High-capacity preconscious processing in concurrent groupings of colored dots

Peng Sun[a,1], **Charles Chubb**[a], **Charles E. Wright**[a], **and George Sperling**[a,1]

[a]Department of Cognitive Sciences, University of California, Irvine, CA 92697

Contributed by George Sperling, November 6, 2018 (sent for review August 27, 2018; reviewed by Justin Halberda and Michel Treisman)

Grouping is a perceptual process in which a subset of stimulus components (a group) is selected for a subsequent—typically implicit—perceptual computation. Grouping is a critical precursor to segmenting objects from the background and ultimately to object recognition. Here, we study grouping by color. We present subjects with 300-ms exposures of 12 dots colored with the same but unknown identical color interspersed among 14 dots of seven different colors. To indicate grouping, subjects point-click the remembered centroid ("center of gravity") of the set of homogeneous dots, of heterogeneous dots, or of all dots. Subjects accurately judge all of these centroids. Furthermore, after a single stimulus exposure, subjects can judge both the heterogeneous and homogeneous centroids, that is, subjects simultaneously group by similarity and by dissimilarity. The centroid paradigm reveals the relative weight of each dot among targets and distractors to the underlying grouping process, offering a more detailed, quantitative description of grouping than was previously possible. A change detection experiment reveals that conscious memory contains less than two dots and their locations, whereas an ideal detector would have to perfectly process at least 15 of 26 dots to match the subjects' centroid judgments—indicating an extraordinary capacity for preconscious grouping. A different color set yielded identical results. Grouping theories that rely on predefined feature maps would fail to explain these results. Rather, the results indicate that preconscious grouping is automatic, flexible, and rapid, and a far more complex process than previously believed.

perceptual grouping | summary statistics | centroid judgments | grouping by similarity | preconscious processing

Current models of visual processing assume that low-level vision produces retinotopically organized maps of various "features" that can be thought of as "neural images" (1). For example, one such feature map might reflect the distribution across space of "redness"; another might reflect the distribution of "vertical" pattern elements; yet another might reflect the distribution of pattern elements in a certain spatial frequency band. The concept of multiple feature maps has a long history but is probably best known from studies of visual search (2–6). Multiple predefined feature maps are a core concept of current computational models of subsequent visual functions such as visual attention (7–10). These models make various assumptions about processes that operate on the output of these feature maps to segment the visual field into qualitatively distinct scene components, to establish spatial relations between scene components, and to aggregate some of the scene components into groups (11–15). These processes represent an intermediate level of visual processing that is preliminary to higher-order visual processing such as object recognition and localization that enable visually controlled behavior.

The current study focuses on an intermediate level of visual processing, the grouping process. Here, by grouping we refer to the process in which a scene is divided into different, generally nonoverlapping segments. Each segment as a whole, a group, may be subsequently processed by higher-level visual functions that, among other things, compute summary statistics of the group. This intermediate level of processing is also known as segmentation, figure-ground. Gestalt psychologists proposed various heuristic rules that govern grouping: items tend to be grouped when they are close, connected, similar, move together, and so forth (16). Here, our focus is grouping by similarity; items sharing similar features (e.g., color, shape, size) tend to be grouped together. Grouping by similarity is typically measured through indirect comparisons against other grouping principles (refs. 17–23; see ref. 24 for a review), or through subjective ratings (25). However, we are not aware of studies that directly and quantitatively measure grouping by similarity, that is, grouping by a common property.

Here, we introduce a recently developed paradigm, the centroid paradigm, which is particularly adapted to the study of grouping because it enables a quick, direct quantitative measure of the "weight" that each type of element in the scene contributes to the group—elements that should be in the group (targets) and elements that should not (distractors). With this paradigm, we discover some unexpected properties of the preconscious grouping process that, among other consequences, force a rejection of the concept of predefined feature maps in the context of grouping by similarity.

Outline

The experiments herein deal with subjects' abilities, from observing a brief exposure of an array of 26 dots of eight different colors, to group it into two subsets, a homogeneous subset of 12 identically colored dots and a heterogeneous subset of 14 dots of seven different colors. The subjects' ability to group is demonstrated by

Significance

A critical visual process is segmenting a scene into objects to be processed (foreground) and the remainder (background). Humans are extraordinarily good at segmentation, but how they accomplish this still is not well understood. An important component process in segmentation is grouping by similarity, for example, by color, shape, size, etc., thereby organizing visual information into coherent chunks for subsequent stages in object recognition. That our subjects can also group dissimilar items shows that the grouping process itself is much more complex than previously believed. Furthermore, we present a fully quantitative account of the inclusiveness/exclusiveness of the grouping process and of its extraordinary perceptual capacity. The amount of information preconsciously utilized to form a group is many times greater than is consciously available.

Author contributions: P.S., C.C., C.E.W., and G.S. designed research; P.S., C.C., C.E.W., and G.S. performed research; P.S., C.C., C.E.W., and G.S. contributed new reagents/analytic tools; P.S. and G.S. analyzed data; and P.S., C.C., C.E.W., and G.S. wrote the paper.

Reviewers: J.H., Johns Hopkins University; and M.T., Oxford University.

The authors declare no conflict of interest.

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

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1814657115/-/DCSupplemental.

Published online December 13, 2018.

their accuracies in judging the "centroid" (the center of mass) of each group. The method and the accompanying analysis reveal the weight assigned to each type of target item, and the weight assigned to each type of distractor item, thereby revealing in detail how well subjects have excluded all distractors. We use the term "selectivity" to refer to a subject's ability to exclude distractor items, an indicator of subjects' ability to group the desired items. Another indicator is the amount of stimulus information utilized in forming a group. We measure the lower bound for the proportion of target items that must have been included to account for the accuracy of centroid judgments. We use the term "efficiency" to refer to this lower bound.

Experiment 1 shows that subjects can, with high selectivity and high efficiency, judge the centroids of (*i*) all dots in the display, (*ii*) the homogeneous set only, and (*iii*) the heterogeneous set only. Experiment 2 shows that, with the same selectivity and only a slight efficiency loss, subjects can, on the same trial, simultaneously judge the centroids of both the homogeneous and the heterogeneous sets.

Subjects achieve their performance in experiments 1 and 2 without knowing in advance the color of the homogeneous set before each trial. That subjects can group by both similarity and by dissimilarity, demotes similarity from an explanatory to a merely descriptive principle. In a control experiment, we show that changing to a different set of eight colors yields data that are equivalent to the original set. That virtually any arbitrarily chosen color can serve as a feature demonstrates that the simple concept of dividing and aggregating predetermined color-feature maps is too simple. An ideal observer requires the exact locations of minimally 8–9 dots (of 12 and 14) to match the accuracy of the subjects' judged centroids for homogeneous or heterogeneous sets, and minimally 17 dots to match judged centroids for all 26 dots.

Experiment 3 confirms that the computation of a heterogeneous centroid in experiments 1 and 2 is the result of a faster, preconscious grouping process that could not be achieved by computing the centroid of all and subtracting the centroid of the homogeneous dots. Experiment 4 utilizes two change blindness paradigms to determine how many dots are consciously available to subjects following the brief exposures of experiments 1 and 2. The number of consciously available dots is less than 2—vs. 15 available to centroid judgments.

Together, these four experiments enable a quantitative characterization of both the high selectivity and the capacity of a preconscious perceptual grouping process that makes available to consciousness precise statistics, centroids, distilled from enormously more input information than could have been consciously processed.

Methods

The centroid judgment paradigm adopted in this study was originally developed by Drew et al. (26) to study feature-based attention, and was considerably enhanced by Sun et al. (27, 28). On all trials, subjects viewed a 300-ms flash of a display of 26 dots, 12 dots of one randomly chosen color (homogeneous set), plus 14 dots consisting of 2 each of 7 remaining colors (heterogeneous set; see Fig. 1 for example). The task was to group the dots into different subsets, according to instructions. Subjects' grouping ability is demonstrated by their abilities to accurately judge the centroid of the desired subset. For example, to achieve perfect homogeneous centroid judgments, subjects would have to perfectly group the 12 identical dots while completely ignoring the 14 heterogeneous dots. In a more realistic perceptual grouping process, some heterogeneous dots "slip" into the formed similarity group while some identical dots slip away. In both cases, the location of the judged centroid is influenced by the accuracy of the grouping process. From the precise 2D locations of a subject's centroid judgment errors, we infer the details of the subject's grouping process and the magnitude of the aggregate internal noise.

The subject's task in homogeneous blocks of trials was to move a cursor to the (remembered) centroid—the mean location—of the group of homogeneous dots and, on heterogeneous blocks of trials, to move the cursor to the centroid of the group of heterogeneous dots. On each trial, after the subject's response, there was complete feedback showing the dot display, the true centroid, and the subject's judged centroid.

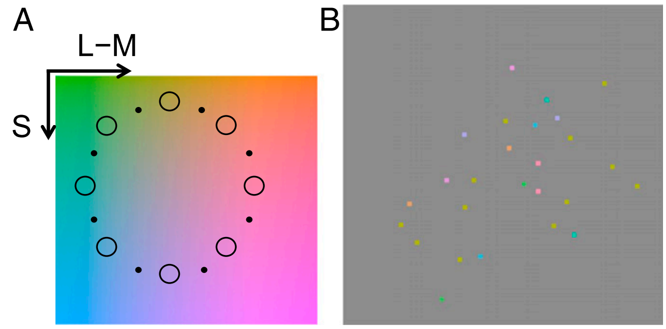

Fig. 1. The locations of the dot colors in cone color space and an example stimulus. (*A*) The colors lie on an equiluminant plane in cone space. The *y* axis represents blue-cone (short wavelength, S) activation in the downward direction. The *x* axis is chosen to be orthogonal to the *y* axis in this equiluminant plane and represents a red–green axis, approximately long (L) minus middle wavelength (M) cone activation. The circles represent the eight colors used in experiments 1 and 2. The dots represent the alternate colors used in control experiment 1a. (*B*) A representative stimulus. The 26 stimulus dots comprise two groups: a homogeneous group of 12 yellow dots (in this particular example) and a heterogeneous group (14 dots, 2 each of the 7 other colors). Background luminance in this reproduction is deliberately much darker than in the actual experimental displays to make the stimulus dots more visible.

A brief control experiment was conducted to demonstrate that the original eight colors did not stimulate unique features, that another arbitrary sets of eight very different colors would produce the same results. Two subjects performed the homogeneous and heterogeneous centroid judgment tasks for a different set of eight equiluminant colors. The new colors were chosen in-between each pair of the original eight test colors (Fig. 1) and yielded statistically identical results.

In experiment 2, instead of reporting the centroid of only one group of dots, subjects reported centroids of both the homogeneous group and heterogeneous group by making two mouse clicks in each trial (a dual-response task). The purpose of control experiment 2 was to show that, even though in the main experiment subjects reported only one centroid, they are in fact able to compute both the homogeneous and the heterogeneous centroid in a single briefly exposed dot display. Because order of report may be important, the order of report was switched between homogeneous centroid reported first and heterogeneous first in alternating blocks of trials.

In experiment 3, subjects also made two mouse clicks. One of them was directed to the centroid of the homogeneous dots, just like in experiment 2. However, the other one was directed to the location that would balance the centroid of the homogeneous dots, around the centroid of *All* dots. Everything else in the procedure was exactly same as in experiment 2.

In experiment 4, subjects viewed two consecutive stimuli that were separated by a 300-ms mask. One of the two stimuli was drawn from the same urn as in previous three experiments. That is, it contained 12 dots of identical colors and 2 dots of each of the remaining colors. The other stimulus was same except had one dot from each set removed. The showing sequence of the two stimulus frames was blocked. In separate blocks of trials, subjects clicked either the two locations of the two missing dots or the two extra dots in the second stimulus frame. The first stimulus frame was shown for 300 ms, followed by a 300-ms mask. The second stimulus frame was shown indefinitely until subjects made their responses. Complete feedback was given after each trial.

Apparatus. The experiment was conducted on an iMac Intel computer installed with Matlab 2012b and Psychtoolbox-2 software (29). Stimuli were displayed on a built-in 23-inch, 60-Hz refresh rate, LED monitor with a resolution of $1{,}920 \times 1{,}080$. The mean luminance of the monitor was 43.8 cd/m^2. Stimuli were viewed binocularly at ~72 cm.

Subjects. The first author (S2), three psychology undergraduate students (S1, S3, and S4), and a high school student (S5) participated in experiment 1a. The five subjects (three females) ranged in age from 17 to 33 y. All subjects except

S2 were naive to the purpose of the experiment. S1 and S3 had participated in experiments conducted in the laboratory before and thus were considered experienced subjects. S4 and S5 had never previously participated in any psychophysical experiments. S1, S2, and S5 participated in experiment 2. S2 and S5 also participated in experiments 3 and 4. All subjects reported having normal or corrected-to-normal vision. Methods were approved by the University of California, Irvine (UCI) IRB, and all subjects provided signed informed consent forms. The protocol and signed consent forms were approved by the UCI IRB.

Stimuli. A stimulus contained 26 dots displayed within a 512×512-pixel-wide, invisible square that spanned a visual angle of 12.1 degree of visual angle (dva). Dots were 9-pixel-wide squares, each spanning 0.21 dva. Each dot was painted in one of eight colors that was calibrated individually for each subject before the main experiment to be equiluminant to a grayscale background of 52.1 cd/m^2.

Of the 26 stimulus dots, 12 were the same color—the homogeneous group. The remaining 14 dots consisted of 2 dots of each of the other 7 equiluminant colors—the heterogeneous group. The eight colors were drawn from an equiluminant hue circle defined in standard cone space (30) with equal distance to the background gray. The first color was chosen to be the maximum saturation in the blue-cone direction. The other colors were chosen to be separated by equal 45° angles around a cone-defined hue circle.

Locations of homogeneous and heterogeneous dots were determined by a two-step process: First, a nominal centroid for each group was determined. The two nominal centroids were drawn from two independent bivariate Gaussian distributions with different means but the same x and y SDs of 3 dva. The means of the two distributions from which nominal centroids were chosen roved independently, both with SD of 0.70 dva. Then, dots were distributed around each nominal centroid. The x, y location of each dot was drawn independently from a Gaussian distribution with x and y SDs of 3 dva centered around its nominal mean. The expected locations of the centroids of the homogeneous dots and of the heterogeneous dots were completely independent of each other. The root-mean-square (rms) trial-to-trial variations of the actual stimulus centroids of the homogeneous, the heterogeneous, and all of the dots, were 1.41, 1.33, and 1.09 dva, respectively. The rms difference between homogeneous and heterogeneous centroids was 1.94 dva. Fig. 1 shows the stimulus colors and an example stimulus.

Procedure. Before starting experiment 1, subjects were trained in the centroid task with the same 300-ms exposure as in the main experiment. Stimuli were either homogeneous or heterogeneous sets of dots chosen from the experimental colors. The numbers of stimulus dots were 1, 2, 6, 12, and 18. The instruction in the training experiment was to click on the centroid of all dots, that is, no distractor dots. The subject had to produce centroid judgments accurate to within an average 20 pixels away from the true centroid position with stimuli of each number before progressing to the next higher number stimulus. Subjects started with experiment 1 after they completed the training session, which typically totaled 400–500 trials.

Fig. 2 shows an example of a typical trial for experiment 1. The trial started with a 1-s blank frame (foreperiod), followed by a 300-ms stimulus frame. After the offset of the stimulus frame, for three subjects, a masking frame containing a grid of randomly colored dots appeared for 300 ms. For three subjects, the masking frame was omitted. The subject used a mouse-controlled cursor to indicate the centroid of a subset, or of all of the dots, according to the attention instruction at the beginning of a block.

One of three different attention instructions was given at the beginning of a block of trials: (*i*) Locate the centroid of the most populous dots (homogeneous dots) or (*ii*) locate the centroid of all of the dots other than the most populous ones (heterogeneous dots) or (*iii*) locate the centroid of all dots. Blocks of the three conditions were counterbalanced. No additional attention instructions were provided during a block of trials as the attention condition was implicit in the feedback.

Within a block, there were 20 trials for each of the 8 possible homogeneous colors, totaling 160 trials that were randomly interleaved. Each block occurred in each of 3 attention conditions; this block was repeated 6 times in a Latin square counterbalanced order over subjects. In total, there were 18 blocks equal to 2,880 experimental trials for each subject. Note that the stimulus construction was same for all three attention conditions. Therefore, attention instructions and feedback were the only experimental factors that could contribute to different outcomes for different attention conditions.

The procedure for experiment 2, dual-response task, was same as for experiment 1, except that, in each trial, subjects now reported two centroids, one for each of the two groups, with two separate mouse clicks. A single feedback display showed the true centroids of both groups plus the subject's two clicked locations. The order of report of the two centroids was blocked. That is, in alternate blocks of trials, subjects reported the homogeneous centroid first and in the other blocks, the heterogeneous centroid first. Subjects completed 12 blocks of 160 trials for the dual-response task (6 homo-first blocks alternated with 6 hetero-first blocks). In a pilot analysis, we found, to our surprise, that subjects' performances in the dual-response task closely matched their performances in the single-response task in experiment 1. As subjects' ability to segregate the two groups might have greatly improved during the thousands of trials in experiment 1, subjects repeated the attend-to-homogeneous and attend-to-heterogeneous tasks (single-response tasks) following completion of the dual-response task. Only the second set of single-response date was used as the control for the dual task, thereby removing any possible advantage the dual-response task might have gained through practice.

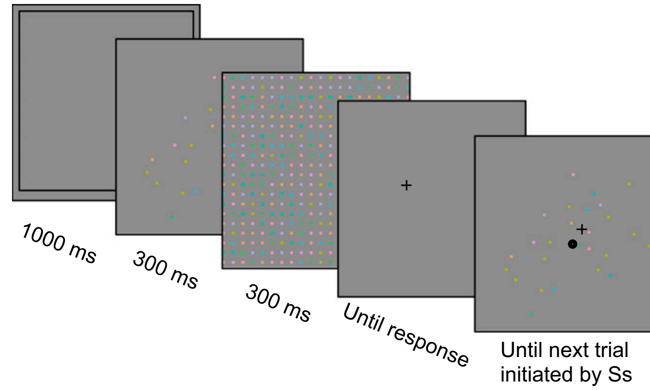

Fig. 2. The sequence of stimuli in a trial: blank field; stimulus; poststimulus mask; blank field with movable cursor (+ sign) for the subject to click on the perceived centroid; feedback showing the stimulus, the target centroid (black disk), and the subject's response (+).

Results

Mean Response Errors: Experiment 1. The most direct measurement of a subject's performance is the response error—the distance between the true centroid location of the target group and subject's mouse-click response. Fig. 3 shows subjects' mean response errors for the three attention conditions. The errors range from 0.4 to 0.8 dva (=5–10 mm) for all subjects. Subjects were most accurate when attending to all dots, they were less accurate when attending to the homogeneous group, and they made the largest errors when attending to the heterogeneous group.

Perceptual Color Filters. Response errors only reflect subjects' relative performances between the three attention conditions. They do not provide quantitative measurement on the degree to which the subjects have isolated the target group and excluded the distracting group. We compute an implicit "perceptual color filter" for each subject in each condition. The filter is created on-the-fly according to the current instruction about which subset to group and the colors of the dots that happen to occur in the immediate stimulus. The perceptual color filter reflects the relative weight of each color of dot in determining the subject's response, that is, the Selectivity (*Sel*) of the grouping process.

A given experimental condition ϕ is characterized by (*i*) the specific histogram of colors that occur in all of the stimuli used in that condition and (*ii*) the filter that the subject is instructed to achieve, that is, the "target filter" for the condition. For any one of our eight colors C, there are three conditions that use displays comprising 12 dots of color C and 2 dots of each of the other seven colors. In the Homogeneous (*Hom*) condition, the target filter assigns weight 1 to C and weight 0 to the other seven colors; in the Heterogeneous (*Het*) condition, the target filter assigns weight 0 to C and weight 1/7 to the other seven colors. In the *All*

PSYCHOLOGICAL AND COGNITIVE SCIENCES

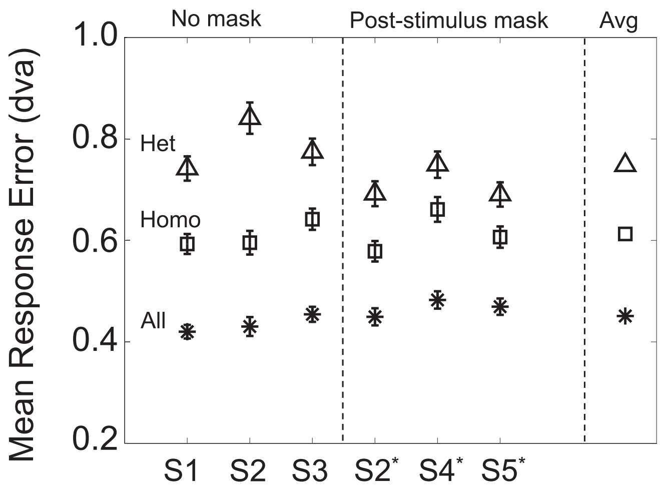

Fig. 3. Mean response errors (distance from the subjects' responses to the target centroid in degrees of visual angle) for five subjects in three attention conditions, and for trials with and without poststimulus masks. Symbols represent the attention condition: squares, homogeneous centroid; triangles, heterogeneous centroid; asterisk, centroid of all. Error bars represent 95% confidence intervals. "Avg" is the average of the mean response errors across six subjects for the three attention conditions. The asterisk (*) after a subject identifier indicates poststimulus mask conditions.

(attend equally to all dots) condition, the target filter assigns weight 1/8 to all eight colors.

Definitions. Let x_n, y_n be the location of the nth dot in a dot cloud, r_x, r_y be the location of the subject's response, and C_n be the color of the nth dot. (The bold-type symbols represent vectors with the trial number as the implicit vector dimension.) We assume that the weight exerted by dot n on the subject's response in condition ϕ is $f_\phi(C_n)$.

Perceptual filter model. To compute f_ϕ for a particular subject, in condition ϕ, with a particular set of color stimuli that has I different colors, we follow the analysis of Sun et al. (28). For a stimulus of N dots, the filter model's centroid response (r_x, r_y) is as follows:

$$r_x = \sum_{n=1}^{N} f_\phi(C_n)x_n + Q_x \quad \text{and} \quad r_y = \sum_{n=1}^{N} f_\phi(C_n)y_n + Q_y. \qquad [1]$$

That is, r_x, r_y is the centroid of all dots weighted by their color weights $f_\phi(C_n)$ plus Q_x and Q_y, normally distributed random variables reflecting residual error. It is more convenient to express the color weights in terms of the color ID i instead of the dot number n, that is, $f_\phi(C^i)$ is the filter weight of all dots that have color C^i. For convenience, we let the sum of a perceptual filter's color weights be 1, $\sum_{i=1}^{I} f_\phi(C^i) = 1$. Eq. **1** implies a simple linear model in which the color-filter weights $f_\varphi(C^i)$ of the best-fitting model (i.e., the model that minimizes Q_x, Q_y) are obtained by standard multivariate linear regression (e.g., ref. 28). [Note: Whereas the perceptual filter function $f_\phi(C)$ describes the weight exerted on the subject's responses by dots of different colors C, the total amount of information that passes through the filter is captured by another statistic, Efficiency (*Eff*), to be described later.]

Perceptual color filters. To illustrate the selectivity analysis, consider subject S1's perceptual color filters f_ϕ for the three attention conditions in which "green" is the homogeneous color (Fig. 4 *A*–*C*). The dashed lines indicate the target filter, that is, weights an ideal observer would achieve with the target filter. In Fig. 4*A*, the large filter weight associated with the green compared with the small filter weights of the other seven colors makes the perceptual green-selective color filter f_ϕ appear like a physical color filter through which one color (the green color of the homogeneous dots) is transmitted and the other seven colors are severely attenuated. Note that homogeneous colors were randomly interleaved within a block; no color cues were ever given to prepare subjects for tuning their "color filters" before a stimulus was shown. It is quite remarkable that, based simply on color homogeneity or heterogeneity, which becomes apparent only during a brief 300-ms exposure, a subject spontaneously achieves a perceptual color filter that either closely approximates the target filter that selectively transmits only that particular trial's homogeneous color (Fig. 4*A*) or selectively blocks the homogeneous color (Fig. 4*B*) and transmits all of the others (Fig. 4*B*). Fig. 4*C* also shows that, when attempting to give equal weight to all dots, S1 achieves a perceptual color filter that indeed gives approximately equal weight to all eight dot colors.

S1's three different task-dependent perceptual color filters for each of the eight different color stimuli (3 × 8 = 24 filters) are shown in Fig. 4 *D*–*F*. Fig. 4 *D* and *E* demonstrates that the general pattern of sharply tuned f_ϕ is very similar for each of the eight stimulus colors. Perceptual color filter weights are sharply peaked or sharply dented according to whether the homogeneous group or heterogeneous group is attended. Under equal-attention instructions (the *All* conditions), each dot was to be weighted equally independent of its color, and the perceptual filter weights indicate that subject S1 was able to achieve a close approximation to equal weighting.

The similarity of the filter shapes for the eight different kinds of stimuli (the eight different homogeneous colors) justifies computing the average filters for each of the three attention conditions. The averaging computation implicitly assumes that the colors are equally spaced; the shape similarity of the filters suggests that this is a reasonable approximation. The average perceptual color filters were quite similar across subjects, so these too are averaged. Because filtering is a multiplicative operation, to better enable comparisons, the three perceptual color filters, averaged over subjects and colors, are shown on logarithmic coordinates (Fig. 4*G*). Data for individual subjects and colors are available in *SI Appendix*.

Quantifying Perceptual Grouping Performance.

Perceptual grouping. To perform the task, subjects have to segregate the target group from the distractor group and compute a summary statistic (the centroid in this case) of only the target group (*Discussion, Summary, and Conclusions*). Good performance requires that most distractors be excluded from, and most targets be included in, the centroid computation. We introduce two measures to quantitatively assess two attributes of the performance, selectivity and efficiency.

Selectivity. *Sel* of a perceptual filter is defined here as the mean filter weights of all of the target color dots divided by the mean filter weights of all of the distractor color dots. *Sel* expresses a filter's preference for targets over distractors for a particular attention instruction and for a particular class of stimuli. *Sel* is computed from the perceptual color filter f_ϕ as follows. Let $[C_\phi^{tar}]$ and $[C_\phi^{dis}]$ be the sets of target and distractor colors in a condition ϕ. Define the total mean weights of all of the targets and of all of the distractors as follows:

$$F_{tar} = \frac{\sum_{C\in[C_\phi^{tar}]} M(C) f_\phi(C)}{\sum_{C\in[C_\phi^{tar}]} M(C)},$$
$$F_{dis} = \frac{\sum_{C\in[C_\phi^{dis}]} M(C) f_\phi(C)}{\sum_{C\in[C_\phi^{dis}]} M(C)}, \qquad [2]$$

where $M(C)$ is the number dots in the displays in condition ϕ that have color C. (F_{dis} includes all of the colors that were not in F_{tar}.)

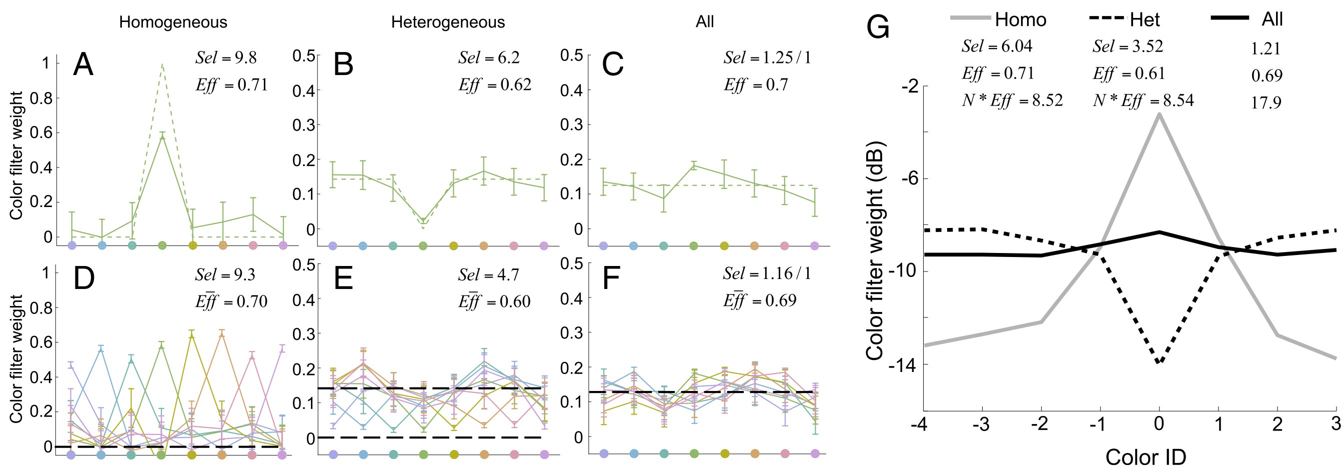

Fig. 4. Perceptual filter weights for three attention conditions and eight stimulus types for a naive subject, S1, and for the subject-population average. (*A*–*C*) S1's perceptual color filter weights as a function of the eight dot colors indicated on the abscissa when viewing stimuli in which the homogeneous color was green. The three panels correspond to the three different attention conditions. The dashed lines indicate the target filter, that is, what an ideal observer, performing perfectly according to each instruction, would have produced. (*D*–*F*) S1's eight perceptual color filter weights plotted simultaneously in each panel, as in *A*–*C*, for each of the eight homogeneous colors. Error bars throughout Fig. 4 represent 95% confidence intervals. The color of each solid line indicates the homogeneous color. The lower black dashed lines in *D* and *E* indicate a weight of zero, the weight that the target filter assigns to distractors. The upper dashed lines in *E* and *F* indicate the target filter's weight for targets. *Sel* (Selectivity) is the summed perceptual filter weight of target colors divided by the summed weight of distractor colors. *Eff* (Efficiency) is the proportion of, and *N*Eff* is the number of, target dots that an ideal observer requires to match the subject's accuracy. Together *Sel*, *Eff*, and *N*Eff* quantify how well the subject has grouped the target dots. See text for details. (*G*) Perceptual color filter weights of the three attention conditions averaged across all eight colors and six subjects. Abscissa value 0 represents the color of the homogeneous dots. Adjacent numbers represent the relative clockwise locations of adjacent colors around the color circle (Fig. 1*A*). The ordinate in *E* is 10*log10 of the average of eight color-filter weights, $\log_{10}(\sum(f_\phi(C)/8))$. At each abscissa value, the stimuli are from the same eight urns; the only difference between the three curves is the three different attention instructions and the feedback.

Sel is the weight ratio: $Sel = F_{tar}/F_{dis}$. *Sel* indicates the degree to which dots of target group vs. dots of distractor group influence a centroid judgment. For example, the selectivities of S1's perceptual color filter when green is the homogeneous color are 9.8 and 6.2 for the attend-to-homogeneous and the attend-to-heterogeneous instructions, respectively (Fig. 4). For S1, averaged across all eight color conditions, the mean selectivities are 9.3 for the attend-to-homogeneous, and 4.7 for the attend-to-heterogeneous instructions. This means that for subject S1, attending to the homogeneous (or the heterogeneous) group, gives the target group colors an advantage of 9.3:1 (or 4.7:1) over the distractor group colors in determining the judged centroid. Averaged for all colors and subjects (Fig. 4*G*), the perceptual filter selectivities for the homogeneous and heterogeneous groups are 6.04 and 3.52, respectively.

When asked to give equal weight to all colors, subjects also succeed remarkably in doing this (Fig. 4 *C*, *F*, and *G*), resulting in the average *Sel* across subjects and colors equal to 1.16. In attend-to-all condition, there are no segregated target and distractor groups, so we arbitrarily set $Sel = F_{Hom}/F_{Het}$, where

$$F_{Hom} = f_\phi(C_{Hom}) \quad \text{and} \quad F_{Het} = \frac{\sum_{C \neq C_{Hom}} M(C) f_\phi(C)}{\sum_{C \neq C_{Hom}} M(C)}, \qquad [3]$$

where F_{hom} is the average weight of the 12 homogeneous dots and F_{het} is the average weight of the 2 × 7 heterogeneous dots. Neither the slight bump in the equal-attention perceptual filter for the homogeneous color in Fig. 4*G* nor any of the differences from the null hypothesis—equal filter weights for all colors—reach statistical significance: Under equal attention instructions, subjects' color weights are indifferent to the group to which colors belong. Fig. 5*A* shows Sel for different subjects under different instructions.

Whereas *Sel* is useful in assessing how well the distractor group is excluded relative to the target group, it does not measure how much of the target group actually is utilized. To see this, suppose the subject always clicks on one dot from the target group and ignores all of the other dots. This high-miss rate strategy would yield a *Sel* of infinity, because all distractor dots are completely excluded and therefore the denominator of *Sel* is zero. However, this strategy does not involve perceptual grouping—the aim of this study.

Efficiency. To determine whether a high value of *Sel* was achieved at the cost of a high miss rate, we introduce an efficiency measure, *Eff*. *Eff* is an estimate of the fraction of dots in the target subset that are utilized in the subject's centroid judgment according to an all-or-none model in which a dot and its x, y location are either perfectly remembered or the dot is entirely forgotten. *Eff* was estimated via a Monte Carlo simulation. Specifically, to estimate *Eff* in a given condition ϕ, a proportion p of all of the dots was randomly deleted from the stimulus, and the target filter for condition ϕ was applied to the remaining dot cloud. (Using the estimated perceptual filter of the observer would seem more natural in computing efficiency. However, using the target filter gives a conservative estimate of *Eff*, which has an expectation that is lower than *Eff* computed with the perceptual filter unless the two filters are equal.) Response error of an ideal detector supplied with these depleted dot-clouds increases monotonically with p; p_{miss} is determined by interpolation; *Eff* is $1 - p_{\text{miss}}$.

Attributing response error entirely to missing dots is a simplification. Other factors such as motor error in placing the cursor undoubtedly contribute to response error. However, *Eff* places an absolute lower bound on the percentage of dots that must be utilized by an ideal detector to match the subject's performance. If the ideal observer receives dots that are already perturbed by other errors, such as an imperfect perceptual filter or motor response error, then *Eff* would be even larger to match a subject's performance. For S1, when green is the homogeneous color (Fig. 4), *Eff* = 0.71, 0.62, 0.70 for attending to the homogeneous set, the heterogeneous set, and all of the dots, respectively. Averaged across all eight color conditions, S1's mean

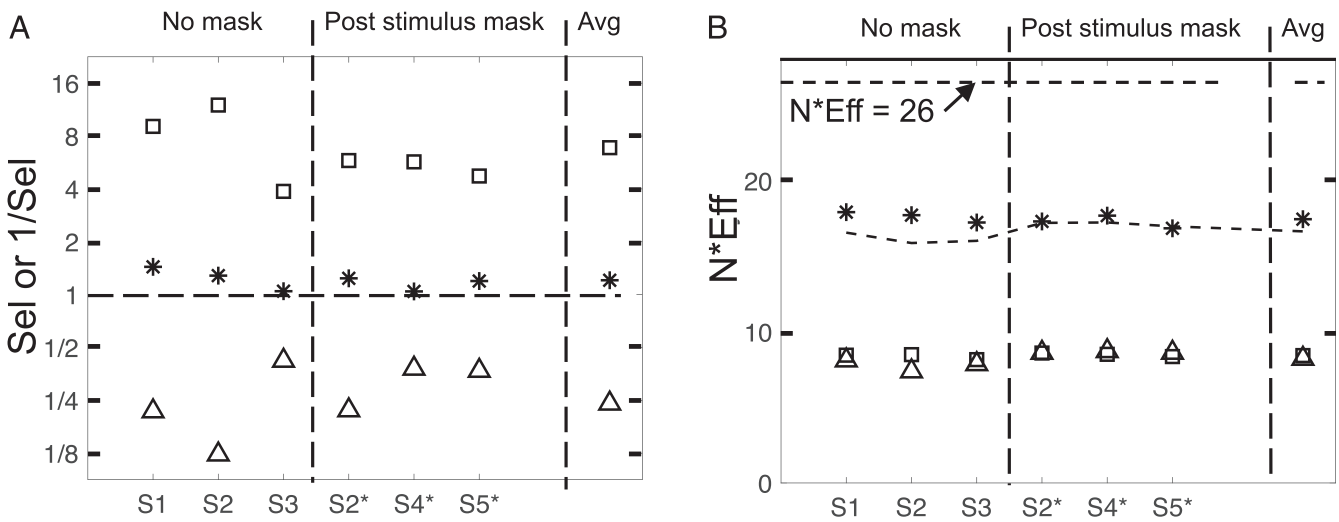

Fig. 5. Selectivity and efficiency of judged homogeneous, heterogeneous, and all centroids. (*A*) (Average homogeneous filter weight)/(Average heterogeneous filter weight) = Selectivity (*Sel*) for homogeneous and 1/*Sel* for heterogeneous centroid judgments. Data points are based on the average filter weights over the eight homogeneous-color stimuli shown individually for the six subjects. Selectivities are so graphed to illustrate the enormous response difference to the same stimuli under different attention instructions. The ordinate is a log scale. Symbols: □ , attend homogeneous; Δ, attend heterogeneous; *, attend all. (*B*) *N*Eff*, the number of stimulus dots an ideal detector requires to match centroid judgments shown individually for six subjects, for three attention conditions, each averaged over the eight homogeneous stimulus colors. *N* is the number of target dots, *Eff* is the estimated efficiency, that is, the lower bound on the proportion of dots utilized in the response. The dotted curve is $N{*}Eff_{hom} + N{*}Eff_{het}$; Avg is the average across all subjects and mask conditions for the three attention instructions. The asterisk (*) after a subject identifier indicates a poststimulus mask condition.

Eff values are 0.70, 0.61, and 0.69 for the three attention instructions. The estimated efficiency measures indicate that an ideal observer requires at least 70%, 61%, and 69% of the 12, 14, and 26 target dots to match the average of the six subjects' performances.

Let *N* be the number of target dots. Then *N*Eff* is the lower bound on number of dots the subject must have utilized. Fig. 5*B* shows *N*Eff* averaged across all color conditions for each subject. According to this all-or-none model, subjects utilized at least eight target dots (or 60~70% of all target dots) when grouping the homogeneous or the heterogeneous dots, and at least 18 dots (or ~70% of all 26 dots) when attending to all dots.

Fig. 5*B* reveals interesting properties of the data. (*i*) *Eff* is slightly lower for the heterogeneous groups than the homogeneous groups. However, when the different number of dots (14, 12) in the two groups is taken into account, the same number *N*Eff* of dots (within measurement error) is processed in each group (*Hom* = 8.52, *Het* = 8.54). (*ii*) When subjects attend to all dots, the number of dots utilized in their response *N*Eff* is 17.9, which is slightly but not significantly greater than the sum of the number of dots utilized in each group (17.1) when that group was the only group tested. (*iii*) *N*Eff* = 17.9 means that to match the accuracy of a subject's centroid-of-all response, an ideal detector needs at least 17 of 26 dots with dot location known exactly—even more if there were dot position errors, an imperfect perceptual color filter, an imperfect centroid computation, or motor response error. Utilizing 17 dots indicates an extraordinary efficiency for a preconscious centroid computation. (*iv*) In judging the centroid of *All*, if a subject formed two subgroups and then combined the information, we would expect the error variance of *All* centroids to be equal to the sum of the individual group error variance, certainly greater than either. In fact, the error variance for *All* centroids is smaller than either *Hom* or *Het* centroid error variances. This implies that the processing resources for *All* are grouped before any subdivision of these resources. We return in *Discussion, Summary, and Conclusions* to consider the significance of these results; here, we consider several related issues.

Control experiment 1.1: Is there something special about the particular set of eight colors used in experiment 1? To explore this issue, the full experiment was repeated with a new set of eight different colors chosen to lie midway between the original ones (Fig. 1). For the first two subjects tested, as expected, mean response errors for the two different color sets were statistically indiscriminable, as were *Sel* and *Eff* for the two color sets. We decided it was not necessary to further pursue this side issue.

Control experiment 1.2: What is the effect of poststimulus masks? The purpose of the mask was to prevent the continuing acquisition of visual information after stimulus termination from a persisting sensory image of the display, that is, to ensure that the acquisition of stimulus information was confined to the premask display duration. Logically, the postexposure mask would be expected to at least slightly impair performance. However, the aim was not to measure the degree to which a poststimulus mask might impair performance but rather to demonstrate that the 300-ms exposure duration was sufficiently long that there are no important parameter differences between masked and unmasked performances. The similarities of the left and right sides of Figs. 3 and 5 *A* and *B* (no mask vs. mask) make it unlikely that presence or absence of poststimulus masks will affect any conclusions concerning the grouping phenomena under investigation here.

Control experiment 1.3: Double-pass procedure. To estimate the internal noise that limits subjects' performance in the centroid judgment task, several weeks after the initial sessions, we reran the first 320 trials of a previous experimental session using the identical stimuli, the same poststimulus mask, and the same sequence of trials, including all three attention tasks: homogeneous, heterogeneous, and equal. S2, the most experienced, and S5, the least experienced, participated. The principle of the double-pass procedure is that, insofar as the centroid judgments differ between the two runs, this is due to internal noise within the subject; everything that can be controlled externally has been kept unchanged.

Results. S2's estimated internal noise variance is 11.5% of the total response variance. The perceptual color filter model (i.e., the model of Eq. **1**) accounts for 90.4% of the residual response variance, that is, it accounts for 90.4% of the predictable component of the data. S5's internal noise variance is 21% of the response variance. The perceptual color filter model accounts for 82.2% of the residual response variance. These results confirm that the perceptual color filter model indeed provides an excellent description of performance.

Experiment 2: Dual Centroid Responses. Can subjects make two different centroid judgments—both the homogeneous dot and the heterogeneous dot centroids—based on the same 300-ms stimulus exposure? This is an extremely interesting issue. If forming the homogeneous group and the heterogeneous group and recording these representations within 300 ms share the same resources and compete for them, one would expect a substantial decline in performances in the dual-response experiment (31).

To explore concurrent centroid judgments, subjects viewed masked (or unmasked) stimuli and performed both single and dual centroid judgments. For procedural details, see *Methods*. We first compare performances between single and dual centroid judgments on a completely model-free dimension—the centroid error—the mean distance of subject's judged centroids from the true centroid. The only statistically significant main effect was that dual-response error (first or second response) was 18% greater than single-response error.

The most interesting results of the dual-response task are the attention operating characteristics (AOCs) in Fig. 6. For AOCs, it is more appropriate to use the weights assigned to each set directly (i.e., F_{Hom} and F_{Het} in Eq. **3**) than to use the *Sel* values. Weights are constrained to be between 0 and 1, whereas Sel values are ratios and therefore unconstrained. The filter weights (averaged over eight color conditions) derived from the dual responses are, within measurement error, the same as the weights derived from the single-response control data. The first and second dual-response *N*Eff* values also do not differ significantly in accuracy. The average (over first and second response, over subjects, and over postmask conditions) dual-response *N*Eff* is 7.5 dots compared with the control single-response *N*Eff* of 8.6 dots, a dual-response loss of ∼1 item. This 13% decline in efficiency (*N*Eff*) but not in selectivity of dual responses accounts for the 18% greater error of dual-response centroids vs. single-response centroids. Overall, however, the dual-response data are remarkably similar to the single-response data, thereby demonstrating that subjects can accurately report two centroids from a single brief exposure.

Can subjects consciously compute a heterogeneous centroid by subtracting the homogeneous centroid from the centroid of all, i.e., by implicit vector subtraction? We know subjects can simultaneously form multiple homogeneously colored groups of dots and estimate their numerosities (32) or their centroids (33). Here we ask: Can subjects indirectly find the centroid of the heterogeneous dots by reflecting the centroid location of the homogeneous dots 180° around the centroid location of all of the dots, never directly forming a heterogeneous group nor computing its centroid. We refer to this procedure as "vector subtraction." We analyzed the residual variance of each subject's centroid judgments for the homogeneous V_{hom}, the heterogeneous V_{het}, and all of the dots V_{all} in experiment 1. Vector subtraction implies that the variance of the resulting vector, that is, the variance of the indirectly computed heterogeneous vector equals the sum of the variances of the two vector components, plus the variance V_{vs} due to an imperfect vector subtraction, $V_{het\_indirect} = V_{hom} + V_{all} + V_{vs}$. When the residual variance of a subject's heterogeneous centroid judgment is less than the sum, $V_{het} < V_{hom} + V_{all}$, it implies a different process than vector subtraction. The data show that the residual variance of heterogeneous centroid judgment is substantially smaller than the vector sum variance $V_{het} \ll V_{hom} + V_{all}$ for four of six subjects and $V_{het} \approx V_{hom} + V_{all}$ for the other two subjects. (We estimate motor and other non–centroid-related noise components as the variance of locating a single dot measured during the training session. To avoid counting this motor noise twice in the heterogeneous vector subtractions, we factor it out of all the $V_{het\_indirect}$ variances.) Based on the error analysis, we conclude that the heterogeneous

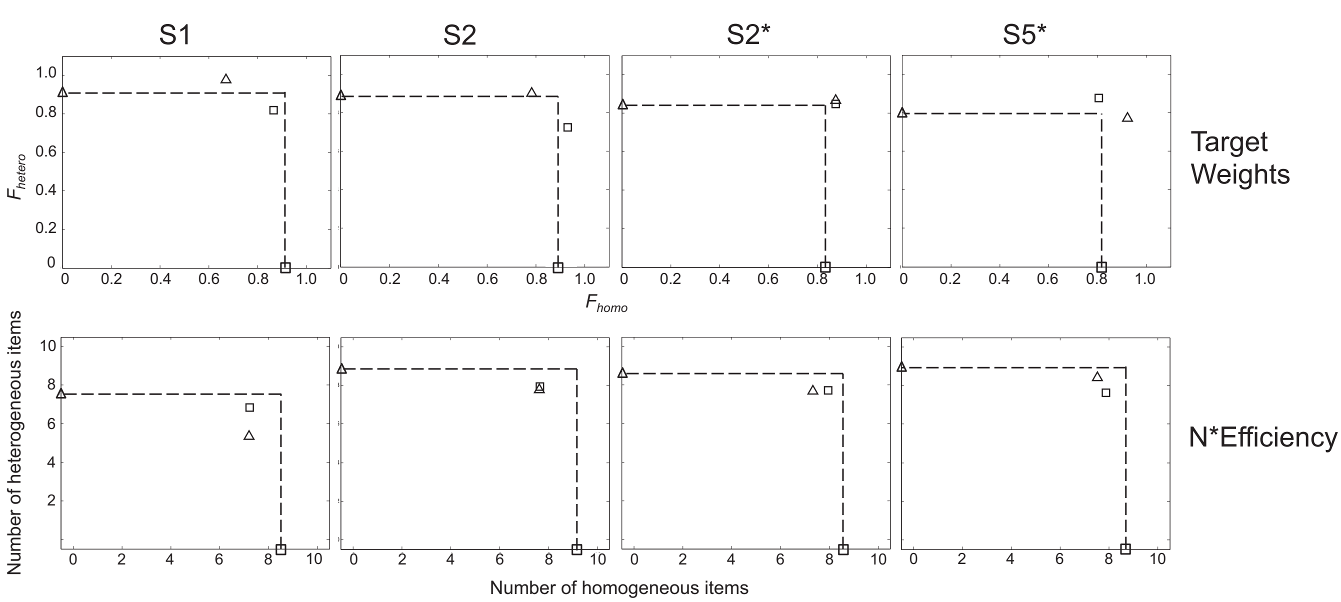

Fig. 6. Attention operating characteristic (AOC) graphs of filter weights (a measure of the judged selectivity for the reported group) and *N*Eff* (lower bound on number of target dots required by an ideal detector to match the subject's performance) for two subjects without poststimulus masks and two subjects with poststimulus in the dual centroid judgment task. Performances for the single-response (control) tasks are indicated on the axes and by the dashed lines. The intersection of these lines represents the independence point where each centroid in the dual-response task is reported as accurately in terms of selectivity (target vs. distractor weights) and number of items processed (*N*Eff*) as the corresponding centroid in the single-response task. The area inside the independence point represents dual-task deficits. Data points represent the average first and average second response made on each trial. Symbols:□ , first response is the homogeneous centroid (*x*); Δ, first response is heterogeneous centroid (*y*); the asterisk (*) after the subject identifier indicates poststimulus mask conditions.

centroid in experiment 1 was not the result of vector subtraction. We will also test this empirically in experiment 3.

Experiment 3: Heterogeneous Centroids by Explicit Vector Subtraction. In experiment 3, subjects were instructed to make two mouse clicks. The first aimed for the centroid location of the homogeneous dots. The second click aimed for the mirror-opposite location that would balance the homogeneous centroid about the centroid location of all of the dots, that is, the operation of vector subtraction. If the estimation of the heterogeneous centroid in experiment 2 were indeed accomplished through vector subtraction, then we would expect performance in experiment 3 to be comparable to performance in experiment 2.

Procedure. S2 and S5 participated in this experiment. We explained to the subjects that the second click was to be aimed at the point that would balance the homogeneous centroid location around the centroid location of all dots. Everything else in the procedure was exactly same as in experiment 2, including complete feedback after each trial, except that subjects completed only two instead of six blocks of 160 trials. Experiment 3 was shortened because subjects found this task extremely exhausting and had to interrupt many times within a block to rest. However, the two blocks produced clear results.

Results: Performances in experiment 3 vs. experiment 2. We compare the two blocks of experiment 3 to the first two (least accurate) blocks in experiment 2. Even so, for S2 and S5, mean response errors were 40% and 15% greater in experiment 3 (vector subtraction heterogeneous centroids) than in experiment 2 (directly estimated heterogeneous centroids). The difference in RT (reaction time between termination of the stimulus display and the time when the second mouse click is recorded) is dramatic. Mean RTs for S2 and S5 in the vector subtraction experiment are 3.42 and 2.84 s. In experiment 2, the mean RTs for the same two subjects were 2.23 and 2.05 s for the first two blocks and diminished to 1.56 and 1.34 s for the last two blocks. All of the RT differences between vector subtraction in experiment 3 and direct estimation of the heterogeneous centroid after estimating the homogeneous centroid in experiment 2 were highly significant ($P < 0.01$). In summary, there are four major differences between vector subtraction and direct estimation of the heterogeneous centroid: Vector-subtraction centroid judgments (*i*) have greater error, (*ii*) have greater variance, (*iii*) have hugely greater reaction times, and (*iv*) require enormously more subjective effort than directly estimated heterogeneous centroids. We conclude that directly estimated heterogeneous centroids do not rely on explicit vector subtraction.

Experiment 4: Conscious Memory for Dots and Their Locations. When judging the centroid of a very briefly flashed display that has 26 dots, the subjective feeling is not of making a complex judgment involving many dots but of carefully attending to a feeling that the centroid is "here." An alternative view might be that, although the exposure is brief, many dots and their positions are remembered and that the centroid computation is based on this persisting visual memory of the stimulus. This would be postexposure visual processing. Consider an extreme example of poststimulus processing: Ask a subject whether a brief exposure of a display of three numerals, for example, 323, represents a prime number. If the three numerals are perceived correctly, the computations to arrive at a correct answer to the prime number question might take several minutes. Is the centroid judgment based on an analogous postexposure computation based on a remembered image of the stimulus vs. a rapid implicit process?

Procedure. It is not easy to determine how much information is retained in memory from a briefly exposed dot array. Among the various less-than-ideal procedures available, we chose a change detection paradigm. The same type of 26-dot stimulus was presented with a postexposure mask exactly as in all of the experiments reported so far. However, instead of showing a blank screen for a centroid judgment, a copy of the stimulus was presented in which one randomly chosen dot was removed from among the homogeneous dots and one dot from the heterogeneous dots. The task of the subject was to click on the locations of both missing dots. The missing-dot stimulus remained visible until the subject responded, that is, no mask was presented after the missing-dot stimulus. As in the previous experiments, following the response, there was complete feedback. There is a complementary experiment in which the stimulus sequence is reversed: The second dot array has two extra dots. The subject's task was to click on the location of the dots that were not present in the first array. S2 and S5 served in these experiments. These procedures were intended to be analogous to experiment 2 in which subjects estimated two centroids.

Scoring procedure. The stimulus field was divided up into 26 areas according to which was the nearest dot. If the subject's response fell within the area in which the nearest dot was the target missing dot, the response was considered correct; otherwise, it was incorrect. There was a standard correction for chance guessing. [There were 12 homogeneously colored dots and 14 heterogeneously colored dots in the stimulus. Therefore, the probability of chance guessing, g, for homogeneous responses was 1/12, and for heterogeneous responses was 1/14. Let p_o be the observed probability correct, and let p_t be the true probability correct. Then $p_t = (p_o - g)/(1 - g)$.]

Results. The guessing-corrected probability p_t of correctly detecting the changed dot location in the change paradigm trials was shockingly poor. (*i*) The second-response data were so close to chance performance that they are not further considered here. (*ii*) The extradot procedure was much more difficult than the missing-dot procedure (in which performance was close to chance) that it not further considered here. The estimation of the heterogeneous missing dot was much worse than the estimation of the homogeneous missing dot. (*iii*) The best performance, the only one considered here, is trials in which the first response click was on the location of the missing homogeneous dot. The estimated numbers of homogeneous stimulus dots available for making the change detection response p_t*12 for the two subjects were 1.7 and 0.17. This compares with *N*Eff* (the estimated minimum number of dots with their exact location required by the ideal observer to match the same subjects' centroid judgments) that is greater than 7 of the 12 homogeneous dots plus 7 of the 14 heterogeneous dots in the dual judgment task (and >8 in single centroid judgments). The extremely efficient preconscious centroid judgment utilizes enormously more information about the dots and their location than is consciously available to subjects as indicated by their performance in change detection.

Discussion, Summary, and Conclusions

To compute a particular scene statistic such as a centroid or to identify a particular part of the visual field as an object such as a tree or a person, it is necessary to divide the scene into those elements that are included in the centroid or in the object identification computation and those parts that are excluded. This process is called grouping, also called segmentation or figure-ground. Here, from subjects' judgments of the centroid locations of interleaved groups of dots, we characterize their grouping performances by a filter that passes some scene components (targets) and rejects others (distracters). The filter is a detailed characterization of what defines membership in the group upon which a centroid is computed. We characterize the quality of filter selectivity with a single statistic, *Sel*. A second statistic *N*Eff* is the lower bound on the number of dots that must have been included in a subject's centroid computation to have achieved the judged centroid accuracy. *N*Eff* measures the combined efficiency of the grouping and the centroid computation. We summarize our findings and discuss their implications below.

Grouping Processes and the Centroid Computation. Among the Gestalt psychologists, grouping referred not only to a selection process but also to the dependent variable—the computation made on the selected item—the subject's impression of "belonging together." Because the impression of subjective belongingness was difficult to quantify, subsequent experiments on grouping used other dependent measures, typically ambiguous arrays. For example, the perceived orientation of lines in an array of dots might be horizontal if "grouping" was according to similar color or vertical if grouping depended more on the spacing between dots or some other property.

The extent to which the centroids measured in the current experiments measure the same processes as other grouping processes is not answered here. However, we can say (*i*) the centroid data are incredibly richer and offer enormously more detailed information about both the selectivity and efficiency of the centroid grouping process than the methods of prior grouping experiments. (*ii*) That subjects can group by both physical similarity and by dissimilarity voids similarity as an explanatory principle: (*iii*) Similarity itself within limits yet to be determined, becomes whatever the experimenter defines as similarity. Physical similarity is only one aspect of a vastly more complex concept. (*iv*) A remarkable amount of information is preconsciously processed to form a group and to compute its centroid. The great efficiency of intertwined process of group formation and of centroid computation was totally unexpected given the prior literature on grouping.

Combining Resources to Group by *All*. Subjects can form three different groups of dots: a subgroup of similar dots without any prior knowledge of what the color of those dots will be, a group of dissimilar dots, or a group of all of the dots. We find that $N{*}Eff(All) \sim N{*}Eff(Hom) + N{*}Eff(Het)$, that is, the number of utilized dots in the homogeneous centroid judgment plus the number of dots used in the heterogeneous centroid judgment is approximately equal to the number utilized when judging the centroid of *All*. Error analysis showed that when judging the centroid of *All*, subjects are deliberately indifferent to dot properties vs. implicitly forming groups of *Hom* and *Het* and combining the information. Rather, when only one group is judged, there is a prior implicit combination of the resources that, under other instructions, are separately available to each of two grouping processes.

Obligatory Formation of Two Groups in the Same Brief Exposure. When subjects make two centroid responses on the same trial, the total number of stimulus dots utilized nearly equals the number utilized when judging the centroid of *All*. Apparently, the capacity of the perceptual grouping process can be flexibly allocated to either one or to two separate subgroups. However, when judging the centroid of only the homogeneous or only the heterogeneous items, subjects have no more resources available than when they judge both on the same trial. This implies that to group a target set, for example, homogeneous items, it is obligatory to also group distractor items, and vice versa.

Heterogeneous Dots: Selected or Merely Remainders? (*i*) Early selection: Is a heterogeneous group formed by positively selecting heterogeneous elements [e.g., by tuning the visual system for high-variance information (34)]? Or (*ii*) late selection: Is a heterogeneous group formed as a remainder by selecting all elements and, after computing the centroid of the homogeneous elements, removing them [e.g., "visual marking" (35)], computing the centroid of the remaining heterogeneous elements? Obviously, homogeneous dots are removed from the heterogeneous centroid computation; that is what the heterogeneous filters in Fig. 4 demonstrate. The question is whether the heterogeneous group is formed and its centroid is computed analogously and in parallel with the homogeneous centroid—as has been implicitly assumed—or whether the homogeneous dot group is formed first and then reused to subtract from a grouping of ALL in the heterogeneous centroid computation. Although this complex issue cannot be definitively decided on the basis of available data, there are very good reasons in favor of positive early heterogeneous dot selection (*SI Appendix*). We repeat the conclusion here. Because a remainder process is unlikely in a salience process of centroid computation, because the postmask truncated exposure time is too brief for successive homogeneity followed by heterogeneity centroid computations, because heterogeneous judgments can easily be made to be the primary foreground judgments in situations that differ only slightly from the current experiments, and because complex top–down reweighting selection mechanisms act similarly to simpler bottom–up mechanisms, there is no need to invent an unlikely special mechanism (high-speed marking and elimination) to account for the subjects' abilities to make heterogeneous centroid judgments in the current displays.

Extreme Amount of Information in Preconscious Centroid Judgments. Subject accuracy in *All* centroid judgments is such that an ideal detector working with the perfect *All* target filter requires knowledge of the exact location of 17.9 of 26 dots to match subject accuracy. Had the ideal detector used the subjects' imperfect perceptual filters instead of the perfect target filter, or if dot-location representation were imperfect, or if there were any motor response error, the ideal detector would require even more dots.

In trials where subjects make two centroid judgments following the same 300-ms stimulus exposure (experiment 2), an ideal detector needs 8 + 7 = 15 of 26 dots to match subject accuracy. The process of extracting the information from at least 17 (or 15 in two response tasks) of the 26 scattered dots is extraordinary; an enormous amount of information has to be distilled to account for the accuracy of centroid judgments.

Conscious Memory of These Low-Contrast Color Dot Displays Is Astoundingly Poor. The change detection data show that subjects cannot consciously remember the approximate location of even two dots in these displays under precisely the same conditions in which the unconscious grouping process utilizes at least 15 dots. However, in a change detection experiments with high contrast black dots on a white background (no distractor dots), we find that these same subjects can remember about 4∼5 dot locations (*SI Appendix*). This indicates that the unconscious grouping process can operate at the much lower contrast levels of the color displays, whereas the less-efficient conscious visual memory computation requires distinctive, high-contrast dots. By analogy, it is reasonable to compute the centroid of all of the leaves on a bush. However, there is no reason to form a conscious memory of individual leaves unless they are distinct in some way. Our 26 low-contrast dots did not pass our subjects' memory-distinctiveness threshold.

Exclusive Feature Maps for Grouping Are Not Feasible. "Feature map" refers to an array of detectors for a particular visual feature that tile a large area of visual space and the contents of which are available without contamination from other maps for inspection by higher-level visual processes. This idea was formulated in the context of pop-out visual search (2); searching for a red dot embedded in a sea of green dots is easy, because the red dot marks the only entry in a red-feature map. Searching one map is easy. Searching for a red round dot in sea of red square dots and green round and square dots is more difficult. Searching two maps (red and round) and combining the information is much slower and more difficult.

Whereas feature maps are undoubtedly a critical component at the earliest stages of visual processing, at the level of complexity of separating target colors from distractor colors, predefined feature maps are no longer useful. In experiment 1, each of the eight colors served as the homogeneous color on 1/8 of the trials. That requires eight unique color maps. However, the ability of subjects to compute the centroid of the heterogeneous features, requires that seven feature maps, each activated by two dots, be combined to enable the heterogeneous centroid computation. Those eight different combinations (depending on the homogeneous color) of seven distractor features are computed with almost the same *N*Eff* (minimum number of dots utilized by the centroid computation) as the single-feature map computation for the homogeneous feature. There was very little cost for combining seven maps.

Control experiment 1a used a different set of eight colors. To deal with experiment 1a, unique-feature theory, requires eight more unique feature maps, making total of 16 feature maps. Additionally, it is quite obvious that, had we chosen yet another set of eight disparate colors in the same color circle, that set also would have yielded equivalent results. There is no reasonable upper limit on the number of color-feature maps if every discriminable color requires its own map. This reveals another problem: A feature map is neurally expensive; it requires detectors at all spatial locations—enough detectors to enable fairly precise spatial localization.

Once there is a fixed set of color-feature maps, it is inevitable that some stimulus colors will fall between fixed features and stimulate more than one map. If the brain must and can utilize multiple maps, even for a homogeneous feature that happens to fall between two fixed color maps, and to group heterogeneous colors that involve many maps, then why have so many expensive maps? It is more logical to use the smallest number of maps needed to discriminate colors, that is, three for humans with 3D color vision, and to utilize the genetic hardware resources to construct better computational rules for combining maps rather than for constructing more and more maps. Instead of a very large number of predefined feature maps, an architecture that segregates visual elements into a limited number of layers (36, 37) according to visual regularity (21) is much more efficient.

These considerations about the brain mechanisms of grouping are analogous to the organization of the brain more generally. At early stages of neural processing, single neurons code simple local features. However, in the temporal lobe, at least hundreds, and probably many thousands of neurons are involved in coding a visual concept, and these same neurons are involved in equally many other concepts (38). Grouping seems to fall somewhere in the middle of this processing continuum.

ACKNOWLEDGMENTS. We thank Adam Reeves and reviewers Justin Halberda and Michel Treisman for their helpful comments and suggestions, and Pauline Ton for data collection in experiment 1.

1. Robson JG (1980) Neural images: The physiological basis of spatial vision. *Visual Coding and Adaptability*, ed Harris CS (Erlbaum, Hillsdale, NJ), pp 177–214.
2. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12:97–136.
3. Treisman A (1982) Perceptual grouping and attention in visual search for features and for objects. *J Exp Psychol Hum Percept Perform* 8:194–214.
4. Treisman A (1985) Preattentive processing in vision. *Comput Vis Graph Image Process* 31:156–177.
5. Quinlan PT (2003) Visual feature integration theory: Past, present, and future. *Psychol Bull* 129:643–673.
6. Wolfe JM (2007) Guided search 4.0: Current progress with a model of visual search. *Integrated Models of Cognitive Systems*, ed Gray W (Oxford Univ Press, New York), pp 99–119.
7. Koch C, Ullman S (1985) Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum Neurobiol* 4:219–227.
8. Milanese R, Wechsler H, Gill S, Bost JM, Pun T (1994). Integration of bottom-up and top-down cues for visual attention using non-linear relaxation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 1994* (IEEE, New York), pp 781–785.
9. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 20:1254–1259.
10. Itti L, Koch C (2001) Computational modelling of visual attention. *Nat Rev Neurosci* 2: 194–203.
11. Beck J (1966a) Perceptual grouping produced by changes in orientation and shape. *Science* 154:538–540.
12. Beck J (1966b) Effect of orientation and of shape similarity on perceptual grouping. *Percept Psychophys* 1:300–302.
13. Huang L, Pashler H (2007) A Boolean map theory of visual attention. *Psychol Rev* 114: 599–631.
14. Levinthal BR, Franconeri SL (2011) Common-fate grouping as feature selection. *Psychol Sci* 22:1132–1137.
15. Grossberg S, Mingolla E, Ross WD (1997) Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends Neurosci* 20:106–111.
16. Wertheimer M (1923) Untersuchungen zur Lehre von der Gestalt, II. *Psychol Forsch* 4: 301–350.
17. Rush GP (1937) Visual grouping in relation to age. *Arch Psychol* 31:1–95.
18. Hochberg J, Silverstein A (1956) A quantitative index of stimulus-similarity proximity vs. differences in brightness. *Am J Psychol* 69:456–458.
19. Ben-Av MB, Sagi D (1995) Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Res* 35:853–866.
20. Kubovy M, van den Berg M (2008) The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychol Rev* 115:131–154.
21. van den Berg M, Kubovy M, Schirillo JA (2011) Grouping by regularity and the perception of illumination. *Vision Res* 51:1360–1371.
22. Schmidt F, Schmidt T (2013) Grouping principles in direct competition. *Vision Res* 88: 9–21.
23. Tannazzo T, Kurylo DD, Bukhari F (2014) Perceptual grouping across eccentricity. *Vision Res* 103:101–108.
24. Wagemans J, et al. (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol Bull* 138:1172–1217.
25. Quinlan PT, Wilton RN (1998) Grouping by proximity or similarity? Competition between the Gestalt principles in vision. *Perception* 27:417–430.
26. Drew SA, Chubb CF, Sperling G (2010) Precise attention filters for Weber contrast derived from centroid estimations. *J Vis* 10:20.
27. Sun P, Chubb C, Wright CE, Sperling G (2016) The centroid paradigm: Quantifying feature-based attention in terms of attention filters. *Atten Percept Psychophys* 78: 474–515.
28. Sun P, Chubb C, Wright CE, Sperling G (2016) Human attention filters for single colors. *Proc Natl Acad Sci USA* 113:E6712–E6720.
29. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
30. Stockman A, Sharpe LT (2000) The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Res* 40:1711–1737.
31. Sperling G, Dosher BA (1986) Strategy and optimization in human information processing. *Handbook of Perception and Human Performance: Sensory Processes and Perception*, eds Boff K, Kaufman L, Thomas J (Wiley, New York), Vol 1, pp 2-1–2-65.
32. Halberda J, Sires SF, Feigenson L (2006) Multiple spatially overlapping sets can be enumerated in parallel. *Psychol Sci* 17:572–576.
33. Sperling G, Chu V, Sun P (2016) Multiple salience maps? *Abstr Psychon Soc* 21:34.
34. Michael E, de Gardelle V, Summerfield C (2014) Priming by the variability of visual information. *Proc Natl Acad Sci USA* 111:7873–7878.
35. Watson DG, Humphreys GW (1997) Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychol Rev* 104:90–122.
36. Grossberg S, Mingolla E, Ross WD (1994) A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychol Rev* 101:470–489.
37. Grossberg S, Zajac L (2017) How humans consciously see paintings and paintings illuminate how humans see. *Art Percept* 5:1–97.
38. Stevens CF (2018) Conserved features of the primate face code. *Proc Natl Acad Sci USA* 115:584–588.