

# UC Irvine

## UC Irvine Previously Published Works

### Title

Saliency maps for judgments of frontal plane distance, centroids, numerosity, and letter identity inferred from substance-invariant processing.

### Permalink

<https://escholarship.org/uc/item/35t7f4mq>

### Journal

Journal of Vision, 25(1)

### Authors

Gan, Lingyu

Sperling, George

### Publication Date

2025-01-02

### DOI

10.1167/jov.25.1.8

Peer reviewed

# Saliency maps for judgments of frontal plane distance, centroids, numerosity, and letter identity inferred from substance-invariant processing

Lingyu Gan

Department of Cognitive Sciences,  
University of California, Irvine, California, USA



George Sperling

Department of Cognitive Sciences and Neurobiology and  
Behavior, University of California, Irvine, California, USA



A saliency map is a topographic map that has inputs at each  $x,y$  location from many different feature maps and summarizes the combined saliency of all those inputs as a real number, saliency, which is represented in the map. Of the more than 1 million Google references to saliency maps, nearly all use the map for computing the relative priority of visual image components for subsequent processing. We observe that saliency processing is an instance of substance-invariant processing, analogous to household measuring cups, weight scales, and measuring tapes, all of which make single-number substance-invariant measurements. Like these devices, the brain also collects material for substance-invariant measurements but by a different mechanism: saliency maps that collect visual substances for subsequent measurement. Each saliency map can be used by many different measurements. The instruction to attend is implemented by increasing the saliency of the to-be-attended items so they can be collected in a saliency map and then further processed. Here we show that, beyond processing priority, the following measurement tasks are substance invariant and therefore use saliency maps: computing distance in the frontal plane, computing centroids (center of a cluster of items), computing the numerosity of a collection of items, and identifying alphabetic letters. We painstakingly demonstrate that defining items exclusively by color or texture not only is sufficient for these tasks, but that light–dark luminance information significantly improves performance only for letter recognition. Obviously, visual features are represented in the brain but their saliency alone is sufficient for these four judgments.

## Introduction

The concept of a saliency map was first proposed by Koch and Ullman (1985). Their saliency map was a topographic map that had inputs from many different feature maps and summarized those inputs at each  $x,y$  location as a single real number, now commonly called saliency, which was represented in the map. The concept of a saliency map was initially elaborated by Itti, Koch, and Niebur (1998) and Itti and Koch (2000). Of the more than 1 million subsequent Google references to saliency maps, nearly all use the map to compute the priority of components of visual images for subsequent processing. In contrast, Lu and Sperling (1995) proposed that a saliency map was used to compute the motion direction of their complex visual stimuli because all simpler computations were excluded. Subsequently, visual centroid (center of mass) judgments (Sun, Chu, & Sperling, 2021; Gan, Sun, & Sperling, 2023) and distance judgments (estimates of the distance between two objects, both seen in the frontal plane (Gan, Sun, & Sperling, 2021) were reported to use saliency maps.

Here we aim 1) to clarify the concept of a saliency map, 2) to demonstrate that perceptual judgments of distances in the frontal plane, of centroids, and of numerosity not only utilize saliency maps, but there is no better measurement system available to these judgments, 3) to show that to a lesser extent letter shape also can utilize saliency maps, and 4) to lay to rest the notion that luminance information (light–dark) is required for any of these tasks versus variations only in color or texture. The conclusion is that saliency is a critically important brain process for representing information that is independent of the particular features that happen to carry the information to the

Citation: Gan, L., & Sperling, G. (2025). Saliency maps for judgments of frontal plane distance, centroids, numerosity, and letter identity inferred from substance-invariant processing. *Journal of Vision*, 25(1):8, 1–19, <https://doi.org/10.1167/jov.25.1.8>.



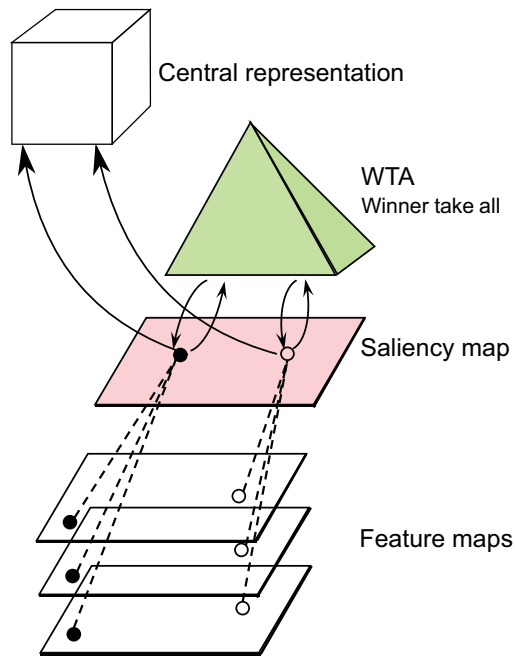


Figure 1. The original saliency processing system of Koch and Ullman (1985), colors added.

brain is therefore available for previously unencountered features.

## Substance invariance

We propose the term substance-invariant measurement for a measurement that is invariant to the substance being measured.<sup>1</sup> Consider three common kitchen substance-invariant measuring tools: measuring cups, weight scales, and measuring tapes (Figure 2). A measuring cup tells us how much substance it contains, but a measuring cup is invariant to what the substance is. A cup could contain water or milk, rice, or sugar,

sand, or nails, or any mixture. The defining feature of measuring cups, weight scales, and tape measures is that they do not know what they are measuring, and they provide a positive real number output that describes the amount.

For the visual system, substance is the composition of visual input that is represented as figure versus ground, or in more contemporary terms, the feature composition of an area or areas of the visual field that are represented as salient. Because the brain will encounter an indefinitely large number of different visual substances, it needs to be able to make substance-invariant measurements. An important difference between the brain and the three kitchen measuring devices is that the brain seems to be able to use the same or very similar representation for many different measurements. The analogy would be that we collect the substance to be measured in a container, the saliency map, and then empty it onto a cup, weight scale, or a ruler for the computation of volume, weight, or length. Figure 2d shows pin art, a physical embodiment of saliency map architecture. This distinction between the representation of saliency in a saliency map and the subsequent computation of priority is essential in computational models of priority processing (Itti & Koch, 2000), has been observed in the brain (Bogler, Bode, & Haynes, 2011), and it recurs in computational models that propose saliency processing for many different tasks (Lu & Sperling, 1995; Gan et al., 2021; Sun et al., 2021; Gan & Sperling, 2022; Gan et al., 2023).

For the measurement tasks considered here, the brain's substance-invariant measurements require 1) isolating the particular substance to be measured, a process called grouping, 2) representing the group in a saliency map so that it can be measured, 3) making the measurement, and 4) associating the measurement with the group identity. These processes are considered in detail elsewhere (Sun et al., 2021;

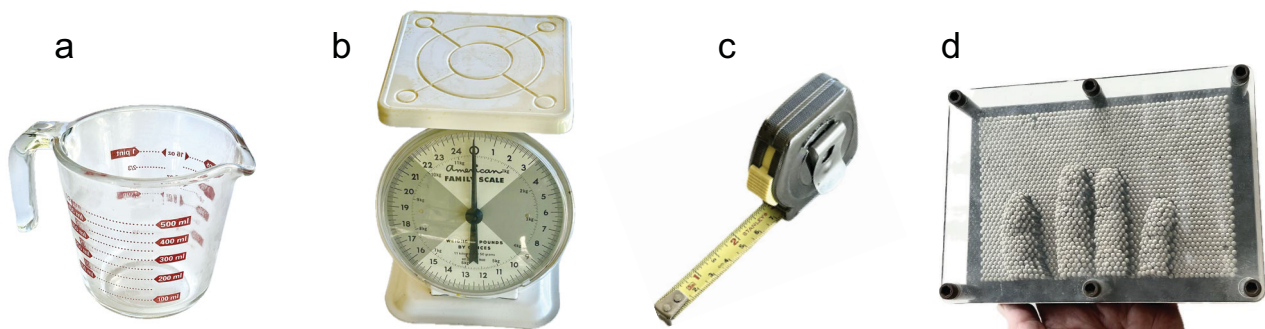


Figure 2. Three household substance-invariant measuring devices, and pin art for representing spatial patterns: (a) Measuring cup. (b) Weight scale. (c) Tape measure. (d) Pin art. Measuring cups, scales, and tape measures deliver a single non-negative number to describe their measurement; pin art delivers a spatial array of non-negative numbers (like a saliency map) to represent its measurements.

Gan et al., 2023). Here we are concerned primarily with demonstrating that four additional perceptual tasks, beyond calculating processing priority or motion direction, are solved by substance-invariance processes, and therefore a salience map is the likely intermediate step in each of the four solutions. We take great care to also demonstrate a version of each task that is completely impervious to luminance processing and therefore requires an alternative—a higher-level salience process.

### Four substance-invariant judgments

1) Compute the distance between two target items in the frontal plane; 2) estimate the center of mass of a set of spatially distributed target items; 3) estimate the number of target items; and 4) identify a letter image. For these tasks, the only information required is the location occupied by each target item. Insofar as these judgments utilize salience maps that record the locations of targets, the accuracy of these judgments should be independent of the particular features used to define targets, provided that the features contain sufficient location information.

## General methods

### Substance-mixed paradigm

To demonstrate substance invariance, we utilize two versions of a substance-mixture paradigm.

In version 1 substance mixture paradigm, target items range from being identical to being defined by a great variety of different colors, textures, or shapes. The required judgment has to be made on the basis of possibly unknown-in-advance features that may be different for each target item. Because there is a virtual infinity of potential distinguishing features, the task requires a salience map representation that is independent of the particular critical features. All experiments use version 1 for most conditions.

In version 2, there are both target items and foils. Foils are nontarget items that are targets on other trials. For example, in a stimulus with both red and green items, subjects are told in advance to respond only to the red (or the green) items and to ignore the others. Experiments 2b, and 3b used version 2 trials. In isoluminant version 2 trials, selective attention to a particular color is necessary to select target items that are equal in luminance not only to some proportion of the gray distracters that fill the background but also to differently colored isoluminant foils that are targets on other trials. The reason for some of these complications is that, in addition to demonstrating

substance invariance, there is at least one condition for each of the tasks that absolutely rules out the possibility that luminance information could be used to solve that task. Full details of the apparatus and procedures are given in [Appendix A](#).

## Experiment 1

### Experiments 1a and 1b: Frontal plane distance judgments

Are frontal plane distance judgments substance invariant and therefore based on a salience map? This is not an issue that has been specifically addressed in the very few published reports of frontal plane judgments. Cook (1978) tested the scalability of interval-ordering judgments of distance along a planar horizontal surface and found that the form of the distance scale was a power function. An incidental finding from Burbeck (1987) and Burbeck (1988) was that the spatial frequency composition of two parallel bars did not influence the accuracy of judging the distance between them. Gan et al.'s (2021) preliminary study that is greatly elaborated here was the first to consider a salience mechanism to account for distance judgments in the frontal plane. Experiment 1a elaborates on that study to demonstrate that in judging the distance between two disks in the frontal plane, it does not matter whether the disks are similar or different; all that matters is knowing that they are different from the background and knowing where they are. Experiment 1b extends these results to disks that are a priori guaranteed to be invisible to the visual luminance system.

#### Procedure and stimuli

In both Experiments 1a and 1b, subjects viewed a computer screen that displayed a fixation cross. A key press was followed in 0.5 seconds by a 200-ms exposure of a stimulus containing two target disks followed immediately by a random masking field to terminate visual persistence. Subjects typed their estimate of the separation of the two targets in tenths of inches, and full feedback was provided after each trial. The background was either filled with 142 distracter disks equal in size to target disks colored in various shades of gray to camouflage colored targets for a luminance-dependent system, or in control conditions, the background was uniform gray to maximally expose the two targets.

In Experiment 1a, 15 different substance compositions were tested in a mixed list design. Seven target pairs were identically composed, and eight pairs were differently composed, as shown in the abscissa of [Figure 3f](#), with each column representing one target

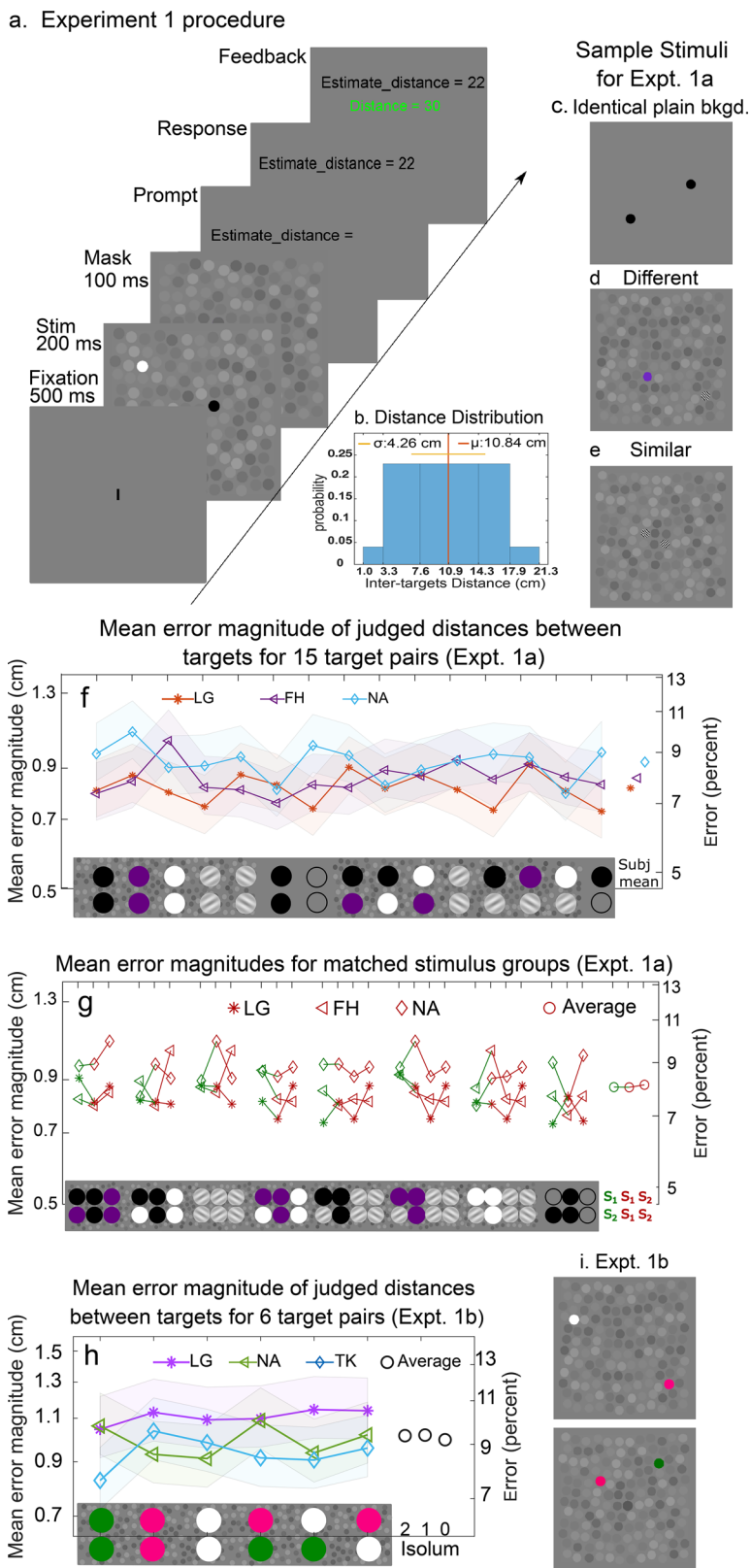


Figure 3. Procedure, experimental conditions, sample stimuli, and results for the distance judgments in Experiments 1a and 1b. (a) Procedure: Every trial began with a 500-ms blank field with a fixation bar, a 200-ms stimulus display, and a 100-ms postexposure masking field. Subjects were then prompted to enter on the keyboard their estimate of the distance between the two targets. Feedback was provided after each trial. (b) The distribution of distances between the two targets; 1 cm = 1.0 degree of visual angle. (c, d, e) Sample stimuli for the distance judgments of Experiment 1a. (c) Two black disks. (d) A grating patch and a purple disk. (e) A clockwise-oriented grating patch and a counterclockwise grating patch. (f) Mean error magnitudes of three subjects' judgments of the

← distance between target pairs of 15 types (shown at the bottom). The subjects' overall mean error magnitude of 0.87 cm corresponds with an average Weber Fraction of 0.080. The colored area around the data represents a 95% confidence interval. The right-side ordinate is the percentage error of the estimated distances. (g) Groups of three and of four matched pairs of targets to compare distances estimated between identical-targets pairs (homogeneous, red) with different-targets pairs (heterogeneous, green). The last group (extreme right) represents the data averaged over the eight stimulus groups and the three subjects. (h) Three subjects' mean judged distance errors for target pairs shown at the bottom. Targets in red and green were approximately isoluminant to the background. (Distracter disks shown small here actually were the same size as stimulus disks.) The open circles above 2, 1, and 0 represent data averaged over the three subjects over target pairs that contained 2, 1, or 0 isoluminant target disks. (i) Two sample stimuli for Experiment 1b.

pair. Of these pairs, 3 were presented without any distractors, and the remaining 12 were accompanied by 142 distractors. Every target pair was tested 100 times in a mixed list design that included all 15 pairs of targets. The distribution of intertarget distances (but not the physical locations) was the same for each pair type (Figure 3b)

In Experiment 1a, only the purple disks were isoluminant with some background disks. In Experiment 1b, to further demonstrate that luminance is not required for frontal plane distance judgments, targets that were isoluminant to the background were directly compared with targets that maximally differed in luminance from the background. More specifically, three colors were used to define targets: red, which is approximately isoluminant with the background; green, which is approximately isoluminant with the background; and bright white, which differs greatly in luminance from the background. These three colors resulted in six different pairs of targets (Figure 3h). In five of the six conditions, at least one target was isoluminant to the background. Figure 3i shows two sample stimuli. If distance judgments require luminance, it would be impossible for subjects to find the isoluminant red or green targets. The isoluminant colors were calibrated individually for each subject before Experiment 1b. Also, the 142 background disks varied in luminance so that luminance was useless even for identifying just nominally isoluminant red or green disks. However, luminance was intentionally a vivid cue for identifying the bright white disks. Each target pair was tested 100 times, and the distribution of intertarget distances for each pair is depicted in Figure 3b.

## Results

Subjects' performances were measured in terms of the response error — the absolute difference between the true intertarget distance and the subject's judged intertargets distance. The root mean square (RMS) error averaged over three subjects and conditions was 1.10 cm, accounting for 93.0% of the variance of the stimuli. The mean (versus RMS) error magnitude averaged over subjects and conditions was 0.87 cm. The

obvious overall result is that subjects' performances are both very accurate and remarkably invariant over the many conditions of the two experiments. A one-way repeated measures analysis of variance was performed on the subjects' responses to examine the extent to which their performances varied in Experiment 1.<sup>2</sup> No statistical difference in mean error magnitude was found among the 15 experimental conditions,  $F(14, 28) = 0.9824, p = 0.4944$ , demonstrating that accuracy was statistically indistinguishable with 100 trials per condition. Even in the most difficult trials (condition 11), when the target pair was composed of two differently orientated gratings and the mean luminance of the two target gratings was statistically the same as the mean luminance of the 142 distractors, judgments were statistically no less accurate than in the easiest trials, where the target pairs were two solid black discs on a uniform gray background (condition 6),  $F(1, 2) = 2.5291, p = 0.2527$ .

To further examine whether there was a difference in response accuracy for identically composed disks versus two differently composed disks, 15 experimental conditions were grouped into eight groups (Figure 3g). Five of the eight groups had three pairs of targets, two of which were identical target pairs; the remaining pair was a mixed target pair, one target from each of the identical pairs. The remaining three groups each had four pairs of targets. Three of the four pairs were identical target pairs, the fourth pair was a different target pair in which one target was a randomly chosen orientation of grating, and the other target was from the nongrating pair of identical targets. No statistical difference in mean error magnitude was found between identical target and matched different target pairs of elements within each group, indicating that there is no advantage in judging identically composed target pairs.

Results for Experiment 1b are shown in Figure 3h. There was no statistical difference in mean error magnitude among the six experimental conditions,  $F(5, 10) = 0.4990, p = 0.7708$ . Luminance (vs. color or pattern) information was neither necessary nor advantageous, nor was there any advantage for color or pattern similarity for these frontal plane distance judgments: perfect substance invariance within resolution error.

## Discussion

The statistical comparisons between the various conditions are the conventional way of describing how similar distance judgments with various differently constituted stimulus pairs are to each other. However, what is important here is not that the judged distances between similar versus differently composed stimulus pairs, or differences between luminance-defined versus isoluminant stimulus pairs are statistically indistinguishable with only 100 trials. It is that the accuracy differences are, in fact, very small by any reasonable measure, which suggests all judgments are made by a single system—a salience system. We consider and reject two alternative hypotheses:

1) The null hypothesis. Subjects have only a luminance mechanism for distance judgments. If a subject did not have a visual system and brain mechanism that could detect and then compute the distance between two isoluminant items, then the subject's isoluminant distance judgments would be uncorrelated with the stimuli distances. This would be evident after just a few trials, and lead the experimenter to interrupt and ask “What’s the problem?” and the subject to reply “The task was easy on the trials with dark gray and light gray targets but I can’t do it now because I don’t see any targets that stand out from the background” or (less probable) “I see the targets but for some reason I can’t judge how far apart they are.” We do not know how to resolve the answers to these questions because, for our subjects and for normally sighted persons who viewed our displays, these events never occurred; trials with the highly discriminable isoluminant stimuli were not judged differently from other trials.

2) A second hypothesis: There are two distance judgment systems, one for luminance stimuli and another for isoluminant stimuli. There are two problems with this formulation. First, isoluminance simply means luminance is not a distinguishing cue; it does not specify what the other system or systems are; whether items are identified on the basis of color or texture or form or of any other set of features. Second, given the costs for two systems versus one system in terms of genetic code, anatomical requirements, and metabolic costs, it seems so much more likely to us that one system will have evolved to perform both tasks that we do not further consider here the unlikely possibility that multiple similar but separate systems exist for estimating the frontal plane distances studied herein.

### Statistical versus functional differences

Although the luminance and isoluminance data are statistically indistinguishable in the present dataset, the neural processing cannot be exactly the same. Different stimuli are processed—at least to some extent—by different neurons so it is remarkable that the results

are so similar. A useful analogy is the speed of light in air and in a vacuum. Based on physical principles, the speeds cannot be the same, but they differ only in the sixth decimal place. Psychophysical measurements do not allow that degree of precision. Although responses to different stimuli may be very similar, they are never exactly the same, so we assert here only that differences are small or unimportant. That said, the appendix contains statistical power analyses to further inform all the data analyses herein.

## Conclusions

The great similarity between luminance and isoluminance results strongly suggests that there is only one brain system for computing distances in the frontal plane, and that it was used equally by all the isoluminant and luminance stimulus variants. We designate it as a salience system because that is what all the judged stimuli have in common and what, in Experiments 2 and 3, distinguishes attended stimuli from to-be-ignored stimuli.

## Experiment 2

### Experiments 2a and 2b: Centroid

A centroid is the center of gravity of a group of items. In Experiment 2a, all items have the same weight. Experiment 2a seeks to demonstrate that judgments of the centroid of a group of items can be substance invariant, that is, invariant to the variety of features of which the items are composed. Experiment 2b compares luminance with isoluminant stimuli crossed with conditions that require or do not require attention to extract the target items for centroid computation. In the most difficult condition, subjects must distinguish targets from distracters of various gray levels and also from foils of the same luminance as the target but of a different color. It is inconceivable that a luminance-based system could perform above chance in any of the isoluminant conditions but especially impossible in the isoluminant distracters plus isoluminant foils condition.

### Procedure and stimuli

In Experiment 2a, substance-mixed paradigm version 1 was used. [Figure 4a](#) depicts the trial procedure. Subjects were presented with a stimulus display containing 16 stimulus items for 300-ms, immediately followed by a 50-ms blank field and a 100-ms masking field. Subjects were instructed to judge the centroid of all the stimulus items, and feedback was provided after each trial ([Figure 4a](#)). Stimulus items were either identically composed or differently composed. In total,

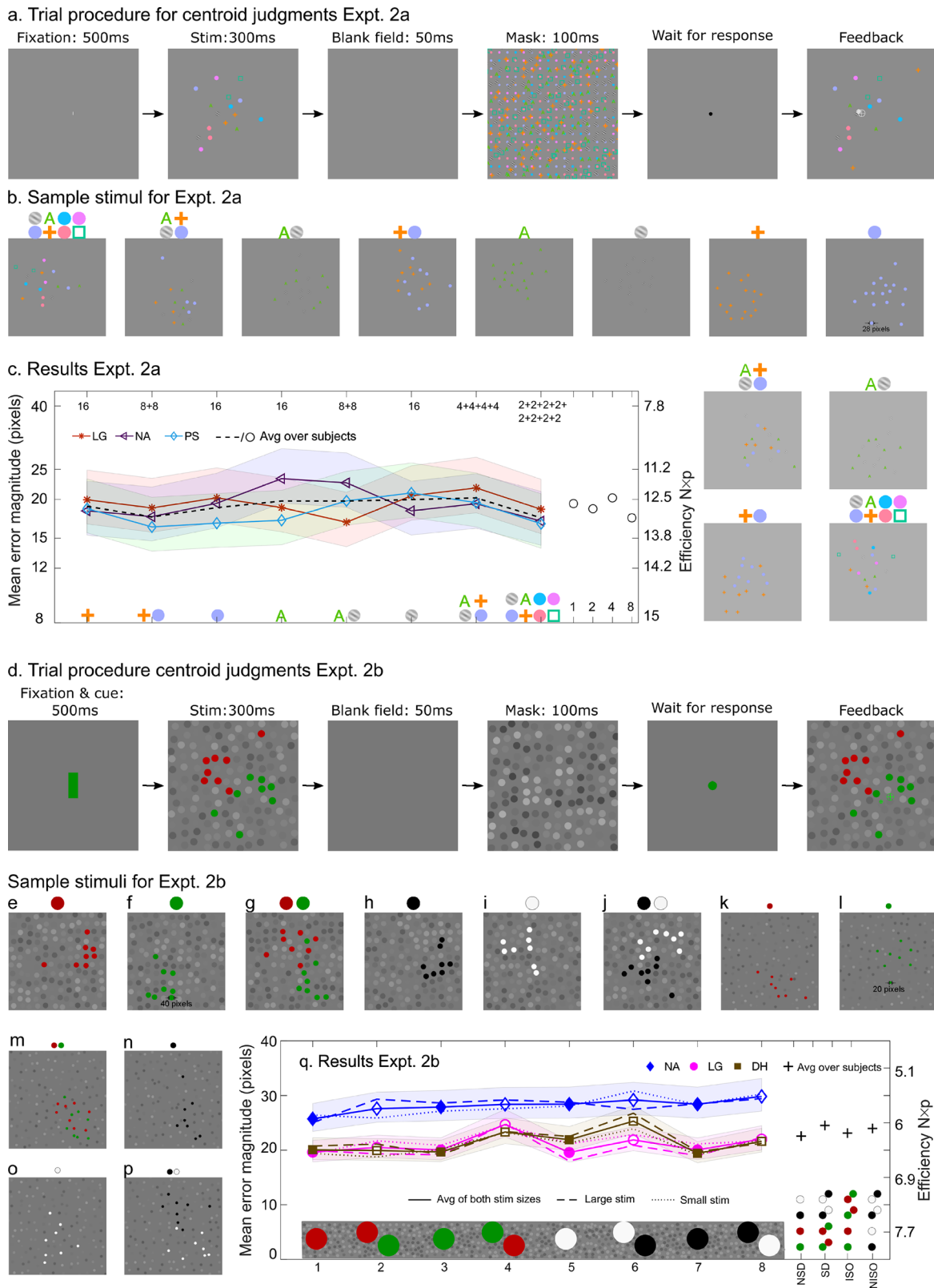


Figure 4. Procedure, sample stimulus displays, and results for centroid judgments, Experiment 2a and 2b. (a) Trial procedure for Experiment 2a. Every trial began with a 500-ms blank field that contained a fixation point that indicated the to-be-attended color. It was followed by a 300-ms stimulus, a 50-ms blank field, a 100-ms masking field, a blank field with a movable cursor that the subject moved to the judged centroid location (centroid of all items), and finally, a feedback display. Feedback showed the stimulus, the centroid of all stimulus items (a large gray plus sign inscribed in a gray open circle), and the subject’s response (a smaller gray plus sign



←  
 inscribed in a gray circle). (b) Sample stimuli for Experiment 2a. To make these sample stimulus items more visible, the gray level of the background of sample stimuli above panel (c) is approximately 77% darker than the actually presented background. The four sample stimuli to the right of (c) show the actual stimulus gray levels. (c) Results Experiment 2a: Three subjects' mean error magnitudes of judged centroids of the 16 target items for the eight experimental conditions shown below. The numbers at the top of the figure are the number of items per substance in each experimental condition (shown at the bottom). The colored area around the data represents a 95% confidence interval. The right-side ordinate is the corresponding efficiency—the number of stimulus items that an ideal observer has to perfectly process in order to match a subject's performance. (d) Trial procedure for Experiment 2b. Everything was the same as Experiment 2a, except that the fixation bar also served as a pre-cue, indicating which centroid to report. (e–j) Sample stimuli of large disks in Experiment 2b. (k–p) Sample stimuli of small disks in Experiment 2b. (q) Three subjects' mean error magnitude in their centroid judgments for the 8 experimental conditions in Experiment 2b shown at the bottom. Targets in red and green were isoluminant with the background. On the bottom, the left symbol in a pair was the target. Dashed lines represent small stimuli, dotted lines represent large stimuli, solid lines represent the average of small and large stimuli. The colored area around the data represents a 95% confidence interval of the average of small and large stimuli. Each + at far right of panel q represents the average data of three subjects for the four conditions indicated underneath: "NSD" (no similar distracters), "SD" (similar distracters = foils), "ISO" (isoluminant), and "NISO" (not isoluminant) represent data averaged over three subjects and the four indicated conditions.

50 trials of 8 different compositions of stimulus items were tested, all in a mixed list (Figure 4b).

In Experiment 2b, to exclude the confounding variable of luminance, both substance-mixed paradigm versions 1 and 2 were used. Specifically, each stimulus display contained 144 same-size disks, 8 targets, 0 or 8 foils, and the remainder were distracters. Distracters were of the same size as targets and of varying luminance so that colored targets could not be distinguished from distracters on the basis of luminance (Figures 4e–p). Eight combinations of targets and distracters were tested. In four pre-cued version 1 conditions, one of four colors was used to define targets (Figure 4q): vivid red that is approximately isoluminant to the background, vivid green that is approximately isoluminant to the background, maximally white, and maximally black (Figures 4e, f, h, i, k, l, n, and o). In four pre-cued version 2 conditions, each stimulus display contained 128 varied gray-level disks, 8 paired foils, and 8 targets (isoluminant red and isoluminant green was a pair of targets and foils as was maximally white and maximally black) (Figures 4g, j, m, and p). Subjects had to ignore the foils (which were targets on other trials), that is, treat them like the distracters. All 8 experimental conditions were tested in a mixed list of 100 trials per condition. Two different sizes of disks (large  $40 \times 40$ , small  $20 \times 20$  pixels) were examined in a blocked design.

## Results

Subjects' performance was measured in two ways: *response error*—the difference between the true centroid location of the target feature and the subject's mouse-click response, and *efficiency*—the minimum number of target items that an ideal detector would need to detect, to locate perfectly, and to perfectly compute a centroid to match a subject's mean error.<sup>3</sup>

In Experiment 2a, if centroid judgments were not substance invariant, then the compositions in which

all targets have the same feature would be the easiest trials; whereas composition with eight different features would be the most difficult. In fact, all three subjects performed both very well and remarkably similarly in all eight experimental conditions (Figure 4c). The overall mean error magnitude (the distance between the true centroid location of the stimulus items and subject's mouse-click response) averaged over three subjects was 19.06 pixels, which was less than the diameter of a single stimulus item (28.00 pixels) in the stimulus array. The judged centroids account for 97% of the variance of stimulus centroid variance. An ideal detector would require perfect position knowledge of 12.5 of the 16 stimulus items to match the subjects' average accuracy. No statistical difference in mean error magnitude was found among the eight experimental conditions,  $F(7, 14) = 0.9879, p = 0.4778$ , demonstrating that the differences in accuracy were too small to measure with 50 trials. Also, no statistical differences in mean error magnitude were found between identically composed and matched differently composed stimuli, indicating that there was no statistically significant advantage in judging identically composed stimulus items. For plus signs alone, purple disks alone versus plus signs mixed with purple disks,  $F(2, 4) = 3.0908, p = 0.1543$ . For letter A alone, Gabor patches alone versus letter A mixed with Gabor patches,  $F(2, 4) = 0.0056, p = 0.9942$ .

In Experiment 2b, as shown in Figure 4q, the two sizes of stimulus items produced statistically indistinguishable data, so the data of the large and small stimulus items were combined for the following analyses. All three subjects performed very well (Figure 4q). The mean error magnitude averaged over three subjects and all conditions was 23.52 pixels. An ideal detector would require perfect position knowledge of 6.14 of the 8 stimulus items to match this accuracy. The result of this experiment is that the difference in judging centroids between the four isoluminant and the four maximally obvious luminance conditions is too small to measure with 3 subjects and

100 trials for each of 8 conditions,  $F(1, 2) = 2.0227$ ,  $p = 0.2909$ .

The difference in judging centroids between the four conditions with foils and the four conditions without foils was borderline statistically significant,  $F(1, 2) = 22.5446$ ,  $p = 0.0416$ . The estimated number of dots processed by subjects in conditions requiring attention to disregard foils was 6.04 versus 6.24 dots in the foil-free condition. This indicates that the data are very accurate, that subjects' attention filters (Sun, Chubb, Wright, & Sperling, 2016) were highly selective, and this extreme method for excluding luminance processing produces only a 3% decline in performance.

## Experiment 3

### Experiments 3a and 3b: Numerosity

There are several approximate regimes of number estimation (e.g., Anobile, Cicchini, & Burr, 2016): 1) subitizing, for less than approximately four items; 2) number estimation, for more than four items when items are distinguishable as unique items; and 3) texture mechanisms for items that are so densely packed that items cannot be segmented from each other. Experiment 3a seeks to demonstrate that in a middle range of 9 to 27, estimates of the number of items in a briefly flashed display are substance invariant—numerosity judgments do not depend on whether the items are similar or different; all that matters is knowing items are present. Experiment 3b further compares numerosity estimations of stimuli that are invisible to the luminance system with numerosity estimates of highly visible black- or white-on-gray stimuli.<sup>4</sup>

#### Procedure and stimuli

The number of target items ranged from 9 to 27 for Experiment 3a and from 5 to 13 for Experiment 3b. For both Experiments 3a and 3b, the overall area occupied by stimulus items was kept constant for different numbers of items; that is, the density of items covaried with numerosity. He, Zhang, Zhou, and Chen (2009) and Ross and Burr (2010) report that perceived numerosity is mostly independent of the density of stimulus items. The procedures and stimuli in Experiment 3a were similar to Experiment 2a, except for the following: 1) the task in Experiment 3a was to estimate the numerosity, always of ALL the stimulus items; and 2) the number of items of a particular feature was randomly determined for each trial with the constraint that each feature occurred at least once. Within a session, for each number composition, the total number of items per feature was the same for every feature. The approximately rectangular distribution of the number of items was identical across all experimental conditions (Figure A2a).

Experiment 3b was identical to Experiment 2b except that 1) only the large item size was tested; and 2) the number of items per feature was randomly drawn from an approximately rectangular distribution between 5 and 13 (Figure A2b), with each feature having the same distribution of items across all experimental conditions. As in Experiment 3a, the task was to estimate the numerosity only of the items being cued. Eight experimental conditions were tested in a block design.

#### Results

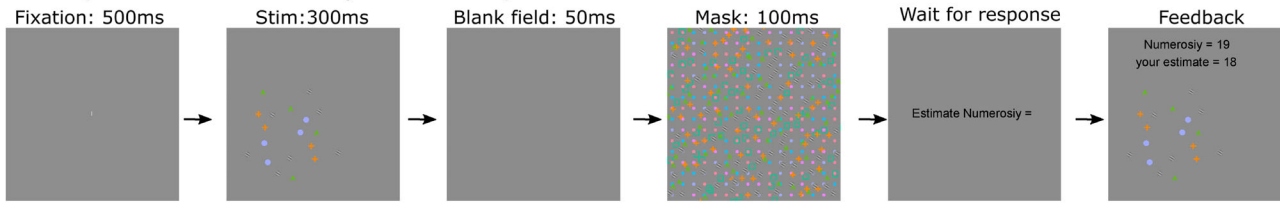
Response error magnitude—the absolute value of the numerical difference between subjects' estimate and the true numerosity—was used to evaluate subjects' performance. In Experiment 3a, all three subjects performed very well in all seven experimental conditions with an average error of less than one item (Figure 5c). With an average of 220 trials per condition for each subject, no statistically significant difference in mean error magnitude was found among the seven experimental conditions,  $F(6, 12) = 1.3929$ ,  $p = 0.2938$ . Responses accuracy of two conditions with identically composed targets were almost statistically different from five conditions with differently composed targets, 4608 trials,  $F(1, 2) = 16.3073$ ,  $p = 0.0562$ . There was a trend toward slightly better numerosity judgments with diverse versus identical stimuli. Very small differences like this are of great interest in competitive sports, but here we are concerned with the overall picture. Numerosity estimates are extremely similar for extremely different stimulus compositions. Therefore, we conclude numerosity are salience based, not feature specific.<sup>5</sup>

#### Results, Experiment 3b

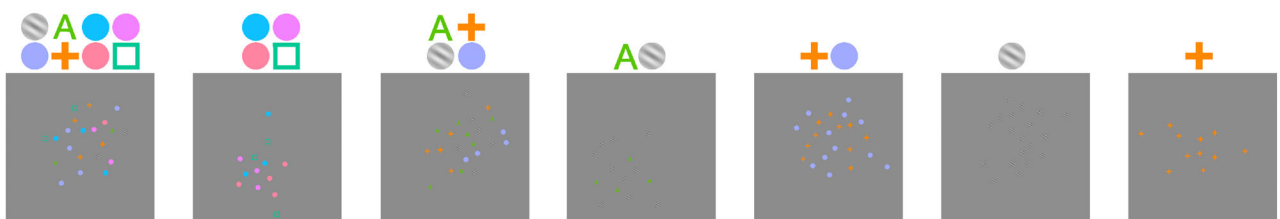
Overall, numerosity judgment accuracy was very good and very similar for all eight conditions,  $F(7, 14) = 2.6551$ ,  $p = 0.0569$ , including luminance and isoluminant stimuli. There was a trend for the version 2 conditions (which required subjects to attend only to pre-cued targets and to ignore foils that were targets on other trials) to be slightly less accurate, 0.76 vs. 0.67 (Figure 5k SD vs. NSD). Based on 100 trials per condition times 3 subjects, the cost of attention-required versus not-required was not statistically significant,  $F(1, 2) = 9.3233$ ,  $p = 0.0926$ .

There was a small difference in numerosity estimation errors between the four isoluminant (0.73) and the four luminance conditions (0.70). Again, this would be inconsequential even if it were statistically significant, which it is not for this sample size of 100 trials per 8 conditions times 3 subjects:  $F(1, 2) = 0.2040$ ,  $p = 0.6957$ . Because the accuracy of numerosity judgments is essentially similar independent of the diversity of the

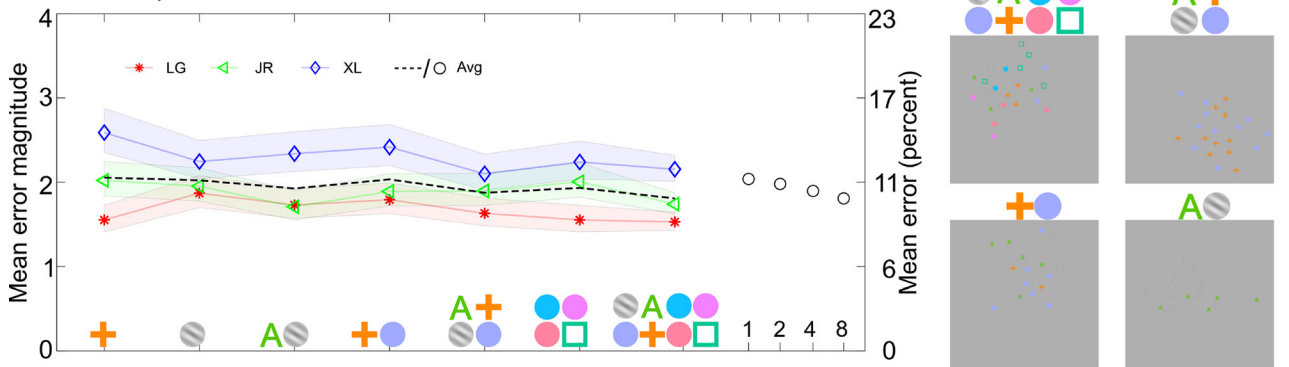
a. Trial procedure for numerosity estimates Expt. 3a



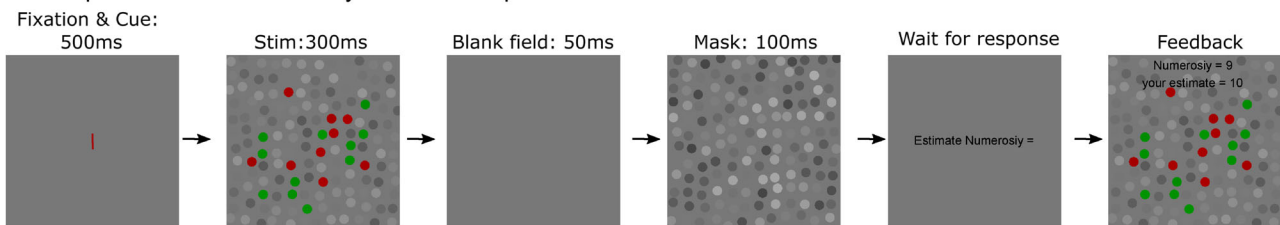
b. Sample stimuli for Expt. 3a



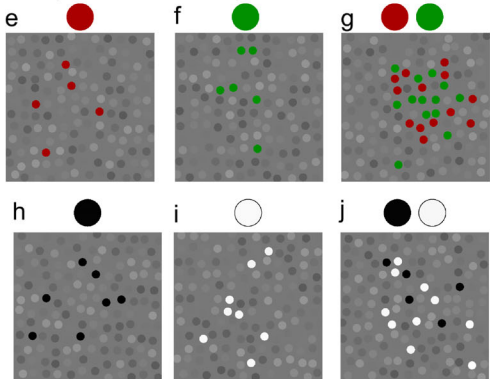
c. Results Expt. 3a



d. Trial procedure for numerosity estimates Expt. 3b



e. Sample stimuli for Expt. 3b



k. Results Expt. 3b

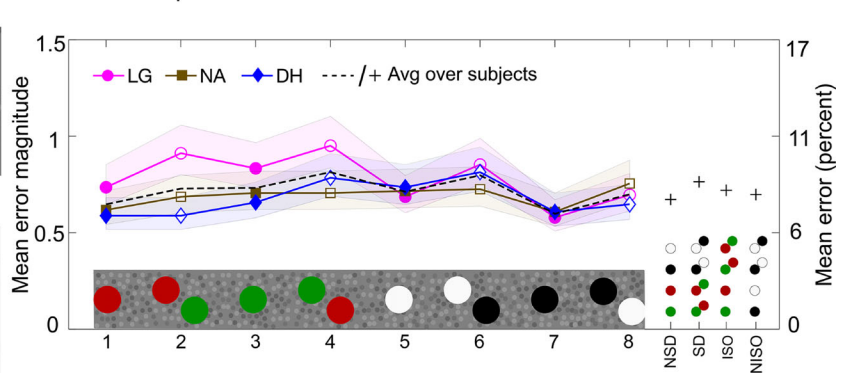


Figure 5. Procedure, sample stimulus displays, and experimental results for numerosity estimation, Experiments 3a and 3b. (a) Trial procedure for Experiment 3a. Every trial began with a 500-ms blank field with a fixation point, followed by a 300-ms stimulus, a 50-ms blank field, a 100-ms masking field, a field that prompted the subject to type a one- or two-digit estimate of the total number of stimulus items, and finally, a feedback display. The feedback display showed the stimulus, the subject’s estimate, and the numerosity of the stimulus. (b) Sample stimuli. To make the stimulus items more visible in this reproduction, the gray level of the background of the sample stimuli is darker than the actual background. The four sample stimuli with lighter gray backgrounds show what the

← stimulus items look like in the experiment. (c) Results, Experiment 3a. Mean error magnitude of the judged numerosity of stimulus items in seven experimental conditions for three subjects. “1, 2, 4, and 8” at the far right of the abscissa indicate the experimental conditions in which target items are composed of 1, 2, 4, or 8 different features. The colored area around the data represents a 95% confidence interval. The right-side ordinate is the corresponding error fraction: (mean error magnitude)/(mean number of stimulus items). (d) Trial procedure for Experiment 3b. The task was to estimate the number only of those items that have the same color as the fixation bar. (e–j) Sample stimuli for Experiment 3b. (k) Results: Mean error magnitude of numerosity estimates for three subjects in the eight experimental conditions shown at the bottom. On the bottom, the left symbol in a pair was the target. “NSD” (no similar distracters), “SD” (similar distracters = foils), “ISO” (isoluminant), and “NISO” (not isoluminant) represent data averaged over three subjects and the four indicated conditions. The colored area around the data represent a 95% confidence interval. The right-side ordinate is the corresponding error fraction: (mean error magnitude)/(mean number of stimulus items).

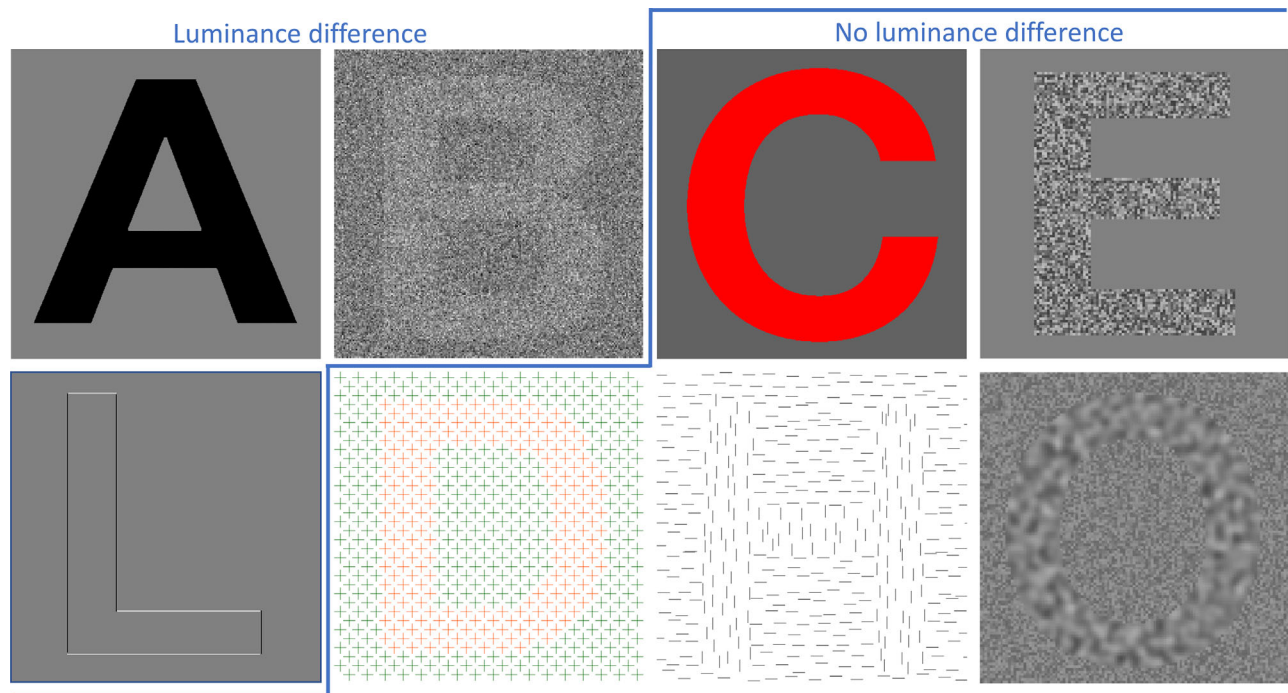


Figure 6. Eight letter examples. Letters A, B, and L have luminance cues to shape. In a calibrated display of the other five letters, any patch within a letter will have the same expected luminance as any patch within the background.

items being judged and of whether or not the items are isoluminant with the background, we conclude that numerosity judgments are based on salience maps.

### Letter identification can use the salience of the substance of which letters are composed

This section displays discriminable letters composed of isoluminant colors and of isoluminant textures to demonstrate a degree of substance invariance in letter identification.

Figure 6 shows eight letters, each of which is composed of a different substance. The luminance of letters A, B, and L is different from the background whereas, on a calibrated screen, the luminance of letters C, E, D, H, and O equals the luminance of the background. All eight letters are identifiable

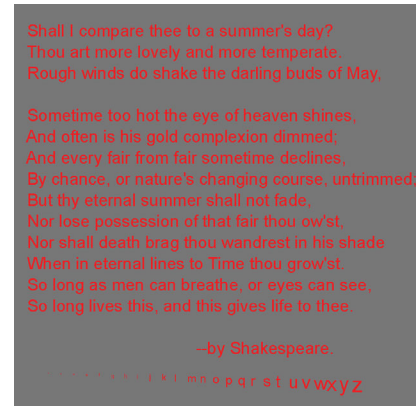
but not equally so. The examples are chosen to illustrate different substances. In Figure 6, A has a big luminance difference from the background, B has a small luminance difference, and only the edges of L differ from the background luminance. Among the nominally isoluminant letters, the red C is as obvious as the black A. As will be demonstrated below, big red letters remain obvious even at isoluminance. D is a low-contrast isoluminant counterpart for B. H relies on texture orientation differences, and E and O differ in texture spatial frequency composition from the background.

Figure 7a shows letters of 14 different colors and of 6 different sizes on a gray background. Letters in the same column have the same color. Except for the black and white letters (the leftmost and the rightmost columns), on a calibrated display, other letters are more-or-less isoluminant with the background. On a dynamic

## a. Letters of 14 different colors and of 6 different sizes



## b. Shakespeare's Sonnet 18



## c. Letters of 25 different contrasts and of 6 different sizes

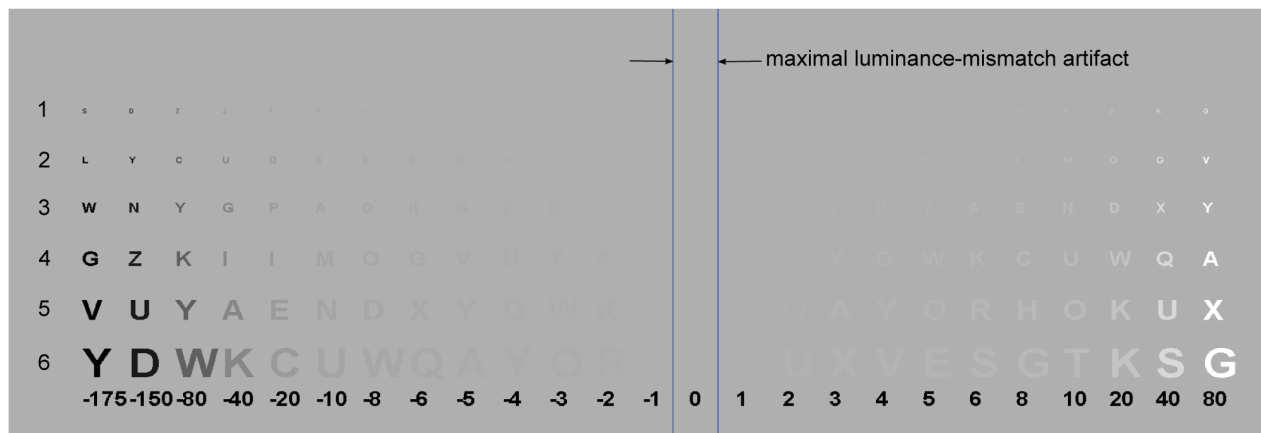


Figure 7. (a) Letters of 14 different colors and of 6 different sizes on a gray background. Letters in the same row have the same size, and letters in the same column have the same color. Except for black letters in the first column and white letters in the last column, other letters are isoluminant with the background on a calibrated display. In demonstration (<https://github.com/Lingyu-Gan/Saliency-maps-for-judgments-of-frontal-plane-distance-1-centroids-numerosity-and-letter-identity-.git>) 1 on a display monitor, the background intensity of Figure 7a can be varied to find the background intensity value that produces isoluminance (minimum visibility) for any particular letter for that particular observer. Most, if not all, larger letters typically remain visible in this demonstration. (b) A nominally isoluminant text version of Shakespeare's Sonnet 18. (c) Letters of 25 different luminances and of 6 different sizes on a background with a gray level of "175" (range, 0–255). Letters in the same row have the same size and letters in the same column have the same luminance. The numbers in the bottom row represent nominal differences in luminance between letters and the 175 background. The actual contrasts in Figure 7b depend on the Gamma function of the monitor on which the figure is viewed. For typical Gammas of 2.0 to 2.4, the contrast of one unit (bottom row of Figure 7b) is 1.15% to 1.38%.

version of this display, (<https://github.com/Lingyu-Gan/Saliency-maps-for-judgments-of-frontal-plane-distance-1-centroids-numerosity-and-letter-identity-.git>) the background intensity is variable so that the viewer can find the intensity that best conceals any particular letter color. The main observation is that, from viewing monitors or journals at a distance of 1 m or so, appropriate adjustment of the background luminance can conceal the smallest letters but for most colors, the larger size letters remain clearly visible even in isoluminant conditions.

Figure 7b shows a nominally isoluminant text version of Shakespeare's Sonnet 18. The journal version may or may not be isoluminant on any

particular display. However, in a dynamic version of this sonnet, (<https://github.com/Lingyu-Gan/Saliency-maps-for-judgments-of-frontal-plane-distance-1-centroids-numerosity-and-letter-identity-.git>) the background intensity is continuously variable to enable search for a background intensity that maximally conceals the text. In our experience, there is no intensity that fully conceals the text, although some viewers may have to approach closer to perceive a fully readable display. Isoluminant perception, which depends on photon differences between quite similar long- and medium-wavelength receptors, requires many more photons than luminance perception which depends on the sum,

long- plus medium-wavelength receptors, versus zero.

Figure 7c shows letters of 25 different luminance levels and of 6 different sizes on a gray background. The luminance table that generated this display contains 256 gray levels, ranging from 0 (maximally black) to 255 (maximally white). The numbers in the bottom row indicate the nominal differences in luminance between letters and the reference background. The actual contrasts depend on the Gamma functions of the viewing device. Figure 7c demonstrates, as noted by Legge, Rubin, and Luebker (1987), that even low-contrast black letters are surprisingly difficult to decipher. The difficulty in perceiving low-contrast shapes also applies to the low-contrast luminance artifacts that are occasionally alleged to account for the perception of nominally isoluminant chromatic letters (e.g., Knoblauch, Arditi, & Szlyk, 1991).

Whereas the proposal that salience is an alternative mechanism for letter recognition is new, the fact that luminance is not necessary has a long history. Two examples (Legge, Parish, Luebker, & Wurm, 1990) found that reading speed increased quite similarly as a function of luminance and of color contrast and concluded, obviously, that luminance is not necessary for letter recognition. Wurm, Legge, Isenberg, and Luebker (1993) found that both normal and low vision subjects, had faster reaction times to recognize objects when the objects were displayed in color compared to the grayscale version. And Legge et al. (1990) note that large (6 degree) letters defined by isoluminant color are as easy to read as black-on-white letters when matched for just noticeable differences above threshold contrast.

Unlike the novel stimuli in the prior three experiments, letters are highly overlearned objects. We do not know the extent to which substance invariance applies to small, low contrast, or to noise obscured letters. More generally, the extent to which the overlearned brain representations of highly familiar objects such as letters and faces are as substance independent as the perceptual representations of the novel stimuli in the distance, centroid, and numerosity experiments are intriguing unanswered questions. To summarize what we do learn about letters from the demonstrations in Figures 6 and 7 is that either luminance or color information can be sufficient for the perception of letter shape and reading and therefore that a salience representation, to a certain extent, serves letter recognition.

## Discussion

### *Cue-invariant activation versus substance invariance*

Substance invariance is an instance of cue invariance observed in neurophysiological studies. The term “substance invariance” is used here because

cue invariance also applies to shape and to other invariances, each of which requires a different computation carried out in different neurons. Four of many examples possible of substance-invariant neural cue invariances and two examples of other cue invariances follow. Chaudhuri and Albright (1997) report that the majority of sampled neurons in macaque monkey cortical area V1 that respond to luminance motion also respond to texture-defined motion. However, in comparison to the responses to luminance stimuli, the responses to texture motion were weaker and in many instances not direction sensitive. On the other hand, recording from neurons in macaque inferior temporal cortex (Sary, Vogels, & Orban, 1993) found many neurons that showed very similar responses to luminance, motion, and texture-defined stimuli. Using magnetoencephalography in human adults (Okusa, Kakigi, & Osaka, 2000) found that stimuli defined by flicker, texture, and luminance activated the same localized site in the extrastriate cortex. Using functional magnetic resonance imaging in human lateral occipital complex (Grill-Specter, Kushmir, Edelman, Itzhak, & Malach, 1998) found that luminance and motion-defined object silhouettes produced similar activation. Cue invariance also applies to higher order processes. In monkey inferior temporal cortex (Schwartz, Desimone, Albright, & Gross, 1983) found neurons that were similarly activated by shape stimuli that differed in size, retinal position, and the sign of contrast. Quiroga, Reddy, Kreiman, Koch, and Fried (2005) found an even higher-order invariance in single neurons of human medial temporal lobe. They found neurons that responded similarly to strikingly different pictures of particular individuals, landmarks, or objects.

In summary, the single-neuron studies in monkeys show partial cue-invariant brain activation in occipital cortex and instances of complete invariance in temporal cortex. Human imaging by MEG shows partial cue-invariant activation in the occipital lobe, fMRI shows it in the post-V1 occipital lobe, and single neuron recordings show cue-invariant responses in temporal lobe. Although these studies observe cue invariance and believe it must be useful, nowhere is there a clear theory of how it is achieved and utilized. In the brain, achieving cue invariances appears to involve a long sequence of complex, interacting brain areas. Whereas, in a computational model (Gan et al., 2023, page 9) (Figure 7), after five clearly defined steps, substance invariance occurs when content is loaded into a salience map. That content subsequently is passed to a centroid computation but otherwise could have been passed to many other computations.<sup>6</sup>

### **Logic requires salience maps**

Although it seems effortless for us to judge the distance between two target items, it requires quite complex and extensive neural circuitry to compute the

distance between items that are arbitrarily located in the visual field. It would be overwhelmingly expensive to have a neural circuit for estimating the distance between two black disks, a different neural circuit for estimating the distance between two Gabor patches, or yet another neural circuit for estimating the distance between a Gabor patch and a black disk. The distance computation between two target items has to be made on an internal representation of the target items—a representation that depicts the existence of things at various locations—that is distinct from the representation of the things themselves. The same logic applies to centroid judgments, numerosity, and to a component of letter recognition (vs. simply letter location). These tasks were chosen to illustrate a flexible spatial representation that is not dependent on any particular feature, neither luminance nor color nor texture—just something that distinguishes the targets from the background. Saliency that is recorded in saliency maps is the name assigned to this representation, it is a quantifiable upgrade of figure ground.

## Summary and conclusions

Household substance-invariant measuring devices like measuring cups, weight scales, and measurement tapes can measure an infinity of previously unknown substances and represent the result as a real number: weight, volume, or length. A saliency map (analogous to pin art) is a representation of a real number, saliency, as a function of space and time,  $x, y, t$ . A saliency map delivers a simplified representation of the substance to be measured to the actual neural measurement processes, and a saliency map offers both an efficient representation and a unique way of dealing with potentially infinitely many different stimuli. Substance invariance (similar to physiologists' cue-invariant activation) is a way to prove that the brain uses processes that function equivalently to saliency maps to collect the material to perform a measurement. Previously, saliency maps have been proposed as a component in the mechanisms for measuring processing priority, the direction of higher-order motion, and for centroid computations, but no proofs were offered. Here, three formal substance-invariant experiments confirmed that the brain uses a saliency measurement system to measure distances in the frontal plane, to estimate the center of a cluster of items, and to estimate the number of items in a cluster, no other measurement system is needed. We demonstrate that the saliency system also can be used for determining letter identity, but it clearly is not the only brain system available for that purpose.

*Keywords:* saliency map, substance invariance, centroid, numerosity, frontal plane distance, letter identification

## Acknowledgments

The authors thank Jack Gallant, University of California, Berkeley; Michael Morgan, City University of London; and Kenneth Knoblauch, Stem-cell and Brain Research Institute, Bron, France, for helpful suggestions. Publication costs were supported by the ARVO Publication Financial Assistance Program.

Commercial relationships: none.

Corresponding author: George Sperling.

Email: sperling@uci.edu.

Address: Department of Cognitive Sciences & Department of Neurobiology and Behavior University of California, 2181 Social Science Plaza A, Irvine, CA 92617, USA.

## Footnotes

<sup>1</sup>Neurophysiologists use the term cue invariance for a related concept. See Discussion section.

<sup>2</sup>All the statistical tests reported in this study were conducted using a one-way repeated measures analysis of variance.

<sup>3</sup>For details of the ideal detector model, see figure 4 in Gan et al. (2023).

<sup>4</sup>Appendix Figures A2a and A2b show the distribution of the number of items for each experimental condition in Experiments 3a and 3b, respectively.

<sup>5</sup>Appendix Figure A1 shows three subjects' pooled numerosity estimates versus the presented number of stimulus items in each experimental condition.

<sup>6</sup>Gan et al. (2023), p. 11, figure 9.

## References

- Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception*, 45(1–2), 5–31.
- Bogler, C., Bode, S., & Haynes, J.-D. (2011). Decoding successive computational stages of saliency processing. *Current Biology*, 21(19), 1667–1671.
- Burbeck, C. A. (1987). Position and spatial frequency in large-scale localization judgments. *Vision Research*, 27(3), 417–427.
- Burbeck, C. A. (1988). Large-scale relative localization across spatial frequency channels. *Vision Research*, 28(7), 857–859.
- Chaudhuri, A., & Albright, T. D. (1997). Neuronal responses to edges defined by luminance vs.

- temporal texture in macaque area v1. *Visual Neuroscience*, 14(5), 949–962.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Routledge.
- Cook, M. (1978). The judgment of distance on a plane surface. *Perception & Psychophysics*, 23(1), 85–90.
- Gan, L., & Sperling, G. (2022). Centroid judgments are substance indifferent and therefore based on a salience map. *Journal of Vision*, 22(14), 3832–3832, doi:<https://doi.org/10.1167/jov.22.14.3832>.
- Gan, L., Sun, P., & Sperling, G. (2021). Frontal-plane distance judgments between two equal-size items are made on the basis of a salience map. *Journal of Vision*, 21(9), 2828.
- Gan, L., Sun, P., & Sperling, G. (2023). Deriving the number of salience maps an observer has from the number and quality of concurrent centroid judgments. *Proceedings of the National Academy of Sciences of the USA*, 120(21), e2301707120.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, 21(1), 191–202.
- He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review*, 16(3), 509–517.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10–12), 1489–1506.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259.
- Knoblauch, K., Arditi, A., & Szlyk, J. (1991). Effects of chromatic and luminance contrast on reading. *Journal of the Optical Society of America A*, 8(2), 428–439.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219–227.
- Legge, G. E., Parish, D. H., Luebker, A., & Wurm, L. H. (1990). Psychophysics of reading. xi. Comparing color contrast and luminance contrast. *Journal of the Optical Society of America A*, 7(10), 2002–2010.
- Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading—v. The role of contrast in normal vision. *Vision Research*, 27(7), 1165–1177.
- Lu, Z.-L., & Sperling, G. (1995). Attention-generated apparent motion. *Nature*, 377(6546), 237–239.
- Okusa, T., Kakigi, R., & Osaka, N. (2000). Cortical activity related to cue-invariant shape perception in humans. *Neuroscience*, 98(4), 615–624.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102–1107.
- Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, 10(2), 10.
- Schwartz, E. L., Desimone, R., Albright, T. D., & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences of the USA*, 80(18), 5776–5778.
- Sun, P., Chu, V., & Sperling, G. (2021). Multiple concurrent centroid judgments imply multiple within-group salience maps. *Attention, Perception, & Psychophysics*, 83(3), 934–955.
- Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). The centroid paradigm: Quantifying feature-based attention in terms of attention filters. *Attention, Perception, & Psychophysics*, 78, 474–515.
- Sáry, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, 260(5110), 995–997.
- Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human perception and performance*, 19(4), 899.

## Appendix A: Detailed materials and methods

For all experiments, all subjects had normal or corrected-to-normal visual acuity. All subjects gave informed consent for participation in the study. All methods were approved by the University of California, Irvine Institutional Review Board.

### Experiment 1a

#### Subjects

Two naive subjects FH and NA were unaware of the purpose of the study. The other subject, LG, was the junior author.

#### Apparatus

The experiment was conducted on an iMac intel computer running MATLAB with a Psychtoolbox package. A 60-Hz refresh rate CRT monitor with 1,280 × 1,024 resolution was used to display the stimuli. Stimuli were viewed at a fixed distance of 58 cm.



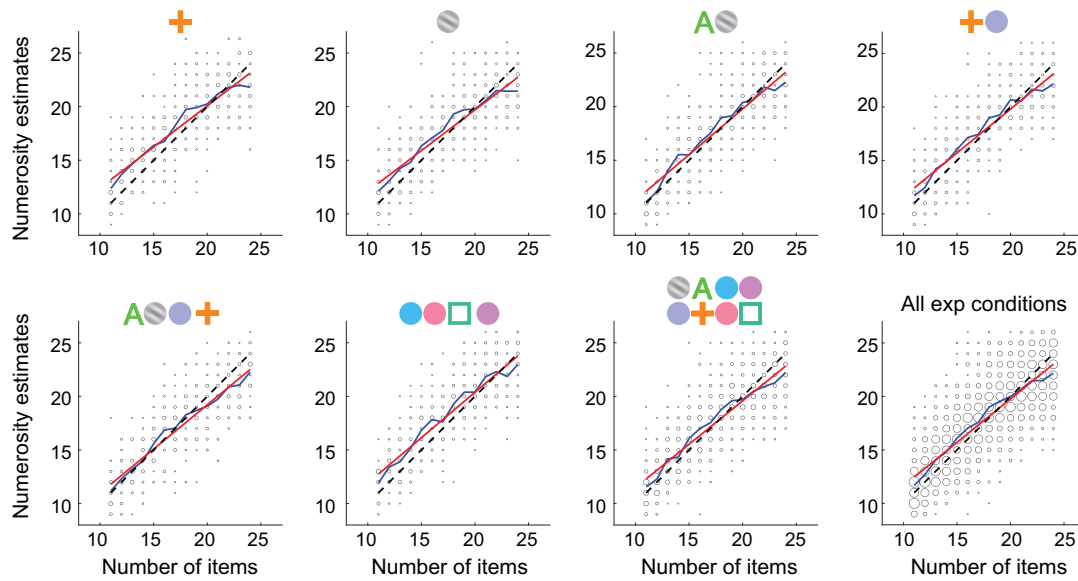


Figure A1. Experiment 3a: Three subjects' pooled numerosity estimates versus the presented number of stimulus items in each experimental condition. The size of open circles is proportional to the frequency of responses. The blue curve is the mean response for each number of items. The dashed black line is the presented number of items. The red line is the best-fitting line to the data.

### Stimuli

The stimulus display was  $720 \times 720$  pixels (visual angle  $17.25^\circ$ ) centered on a display of  $1,280 \times 1,024$  pixels. In total, 15 conditions, each with a different pair of targets, were tested. For 12 conditions, 144 same-size disks were presented in every stimulus display. The diameter of the disks was 50 pixels (visual angle  $1.29^\circ$ ). The gray levels of 142 disks were drawn from a uniform distribution of intensities  $U(-0.35, +0.35)$ , where 0 represents mid-gray, and  $-1.0, 1.0$  represent the lowest and highest intensities available on the monitor. Two disks were significantly different from the rest of the disks. Their interiors were randomly drawn from one of 15 pairs of disks described in the abscissa of Figure 3f. Two disks were targets and the other 142 disks were distracters. For the remaining three conditions, only the two targets (no distracters) were presented. Figures 3c, d, and e illustrate stimulus displays. A poststimulus masking field (Figure 3a) consisting of 144 disks whose gray levels were sampled from a uniform distribution of intensities  $U(-0.5, 0.5)$  immediately followed the stimulus exposure.

### Procedure

Figure 3a depicts the procedure of each trial. Each trial started with a blank screen with a fixation bar for 500 ms, followed by a stimulus display shown on the screen for 200 ms. A 100-ms-duration postexposure masking field immediately followed the stimulus display to strictly control the duration for which information was visually available. Then subjects were prompted to enter their estimate of the distance between the

two target disks on the keyboard. Distance estimates consisted of two integers, the digit was the number inches, the second number was the number of tenths. Subjects were allowed to change their answer before submitting it by pressing return. Feedback was provided after each trial.

### Number of trials

Each subject participated in three sessions of the experiment, one session per day. Each session was divided into a training section and a formal experiment. In the training section, to facilitate learning, the duration of the feedback was controlled by the subjects. The training section consisted of 20 trials on each of the three training conditions, at the end of which subjects were able to estimate the distance between the targets quite accurately and without demonstrating additional improvement. In the formal experiment, feedback was shown on the screen for 1 second and the intertrial interval was 1 second. The first two sessions contained 30 trials of each of the 15 conditions and the last session contained 40 trials of each of the 15 conditions. Experimental conditions were mixed in each session. In Experiment 1a, for three subjects, 37, 47, and 38 outlier trials—those beyond 2.5 times the standard deviation of the response error—were discarded from a total of 1,500 trials per subject.

### Results: Power analysis

With a desired significance level of 0.05, an observed effect size ratio ( $\frac{\sigma_{between}}{\sigma_{within}}$ ) of 0.0802, for 15 conditions,

3 subjects, and 100 trials per condition, the estimated power is 0.0870 (8.70%). To achieve a power of  $\beta = 0.80$ , an estimated 1381 trials per condition would be required (Cohen, 2013) (or an effect size of 0.2982).

## Experiment 1b

### Subjects

Subjects LG and NA from Experiment 1a, plus a new subject participated TK.

### Stimuli, procedure, and number of trials

The stimulus and the procedure of Experiment 1b was the same Experiment 1a except for the following: 1) The composition of the target pairs. In total, 6 target pairs were tested in Experiment 1b (shown in the abscissa of Figure 3h). The red and green were isoluminant to the background and were tested for each subject before the formal experiments. 2) Each subject did two sessions of the experiment, one session per day. Each session contained 50 trials of each of the six conditions. Experimental conditions were mixed in each session.

### Results: Power analysis

With a desired significance level of 0.05, an observed effect size ( $\frac{\sigma_{\text{between}}}{\sigma_{\text{within}}}$ ) of 0.0401, for all 6 conditions, 3 subjects, and 100 trials per condition, the estimated power is 0.0636 (6.36%). To achieve a power of  $\beta = 0.80$ , an estimated 4711 trials per condition would be required. or an effect size of 0.2691.

A more detailed power test was conducted to just compare conditions in which both targets were isoluminant with the condition in which both targets contained luminance information. With a desired significance level of 0.05, an observed effect size ( $\frac{\sigma_{\text{between}}}{\sigma_{\text{within}}}$ ) of 0.0349, the estimated power is 0.0668 (6.68%). To achieve a power of  $\beta = 0.80$ , an estimated 8751 trials per condition would be required (or an effect size of 0.3269).

## Experiment 2a

### Subjects

Subjects LG and NA from Experiment 1, plus a new subject PS, all were experienced in centroid tasks.

### Apparatus

This experiment was conducted on an iMac intel computer installed with MATLAB 2018b and Psychtoolbox-3 software. For LG and NA, the stimuli

were presented on an ASUS ProArt Display monitor with  $1,920 \times 1,200$  resolution at a refresh rate of 60 Hz. The monitor screen was 51.8 cm wide  $\times$  32.4 cm high. Each pixel was 0.27 mm  $\times$  0.27 mm. For PS, the stimuli were presented on a SamSung Syncmaster Display monitor with  $1,680 \times 1,050$  resolution at a refresh rate of 60 Hz. Each pixel was 0.282 mm  $\times$  0.282 mm.

### Stimuli and number of trials

The stimuli were displayed within an  $800 \times 800$  pixel-wide square (Figure 4b) that spanned 20.4 deg of visual angle (dva) for LG and NA, and 22.0 dva for PS. The stimulus display contained 32 stimulus items. Each stimulus item was inscribed inside invisible circles of 28-pixel diameter, spanning 0.72 dva, that were prohibited from overlapping. The 32 stimulus items had either the same feature or two, four, or eight different features. Items varied in color, shape, and luminance, which could be less than, equal to, or greater than the background gray level (175). In total, eight different compositions of stimulus items were tested in a mixed-list design, 50 trials per composition. Figure 4b shows sample stimuli for each of the 8 compositions.

### Procedure

Figure 4a depicts the procedure of each trial in Experiment 2a. Every trial began with a 500 ms blank field with a fixation point, followed by a 300 ms stimulus array, a 50 ms blank field, a 100-ms masking field, a blank field with a movable cursor that the subject moved to the judged centroid location, and finally, a feedback display. Feedback displayed the stimulus, the centroid of the target set as the larger gray plus sign inscribed in a gray open circle, and the subject's response as the smaller gray plus sign inscribed in a gray open circle. In Experiment 2a, For the three subjects, 7, 4, and 3 outlier trials—those beyond 2.5 times the standard deviation of the response error—were discarded from a total of 400 trials per subject. In Experiment 2b, 7, 9, and 10 outlier trials were discarded from each subject's total of 1,600 trials.

### Results: Power analysis

With a desired significance level of 0.05, an observed effect size ( $\frac{\sigma_{\text{between}}}{\sigma_{\text{within}}}$ ) of 0.1421, for 8 conditions, 3 subjects, and 50 trials per condition, the estimated power is 0.1368 (13.68%). To achieve a power of  $\beta = 0.80$ , an estimated 381 trials per condition would be required (or an effect size of 0.3901).

## Experiment 2b

### Subjects

Subjects LG and NA as in Experiment 2a, plus a new subject DH.

### Apparatus

This experiment was conducted on an iMac intel computer installed with MATLAB 2018b and Psychtoolbox-3 software. For all subjects, the stimuli were presented on the built-in Retina Display with 1280 × 800 resolution at a refresh rate of 60 Hz. The monitor screen was 30.41 cm wide × 21.24 cm high. Each pixel was 0.22 mm × 0.22 mm.

### Stimuli and number of trials

The stimuli were displayed within an 800 × 800 pixel-wide square that spanned 17.86 dva for all subjects. The stimulus display contained 144 same-size stimulus items. Each stimulus item was inscribed inside invisible circles of 28-pixel diameter, spanning 0.72 dva, that were prohibited from overlapping. Eight different combinations of targets and distracters were tested. For four conditions, each stimulus display contained 136 varied gray-level disks and 8 target disks. In each condition, one of the four colors was used to define targets, one most red that is isoluminant to the background, one most green that is isoluminant to the background, the maximally white, and the maximally black (Figures 4e, f, h, i, k, l, n, and o). For the remaining four conditions, each stimulus display contained 128 varied gray-level disks, 8 paired salient distracters, and 8 targets (isoluminant red and isoluminant green was a pair; maximally white and maximally black was a pair) (Figures 4j, i,m, and p). These eight experimental conditions were tested in a mixed list with each condition comprising 100 trials. Additionally, two different sizes of disks were examined in a blocked design, yielding a 2 × 8 experimental design. The large disks were 40 × 40 pixels and the small disks were 20 × 20 pixels.

### Procedure

Figure 4d depicts the procedure of each trial in Experiment 2b. Every trial began with a 500 ms blank field with a fixation point that had the same feature as the target, followed by a 300-ms stimulus array, a 50-ms blank field, a 100-ms masking field, a blank field with a movable cursor that the subject moved to the judged centroid location, and finally, a feedback display. Feedback displayed the stimulus, the centroid of the target set as the larger gray plus sign inscribed in a gray open circle, and the subject's response as the smaller gray plus sign inscribed in a gray open circle.

### Results: Power analysis

With a desired significance level of 0.05, an observed effect size ( $\frac{\sigma_{between}}{\sigma_{within}}$ ) of 0.1493, for 8 conditions, 3 subjects, and 200 trials per condition, the estimated power is 0.5339 (53.39%). To achieve a power of  $\beta = 0.80$ , an estimated 341 trials per condition would be required (or an effect size of 0.1953).

## Experiment 3a

### Subjects

LG plus two new subjects JR and XL.

### Apparatus

This experiment was conducted on an iMac intel computer installed with MATLAB 2018b and Psychtoolbox-3 software. For all subjects, the stimuli were presented on the Acer Predator XB321HK monitor with 3,840 × 2,160 resolution at a refresh rate of 60 Hz. Each pixel was 0.19 mm × 0.19 mm.

### Stimuli and number of trials

The stimuli were displayed within an 800 × 800 pixel-wide square that spanned 14.19 dva. The stimulus display contained 9 to 26 stimulus items. Each stimulus item was inscribed inside invisible circles of 28-pixel diameter, spanning 0.50 dva, that were prohibited from overlapping. The stimulus items had either the same feature or two, four, or eight different features. Items varied in color, shape, and luminance, which could be less than, equal to, or greater than the background gray level (175). In total, seven different compositions of stimulus items were tested in a mixed-list design. Figure 5b shows sample stimuli for each of the seven compositions. For a single session, except for the composition of 8 features that had 96 trials, all other compositions had 48 trials. All compositions have the same distribution of the number of stimulus items. Each subject completed four sessions, conducted on separate days.

### Procedure

On each trial, subjects were instructed to estimate the numerosity of all stimulus items. Figure 5a depicts the procedure of each trial. Every trial began with a 500-ms blank field with a fixation point, followed by a 300-ms stimulus array, a 50-ms blank field, a 100-ms masking field, and a blank field with prompt cueing subjects to enter their estimate of the numerosity of all stimulus items on the keyboard, and finally, a feedback display. Feedback displayed the stimulus, the correct answer, and the subject's response.

### Results: Power analysis

With a desired significance level of 0.05, an observed effect size ( $\frac{\sigma_{between}}{\sigma_{within}}$ ) of 0.0522, for 7 conditions, 3 subjects, and an average 220 trials per condition, the estimated power is 0.1011 (10.11%). To achieve a power of  $\beta = 0.80$ , an estimated 2721 trials per condition would be required (or an effect size of 0.1842).

## Experiment 3b

### Subjects

Same subjects as in Experiment 2b.

### Apparatus

Same as Experiment 2b.

### Stimuli and number of trials

The stimuli in Experiment 3b were identical to Experiment 2b except the following: 1) the number

of items per feature was randomly drawn from a discrete uniform distribution between 5 and 13; 2) only the large size stimulus elements were tested; 3) the eight experimental conditions were tested in a block design, with each condition comprising 102 trials. The order of conditions was randomized for each subject.

### Procedure

The procedure was identical to Experiment 2b except the task was to estimate the numerosity of target stimulus items.

### Results: Power analysis

With a desired significance level of 0.05, an observed effect size ( $\frac{\sigma_{between}}{\sigma_{within}}$ ) of 0.0887, for 8 conditions, 3 subjects, and 102 trials per condition, the estimated power is 0.1165 (11.65%). To achieve a power of  $\beta = 0.80$ , an estimated 971 trials per condition would be required (or an effect size of 0.2735).

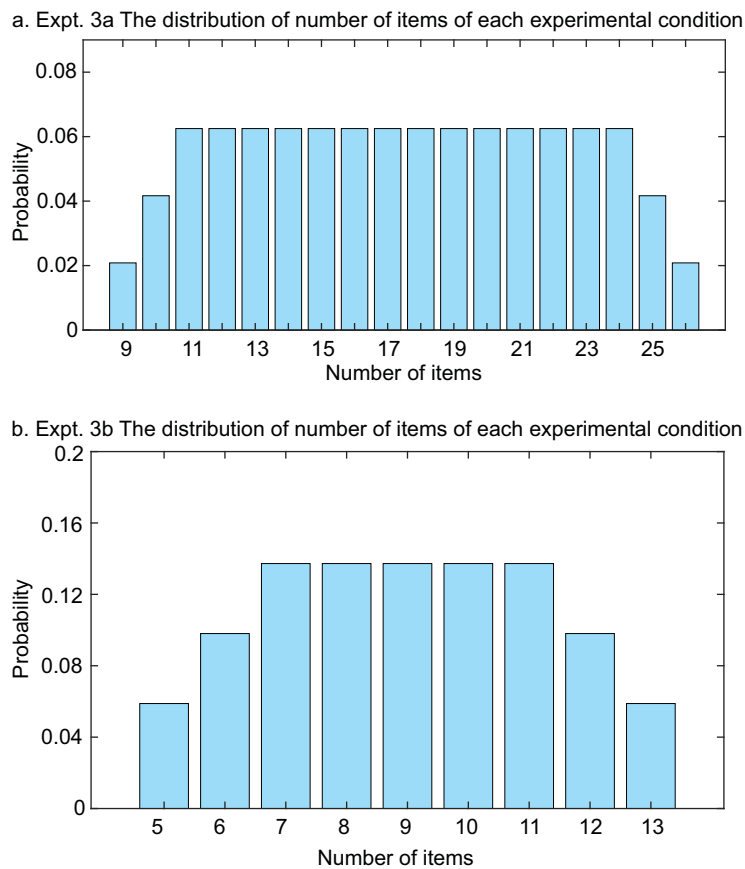


Figure A2. Numerosity estimation. (a) The distribution of the number of items of each experimental condition, with identical distributions across all conditions in Experiment 3a. (b) The distribution of the number of items of each experimental condition, with identical distributions across all conditions in Experiment 3b.