Multiple concurrent centroid judgments imply multiple within-group salience maps

Peng Sun[1] · Veronica Chu[1] · George Sperling[1]

Abstract

Subjects viewed a brief flash of 8–24 dots of either two or three colors randomly arrayed. Their task was to move a mouse cursor to the centroid (center-of-gravity) of each color in a pre-designated order. Conventional and idea-detector analyses show that subjects accurately judged all three centroids utilizing an astounding 13/24 stimulus dots, with only a modest loss of accuracy compared to judging a single-predesignated color centroid. The ability to concurrently compute three centroids is important because it is believed that centroid judgments are made on salience maps that record only salience and are ignorant of the features that produced the salience. Our explanation, instantiated in a computational model of salience processing, is that subjects have three salience maps. Dots are initially segregated into three groups according to color, then each color-group is recorded on a different salience map to compute a centroid. In Part 2, the data are analyzed in terms of Attention Operating Characteristics to characterize impairments in subjects' color-attention filters (mostly insignificant) and encoding efficiency (20% drop for the hardest task) in making multiple versus single centroid judgments. A new, more sensitive analysis measured five sources of subject error variance, four independent, additive sources of error variance: imperfect color-attention filters; a Bayesian-like bias towards a central tendency; storage, retrieval, and cursor misplacement error; a large residual error due mostly to inefficient encoding; and fifth, an interactive source – error in all four components that increases when multiple centroid judgments versus a single centroid judgment are required on each trial.

Significance statement

An important brain process is a salience map, a representation of the relative importance (salience) of the locations of visual space. It is needed to guide where to look next, for computing the center (technically "centroid") of a cluster of items, and for other important computations. Here we show that in a brief flash of dots of three different colors, randomly interleaved, subjects can compute all three centroids. As a single salience map cannot discriminate dots of different colors, accurately reporting three centroids demonstrates that subjects have not just one, as is commonly believed, but at least three salience maps.

Introduction

Humans form accurate summary statistic representations (SSRs) for a wide range of attributes including average orientations (Alvarez & Oliva, 2009; Li, Castañón, Solomon, Vandormael, & Summerfield, 2017; Solomon, May, & Tyler, 2016), average sizes (Ariely, 2001; Chong & Treisman, 2005; Gorea, Belkoura, & Solomon, 2014), average positions, i.e., centroids (Drew, Chubb, & Sperling, 2010; Sun, Chubb, Wright, & Sperling, 2016a), average motion directions (Dakin & Watt, 1997), numerosity (Solomon & Morgan, 2018), and more. The ability to form an ensemble representation of a visual scene and to compute an SSR on that ensemble is a brain mechanism for bypassing cognitive capacity limitations by distilling the sensory information. Many studies show that SSRs may underlie a wide range of well-known phenomena in vision including texture segmentation (Ackermann & Landy, 2015), visual crowding (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, 2011) and visual

✉ Peng Sun
  peng.sun@uci.edu

✉ George Sperling
  sperling@uci.edu

[1] Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697-5100, USA

🖄 Springer

search (Ariely, 2001; Chong & Treisman, 2003, Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Wolfe, Võ, Evans, & Greene, 2011).

Recent studies show that subjects can compute SSRs for selective subsets that are spatially segregated (Attarha, Moore, & Vecera, 2014; Im & Chong, 2014) or are defined by different features (Chong & Treisman, 2005; Drew, Chubb, & Sperling, 2010; Halberda, Sires, & Feigenson, 2006; Inverso, Sun, Chubb, Wright, & Sperling, 2016; Poltoratski & Xu, 2013; Sun, Chubb, Wright, & Sperling, 2016a, 2016b). More impressively, some studies show that subjects can even compute multiple concurrent SSRs (Attarha, Moore, & Vecera, 2014; Chong & Treisman, 2005; Halberda, Sires, & Feigenson, 2006; Im & Chong, 2014; Poltoratski & Xu, 2013; Sun, Chubb, Wright, & Sperling, 2018). The present study investigates multiple SSRs.

Outline We studied a particular SSR – the centroid, the center of mass of a cluster of items. In Part 1, we found that all our subjects could report three different centroids after viewing a single brief flash of a dot array composed of three colors of dots, randomly interspersed. Although post-stimulus-cued partial-report procedures (Sperling, 1960) are usually used to study multiple concurrent tasks, all our subjects could accurately report three centroids that were defined by the colors of the target dots. In Part 2 we took advantage of the centroid task's property that responses are continuous in two dimensions and therefore provide much richer information than conventional binary-response methods. We developed an analytical tool based on error variance to partition subject errors into four different additive components: (1) imperfect grouping, in this case, imperfect color-attention filters that fail to perfectly distinguish target color from distractor color items; (2) a Bayesian-like bias to report the mean location; (3) error due to imperfect motor control in making a continuous response; and (4) failure to encode all target items, which is the largest single error source; and then (5) an interaction, the increased task demand of making multiple instead of single responses, which causes increases in all four error components.

Methods

Overview

The current study used a variant of the centroid paradigm that was originally developed by Drew, Chubb, and Sperling (2010) and was considerably enhanced by Sun, Chubb, Wright, and Sperling (2016a). In all trials, subjects viewed briefly flashed dot arrays containing either two or three colors of dots. In the experimental trials, the subject's task was to judge the centroid of every color subset using a computer mouse to move a plus sign (+) to the judged centroid locations on the display screen and mouse-click (multiple responses). In

separate blocks of control trials, subjects were cued at the beginning of the block to report only the centroid of a single-color subset (single response).

Experiment 1 investigated two-color displays in which two centroids were reported. The results were so encouraging that the procedures were repeated with three-color displays and three centroids reported. Although these experiments were consecutive, the procedures, subjects, results, and analyses were so similar, we report them together.

Apparatus

The experiment was conducted on an iMac intel computer installed with Matlab 2012b and Psychtoolbox-2 software (Brainard, 1997). The stimuli were presented on a Dell UltraScan P990 Trintron CRT monitor with 1,024 x 768 resolution at a refresh rate of 85 Hz. The mean luminance of the monitor was 28.9 cd/m^2. The monitor screen was 36.5 cm wide x 27.5 cm high. Each pixel was a square of 0.357 mm^2 area. A chin rest stabilized head position 54 cm away from the monitor. At 54-cm distance, a pixel subtends $(0.0379°)^2$.

Subjects

The first and second authors (S1, S4) and three psychology undergraduate students (S2, S3, S5) participated in the experiment. The five subjects (three females) ranged from 20 to 35 years in age. S2, S3, and S5 were naïve to the purpose of the experiment, and had little experience running psychophysical experiments. 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.

Stimuli

Each experimental and control stimulus contained two (Exp. 1) or three (Exp. 2) spatially interleaved colors of dots displayed within a 640×640 pixel-wide square (Fig. 1) that spanned 23.3° of visual angle (dva). All dots were 9-pixel-wide squares, each spanning 0.33 dva, all dots were darker than the background. Dots in each subset shared a common color of red (X = 17.11, Y = 9.61, Z = 1.23), black (X = 0.67, Y = 0.55, Z = 0.44), or green (X = 6.18, Y = 12.25, Z = 2.51), and appeared on a gray background (X = 29.56, Y = 28.93, Z = 34.30). The dots were clearly visible on the gray background (28.9 cd/m^2).

Singletons In addition to the above stimuli, there were matched singleton stimuli consisting of just one dot of each color. The positions of the singleton dots where chosen to be
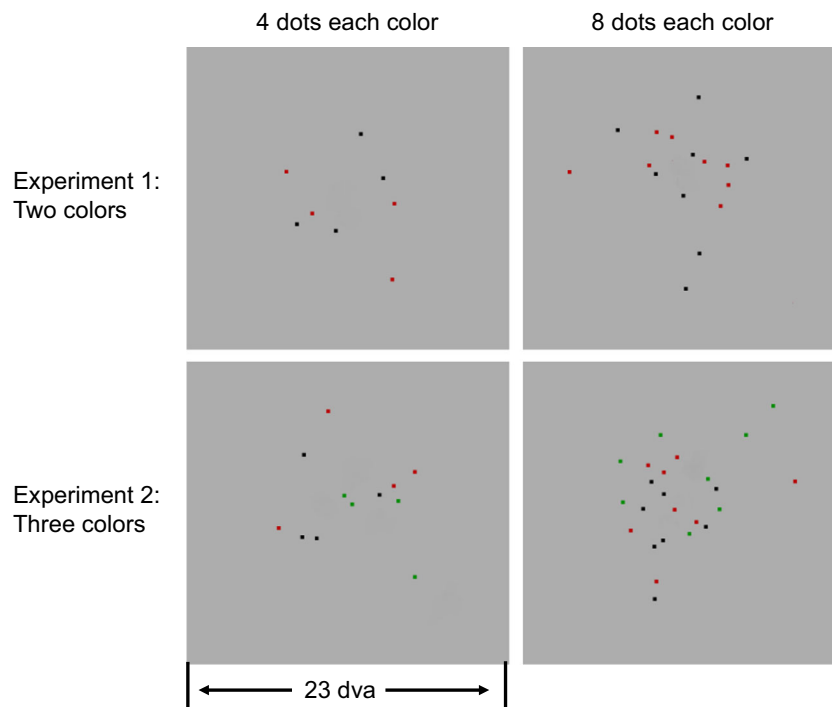
Fig. 1 Sample stimuli from Experiment 1 (**top**) and Experiment 2 (**bottom**). **Top:** Two colors, black and red. **Bottom:** Three colors, black, red,   and green

the same as the positions of the centroids of corresponding multi-dot stimuli.

Dispersion In each stimulus, locations of the dots were determined by a two-step process: First, for a stimulus that contained a total of N dots, N samples were drawn from a bivariate normal distribution. Second, the sample locations were divided by the standard deviation of the N samples and then multiplied by a fixed constant to ensure that the locations of all generated dot arrays with sample size N would have the same standard deviation on every trial. In our experiment, the fixed standard deviation was 110 pixels (4.0 dva for N=16), and increased or decreased by $\sqrt{N/16}$ for other values of N. This manipulation, which followed the procedure of Sun et al. (2016a), was not critical to the purpose of the study. Following the nomenclature of Sun et al. (2016a), we refer to the standard deviation of the dot display as "Dispersion." It measures the size of the physical spread of the entire dot array. Figure 1 shows screen shots of sample stimuli used in Experiments 1 and 2.
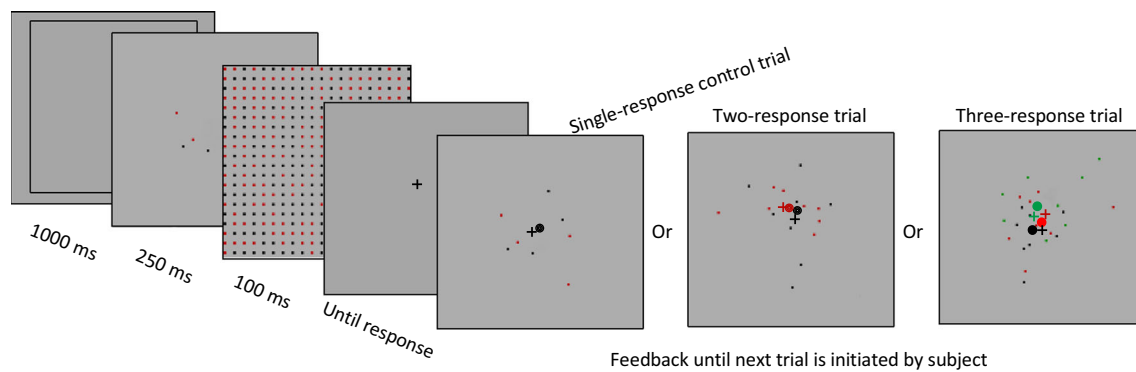
Trial procedures

Prior to starting Experiment 1, subjects were trained for 2–3 days during 1-h sessions (with four 5-min breaks) on various centroid displays: No distractor, distractor dots of one color different from the target color, and, finally, distractors of two colors. In the

training trials, subjects' tasks were to locate the centroid of a single subset of dots whose common color was indicated and fixed within a block – the single-response task. Subjects were considered trained once they reached an asymptote in their performance on the centroid task. Training was concluded when there was no further improvement in mean error magnitude (average distance between judged and true centroid locations) between two consecutive blocks of trials.

The main experiment consisted of two tasks: the single-response task in which only the centroid of a single color of dots was to be reported, and the multiple-response task in which the centroid of every color was reported. Procedures, including stimuli, for both tasks were same except for the number of centroids to be reported.

Figure 2 shows an example of a typical trial. The trial started with a 1-s blank frame followed by a 250-ms stimulus frame. After the termination of the stimulus frame, for two subjects, a masking frame containing a grid of randomly colored dots appeared for 100 ms. For three subjects, the masking frame was omitted. Subjects used a mouse-controlled cursor to indicate the centroid of a single fixed color, or of every color subset, according to the instruction at the beginning of a block.

In the multiple-response task, the sequence in which the centroids of the color subsets were to be reported was fixed within a block. This sequence was indicated at the beginning of a block and was again indicated on each trial by displaying the cursor in

Data analysis

Overview The most basic measurement of a subject's performance is the response error – the Euclidean distance between the true centroid location of the target subset and subject's mouse-click response. Distance is measured in units of display pixels – a pixel is 0.357 mm on the display, 0.0379 dva (2.27 min) at the subjects' eyes. In Part 2, which considers models of performance, both the magnitude and direction of error are critical.

Results

The results of Experiments 1 and 2 are displayed in Figs. 3 and 4. The left side ordinate shows the error (up is bad) for each of the conditions indicated on the ordinate. It is immediately clear that there is a large difference between single and multiple response conditions and that the other factors, color and order of report, have small inconsistent differences. For every subject, there is no panel in which any multiple response falls below (is better than) any single response. The single-better-than-multiple difference is statistically significant at p<0.001

with a t-test. That there is no clear advantage of any color over any other only means that we succeeded in choosing distinctive colors that were statistically equivalent in the context of this experiment. That there were no important differences in order of report was quite unexpected because in nearly all other experiments the first reported items have an accuracy advantage over subsequent ones.

The important result of these experiments is that subjects can make three centroid judgments of approximately equal accuracy following a single brief exposure of three colors of interleaved dots. From the subjects' errors – how far their judgments were from the target centroids – it is not obvious how good their judgments were in terms of how much stimulus information they acquired from the briefly exposed stimulus to enable the response accuracy they achieved. There are various ways that one might measure the quality of centroid judgments: the way that we believe gives the best intuition is the comparison of the subjects to an ideal detector faced with the same task (Fig. 5).

The ideal detector knows the location and color of every dot and computes a perfect centroid. Of course, this ideal detector will never make an error. However, by randomly
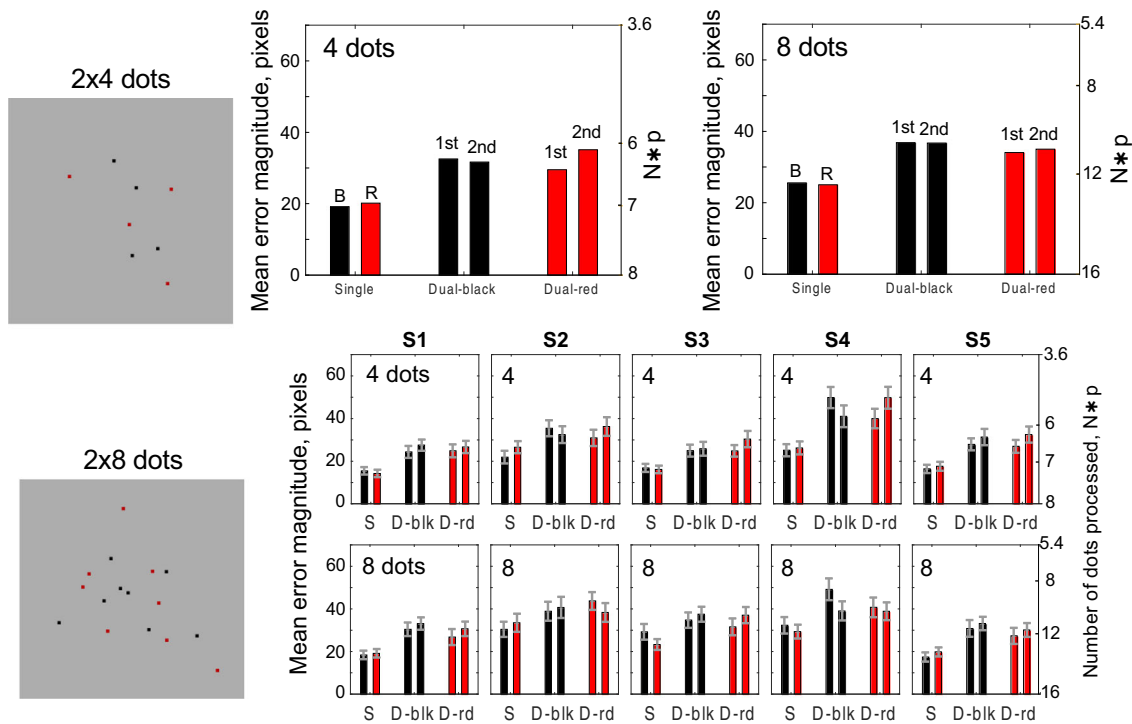
Fig. 3 Experiment 1: Response accuracy in terms of error – the distance of subjects' centroid judgments from the true target centroids – in stimuli with dots of two colors and two required centroid judgments per trial. Far left shows sample stimuli with 2 x 4 and 2 x 8 dots. The upper row shows the average error for all subjects, for each color stimulus, for each condition, and for each report order. The lower rows show the data for each individual subject. Single (S) refers to the control conditions in which subjects judged the centroid of only one predetermined color, black (B) or

red (R). Dual (D) refers to trials in which subjects made two centroid judgments, the bar color (black or red) indicates the color of the dots whose centroid was reported either first or second as indicated. The right-hand ordinate is the absolute minimum number of stimulus dots an ideal detector must process (N*p) in order to reduce its output error to the value indicated on the left-hand ordinate. N is the number of stimulus dots, p is the proportion of stimulus dots processed by the ideal detector (see Fig. 5 and text for details)
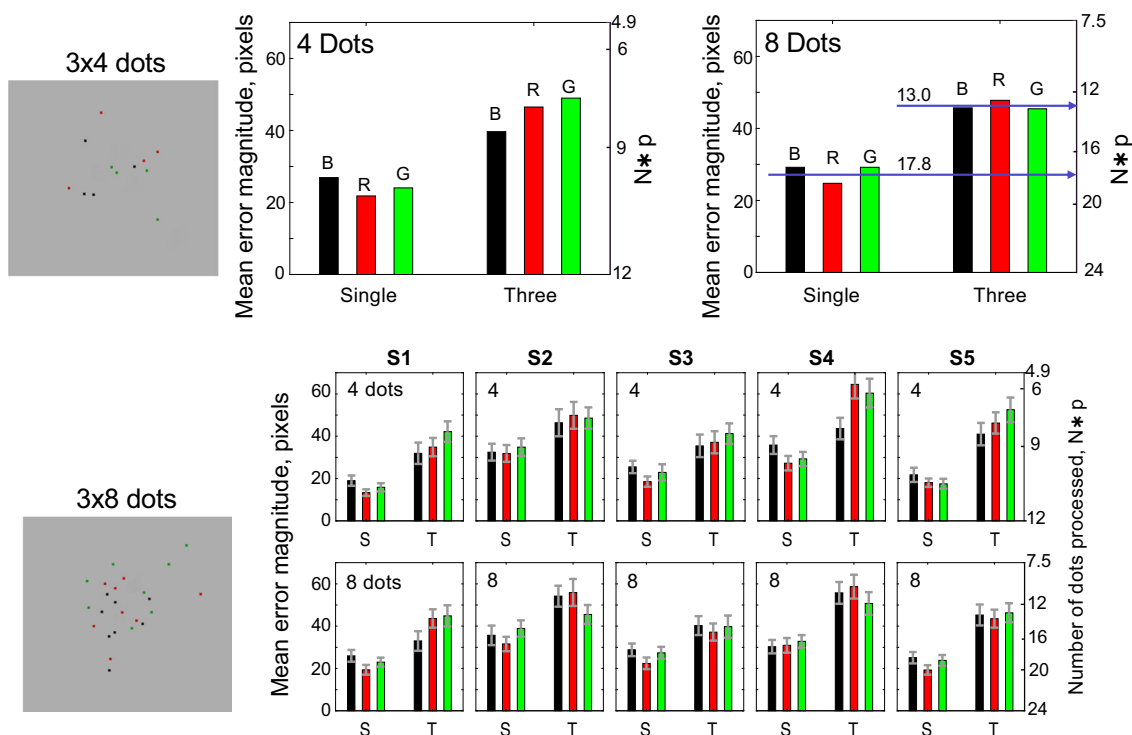
Fig. 4 Experiment 2: Response accuracy in terms of error – the pixel-distance of subjects' centroid judgments from the target centroids – in stimuli with dots of three colors and three required centroid judgments per trial. The caption for Fig. 3 applies except that here, in the experimental conditions, subjects judged three centroids on each trial. Similar stimuli were used in the single-judgment controls as in the experimental conditions. For the multiple centroid judgments, the order of report was always black, red, green
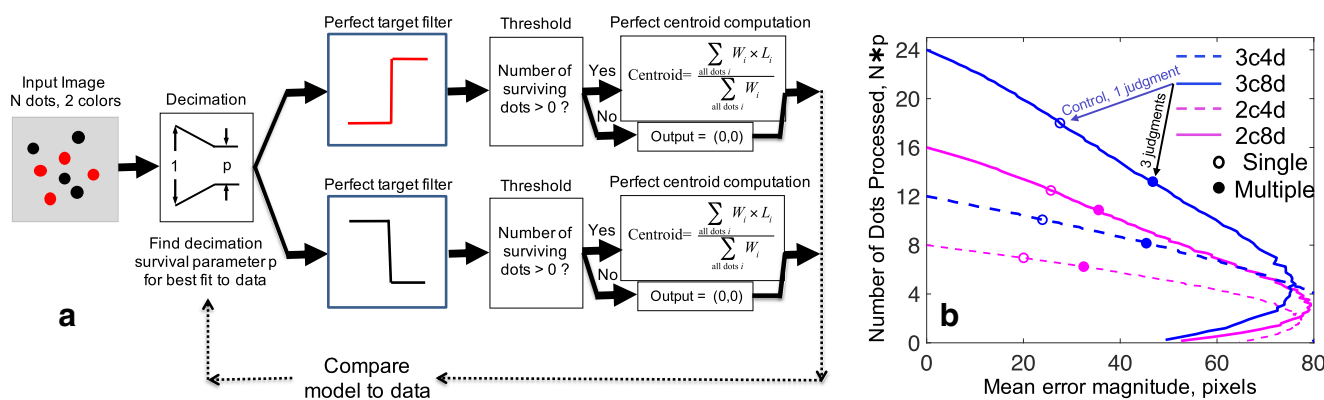
Fig. 5 Ideal Detector model to evaluate the quality of subjects' centroid judgments in terms of the minimum number of stimulus dots that the Ideal Detector must process to match a subject's accuracy, shown for two centroids to be estimated. (**a**) In this example, the input is a 2c4d (two color, four dot) stimulus. A decimation process randomly eliminates stimulus dots, each dot has a probability p of surviving. A perfect color filter for each color selects the surviving target dots for each to-be-estimated centroid. Perfect centroid computations compute the centroid of each color of dots and compare the x,y locations of the computed centroids with the x,y locations of the subject's data, and the decimation process adjusts its parameter p to minimize the model-data difference. When an optimum p is reached, the resulting centroid x,y locations are output. The expected number of dots utilized by the Ideal Detector (N*p) is the number of stimulus dots (N) times the fraction of surviving target dots (p). (**b**) Abscissa: Pixel error of the Ideal Detector in Panel a for each of the four conditions of Experiments 1 and 2 as a function of the ordinate, N*p. On trials in which, by chance, there are zero surviving dots, the ideal detector's response is the best option for that situation, the expected location of target dots and of all dots, (0,0). The curves in (b) represent number of colors, 2 or 3, whose centroids are computed, and the number of target dots, 4 or 8. The filled points in the graph indicate the mean error in pixels achieved by the subjects and the corresponding N*p of the Ideal Detector model at this error level. The unfilled points in the graph represent the mean error on single-response control trials. N*p for these trials represents the performance that would have been achieved if the subject had been able to perform as accurately in multiple response trials as in the corresponding single-response trials

removing some of the stimulus dots (decimation), the performance of the ideal detector is degraded. The minimum proportion p of surviving dots that the ideal detector needs to match the subject's performance is a measure of the fraction of stimulus information that is represented in the subjects' responses. For a stimulus of N items, it is more intuitive to express the ideal detector's performance in terms of N*p, the minimum number of stimulus dots the ideal detector must process to match a subject's performance. This does not imply that the subject processes N*p items. Because there are other sources of error than merely failing to encode and retrieve some fraction of stimulus items, subjects undoubtedly process more dots and quite likely by a quite different algorithm. However, this ideal detector sets an absolute lower bound on the number of stimulus dots that must be processed in order to achieve the observed centroid accuracy.

Figure 5b shows the number of dots processed N*p by the ideal detector (ordinate) for the four kinds of stimuli and corresponding tasks and the resulting centroid error (abscissa) produced by varying decimation parameter p. Each of the closely spaced points that comprise the curves in Fig. 5b was generated by a Monte Carlo process in which the ideal detector assumed a particular value of p and was presented with 100 of the same stimuli as the subjects viewed. Figure 5b also shows where the subjects' centroid judgments fall on these functions.

For trials in which, by chance, there are zero surviving dots, we assume the ideal detector adopts a Bayesian-like strategy that yields statistically the best possible accuracy for this situation – choosing the overall expected location of all dots (see Part 2 for more details). This strategy yields the unusual non-monotonic function of error as a function N*p. An optimal overall strategy would be to choose the mean location whenever that yields a higher expected accuracy than utilizing the number of surviving dots. For the four experimental conditions, optimal thresholds are: 2c4d, 3c4d: 1.5 dots, 2c8d, 3c8d: 3.5 dots. These thresholds do yield monotonic decreasing functions of N*p as a function of error magnitude but, as will be demonstrated in Part 2, do not represent our subjects' strategies. Our subjects, as they were trained and directed, base their responses overwhelmingly on stimulus information, and the influence of alternative Bayesian-like strategies is very small, even when such strategies offer a significant advantage.

The N*p derived from the curves in Fig. 5b are shown on the right-hand ordinate labels in Figs. 3 and 4. In the case of multiple responses, because these responses are of statistically equivalent quality, the error shown is the average error of the responses. The N*p shown is based on the total number of stimulus dots because, in their multiple responses, the subjects judge all the stimulus dots. For the control trials in which subjects judge the centroid of only one color of dot, the indicated N*p represents the performance subjects would have exhibited if, on the multiple-response trials, they had been able to perform as well as on the single-response trials.

The most significant datum in Fig. 5a is the N*p of 13.0 for subjects judging three 8-dot centroids on each trial. Each of the individual centroid judgments would have required the ideal detector to process 4.3 of 8 dots to equal the subject's centroid accuracy. When subjects devote all their resources the processing a single centroid in the control condition, N*p is 17.8, indicating an average requirement of processing 5.9 of 8 dots per judged centroid. The accuracy difference is significant but represents a much smaller percentage change when considered in terms of equivalent number of items processed versus the difference in error magnitude. There are many other interesting aspects of the data in Figs. 3 and 4, but these are better considered in the context of the more complex analyses and the model in Part 2.

Discussion: Part 1

Three within-group salience maps inferred from three high-quality centroid judgments per single stimulus.

Maps A map is a mathematical construct to explain some aspect of behavior. A map, often called a topographical map, is a function z = f(x,y,t) that gives the value of a variable z, as a function of spatial variables, typically x, y, and time, t. Whereas the independent variable in a color map may be three-dimensional, in all the examples considered here, f(x,y,t) is single-valued (scalar). It is widely assumed that there are single-valued feature maps for elementary features such as particular colors, orientation, motion direction, and other "separable dimensions" (e.g., Treisman & Gelade, 1980, p. 98).

A computational example of a feature map is the so-called luminance map for first-order motion. It gives a real number, positive or negative, that represents the luminous intensity – or, more accurately, the Weber contrast – at a point x,y,t relative to its neighborhood. Although the luminous intensity may be produced by different wavelength combinations of light, the particular wavelength combination is irrelevant because the map represents only a single value at each point. Computations of the direction and strength of first-order motion based on such a map have made extremely accurate predictions of first-order motion and verified many counter-intuitive properties (e.g., van Santen & Sperling, 1984). Second-order motion is assumed to be computed on a map that represents the local variance of intensity. Third-order motion is assumed to be computed on a map that represents local salience, analogous to the strength of the perception of figure versus ground at the point x,y (Lu & Sperling, 1995).

Salience maps This concept was introduced by Koch and Ullman (1985, p, 221) "Thus, the different feature maps code for the conspicuity within a particular feature dimension. In order to assess the global, overall conspicuity of a location, we

will assume the existence of another topographical map, termed the saliency map, which combines the information for the individual maps into one global measure of conspicuity." "Salience map" now has well over 1,000,000 citations on Google, most are for the representation of priority for a sequence of subsequent processes or actions such as visual search, movements of the eyes and of attention, and for segmenting an image for object processing.

Most computational theories of salience are based on bottom-up computations that seek smaller stimulus areas that differ from their larger surrounds in some coherent set of features. But feature-contrast is not necessary. A searched-for color or a familiar face in a classroom of unfamiliar faces will produce high salience at its location. The importance of top-down influences on salience computations was recognized from the beginning (Koch & Ullman, 1985) and will be critical in the procedures considered here.

Multiple salience maps? A basic principle in evolution is that once the genetic code for a complex structure has evolved, it may be re-used with different parameters for related purposes. Once the genetic code for a right foot with 14 phalanges and five metatarsals has evolved, that same code with slight parameter changes is used to produce a left foot and right and left hands, all with the 14–5 anatomy. We consider here something that neither Koch and Ullman (1985) nor their many followers considered: the possibility that the neural architecture that instantiates their saliency map might be repeated in the brain and adapted for related purposes. How would we recognize a second salience map if there were one? What are the computational properties of a salience map that distinguish it from other topographic brain maps?

Four salience-map criteria A salience map is fundamentally different from the feature maps mentioned above. In a luminance map, the values represented in the map are informative about the stimulus luminance. In a color map, the values recorded in the map are highly informative about the relative effectiveness of the light in stimulating the retinal cones (visual color-sensitive receptors), although the color-map values are not very informative about the light spectrum itself. However, in a salience map, the values are completely uninformative about the specific features or substance that produced the salience value. Although the input that determines the salience values in the map can arrive from many different sources, i.e., the input is a complex vector, the output is a scalar function of x,y,t. Here, we generalize that the concept of a salience map to apply to any topographic map that has four properties of the original saliency map: (1) what is represented in the map is a scalar, (2) the output is uninformative about the particular features or substance of the input, (3) the map values are monotonically related to the priority or importance for subsequent processes, and (4) several different processes can utilize the same map.

It is critical to separate the processes that use a salience map from the map itself as they involve very different computations. Insofar as the concept of "salience map" has come to implicitly imply that the map inputs are elementary features and the processes that utilize the map are search sequences, eye and attention movements, and segmentation for object processing, the four criteria are a significant generalization of the concept of a salience map. Based on this generalization, we propose that there exist topographic maps, "within-group salience maps," that satisfy the four salience criteria but differ from the original salience map concept – they have different inputs and are utilized by different subsequent processes.

Three within-group salience maps Figure 6 shows a flow chart of the color segregation and centroid computation processes that we hypothesize occur in the centroid task and that are implemented in the computational model described in Part 2. The computational model is elaborated to show how centroid computations might be embedded in and carried out in a brain that is occupied with much more than merely computing centroids.

Centroids are computed on topographic maps, and, more specifically, on salience maps. Insofar as a salience map, by definition, encodes only salience and not features, one salience map cannot distinguish different colors of dots that have approximately equal salience. A single salience map presented with a multi-colored dot display can compute only the centroid of all the dots. However, different colors of dots may, on average, have different salience weights, as illustrated in Fig. 6. Computing three centroids, one for each color of dots, requires three salience maps, each with a different color weighting. Each of these salience maps satisfies the four criteria of a topographic salience map: it represents a scalar quantity (the salience of each item) that is uninformative about the item features but determines its weight in the subsequent centroid computation, and the map can be utilized by other processes than merely centroid computations.

Figure 6 places the centroid task in the context of other processes that are presumed to occur in the salience system and, of course, in parallel in other systems. In Part 2, the color segregation process is quantitatively described in terms of color-attention filters, filter quality is determined, as well as the quality of other processes shown in Fig. 6. Here, we consider just the most important aspects of Fig. 6.

Centroid judgments are a high-capacity preconscious computation It is important to keep in mind that a centroid judgment is a statistical summary representation, that is, it is a preconscious computation that utilizes much more display information than can be subsequently recalled. For example, in stimuli quite similar to those used here, Sun, Chubb, Wright, and Sperling (2018) found that even with a very generous scoring procedure in a change-detection memory task, the best subject
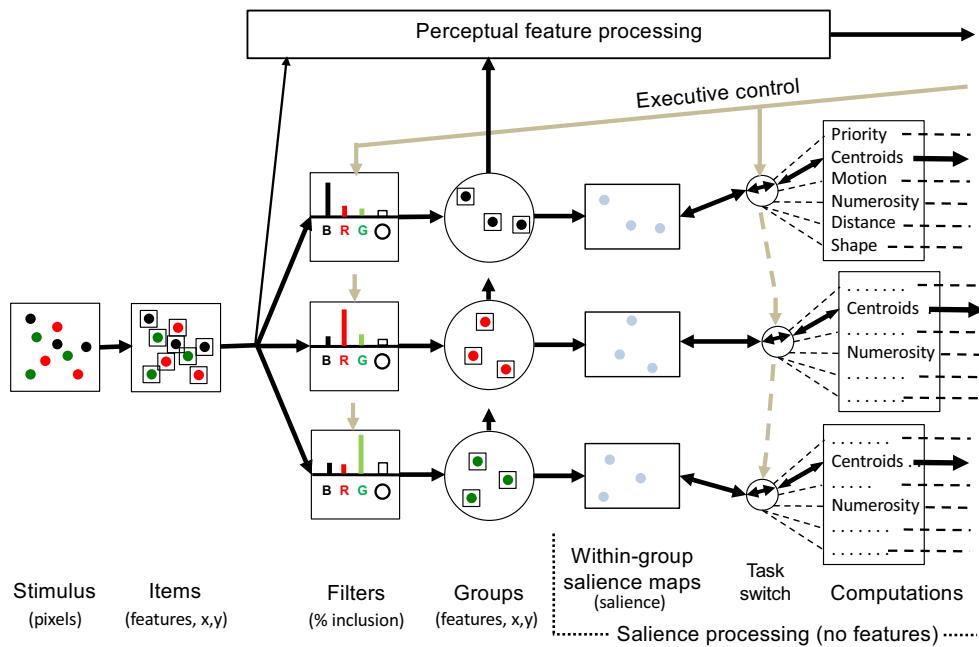
Fig. 6 Flow chart of the essential components of the salience system in a computational model of the processes that transform a stimulus array of dots of three colors into three centroid judgments. Initially, the array dots are represented as pixels in an x,y image. Early image processing transforms the representation into an array of items, represented as small squares. An item consists of a bundle of features at an x,y location. A combination of bottom-up and top-down processes segregate the items into three groups, here defined by color-attention filters. The filters also contain a bias component (the black circle). The content of each group – the spatial and feature information of each dot – is passed to a salience map that computes the salience – a real number – at each location. Although feature information is lost in the salience pathway, it is preserved in feature-processing pathways. A centroid is computed on the contents of each of the three within-group salience maps. Because the feature information is no longer available in the salience path, an executive process keeps track of which centroid has been computed, binds the feature information with the salience-computed centroid location, ultimately this results in x,y centroid responses. The far right indicates other computations that can optionally be carried out on the contents of the salience maps. Dotted lines indicate that the ability of those processes to utilize multiple salience maps has not been demonstrated

remembered the approximate locations of fewer than two of 26 dots. Yet, in the same stimuli, subjects were able to estimate centroids that minimally required the accurate processing of more than 15 of 26 dots. In the present experiments, when subjects make three centroid judgments, they remember primarily the statistical summary of the stimulus, the three centroids. We know from Sun et al. (2018) and other prior work that they remember very little else about the stimulus. All the computations represented in Fig. 6, except for Executive control, are pre-conscious.

Is a salience map really necessary, isn't there a simpler solution? To compute a statistic on some items, such as a mean x,y position (a centroid) or a mean orientation, or a mean size, one needs to first to group the items on which the statistic is to be computed, and then to perform the computation. Computing a centroid requires a reasonably accurate spatial representation of the items in the group, i.e., a map. For example, in the current experiment, dots of different colors are represented on maps. However, what is represented in the salience map is not the color itself, but the weight assigned by the grouping process to that color according to the task being performed. To find the centroid of the black dots, a group is formed in which black dots are given a large weight, red and green dots are given small weights – insofar as the filter has sufficient information to make an accurate assignment.

The more complex aspects of filter-group-salience processing are not exposed in the present experiments. Consider instead two complex examples: In a briefly flashed stimulus consisting of 12 dots, all of the same unknown color, and 14 dots of seven different colors, Sun et al. (2018) demonstrated that subjects could concurrently compute both centroids, the centroid of the homogeneous group and the centroid of the heterogeneous group. There is no predefined map devoted to "heterogeneity versus homogeneity of colored dots" nor any pre-defined color filter that can make such a discrimination. However, a sophisticated grouping process and a salience map solve this problem. The authors called the quantity that is associated with each dot for use in a group's centroid computation "weight."

In Sun et al. (2016b), subjects make centroid judgments of arrays of Gabor patches (small patches of alternating light and dark stripes). Stripes in each patch had one of eight contrast levels ranging from barely discriminable light-gray/lighter-gray stripes to dark-black/bright-white stripes. In different sessions, subjects were able to quite accurately compute

centroids in which patches were assigned values in proportion to patch contrast, in inverse proportion to patch contrast, or of equal value for all patches, independent of contrast. The authors call the perceptual value of each patch "weight," which seems appropriate for a centroid computation. However, Rodriguez-Cintron, Wright, Chubb, and Sperling (2019) and Lu, Wright, Chubb, and Sperling (2019) use the term "influence" for the scalar value used to describe items in their centroid computations.

In the most common examples of salience maps, priority of items for shape and object processing or of spatial areas for visual search, the quantity associated with each item or area is called "priority" or "salience." Rather than having influence maps, weight maps, figure-ground maps, priority maps, and salience maps, which imply a different map for every variation of every possible function that is computed on the map contents, we prefer the more generic term *salience map* to describe a topographic map that satisfies the four criteria of a salience map, and allow flexibility in describing the scalar value that is represented in a particular utilization of the map. A common, implicit assumption (made explicit in Fig. 6) is that the salience maps used by various different computations utilize the same neural substrate, but this issue is unresolved.

Example: The utter indifference of a salience map to features in a motion task With respect to the indifference of salience maps to the features that produced the salience values, perhaps the strongest direct demonstration is Tseng, Gobell, and Sperling's (2004) motion task. Their subjects viewed a dynamic random-dot stereo-depth grating in which alternate stripes appear near and far, and also an isoluminant color grating in which alternate stripes are pink and background gray. Each motion stimulus was adjusted to several JNDs above its threshold where it was clearly perceived to move. Both gratings moved in the same direction. Neither of these gratings stimulates either the first-order (luminance) system or the second-order (variance/texture) system. The motion in each of these gratings involves only the third-order motion system that operates on a salience map. Each stimulus is an alternation of high salience (near, pink) and low salience (far, gray). When these stimuli are added in phase (high + high, low + low), the strength of perceived motion is doubled, when they are added out of phase (high + low, low + high), there is no longer any salience modulation, the salience-motion computation produces zero output, and consequently there is zero perceived motion according to accurate psychophysical measurements. That two gratings moving in the same direction, each of which evokes a strong motion, can nevertheless cancel when they are of opposite phase indicates that the salience system is totally unable to distinguish these two inputs and simply adds their respective salience values. A consequence of salience being a scalar versus a vector is that it is

no more difficult to compute salience, and therefore centroids, on any array of different objects than on an array of homogeneous objects (e.g., Sun et al., 2018). Obviously, other parts of the brain know the difference between a depth grating and a color grating and between heterogeneously and homogeneously colored dots, but not the salience map (or maps) on which motion and centroids are computed.

Is the process of grouping items necessary, cannot items simply be entered directly from filters into salience maps? In Sun et al. (2018) referred to above, subjects were able to concurrently compute the centroid of 12 dots of an unknown homogeneous color and the centroid of 14 dots of seven different colors. There is no color filter that can be assigned in advance to make this distinction, the recognition that there are many identically colored items is a grouping process. And there is no filter to segregate the diverse heterogeneous items in order to compute their centroid. This requires a sophisticated, complex grouping process. Since it is clear that items are segregated by what is traditionally called grouping in some centroid tasks, it seems reasonable to suppose that this is the normal process even when the experimental paradigm does not explicitly expose it.

How many within-group salience maps are there? This study cannot answer that question, it can only assert that there must be at least three. Nor can it answer the question of map specialization, whether the same or different neurons serve maps for centroids, for processing priority, and for third-order motion. Concerning the number of salience maps, Halberda, Sires, and Feigenson (2006), using multi-color dot arrays and post-cued partial reports of numerosity, found that seven of their 30 subjects could estimate the numerosity of more than three colors of interspersed dots (but see follow-up studies cited in the Introduction). As numerosity estimation is a mental computation that likely is carried out on a salience map, their results suggest that, unlike the number of fingers on each hand, there may be significant individual differences in the maximum number of concurrently available within-group salience maps.

The ubiquity of multiple parallel processes There is a long history of related findings that imply multiple parallel perceptual processes (see Introduction). For example, Pylyshyn (1989) proposed a formal computer algorithm FINST to account the ability of subjects to track a small number of target objects moving randomly among a field of similar, similarly moving distracter objects (FINST, finger instantiation, is a process that is analogous to keeping a finger on each moving object). Pylyshyn and Storm (1988) found that subjects can track up to five target objects among ten moving objects. The problem already noted by Pylyshyn and amplified by numerous subsequent tracking studies is that the number of objects

tracked depends on numerous factors, notably the speed and predictability of object movements, whether they cross paths, and many more. Tracking experiments do not yield one specific number of FINSTs.

Sperling ([1967](#)) demonstrated that when subjects viewed a briefly flashed array of five letters, they accumulated information in parallel from all locations and that the net result was the equivalent of three to four reported letters. Similarly, in a rapid serial presentation of briefly flashed letters (Sperling, Budiansky, Spivak, & Johnson, [1971](#)) subjects were able to process three central locations very rapidly and accrue information elsewhere at a much slower rate. Sternberg ([1966](#)) interpreted the search of working memory as a serial exhaustive process. However, McElree and Dosher ([1989](#)) showed that the original reaction-time data and a larger, more diagnostic set of cued-response data were fit only by a parallel search model. Miller ([1956](#)), using quantitative psychophysical methods (following Wundt's ([1912/1924](#)) introspective methods), proposed a magical number 7 plus or minus 2 for slots in working memory. Subsequently, there have been thousands of published articles that follow these threads. The *Introduction* lists some of the most recent studies. The point of this paragraph is to note that models of multiple parallel processes have been proposed for an increasing number of perceptual and cognitive phenomena for more than 50 years, so we should not be surprised to see multiple salience maps joining the historical trend.

Part 2. Parametric Description of the Component Processes in Centroid Judgments and their Contributions to Judgment Error

Overview In Part 1, the costs of making multiple versus single centroid judgments were evaluated in terms of the increases in pixel error. Part 2 describes the cost of making multiple centroid judgments in terms of the costs to the component processes of centroid judgments. A computational model is introduced that quantifies the two most critical component processes: Attention color filters and Efficiency (Eff). Attention color filters give large weights to dots whose centroid is to be computed (targets) and small weights to the remaining dots (distracters). Eff is an estimate of the minimum number of stimulus dots an ideal detector requires to match a subject's performance (e.g., N*p in Fig. [5](#)). Both individual and average subject results are visualized in terms of Attention Operating Characteristics (AOCs). The AOCs show that filter quality is relatively unimpaired in multiple centroid judgments; the main costs are reduced Eff for the more difficult tasks. The second part of Part 2 offers a new, much more detailed analysis of centroid

judgment error in terms of five component processes: (1) imperfect color-attention filters, (2) a Bayesian-like bias, (3) residual error due mostly to inefficient encoding, (4) storage, retrieval, and cursor misplacement errors, and (5) an interaction, the added difficulty of making multiple versus a single centroid judgment on each trial. Unlike all the prior analyses, the five-component error analysis refers only to average data.
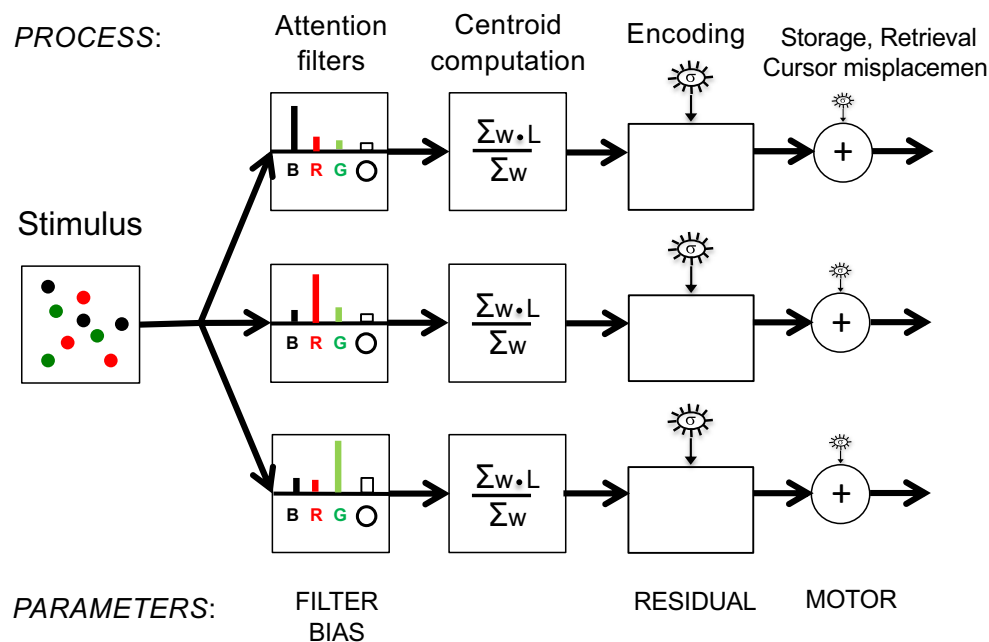
Computational model of the component processes in multiple centroid judgments Response errors are used to derive precise color-attention filters and efficiencies (performance limitations in terms of N*p) to characterize each individual subject's performance in the single- and dual-response tasks. The computations are described in the flowchart of Fig. [7](#), which is simpler than the model. The computations assume that the input is items, not pixels, and the grouping process is implicit in the attention filters. The bias, which may be a higher-level process, is placed with the attention filters because that is computationally equivalent and it simplifies the presentation.

Imperfect color-attention filter, $f_\phi(c)$ We conceptualize the process of excluding irrelevant dots from the centroid computation as color-attention filtering, where $f_\phi(c)$ is the weight given to color c when ϕ is the target color. To estimate a subject's $f_\phi(c)$, we need the following definitions: Let x_n, y_n be the location of the nth dot in a stimulus, r_x, r_y be the subject's location response, and c_n be the color of the nth dot in a display of N dots: c, ϕ = [r (red), b (black), g (green)]. The bold-type symbols represent vectors with the trial number as the implicit vector dimension. Let ϕ be the target color, i.e., the color of the dots intended to be incorporated into the centroid computation. The subject's response, r_x, r_y, can be written as:

$$r_x = \sum_{n=1}^{N} f_\phi(c_n)x_n + f_\phi(b_x)b_x + Q_x \qquad (1)$$

$$r_y = \sum_{n=1}^{N} f_\phi(c_n)y_n + f_\phi(b_y)b_y + Q_y$$

In Eq. [1](#), $f_\phi(c_n)$ is the filter weight exerted by dot n on the subject's response in condition ϕ. $f_\phi(b)$ is the weight exerted by a fixed, default location (b_x, b_y) on the screen. We further constrain that $\sum_{n=1}^{N} f_\phi(c_n) + f_\phi(b) = 1$. Q_x, Q_y are normally distributed random variables representing residual error. For all the experiments herein, the target color-attention filter $f_\phi(c)=1$ when $c = c_\phi$, $f_\phi(c) = 0$ when $c \neq \phi$. In real data, weights are seldom exactly equal to 0 or 1, and an imperfect color filter causes r_x, r_y to be systematically deviated toward centroids of distracter colors. Equation [1](#) implies a simple linear model in which all the color-filter weights $f_\phi(c)$ of the best-fitting model (i.e., the model that minimizes Q_x, Q_y) are obtained by standard multivariate linear regression (e.g., Sun et al., [2016a](#)).

PROCESS:

Fig. 7 Flowchart of the computational model of the visual, memory, and motor processes that transform the visual input stimulus into – in this illustration – three centroid-localization responses with four independent sources of response error (FILTER, BIAS, MOTOR, RESIDUAL). The input stimulus is sent to three color-attention filters, each filter assigns a color weight (illustrated) on a scale of 0 to 1 to every dot according to its color: B, black; R, red; and G, green. The small, black circle represents a Bayesian-like bias – a filter weight BIAS assigned to a virtual dot located at the overall mean session centroid. For a given color-attention filter, the sum of all color weights plus BIAS is 1.0. FILTER is the centroid location error resulting from an imperfect attention-color filter. The process of encoding the stimulus dots and their locations is represented here as adding a random variable, RESIDUAL noise, to the perfectly computed centroid location. The process of storing, retrieving, and converting the retrieved centroid location into a cursor location-click is represented as adding a small amount of MOTOR noise to the computed centroid location. MOTOR noise was estimated from singleton trials in which just one dot of each color was displayed

Cursor misplacement error Misplacement error is the term we use for errors that occur after the centroid computation, i.e., errors caused by "motor noise". Cursor misplacement error was estimated by the singleton trials in which only one dot per subset was shown (vs. four or eight per subset in all other trials) and the subject's task was to click on the location of a single dot in "control" trials or consecutively on the locations of two or three single dots in the multiple-response tasks. We modeled this error component as a normal random variable with mean 0, and variance equal to the variance of subject's response errors in the singleton trials.

Efficiency, Eff Eff is an abbreviation of "efficiency" to make it clear that, although it refers to a specific computation introduced in Fig. 5b, it also is intended to be similar to a common understanding of "efficiency." Eff is the absolute minimum number of stimulus items needed by an ideal detector to match the encoding efficiency of a subjects' implicit centroid error computation. Whereas encoding is all-or-none, weight in the centroid computation is determined by the color-attention filter and is continuous. The great advantage of the all-or-none encoding model is that it permits an estimate of the absolute minimum number of dots that a statistical ideal observer must utilize to match the observed accuracy of a subject's centroid

judgments – given the model assumptions about the color-attention filter and cursor misplacement error. This is an intuitive, meaningful measure of the subject's ability to extract and transmit information from the display.

Computation of Eff, encoding efficiency Eff incorporates one additional assumption beyond the computation of N*p described in Fig. 5b: Cursor misplacement error variance (as measured in singleton trials) is subtracted from observed error variance, therefore Eff is always larger than the corresponding N*p. For each multi-centroid condition, there was a matched singleton condition. For example, for judging three eight-dot centroids, in the corresponding singleton task, the subject judges the locations of three single dots of the same color and placed in the same locations as the corresponding eight-dot centroids. The three singleton responses are made in the same order as in the centroid task. Singleton errors typically are quite small, so the reduction of response error is a small but significant correction. Insofar as remembering and clicking on two or three dot locations is not very different from remembering and clicking on two or three centroid locations, the singleton correction also removes storage and retrieval losses. Because filter error and bias usually are also small, the singleton correction makes Eff a measure primarily of the efficiency of the Encoding component

(Fig. 7). That is, Eff is a measure primarily of encoding efficiency, here indexed by the number of encoded dots as derived from the ideal detector model.

In the computation of Eff, the color-attention filter is assumed to be perfect, as in the computation of N*p. This ensures that there is no overlap of items in different filters, no double-counting of the same stimulus dot, and yields the minimal estimate of Eff. Exactly as in Fig. 5b, Eff can be derived by a Monte Carlo method in which the ideal detector is presented the same type of stimulus and task as the subject. The random decimation differs from trial to trial so, to be accurate, a Monte Carlo simulation to estimate model parameters requires thousands of trials. Alternatively, Appendix 1 presents an algebraic computation of Eff that is very quick and is a very good approximation to Eff in the range of the relevant data.

Model results

Color-attention filters, $f_\phi(C)$ Color-attention filters derived via Eq. 1 measure how well the subject excludes dots from the irrelevant color subset (in his/her centroid computation) compared to the included dots in the target-color subset. However, just describing a filter is not enough, it is also necessary to show that the filter accurately predicts something, i.e., the filter is real, not imaginary. To evaluate filter efficacy, we create a Pure Filter Model that assumes the only predictor of a subject's centroid judgments is the color-attention filter derived by Eq. 1 from that subject's data. The input to the Pure Filter Model consists of the same stimuli as the subject viewed. The dots in each input stimulus are assigned weights according to $f_\phi(C)$, a perfect centroid is computed, and the result is output as an x,y display-screen coordinate.

Evaluating filter efficacy via the Pure Filter Model Evaluating how well the Pure Filter Model predicts the response requires comparing the subject's response and the model prediction, vectors that both originate at the mean stimulus location, approximately (0,0). One vector ends at the subjects' response, the other vector ends at the Pure Filter Model's predicted response. The filter model accounts for 92% and 83% of the total x,y response variance for the single- and dual-response four-dot tasks and 78% and 67% for the eight dot tasks, respectively. Accounting for error variance is more demanding – it is the comparison of two vectors that originate at the true centroid, one ends at the subject's response, the other at the model's predicted response. Whereas the filter model is quite good at predicting where on the screen the subject will click, it accounts for only about 20% of the error variance (Appendix Table 1, column C). This is in part because the error is small in relation to the length of the response, and because, as the following sections demonstrate, subjects have quite accurate

color-attention filters and, therefore, understandably, color-attention filters are not a primary source of error in centroid judgments. Appendix Table 1 provides both the x,y and the centroid-judgment error predictions of the color-attention filters for the eight experimental conditions.

Figure 8 shows that for 2 x 4 dot stimuli, subjects achieved target filter weights greater than 0.9 (and therefore the distracters weights of less than 0.1) in single-response control conditions. These weights did not change significantly when subjects made dual responses. The 2 x 8 target weights are slightly lower but the overall data are basically similar. The two-response tasks did not challenge subjects' color-attention filters.

Efficiency results, Eff Figure 9 shows the AOC graphs for Eff (Encoding efficiency, Fig. 9) in terms of number of dots that an ideal detector requires to match this error for each individual judgment. The Eff results in Fig. 9 are not very different from the filter results in Fig. 8, subjects were not challenged in these conditions. Subjects performed equally well in dual- as in single-response conditions except for the slight drop in dual performance for 2*8 dot displays exhibited by 4/5 subjects.

Figure 10 shows the Attention Operation Characteristic (AOC) graphs for triple judgment color-attention filters. Perfect performance is (1,1,1), which is slightly below the upper rightmost point of the graphs. If subjects were to ignore the stimulus and click on the session's fixed mean centroid location on every trial (Perfect Bayesian Bias), that would be the point (0,0,0), which is the bottom leftmost point. The sum of the filter weights for targets and distracters plus the weight for the Bayesian Bias location is 1.0. On average, subjects achieved filter weights that are above 0.8 for the target subset colors in the single-response task (top row, data points close to axes). Filter weights for the eight-dot subsets condition are slightly lower but still above 0.7. This indicates that subjects can exclude the irrelevant subset quite well while computing and reporting the centroid of the target color subset. Filter weights derived for the triple-response task are almost same as that for the single-response task for the four-dot-each condition. For the eight-dot-each condition (Fig. 10, top right), the triple-response data point is even slightly out of the independence cube formed by the dashed lines. Recall that the filter weight measures the subject's ability to exclude irrelevant dots when computing the centroid of the target subset. Therefore, what Fig. 10 suggests is that, even when computing and reporting three centroids, subjects still show no deficit in this cognitive ability.

Efficiency, Eff Figure 11 shows the AOC graphs for Eff, encoding efficiency, in three-color single- and triple-response trials. For the four-dot condition, on average, subjects perform the triple-response task with a 7% deficit in $N*p$ compared to their performance in the single-response task (Fig. 11, top left). On the other hand, for the eight-dot-each-subset condition, there is a 20% loss in the triple response Eff
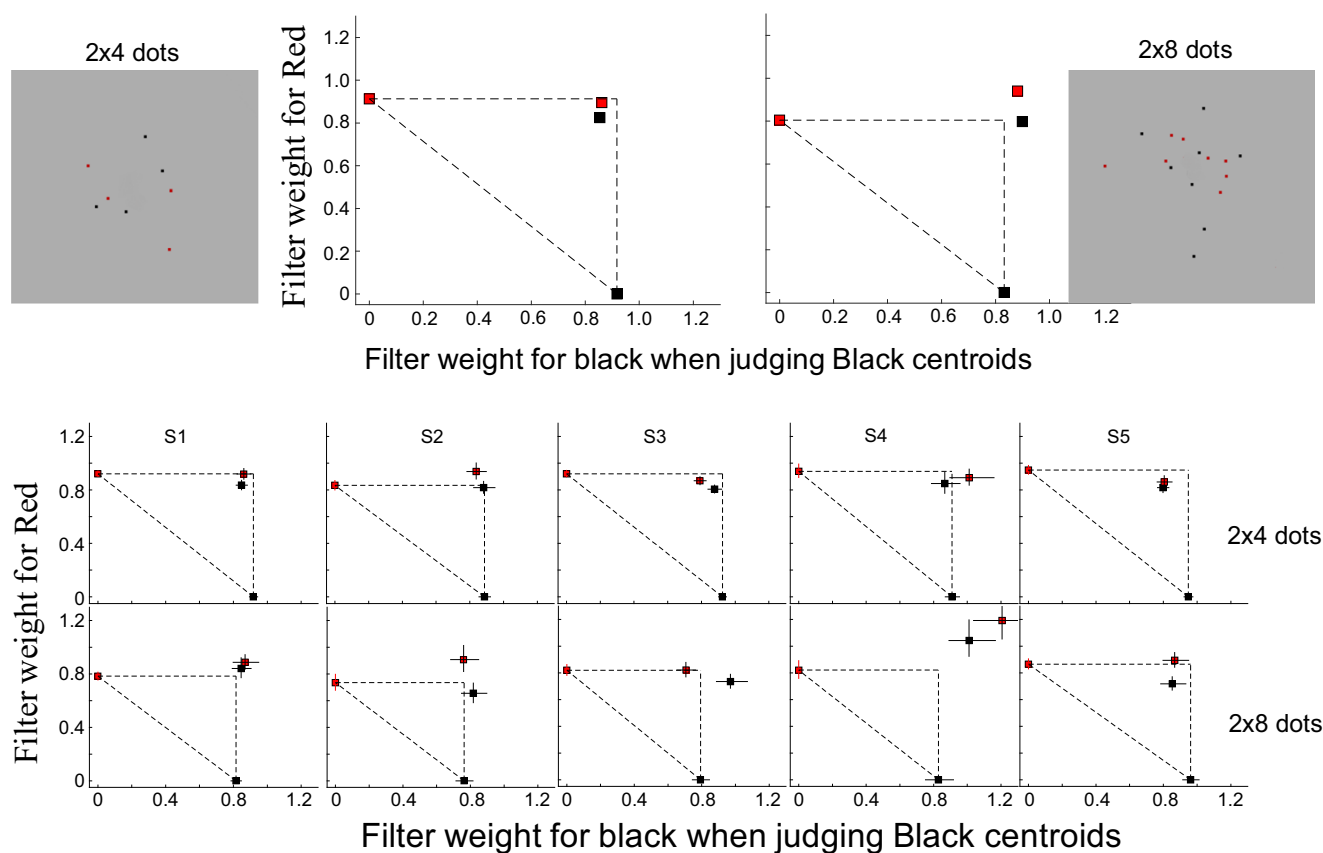
Fig. 8 Filter weights: Attention Operating Characteristic (AOC) graphs for two-color trials showing no significant loss for dual-response versus single-response attention color-filters. **Top row:** Sample stimuli plus AOC graphs of filter target-weights, average of five subjects. **Bottom two rows:** AOC graphs of target weights for each individual subject. The horizontal axis represents the filter's target weight for the Black subset, derived from the subject's judged centroid when the centroid of the Black dots was reported. The vertical axis represents the filter's target weight for the Red subset derived from the subject's judged centroids when the centroid of the Red dots was reported. Filters' target weights for single responses are indicated adjacent to the corresponding axis. The intersection of the lines that are perpendicular to the axes and which emanate from the single-response points represents the independence point where each centroid in the multiple-response task is reported as accurately in terms of filter weights as the corresponding centroid in the single-response task. A point inside the independence point indicates a multiple-task deficit. The dotted diagonal line represents the possible observed filter weights if a subject were to choose in advance to report only one particular centroid on dual-response trials. Symbol color red: first response is the red centroid (x); symbol color black: first response is black centroid. Error bars in individual subject data graphs represent 95% confidence intervals

compared to the single response Eff, and 5/5 subjects show this loss. Nevertheless, that the triple response Eff is 80% of control Eff means that triple-response performance could not have been achieved by sharing single-response resources.

To summarize the AOC analysis There was no significant loss in filter quality in multiple versus single response conditions. The main Eff loss in multiple- versus single-response conditions was 20% in the 3*8 24-dot condition, the most difficult task. The next section offers a more sensitive error analysis.

Variance analysis of the Sources of Error in Multi-Centroid Judgments

The aim in this section is to estimate the quantitative contributions of individual error sources to the total centroid judgment error that subjects make in each condition of the experiments. We initially consider four more-or-less independent error sources outlined in Fig. 7: imperfect color-attention filter, Bayesian-like bias, cursor misplacement error (Motor), and residual, which consists largely of encoding but also includes, by definition, everything else not already accounted for. Insofar as error sources are independent, the component error means and the error variances add linearly to produce the total mean error and total error variance. The mean error (vs. the magnitude of the mean error) is approximately zero for all of the error sources, that is, the errors are statistically centered around the true centroid, therefore, the mean errors are not useful. However, the component error variances are approximately additive and are very informative.

To measure the contribution of component errors in a particular condition we make full use of the assumption of error independence. Consider just one of the 14 experimental
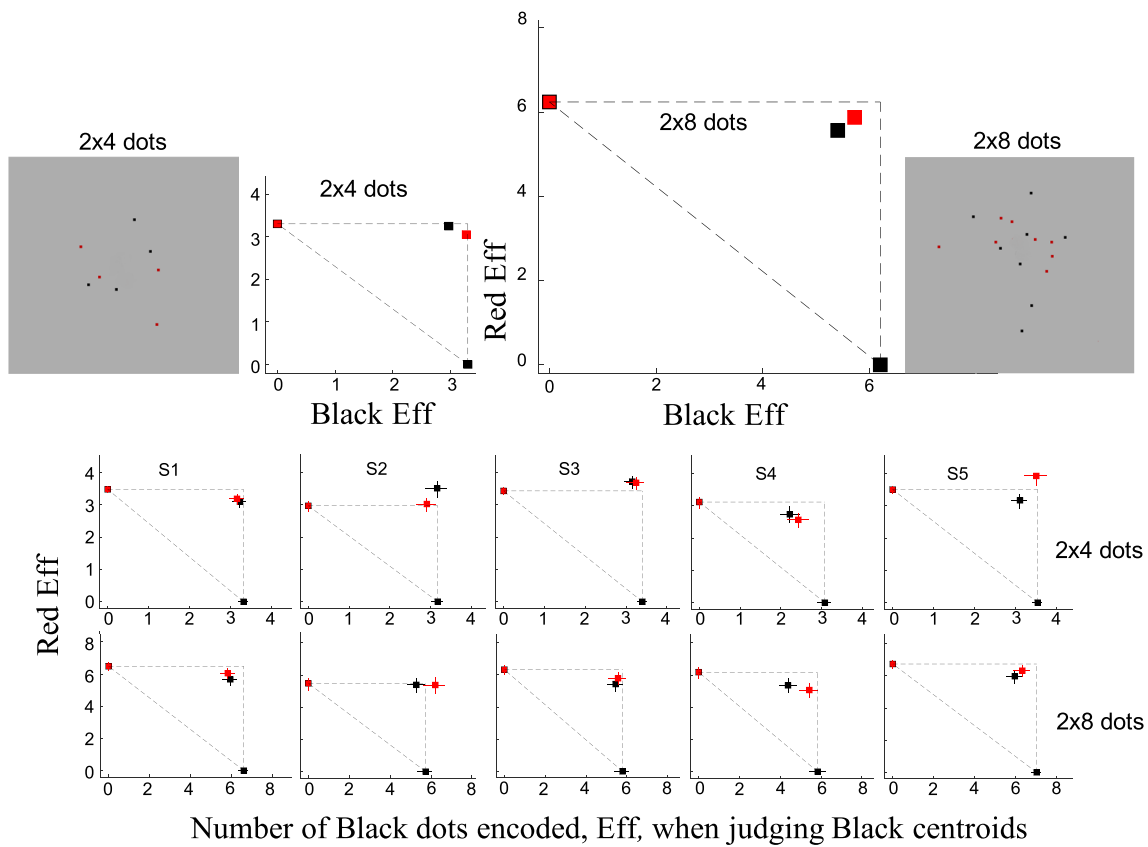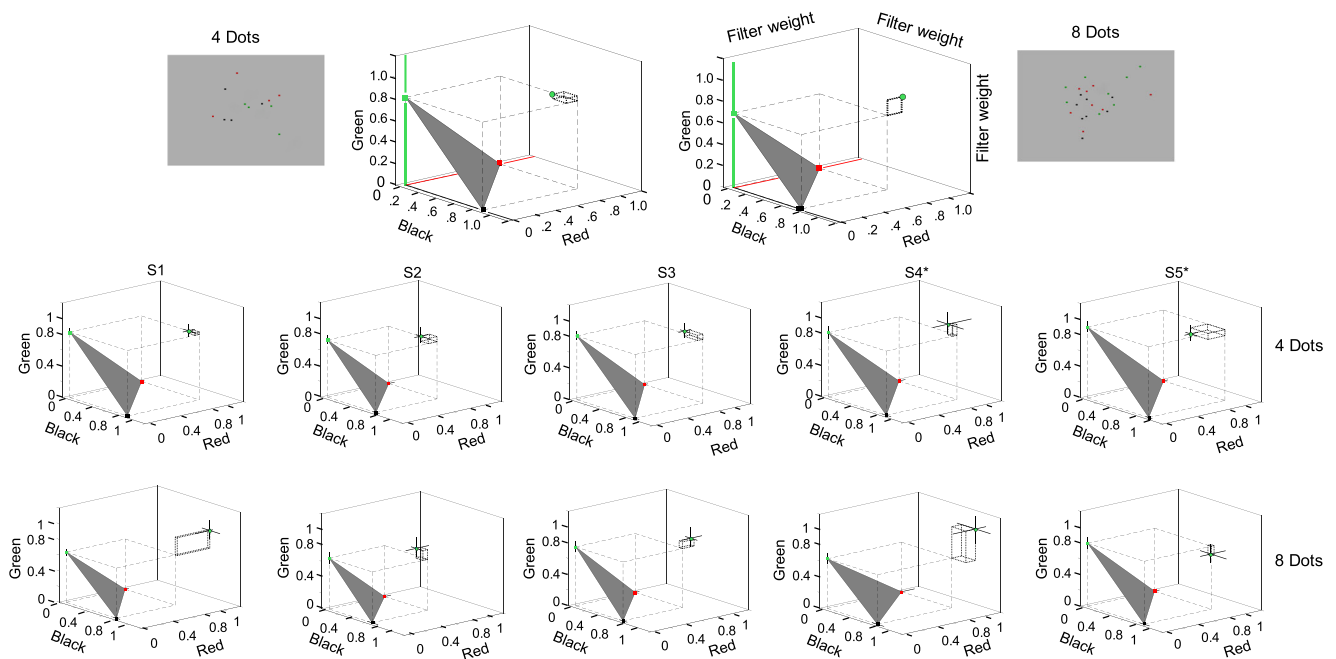
Fig. 9 Attention Operating Characteristics (AOC) for Eff (encoding efficiency) in two-color trials show no dual-response loss in 2*4 dot stimuli and a slight (7%) loss in 2*8 stimuli. Eff is the minimum number of stimulus dots an ideal detector (Fig. 5) needs to match the subject's singleton-corrected performance in concurrent judgments of black-dot and red-dot centroids following the same brief exposure. **Top row:** Sample stimuli plus AOC graphs, average of five subjects. **Bottom two rows:** AOC graphs for each individual subject. The horizontal axis represents Eff, the number of dots utilized, when judging the Black subset. The vertical axis represents Eff when judging the Red subset. Eff for the single response task is indicated adjacent to the corresponding axis The intersection of the lines that are perpendicular to the axes and which emanate from the single-response points represents the independence point where each centroid in the multiple-response task is reported as accurately in terms of Eff as the corresponding centroid in the single-response task. A point inside the independence point indicates a multiple-task deficit. The dotted diagonal line represents the possible observed Effs if a subject were to choose in advance to report only one particular centroid on dual-response trials. Symbol color red: first response is the red centroid (x); Symbol color black: first response is black centroid. Error bars in individual subject data graphs represent 95% confidence intervals

conditions, for example, a particular number of colors; number of target dots per centroid; control, first, second, or third response. If the subject had a perfect color-attention filter, there would be zero error for this component. However, the color-attention filter was not perfect, we know the weight of each color in the subject's centroid computation. Given all the stimuli that the subject judged in this condition, a filter with those imperfect weights and perfect everything else, would produce a distribution of centroid errors; the variance of that distribution is our Estimate1 of filter error variance. Estimate 1 of the error variance produced by an imperfect filter includes both the filter weights and the bias weight. To separate the bias error from the filter error, it's convenient that filtering is a linear operation. We can regard the color-filter weights as applying to only (1-Bias) of the filter error so that (1-Bias)*Estimate1 = Estimate2, the estimate of filter error variance. Alternatively and equivalently, we note that the pure bias error distribution is simply the centroid distribution. The Bias error variance is then Bias*Centroid_Variance. The cursor misplacement error distribution (Motor) is estimated directly from singleton trials. Subtracting these three variances form the total error variance gives the Residual variance. Within the context of the computational model, Residual is entirely due to encoding losses. However, the model certainly is not completely correct, so Residual also includes model error, which encompasses all error sources that are not explicitly considered in the model.

Figure 12 shows the results of the above error analysis for all 14 experimental conditions arranged in order of increasing total error. It's immediately obvious that the main error component is residual, and that generally but not regularly, all errors increase as total error increases. The unexpected total

Fig. 10 Attention Operating Characteristic (AOC) for color-attention filter weights in three-color single- and triple-response trials showing no significant multi-response filter deficits. **Top row:** Sample stimuli plus AOC graphs of filter weights, average of five subjects. **Bottom two rows:** AOC graphs of filter weights for each individual subject. Each axis represents the filter weight for a color when that color subset is the target subset. Filter weights for the single response are indicated close to the axes, and by the dashed lines. These single response points form the vertices of the dark triangle, a surface that represents attention to only one color per trial. The dashed lines form the "independence cube," of which the upright corner represents the independence point where each centroid in the triple-response task would be reported as accurately in terms of filter weights as the corresponding centroid in the single-response task. The space inside the independence cube represents triple-response task deficits. The gray triangular surface represents the possible observed Effs if a subject were to choose in advance to report only one particular centroid on triple-response trials. Error bars in individual subject data graphs represent 95% confidence intervals

error of the three-color four-dot centroid task is most likely due to the greater variance of four- versus eight-dot centroids. The change in centroid variance between different types of stimuli could have been compensated by increasing the spread of dots in eight-dot versus four-dot centroids, or by adding centroid variation that's independent of the dot distribution, and in many other ways, but there is no way to make conditions that differ in only one aspect.

Figure 12 omits a critical error component, the increase in centroid judgment error when multiple judgments – versus a single judgment – are made. Figure 13 shows the complications that are necessary in the simple flow chart of Fig. 7 to deal with single-versus-multiple centroid judgments. Each one of the four error sources is affected. When multiple judgments are required, color-attention filters are impaired, bias is increased, cursor misplacement increases, and residual (mostly encoding), has the largest absolute increase.

Figure 14 shows the results of the more elaborate analysis outlined in the flowchart of Fig. 13. Unlike the four approximately additive error components, the added difficulty of multi-versus-single responses increases every error component. Figure 14 represents 4 x 4 (single-response) +10 x 8 = 96 (double- and triple-response) error components in the 14

conditions, an analysis that was not anticipated when we undertook the experiments. These data are highly complex and specific to the particular experimental conditions. There is much that is suggestive but little that does not require more substantiation. Therefore, we consider only a few of most important overall aspects. First and primary is the demonstration of a method of gaining insight into micro-components of perceptual and cognitive processes by a detailed analysis of error variance. Second, the empirical observation that the difficulty of making multiple versus single centroid judgments adversely affects all the component processes, primarily residual, which is mostly encoding, and to a lesser extent, filter, bias, and motor.

Summary

After training to asymptote in judging centroids, subjects viewed 250-ms flashes of displays with two and three colors of 8–24 randomly placed dots and judged the centroids of each of the colors. In control conditions, centroids of just one particular color were judged for an entire block of trials.
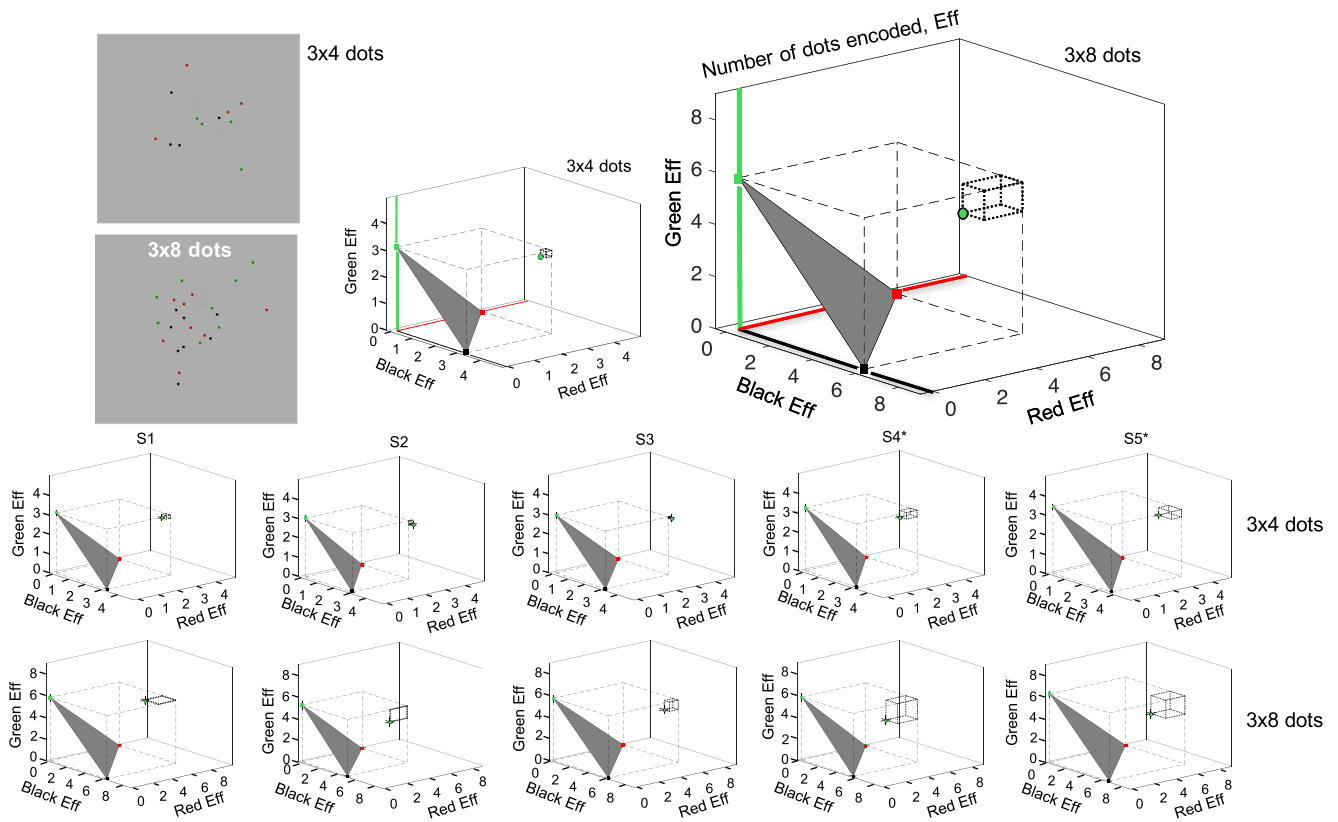
Fig. 11 Attention Operating Characteristics (AOC) for Eff, encoding efficiency, in three-color, single- and triple-response trials show no multi-response deficit for 3 x 4 dot stimuli but a 20% multi-response deficit for 3 x 8 dot stimuli. **Top row:** AOC graphs of the average of the five subjects. **Bottom two rows:** AOC graphs for each individual subject. Axes represent Eff for each of the target color subset, derived from the subject's corresponding centroid reports. Effs for the single response trials are indicated close to the axes, and by the dashed lines perpendicular to the axes. These single response points form the vertices of the dark triangle, a surface that represents attention to only one color per trial. The dashed lines form an "independence cube," of which the upright corner represents the independence point. The space inside the independence cube represents the triple-response task deficits. Distances between the triple-response data point and the vertices of the independence cube are indicated by the dotted lines. The dotted lines form a small cube of which the volume represents the subject's triple-response deficit. The gray triangular surface represents the possible observed Effs if a subject were to choose in advance to report only one particular centroid on triple-response trials. Error bars in individual subject data graphs represent 95% credible intervals

The response errors – the distances of the judged centroids from the true centroids – showed that errors in multiple judgments were always greater than in single-judgment controls, that the stimulus colors used were equivalent, and that order of report was minor factor. All data were reported for each individual subject as well as for the average.

For 2/5 subjects, a very effective masking stimulus was presented immediately after the stimulus exposure. The mask and no-mask data were statistically indistinguishable.

An ideal detector model was used to assess the quality of the responses. The ideal detector was presented with the same stimuli as the subjects, it perfectly encoded the position and color of every dot and it perfectly computed centroids, its output. However, a decimation process just before the centroid computation randomly eliminated dots. A Monte Carlo simulation was used to determine the minimum fraction p of surviving dots in order for the ideal detector's error to match a subject's error. For a stimulus of N dots, N*p is the expected number of processed dots. By implication, the subject must have processed at least the equivalent of N*p dots to achieve his/her judgment accuracy.

In the most difficult condition, three colors, eight dots each color, three judgments, the average N*p for five subjects was 13.0 dots. In similar stimuli, subjects typically consciously remember only the very approximate location of just two dots, so utilization of 13.0 dots is evidence of an extremely efficient pre-conscious computation, often referred to as a statistical summary representation. Had subjects been able to make all three judgment as accurately as their single-centroid judgments in control conditions, their N*p would have been 17.8 dots.

Judging a centroid requires an approximately Euclidean representation of space. This psychophysical judgment is assumed to utilize a salience map that records a real-valued
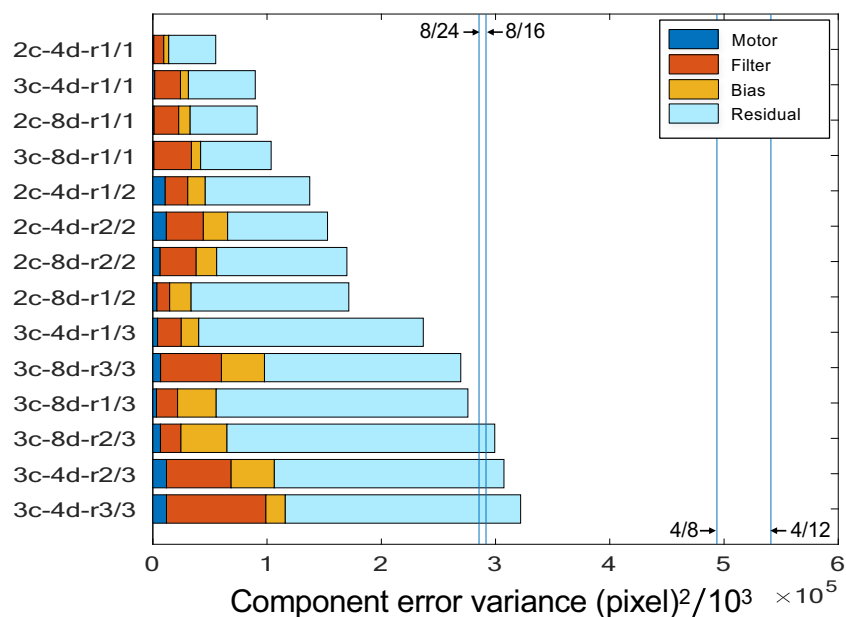
Fig. 12 Results of the error analysis – the proportion of the error variance accounted for in each of the 14 experimental conditions by four processes: An imperfect color-attention filter (Filter), Bayesian-like response bias (Bias), cursor misplacement error that includes memory and retrieval error (Motor), and faulty encoding plus any unknown error sources (Residual). The vertical lines n/N indicate the mean variances of the target centroids of n dots in stimuli of N dots. Notation example: 3c-4d-r2/3, indicates a three-color stimulus, four dots of each color, and the second of three required centroid judgments for each such 12-dot stimulus
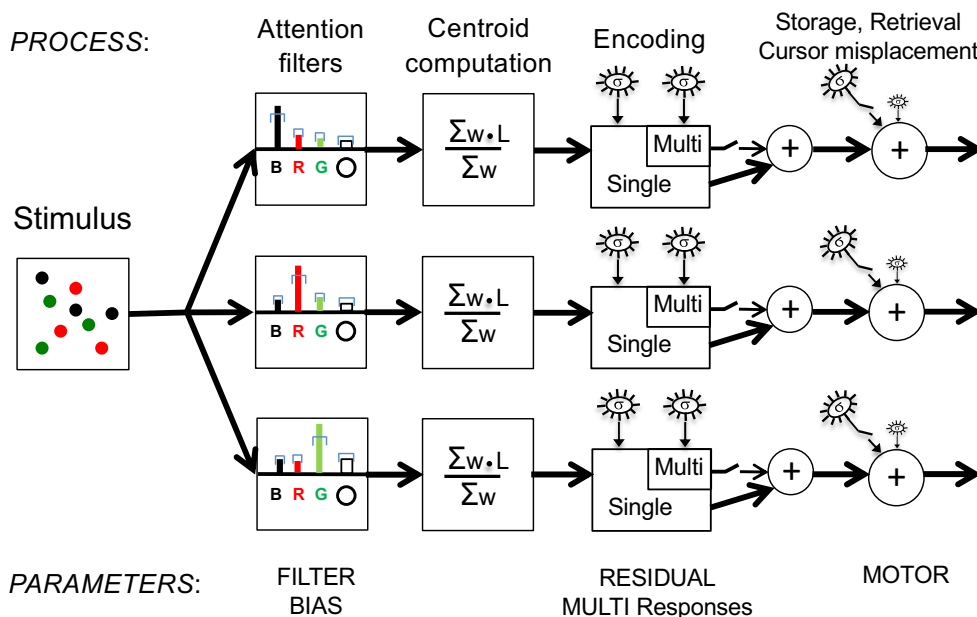
Fig. 13 Flowchart of the computational model that transforms the visual input stimulus into – in this illustration – three centroid-localization responses with four independent sources or response error (FILTER, BIAS, MOTOR, RESIDUAL) and one error source that interacts all four (MULTI vs. single required response). The response impairment in each one of a set of three (in this illustration) multiple responses versus the corresponding single response is represented as a decrement in the selectivity of every attention filter (distracter weights increase while target weights decrease), an increase in bias, plus added noise (small, oval noise source) to every RESIDUAL and MOTOR component. The breaks in the arrows from Multi and the MOTOR noises represent switches that are closed on multi-response trials. For more details, see caption of Fig. 7
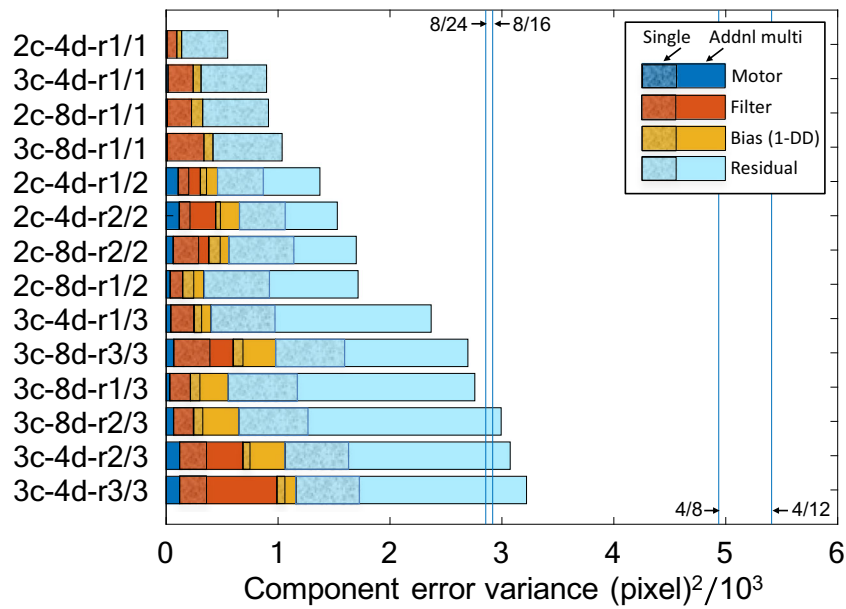
Fig. 14 Results of the error analysis – the proportion of the error variance accounted for in each of the 14 experimental conditions by four additive independent error sources: (1) An imperfect color-attention filter (Filter), (2) Bayesian-like response bias (Bias), (3) faulty encoding, plus all unknown error sources (Residual), (4) storage, retrieval, and cursor misplacement error (Motor), and (5) multi-versus-single responses on reach trial, an error source that interacts with the four additive errors. Colors indicates the error variance component; the left, darker, textured color indicates the variance on the corresponding single-response control trials; the lighter color indicates the additional error variance on corresponding responses in multi-response trials. For more details, see caption Fig. 12

salience at each map location, but with no indication within the map itself of what features may have produced that value. Therefore, a single salience map cannot represent the appropriate salience values for more than one centroid judgment. We conclude that if subjects can compute three accurate centroids, they must have utilized at least three salience maps.

A model of the salience system was proposed. Three color-attention filters segregate the differently colored dots into three groups. Within a group, the precise location and color of the dots is retained. Each group delivers its contents to a salience map, a centroid is computed on the contents of each map. Besides the centroid computation, other process/computations that use a salience map as an input include determining processing priority according to local salience, sequencing locations to be searched, eye and attention movement priority, the computation of the direction of third-order motion, and others.

Part 2 considered the source of centroid judgment errors. Computations were derived to quantitatively describe: color-attention filters in terms of the salience weights assigned to targets and to distracters, Bayesian-like bias, the combination of centroid location storage, retrieval, and cursor misplacement error, and Eff (efficiency) that represents the number of encoded dots available for a centroid computation.

Attention Operating Characteristics of single versus multiple judgments showed that filter weights were virtually unimpaired in multiple versus single judgments, Eff was reduced 20% in three-color 24 dot triple judgments, by 7% in two-color 16-dot dual judgments, by 3% in three-color 12-dot triple judgments, and was almost unimpaired in two-color, eight-dot centroid dual judgments.

A more sensitive analysis of five error sources in centroid judgments used the variance of the centroid judgment errors. Four error sources: imperfect color attention filters, a Bayesian-like bias for the session mean centroid location, cursor misplacement error, and stimulus encoding, storage, and retrieval were assumed to be approximately independent so that their error variances would linearly add. A fifth error source, the greater difficulty of multiple versus single responses, was found to be an interaction that increased all of the four other errors. For the 14 experimental conditions, this analysis yielded 96 error sources, graphically displayed.

Conclusions

Previously it was believed humans had only one salience map; this study shows they have access to at least three. A new error variance analysis enables a quantitative micro-analysis of five component process in judging centroids of selectively attended items.

Appendix 1. Algebraic computation of Eff

Eff is an abbreviation of efficiency to make it clear that it refers to a specific computation introduced in Fig. 6. Eff is the absolute minimum number of stimulus items needed by an ideal detector to match a subjects' centroid error. The number of stimulus items available to the ideal detector is reduced by a random decimation process. Let p be the survival probability of each stimulus element, then the expected number of surviving elements is Eff=N*p. In this section, there are two additional assumptions: (1) The color-attention filter is perfect (always assumed here for ideal detector computations to ensure that there is no overlap of items in different filters. (2) Cursor-misplacement error variance is subtracted from observed stimulus error variance. This is a minor correction. As filter error and bias are also very small, the cursor misplacement correction is intended to make Eff primarily a measure of the efficiency of Encoding, Storage, and Retrieval of stimulus items that survive to be input to the centroid computation.

After subtracting the cursor misplacement error from Q_x, Q_y, the remaining error is designated simply as encoding because the encoding, storage, and retrieval components are inseparable in the present experiments. Note that dot loss by failing to be encoded is an all-or-none process; the weight of an encoded dot in the centroid computation is determined by a subject's color-attention filter, which is a continuous process To find the Maximum A Posterior (MAP) estimator for p, proceed as follows: Let N be the number of stimulus dots, and p be the probability of each dot being preserved (i.e., encoded). Then the probability that the decimated dot cloud contains k dots is given by

$$P_k(p) = \binom{N}{k} p^k (1-p)^{N-k} \tag{2}$$

Let Λ_k be the difference between the true centroid of the original N target dots, and the centroid of a dot set that has preserved k of the N target dots. That is,

$$\Delta_k = \frac{1}{k}\sum_{j=1}^{k} x_j - \frac{1}{N}\sum_{j=1}^{N} x_j = \sum_{j=1}^{k}\left(\frac{N-k}{Nk}\right)x_j - \frac{1}{N}\sum_{j=k+1}^{N} x_j \tag{3}$$

Recall that x is drawn from a Gaussian distribution with mean 0 and standard deviation σ. It follows that Δ_k is normal with mean 0 and variance:

$$\text{var}(\Delta_k) = k\left(\frac{N-k}{Nk}\right)^2 \sigma^2 + \frac{N-k}{N^2}\sigma^2 \tag{4}$$

Now, let σ_m be the standard deviation of the random variable that represents the cursor misplacement error (see the previous section). Assuming the misplacement error is normal with mean 0, then the random variable D_x that gives the difference between x location of the true centroid and of the subject response in a trial in which exactly k of the original N dots are preserved, is normal and has mean 0 and standard deviation

$$\sigma_k = \left(\text{var}(\Lambda_k) + \sigma_m^2\right)^{\frac{1}{2}} \tag{5}$$

The density function of D_x is given by:

$$f_k(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(\frac{-\alpha^2}{2\sigma_k^2}\right) \tag{6}$$

There will be a corresponding random variable D_y. Now consider an experimental trial with an initial number of N dots. Let $D = (D_x, D_y)$ be the difference between the true centroid location and subject response (R_x, R_y). Then the likelihood function of p, the probability that a dot is preserved, given the observation D is:

$$\lambda(p) = P(D|p) = \left(\sum_{k=1}^{N} f_k(D_x)P_k(p)\right) \times \left(\sum_{k=1}^{N} f_k(D_y)P_k(p)\right) \tag{7}$$

Assuming independence between experimental trials, then the likelihood given data of all trials is:

$$\Lambda(p) = \prod_{\text{All trials}} \lambda(p) \tag{8}$$

Assuming a uniform prior on p, then the posterior density of p is:

$$f(p) = \frac{\Lambda(p)}{\sum_{q}\Lambda(q)} \tag{9}$$

The denominator can be approximated by sampling p at small intervals. With the posterior distribution given by Eq. 9, one can obtain an estimate for p and its credible intervals.

Appendix 2. Table of numerical values of estimates of response quality

Table 1. Average over 5 subjects of (A) Response quality, the fraction of target variance accounted for by subjects' responses, (B) Filter prediction quality, the accuracy of the subjects' color- attention filters in predicting the subjects' responses, (C) How much less error the color attention filter has than the subject on a scale where 1 = filter error is equal to subject error, 0 = filter prediction is irrelevant for subject error; (D) the fraction of cursor misplacement error variance relative to subjects' total error variance.

Number of colors	Number of responses	Number of target dots	$1 - \dfrac{E\left(R_{subj} - Target\right)^2}{\text{Target Variance}}$ A: response accuracy	$1 - \dfrac{E\left(R_{filter} - R_{subj}\right)^2}{\text{Target Variance}}$ B: filter predicted responses	$\dfrac{E\left(R_{filter} - Target\right)^2}{E\left(R_{subj} - Target\right)^2}$ C: (filter error)/ (subj error)	$\dfrac{E\left(R_{singleton} - Target\right)^2}{E\left(R_{subj} - Target\right)^2}$ D: (cursor error)/ (subj error)
Two colors	Single	4	0.88	0.92	0.19	0.018
		8	0.69	0.78	0.20	0.014
	Dual	4	0.79	0.83	0.23	0.085
		8	0.62	0.67	0.12	0.028
Tree colors	Single	4	0.80	0.86	0.25	0.020
		8	0.62	0.79	0.27	0.010
	Three	4	0.50	0.63	0.20	0.031
		8	0.35	0.49	0.12	0.021

References

Ackermann, J. F., & Landy, M. S. (2015). Suboptimal decision criteria are predicted by subjectively weighted probabilities and rewards. *Attention, Perception, & Psychophysics,* 77(2):638–658. https://doi.org/10.3758/s13414-014-0779-z

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.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12(2), 157–162.

Attarha, M., Moore, C. M., &Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human perception and performance*, 40(4): 1440-1449.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains. *Journal of Vision,* 9(13). https://doi.org/10.1167/9.12.13

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vison*, 10(4), 433–436.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43, 393–404.

Chong S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research.* 45: 891-900.

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

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

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. Journal of Vision, 14 (9), pp. 1-14. https://doi.org/10.1167/14.9.22

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be numerated in parallel. *Psychological Science*, 17 (7), 572-576.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, 43: 663-676.

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

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology* 4(4): 219-227.

Lu, V. T., Wright, C. E., Chubb, C., & Sperling, G. (2019). Variation in target and distractor heterogeneity impacts performance in the centroid task. *Journal of Vision, 19(4):21,* 1-13. https://doi.org/10.1167/19.4.21

Li, V., Herce Castañón, S., Solomon, J. A., Vandormael, H., Summerfield, C., & Mamassian, P. (2017). Robust averaging protects decisions from noise in neural computations. p*LoS Computational Biology, 13*(8):e1005723. https://doi.org/10.1371/journal.pcbi.1005723

Lu, Z.-L., & Sperling, G. (1995). Attention-generated apparent motion. *Nature*, 377, 237-239.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118*(4), 346–373. https://doi.org/10.1037/0096-3445.118.4.346

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63:* 81-97

Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision ,13*(8): 1-11

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*, 65-97.

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

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,* 1-14.

Solomon, J. A., May, K. A., & Tyler, C. W. (2016) Inefficiency of orientation averaging: evidence for hybrid serial/parallel temporal integration. *Journal of Vision*, 16(1):13, 1-7

Solomon, J. A., & Morgan, M. J. (2018). Calculation efficiencies for mean numerosity. *Psychological Science*, 29, 1824-1831.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11, Whole No. 498).

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica, 27*, 285–292. https://doi.org/10.1016/0001-6918(67)90070-4

Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science, 174*(4006), 307–311. https://doi.org/10.1126/science.174.4006.307

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654. https://doi.org/10.1126/science.153.3736.652

Sun P, Chubb C., Wright C. E., & Sperling G. (2016a). Human attention filters for single colors, *Proceedings of National Academy of Sciences, 113*(43). E6712-E6720

Sun P, Chubb C., Wright C.E., & Sperling G. (2016b). The centroid paradigm: Quantifying feature-based attention in terms of attention filters, *Attention Perception and Psychophysics, 78*(2): 474–515.

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2018). High-capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of National Academy of Sciences, 115*(52), E12153–E1216. Published online 2018 Dec 13. https://doi.org/10.1073/pnas.1814657115

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12: 97–136.

Tseng, C.-h., Gobell, J., & Sperling, G. (2004). Movement of isoluminant red-green gratings and of random-dot stereo depth gratings is perceived by the same salience motion-perception mechanism. *Perception, 33(Suppl.)*, 154.

van Santen, J. P. H., & Sperling, G. (1984). Temporal covariance model of human motion perception. *J. Opt. Soc. Amer. A, 1,* 451–473

Wolfe, J. M., Võ, M. L., 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

Wundt, W. (1912/1924). fAn introduction to psychology. (R. Pintner, translator, from the Second German edition). London: George Allen and Unwin. (Original work published 1912.) Chapter 1, pp. 1–42.

Rosenholtz, R. (2011). What your visual system sees where you are not looking. *Proc. SPIE (Human Vision and Electronic Imaging XVI)* 2011;7865:786510.

Rosenholtz, R., Huang, J., Raj, A., Benjamin J. Balas, B.J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(14). https://doi.org/10.1167/12.4.14