ is the residual after all the other error losses have been accounted, i.e., the total error minus (filter error + centroid error + singleton error):

$$E^2_{1,M} = \sum_{j=1,5} E^2_{j,M} - E^2_{2,M} - E^2_{3,M} - (E^2_{4,M} + E^2_5). \quad [3]$$

Fig. 8 shows the result of the above estimates of the channel, filter, memory, and motor components of response error in the multicentroid post-cued task. The general summary is that channel and working memory are the major sources of error in the multicentroid task according to both the distance and item analyses. In the distance error analysis, there also is appreciable filter loss and negligible motor error. In the item analysis, filter loss was not isolated, it was incorporated into perceptual loss, whereas motor error was appreciable.

Updated Representation of the Salience System. The process by which items that have a similar color or a similar shape are isolated and separated as a group is not simple, but it is simple to describe and its effectiveness is measured by the selectivity of attention filters. The brain process of computing the centroid of the grouped items has to be independent of the identity of the items because there is an infinity of different possible items. This independence of item content is accomplished by representing the contents of a group of similar items on a salience map that records only the location of items and a real number quantity (salience) at that location. Salience maps were originally proposed to represent the importance (salience) of locations. A winner-take-all computation on the map found the most salient location (12), and that location was then used to guide subsequent processes, such as visual, search, eye movements, and pattern identification (Fig. 1 in *Introduction*).

Salience maps have been proposed as the basis of the computations other than processing priority: attention-generated

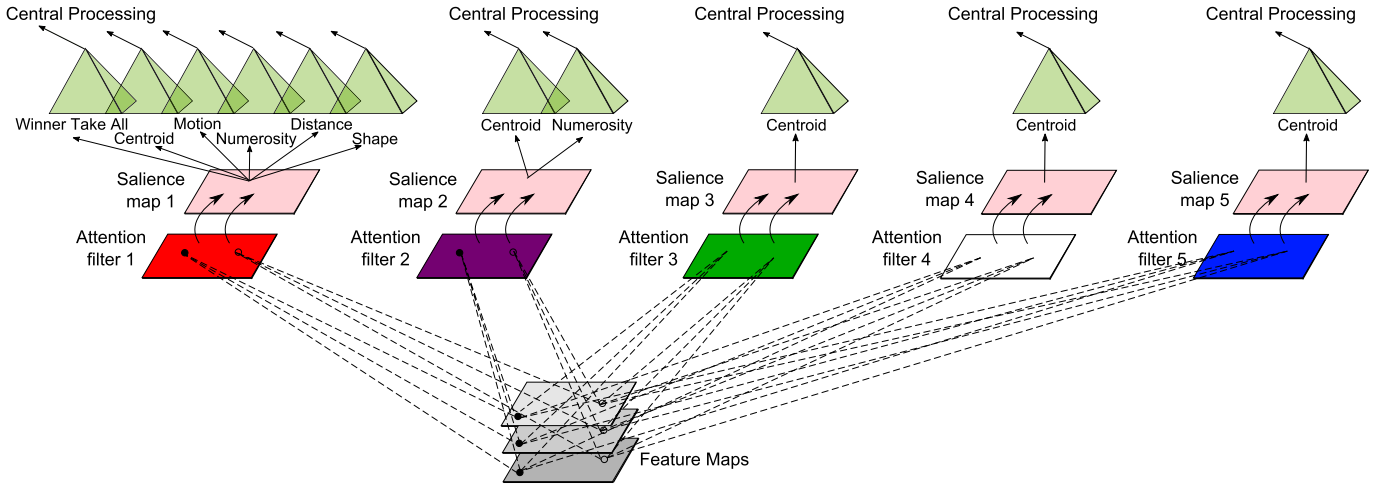

Fig. 9. Updated representation of the human salience system showing five semiindependent salient maps and the processes that until now have been demonstrated to utilize them. Stimulus information is processed by various feature maps, and the outputs of feature processing are directed to attention filters. Five attention filters are illustrated for the conditions of Expt. 3 in which 6 dots in each of 5 different colors were presented. After the stimulus presentation, subjects were cued to report the centroid of just one set (C5, S0, N6). Filter outputs (the locations and strengths of surviving items) are directed to salience maps. In all the experiments described here, the outputs of the salience maps were directed to centroid computations, but the figure also shows the other computations that until now have been demonstrated to utilize salience maps.

motion (25), distance estimation between two points in the frontal plane (26), letter and shape identification (27) and, in the present case, a centroid computation (17). In contrast to Fig. 1, our present understanding of salience maps is represented in Fig. 9 which shows the 5 salience maps that all our subjects demonstrated, and the various uses of salience maps that have been demonstrated by this and the studies cited above.

Conclusions

The previous sections demonstrated how subjects perform when they view brief exposures of displays composed of 28 to 32 items composed of 3 to 8 different features (colors or shapes). The task was to judge the centroid of just the items composed of one randomly designated feature. The feature to be reported was made known to the subject either 0.5 s before the display (precue) or only after a postexposure masking field that followed the stimulus exposure (postcue). The results, centroid estimation distance errors, were interpreted in terms of the number of stimulus items that a statistical ideal observer with perfect knowledge of item identity and location would require to match each subject's performance. In pre-cued trials, as the number of centroids increased from 3 to 8 in displays of 28 to 32 items, performance remained approximately constant, an ideal observer would have to perfectly process 25 items to match the subjects' performances, and about 30 items if motor error in producing a response is taken into account. On the other hand, on post-cued trials, to match the subject's performance on displays with three centroids, the ideal observer required perfect knowledge of 17 items; for displays containing 7 centroids, performance declined to 10 items. Performances on M-centroid trials that exceeded optimal predicted performance (based on subsampling M-1 centroids) indicated that all subjects could compute at least M = 5 centroids and one subject could compute 7 centroids on post-cued trial. When the critical feature of the items whose centroid was to be reported was shown to the subject before a trial (pre-cued condition) their performance was almost the same as on trials in which distracter items were physically removed and only target items were displayed. This near equivalence demonstrates nearly perfect attention filters on pre-cued trials.

A computational model that characterizes the subjects' centroid performances (Fig. 7) consists of perceptual processing (converting pixel images into items composed of features and a location, channeling the items into attention filters, grouping the filter outputs and computing the group centroid), entering the centroids into working memory, and converting the cued memory contents into a motor cursor response. The amount of noise introduced at each of these steps was estimated in terms of response error (pixel distance of response to target) and the number of stimulus items lost at each stage. The two main sources of error were channel loss in processing so many items and memory loss. The distance analysis also found significant attention-filter error, the item analysis found significant motor error, and centroid computation errors were minimal. In the context of the models and the aforementioned error analysis, the accuracy of our subjects in locating the centroids in briefly exposed arrays as accurately as they did requires that the three subjects have processing capabilities equivalent to at least 7,5,5 concurrently available salience maps.

Materials and Methods

Subjects. The first and second authors (Subj1, Subj2) participated in the experiment. Subj3 was naïve to the purpose of the experiments. The three subjects (two females and one male) ranged from 23 to 40 y of age. All subjects reported having normal or corrected-to-normal vision. Methods were approved by the UC Irvine Institutional Review Board, and all subjects signed informed consent forms. The protocol and signed consent forms were approved by the UCI IRB. Subj1 and Subj2 were experienced in doing centroid tasks. Subj3 had no experience in doing centroid tasks; therefore, prior to starting Expt. 1, Subj3 had four 1-h centroid training sessions.

Apparatus. The experiment was conducted on an iMac intel computer installed with MATLAB 2018b and Psychtoolbox-3 software. For Subj1 and Subj3, the stimuli were presented on an ASUS ProArt Display monitor with 1,920 × 1,200 resolution at a refresh rate of 60 Hz, mean luminance 62.4 cd/m2. The monitor screen was 51.8 cm wide × 32.4 cm high, pixels were 0.27 mm × 0.27 mm. For Subj2, the stimuli were presented on a Samsung SyncMaster 2233BW Display monitor with 1680 × 1050 resolution at a refresh rate of 60 Hz, mean luminance 32.5 cd/m2. The monitor screen was 47.6 cm wide × 29.6 cm high, pixels were 0.282 mm × 0.282 mm.

Number of Trials. An experiment consisted of four types of blocked sessions: one pre-cued session in which a precue indicated the target feature, one post-cued session in which the postcue was shown after stimulus and mask, one zero-distracter session in which only the target feature of items was displayed, and one centroid-of-ALL session in which all features were targets. Singleton trials were interleaved in pre-cued, post-cued, and zero-distracter sessions. In both precue and postcue sessions, each feature was cued in 50 multi-item trials + 20 singleton trials; all features were interleaved in a mixed-list design. In zero-distracter sessions, all features were interleaved, and each feature was tested in 25 multi-item trials + 5 singleton trials. Centroid-of-ALL sessions had 50 trials. The 217 data points of Fig. 3 *A* and *B* summarize 32,210 trials.

Stimuli. For pre-cued, post-cued, and centroid-off-ALL trials in each of 8 experiments, the stimulus display contained 28 to 32 interleaved items of *M* different features. Features were either *M* different dot colors, *M* shapes, or a mixture of colored dots and shapes. The colors of dots were red, blue, green, white, purple, and black. The shapes were Gabor patches, plus signs, letter A, and square boxes. All features within an experiment had the same number of items. The number *M* was 3 for Expt. 1a and 1b, 4 for Expt. 2a, 2b, 2c, and 2d, 5 for Expt. 3, 6 for Expt. 4a and 4b, 7 for Expt. 5, and 8 for Expt. 6. (For 3 experiments (Expt 2b, 2c, and 2d) stimuli contained 24 items, composed of either 4 colors or 4 shapes, or a 2+2 mixture.) For pre-cued and for post-cued trials, on each trial, a random one of the features was designated as the target feature. For zero-distracter trials, the stimulus display contained only one randomly selected feature from the *M* features of the corresponding experiment. For centroid-of-ALL trials, all *M* features were targets. In addition to multi-item stimuli, singleton stimuli consisted of just one item from each feature. The position of each singleton item was the centroid of the feature from the corresponding multicentroid trial. Fig. 2 and *SI Appendix*, Fig. S1 show sample stimuli.

Dimensions. Stimuli were displayed for 300 msec within an 800 by 800 pixel-wide square (Fig. 2) that spanned 20.4 deg of visual angle (dva) for Subj1, Subj3 and 22.0 dva for Subj2. The diameter of colored dots (filled circles) was 15 pixels, spanning 0.39 dva for Subj1 and Subj3 and 0.42 dva for Subj2. Stimulus items in any shape other than filled circles were inscribed inside invisible circles of 28-pixel diameter, spanning 0.72 dva for Subj1 and Subj3 and 0.78 dva for Subj2 to prohibit items from overlapping. All stimuli were clearly visible on the gray background.

Centroid standard deviation and item dispersion. In order to produce useful data, different features have to have sufficiently different centroids. To achieve centroid separation, we first randomly chose centroid locations and then varied item locations around the chosen centroid as follows: Stimulus displays contained *M* different features with *N* items per feature for a total of *M* * *N* items. Except for Expt. 1a, nonroving C3S0N*10, in each stimulus display, locations of the items were determined by a three-step process: 1) *M* initial centroid locations were drawn from a bivariate normal distribution centered on the center of the screen and with a SD of 80 pixels. 2) *N* item locations were drawn from a bivariate normal distribution which centered on each of the *M* centroids. The sample locations were divided by the SD of the *N* samples and then multiplied by a fixed constant to ensure that the locations of all generated arrays would have the same SD (dispersion) of pixels on every trial. 3) The location of an item was resampled if its location occurred within 5 pixels of another item. The overall centroid SD that resulted from the initial location of each centroid followed by the repositioning of the centroid due to the sampling and resampling of the items within each individual feature was 170.0 $\pm$ 10.6 pixels for the 7 experiments containing 28 to 32 stimulus items; 121.2 $\pm$ 3.5 pixels for the 3 experiments containing 24 stimulus items. For Expt. 1a, nonroving C3S0N*10 trials, the procedure of generating the locations of items was the same as the procedure described above, except that the 3 initial centroid locations were the center of the screen and the centroid SD was 56.6 $\pm$ 2.6 pixels.

Data, Materials, and Software Availability. Anonymized Matlab files data have been deposited in https://github.com/Lingyu-Gan/Deriving-the-number-of-Salience-maps.git (28). All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Ruth Rosenholtz and Misha Pavel for useful suggestions.

1. D. Ariely, Seeing sets: Representation by statistical properties. *Psychol. Sci.* **12**, 157–162 (2001).
2. D. Whitney, A. Yamanashi Leib, Ensemble perception. *Annu. Rev. Psychol.* **69**, 105–129 (2018).
3. S. C. Chong, A. Treisman, Representation of statistical properties. *Vision Res.* **43**, 393–404 (2003).
4. S. Drew, C. Chubb, G. Sperling, Quantifying attention: Attention filtering in centroid estimations. *J. Vision* **9**, 229–229 (2009).
5. M. Attarha, C. M. Moore, The capacity limitations of orientation summary statistics. *Atten. Percept. Psychophys.* **77**, 1116–1131 (2015).
6. J. M. Harris, S. P. McKee, S. N. Watamaniuk, Visual search for motion-in-depth: Stereomotion does not 'pop out' from disparity noise. *Nat. Neurosci.* **1**, 165–168 (1998).
7. B. Bauer, Does Stevens's power law for brightness extend to perceptual brightness averaging? *Psychol. Rec.* **59**, 171–185 (2009).
8. N. Demeyere, A. Rzeskiewicz, K. A. Humphreys, G. W. Humphreys, Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia* **46**, 2861–2864 (2008).
9. J. Halberda, S. F. Sires, L. Feigenson, Multiple spatially overlapping sets can be enumerated in parallel. *Psychol. Sci.* **17**, 572–576 (2006).
10. J. Haberman, D. Whitney, Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* **17**, R751–R753 (2007).
11. P. Sun, C. Chubb, C. E. Wright, G. Sperling, Human attention filters for single colors. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E6712–E6720 (2016).
12. C. Koch, S. Ullman, Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol.* **4**, 219–227 (1985).
13. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998).
14. N. Bruce, J. Tsotsos, Saliency based on information maximization. *Adv. Neural Inf. Process. Syst.* **18**, 155–162 (2005).
15. L. Itti, P. Baldi, Bayesian surprise attracts human attention. *Vision Res.* **49**, 1295–1306 (2009).
16. H. J. Seo, P. Milanfar, Static and space–time visual saliency detection by self-resemblance. *J. Vision* **9**, 15–15 (2009).
17. P. Sun, V. Chu, G. Sperling, Multiple concurrent centroid judgments imply multiple within-group salience maps. *Atten. Percept. Psychophys.* **83**, 934–955 (2021).
18. G. Sperling, The information available in brief visual presentations. *Psychol. Monogr.: General Appl.* **74**, 1 (1960).
19. G. Sperling, A model for visual memory tasks. *Hum. Factors* **5**, 19–31 (1963).
20. U. Neisser, *Cognitive Psychology* (Appleton-Century-Crofts, New York, NY, ed. Classic, 1967).
21. P. Sun, C. Chubb, C. E. Wright, G. Sperling, High-capacity preconscious processing in concurrent groupings of colored dots. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E12153–E12162 (2018).
22. S. Poltoratski, Y. Xu, The association of color memory and the enumeration of multiple spatially overlapping sets. *J. Vision* **13**, 6 (2013).
23. A. X. Luo, J. Zhao, Capacity limit of ensemble perception of multiple spatially intermixed sets. *Atten. Percept. Psychophys.* **80**, 2033–2047 (2018).
24. B. Efron, Computers and the theory of statistics: Thinking the unthinkable. *SIAM Rev.* **21**, 460–480 (1979).
25. Z. L. Lu, G. Sperling, Attention-generated apparent motion. *Nature* **377**, 237–239 (1995).
26. L. Gan, P. Sun, G. Sperling, Frontal-plane distance judgments between two equal-size items are made on the basis of a salience map. *J. Vision* **21**, 2828 (2021).
27. G. Sperling, L. Gan, Two-dimensional shape perception is based on a salience map. *J. Vision* **22**, 17 (2022).
28. L. Gan, Deriving the number of salience maps. Github. https://github.com/Lingyu-Gan/Deriving-the-number-of-Salience-maps. Deposited 3 May 2023.

PNAS
www.pnas.org

Supplementary Information for

Deriving the Number of Salience Maps an Observer has from the Number and Quality of Concurrent Centroid Judgements

Lingyu Gan, Peng Sun, George Sperling

Lingyu Gan & George Sperling.
E-mail: lingyug1@uci.edu or sperling@uci.ed

This PDF file includes:

Supplementary text
Figs. S1 to S5

Supporting Information Text

Estimating a subject's attention filters

The model for estimating a subject's attention filters assumes that the only source of error in centroid computation is an imperfect attention filter. To compute a centroid response based on an attention filter, each item is weighted according to the filter value of its feature. The predicted response is the centroid of the filter-weighted items as described below.

For a trial j that has M different features and N items per each feature, let $W_T(m)$ be the weight assigned to feature m when T is the target feature, let $X_m(j)$ and $Y_m(j)$ be the x- and y-coordinates of the mean location (centroid) of items whose feature is m, and $P_{x,T}(j)$ and $P_{y,T}(j)$ be the x- and y-coordinates of a predicted response. Equations 1a and 1b describe the subject's predicted response on trial j:

$$P_{x,T}(j) = \sum_{m=1}^{M} W_T(m) \cdot X_m(j) \quad [1a]$$

$$P_{y,T}(j) = \sum_{m=1}^{M} W_T(m) \cdot Y_m(j) \quad [1b]$$

$W_T(m)$ is constrained by:

$$\sum_{m=1}^{M} W_T(m) = 1 \quad [2]$$

In an experimental condition, for a set of J trials that have the same M features and the same target feature T, a subject's attention filter is the M values of $W_T(m)$ that minimizing the sum of the squared distances between the predicted and observed responses. Denote the x- and y-coordinates of a subject's observed response on trial j as $R_{x,T}(j)$ and $R_{y,T}(j)$. The estimated attention filter is the values of $W_T(m)$ as defined in Equations 1,2 that minimize "Loss" in Equations 3 and thereby maximize the likelihood of $W_T(m)$.

$$Loss = \sum_{j=1}^{J} [(P_{x,T}(j) - R_{x,T}(j))^2 + (P_{y,T}(j) - R_{y,T}(j))^2] \quad [3a]$$

$$= \sum_{j=1}^{J} [(\Delta X_{j,T})^2 + (\Delta Y_{j,T})^2] \quad [3b]$$

Minimizing $Loss$ in Equations 3a,b is equivalent to maximizing the likelihood L in Equation 4:

$$L = \prod_{j=1}^{J} \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(\frac{\Delta X_{j,T}}{\sigma})^2 - 2\rho(\frac{\Delta X_{j,T}}{\sigma})(\frac{\Delta Y_{j,T}}{\sigma}) + \frac{\Delta Y_{j,T}}{\sigma})^2]} \quad [4]$$

In Equation 4, ρ is the correlation between ΔX_T and ΔY_T. Taking the log of both sides of Equation 4 yields the log-likelihood ratio $\ln(L)$ which is the normal computational form.

$$\ln(L) = -J[\ln(2\pi) + 2\ln\sigma + 0.5\ln(1-\rho^2)] - \frac{1}{2\sigma^2(1-\rho^2)} \sum_{j=1}^{J} [(\Delta X_{j,T})^2 + (\Delta Y_{j,T})^2 - 2\rho\Delta X_{j,T}\Delta Y_{j,T}] \quad [5]$$

Attention filter weights as estimated by Equation 5 were separately estimated for each experiment, each condition, and each subject to maximize Equation 5 by using a Markov Chain Monte Carlo (MCMC) algorithm constrained by Equation 2.

Sample stimuli

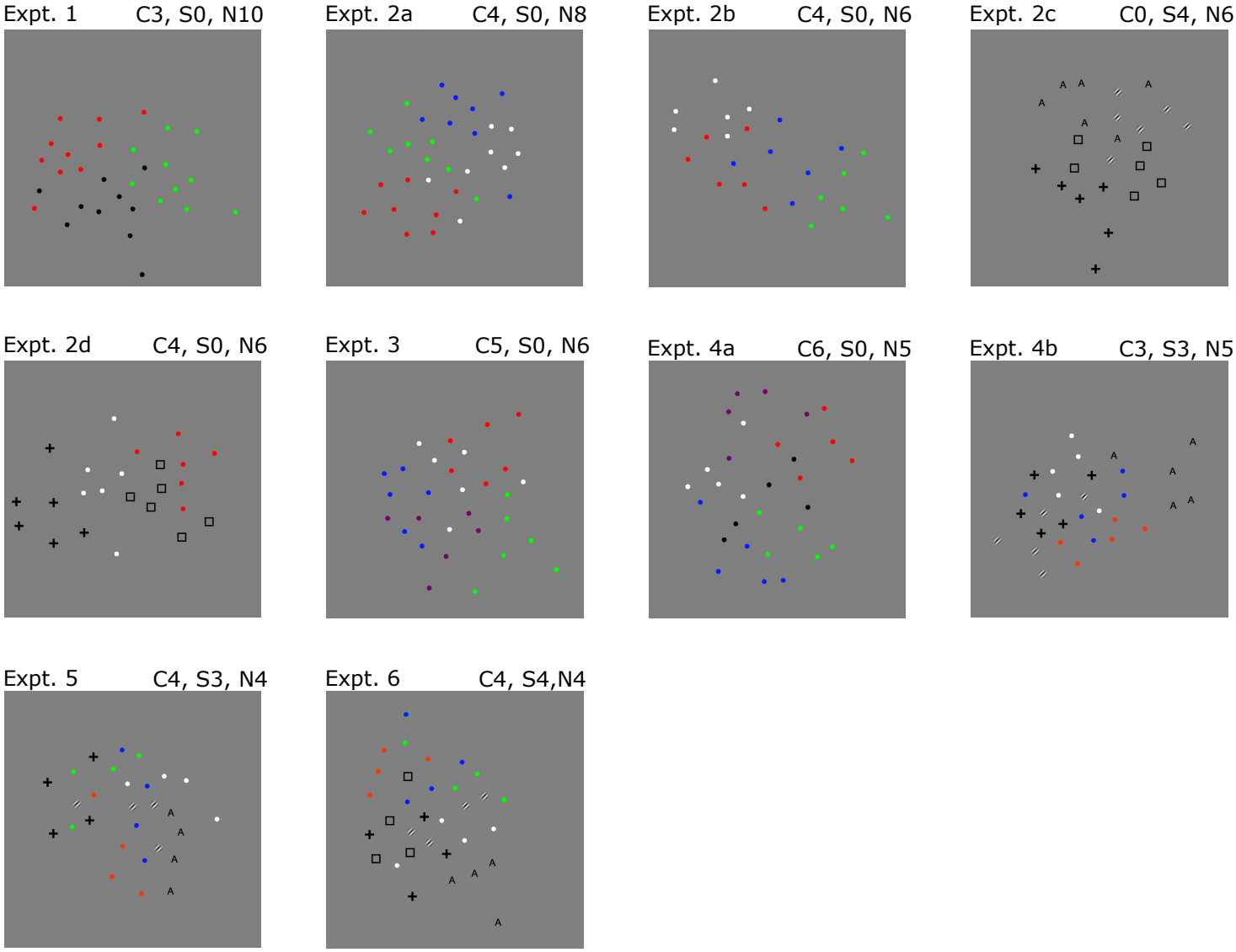

Fig. S1. Sample stimuli for each experiment. 'C' represents the number of different colors of dots in the stimulus display, 'S' represents the number of different shapes, and 'N' represents the number of items in each feature class.

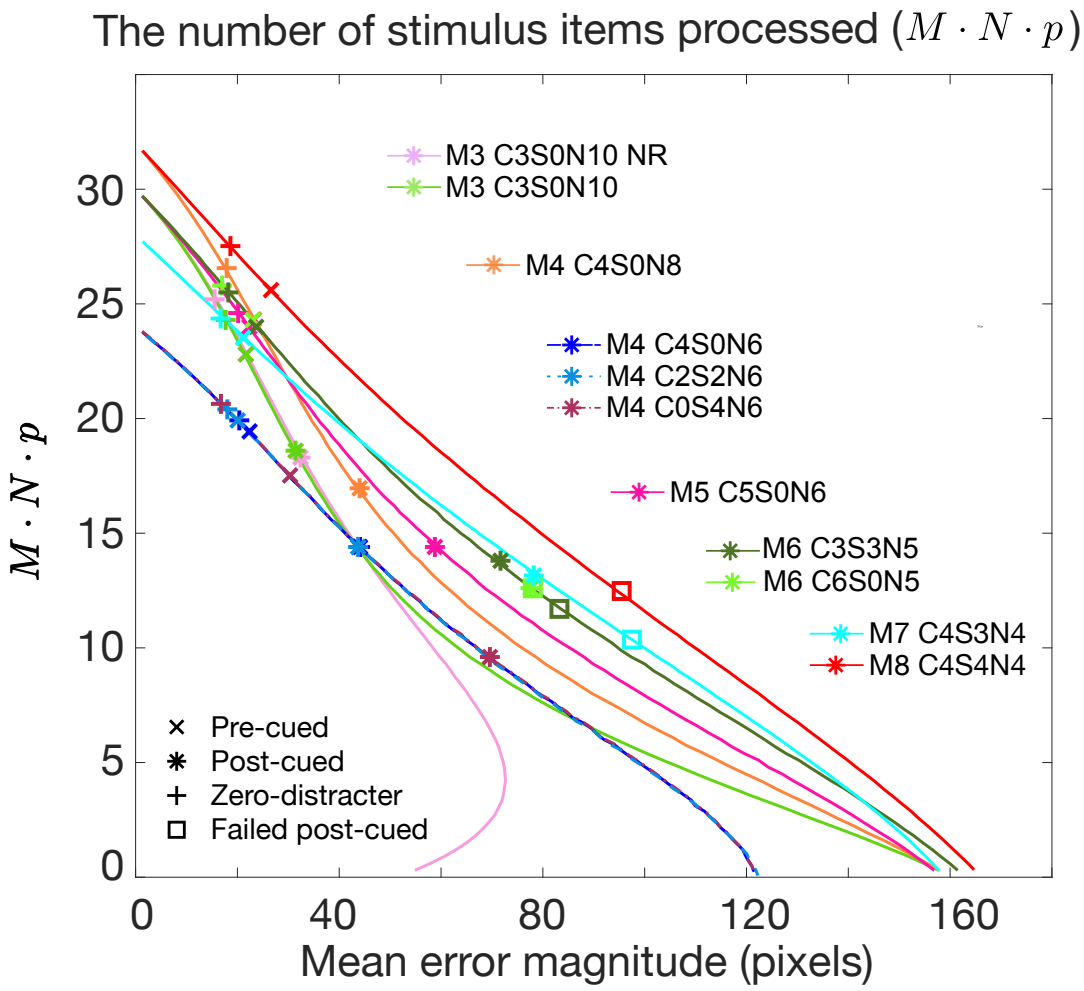

Fig. S2. The number of stimulus items processed by an ideal observer (ordinate) as a function of the mean response error magnitude (abscissa) for 3 kinds of trials in 11 experiments. Text figure 4b shows $N \cdot p$; figure S2 shows $M \cdot N \cdot p$. The data points show $M \cdot N \cdot p$, the least number of stimulus items that must be processed by an ideal observer with a perfect target filter to match a subject's performance. The colored lines represent the number of stimulus items processed by an ideal observer as a function of the mean response error magnitude for each of the 11 experiments that used 8 different stimulus configurations. Data points indicated by 'x', '$*$', and '+' represent the mean error magnitude averaged over subjects' data from an *M*-feature experiment in which performance was better than could be predicted from the corresponding *(M-1)*-feature experiment. Squares represent post-cue data from an *M*-feature experiment that failed to exceed the prediction from the corresponding *(M-1)*-feature experiment.

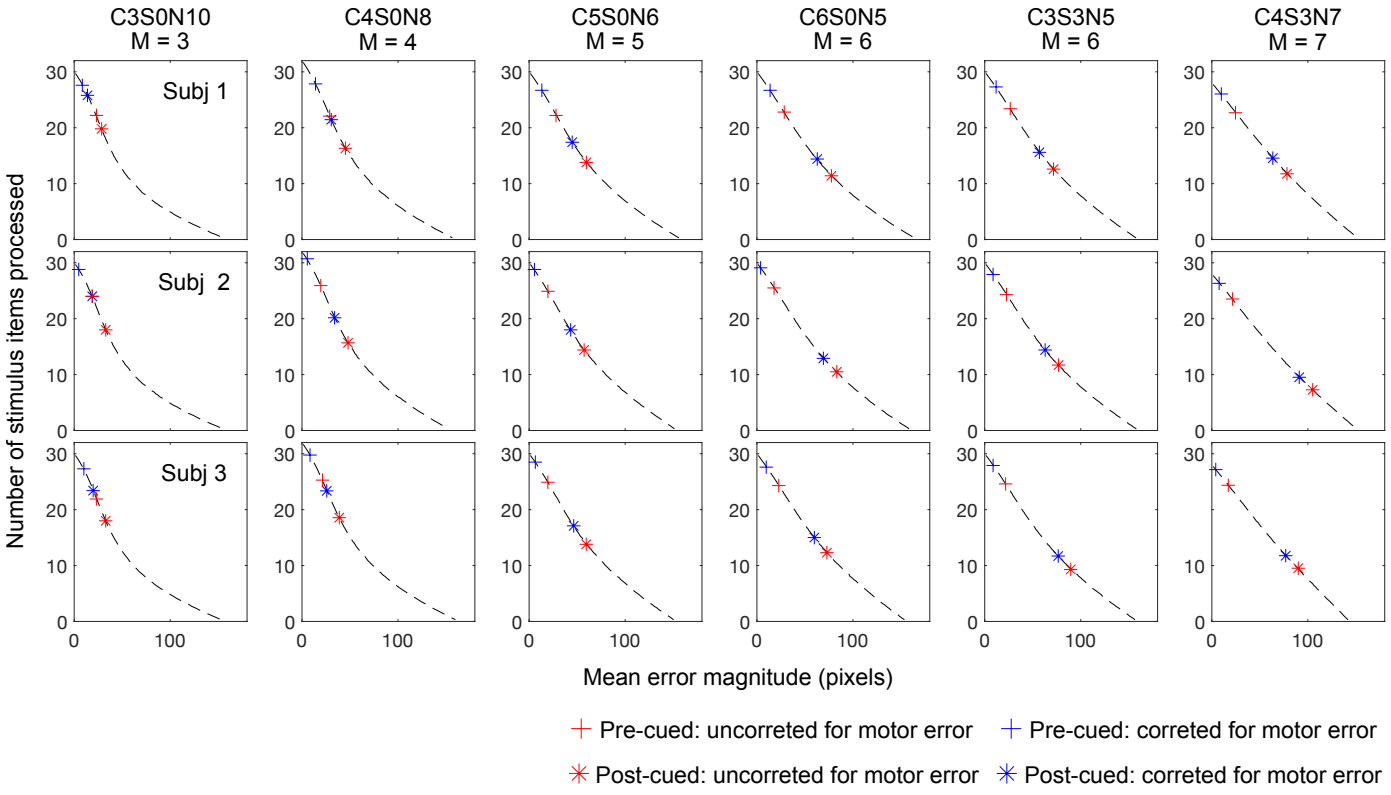

Fig. S3. Number of stimulus items processed in multi-item pre-cued and post-cued trials, uncorrected and corrected for motor error. Each row represents a subject and each column represents one experiment. The dashed black curve in each panel represents the number of stimulus items processed by an ideal observer (ordinate) as a function of the mean response error magnitude (abscissa). The data points show $M \cdot N \cdot p$, the least number of stimulus items that must be processed by an ideal observer with a perfect target filter to match a subject's performance. Red points represent the data uncorrected for motor error Blue points represent data corrected for motor error. '$*$' and '+' represent multi-item post-cued and pre-cued trials.

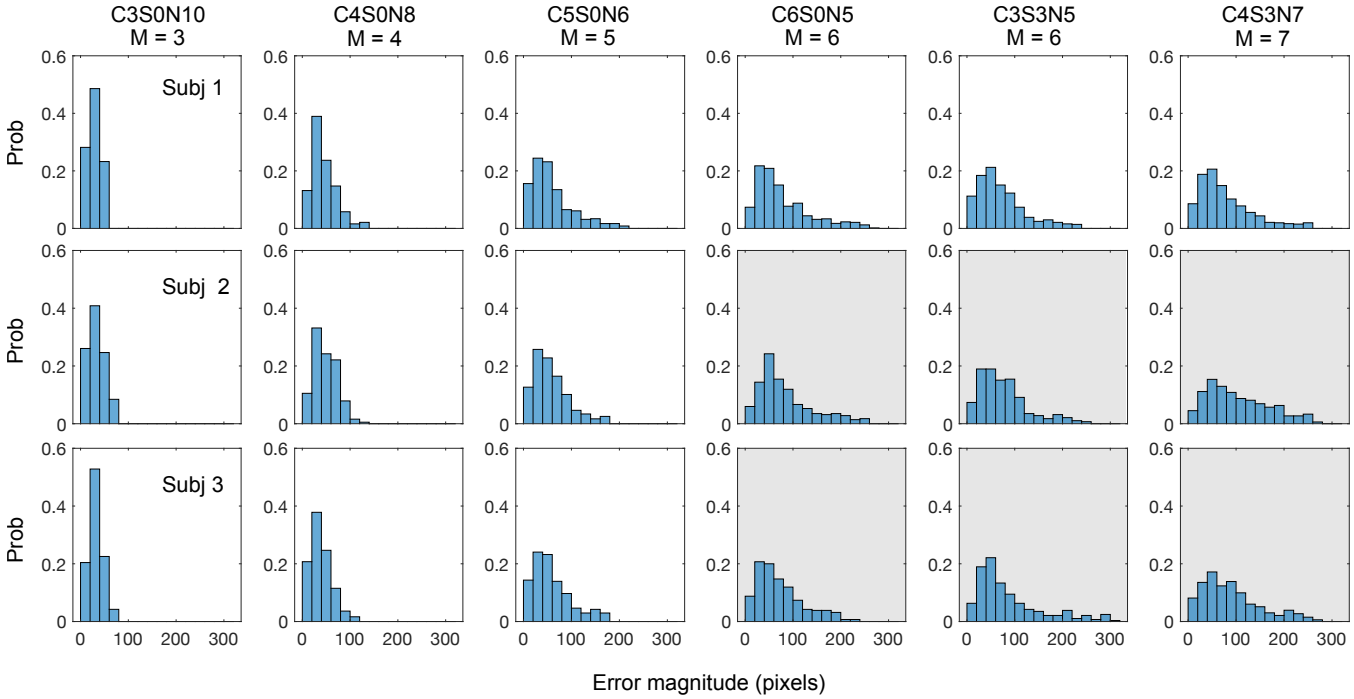

Fig. S4. Histogram plots of error magnitude of multi-item post-cued trials. Each row represents a subject and each column represents one experiment. A gray background indicates data failing to exceed the subsampling strategy prediction.

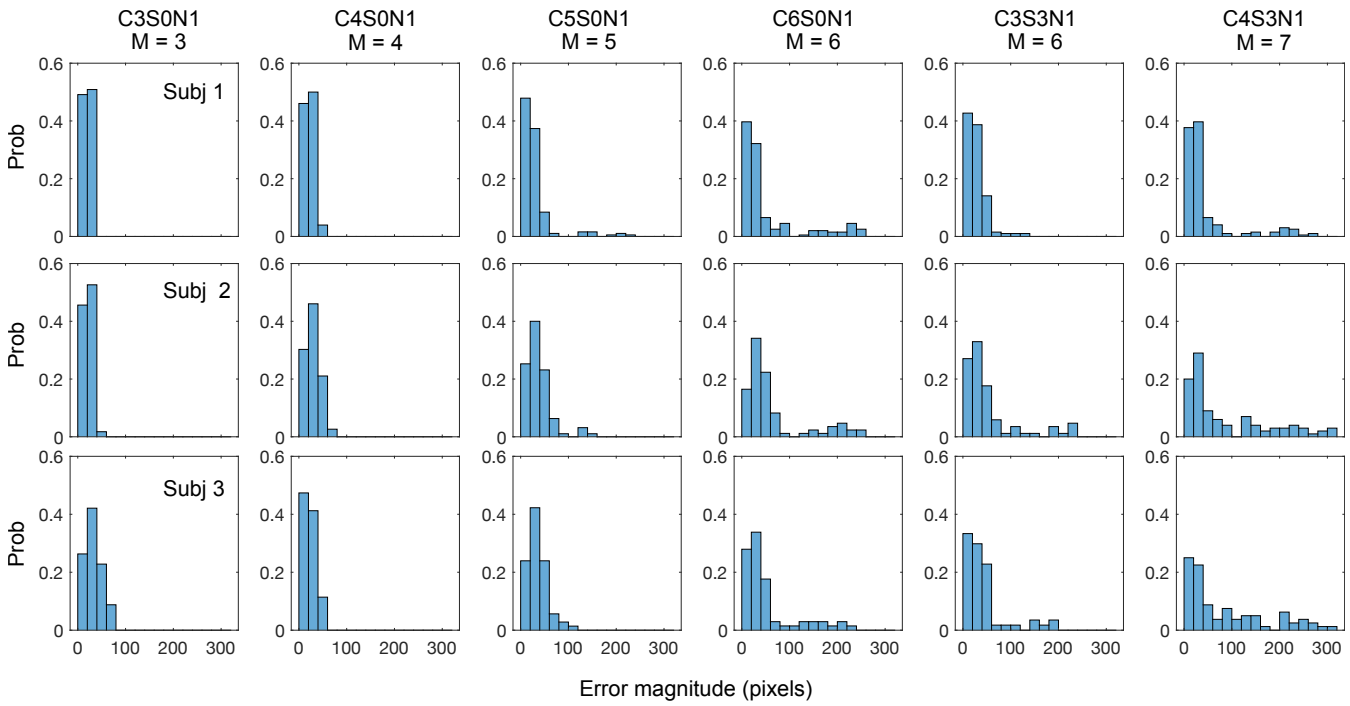

Fig. S5. Histogram plots of error magnitude of singletons post-cued trials for 3 subjects in six experiments. Each row represents a subject and each column represents one experiment.