Journal of Experimental Psychology:
Human Perception and Performance
2017, Vol. 43, No. 9, 1667–1676

0096-1523/17/$12.00 http://dx.doi.org/10.1037/xhp0000314

Redundant Encoding Strengthens Segmentation and Grouping in Visual Displays of Data

Christine Nothelfer
Northwestern University

Michael Gleicher
University of Wisconsin—Madison

Steven Franconeri
Northwestern University

The availability and importance of data are accelerating, and our visual system is a critical tool for understanding it. The research field of data visualization seeks design guidelines—often inspired by perceptual psychology—for more efficient visual data analysis. We evaluated a common guideline: When presenting multiple sets of values to a viewer, those sets should be distinguished not just by a single feature, such as color, but redundantly by multiple features, such as color and shape. Despite the broad use of this practice across maps and graphs, it may carry costs, and there is no direct evidence for a benefit. We show that this practice can indeed yield a large benefit for rapidly segmenting objects within a dense display (Experiments 1 and 2), and strengthening visual grouping of display elements (Experiment 3). We predict situations where this benefit might be present, and discuss implications for models of attentional control.

Public Significance Statement

This study demonstrates that we can more efficiently pay attention to a collection of objects when they differ from other (irrelevant) objects within multiple feature dimensions, such as color and shape, than when they differ by only one feature, such as only color or shape. This result applies broadly to how we attend to objects in our daily environment—it is much more typical that an object will differ from its surrounding objects in multiple feature dimensions than a single feature dimension. These results also apply directly to a common data visualization design technique called *redundant coding*, which differentiates groups of data points by multiple features, such as a scatterplot with red triangles, blue circles, and green squares.

Keywords: visual attention, feature-based attention, data visualization, grouping, segmentation

Supplemental materials: http://dx.doi.org/10.1037/xhp0000314.supp

The world is noisy. To extract or communicate a signal, we often need to integrate multiple sources of information. Airline pilots replace individual letters with words, like Foxtrot, Romeo, or Tango, to introduce redundant information about letters that they want to communicate when transmitting voice messages. Drummers in sub-Saharan Africa use a similar system when sending messages, by relying on sets of familiar "chunks" of pattern that allow noisy messages to be recovered across long distances (Gleick, 2011). Most packets of information sent across a network add additional bits that help detect, or even correct, introduced errors. But while such redundant encoding can strengthen signal among noise, in low-noise environments it might be inefficient or even distracting.

Here we test for potential benefits of redundant encoding in the visual analysis of data. The human visual system is well positioned for data analysis, because its parallel architecture allows broad processing of information and computation of elementary statistics, such as means, maxima, and distributions (Szafir, Haroz, Gleicher, & Franconeri, 2016; Haberman & Whitney, 2012), across data values encoded by visual dimensions, such as position, color, or shape (Munzner, 2014). But like other information processors, the visual system faces noisy representations. Notably, when performing visual statistics on certain sets of values—de-

This article was published Online First April 13, 2017.

Christine Nothelfer, Department of Psychology, Northwestern University; Michael Gleicher, Department of Computer Sciences, University of Wisconsin—Madison; Steven Franconeri, Department of Psychology, Northwestern University.

We thank Satoru Suzuki and Paul Reber for helpful discussion. We are grateful to Zoe Listernick, Kayla Mines, and Kevin Hartstein for assistance in study design and data collection. This work was supported by IIS-1162067, IIS-1162037, BCS-1056730, and the National Science Foundation Graduate Research Fellowship Program.

Correspondence concerning this article should be addressed to Christine Nothelfer, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208. E-mail: cnothelfer@u.northwestern.edu

fined by being red, or circular—isolating the relevant points becomes increasingly difficult in more complex displays (Duncan & Humphreys, 1989).

A common strategy for improving signal among noise is to use redundant features (e.g., red and circular) to encode sets of values. Figure 1 illustrates this practice across visualizations intended for various audiences, such as researchers (Figure 1a), consumers (Figure 1b), and the general public (Figure 1c). Figures 1d–1f present similar examples from psychology articles of the past 3 years, showing the use of redundant encoding across depictions of data with widely varying complexity. Redundant encoding is the default setting for the construction of graphs in Microsoft Excel (Figure 1a), a core part of a software package used by over 1.1 billion people (Microsoft, 2014). Across these examples, redundant encoding might be beneficial—it might help further perceptually segregate different collections, help link legends to data, or enhance memory for relationships.

But redundant encoding also might convey little or no benefit, with the risk of increasing display complexity. Observers might be left confused about which dimension is relevant when linking legends to data, or whether the independent dimensions reflect different aspects of the data. Visual designers strive to strip away unnecessary variation in visual displays, which can lead to confusion and an inelegant appearance (Williams, 2014). In the data visualization literature, influential voices argue that elegant and understandable data presentations should omit unnecessary embellishment as much as possible (Few, 2012; Tufte, 1983), and that redundancy can occasionally be helpful under specific conditions, but is often gratuitous (Tufte, 1990).

Here we investigate whether redundant encoding can confer a benefit when one simultaneously attends to a set of objects. There are a number of related findings that suggest it could be beneficial in this case. Some studies show that classifying a single object is faster or more accurate when redundant information is available. When people are asked to make a speeded key press to indicate whether they are presented with at least one of two possible targets (e.g., Please press a key if you see an asterisk or hear a tone), they are faster when both targets appear (Miller, 1982). When people are asked to classify an object's size, color, or position into a set of predefined magnitude categories (e.g., the "second biggest" type), performance is better when categories can be judged by redundant information from multiple dimensions (Eriksen & Hake, 1955; Lockhead, 1966; Egeth & Pachella, 1969). There are similar redundancy benefits when participants sort values into a dimensional ordering, even when they are instructed to sort along a single dimension (Morton, 1969; Garner, 1969; Biederman & Checkosky, 1970).

While these examples are cited in data visualization textbooks as the best available argument for the benefits of redundant encoding (e.g., Ware, 2013), these tasks do not reflect the demands of judging collections of objects, as is often the case in visual data displays. Previous work requires precise categorization of the value of a single stimulus along a dimension (e.g., Is this the second reddest?), amid closely spaced alternative values (e.g., there might be another possible red with a touch of orange). In contrast, we do not know whether a redundancy benefit would extend to visual data displays requiring selection of the value of one collection of objects (e.g., Pick out the bright ones), with widely spaced alternative values (e.g., red, green, or blue).

Another set of related findings comes from the visual search literature, showing that redundant encoding of target identity can help participants find single targets more quickly. Visual searches

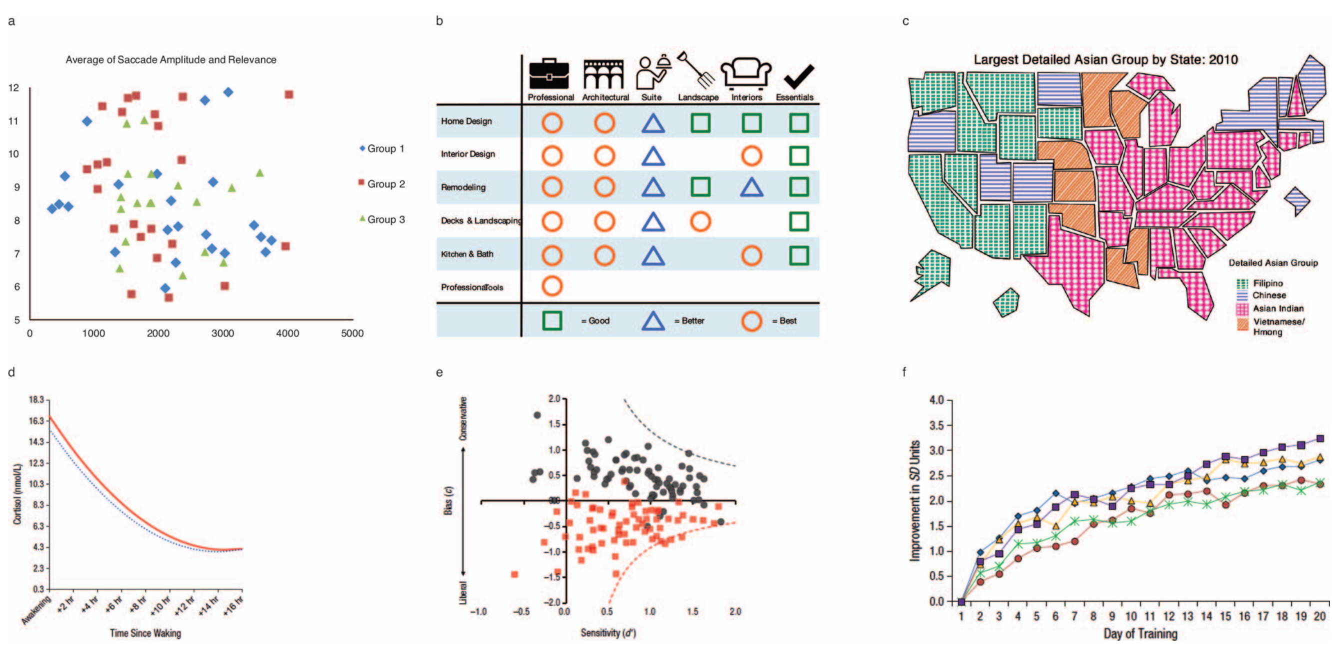

Figure 1. Top row: Redundant encoding is ubiquitous across chart types, such as (a) scatterplots, (b) tables, and (c) choropleth maps. Bottom row: Redundant encoding examples from recent articles in psychology journals. This technique has been used in (d) simple, (e) moderately complex, and (f) dense displays. Examples are taken from (d) Slatcher, Selcuk, & Ong (2015), (e) Lynn and Barrett (2014), and (f) Harrison et al. (2013). See the online article for the color version of this figure.

are faster when pop-out targets are redundantly encoded by color and shape (e.g., Find the red diamond among green squares) than when they are coded by only one dimension (e.g., Find the red square, or the green diamond, among green squares; Krummenacher, Müller, & Heller, 2001). Furthermore, searching for triple conjunctions (e.g., Find a single small red *X* among stimuli that otherwise vary in size, color, and shape) can in some cases be easier than finding double conjunctions of the same features (Wolfe, Cave, & Franzel, 1989), though these particular encodings are not redundant, because each dimension carries additional information about target status. But it is again unclear that these findings reflect the demands of perceiving sets of objects. Critically, both these studies and the previous set of categorization studies require participants to categorize or locate single stimuli. While this might reflect the demands of a subset of tasks (e.g., finding a data point that fits certain criteria), data displays often require observers to select an entire collection of objects in order to segment one data category from another (e.g., What is the shape of the collection of red values? What is its distribution? What is the general location of the collection?). In fact, this type of segmentation of collections is argued to follow different rules compared to parallel visual for single targets (Wolfe, 1992).

These more holistic judgments apply to a wide variety of data displays, such as scatterplots (Figure 1a), choropleth maps (Figure 1c), or matrices of correlations. Such holistic judgments are likely supported by feature-based attention. Color, luminance contrast, shape, orientation, and motion direction are broadly processed across the visual field (Treisman & Gormican, 1988), and the visual system can in many cases selectively filter information from one value along these dimensions (Sàenz, Buracâs, & Boynton, 2002, 2003). In Figure 1, a viewer can estimate the center point of Group 1 by selecting blue (Figure 1a), or the distribution of software features in a table (Figure 1b) or populations on a map (Figure 1c).

Feature-based attention spreads widely across the visual field (Sàenz et al., 2002; but see Leonard, Balestreri, & Luck, 2015), and can amplify a given visual feature within the first 100 ms of the appearance of a display (Zhang & Luck, 2009). Can feature-based attention select two values along different dimensions at the same time? Some results show that second dimensions that are irrelevant, or even interfering, are nonetheless selected. When participants are asked to segment two collections of symbols that differ in one dimension (e.g., color), irrelevant differences along another dimension (e.g., shape) can interfere with performance (Callaghan, 1984, 1989). Another study used brain imaging to show that when participants attended to one of two superimposed fields of dots that differed by task-relevant (color) and task-irrelevant (motion direction) dimensions, there was greater activity to an unattended dot group in the opposite visual field when its motion direction matched the task-irrelevant second dimension within the attended field, suggesting that it was selected anyway (Lustig & Beck, 2012). These results suggest that feature-based attention is capable of selecting two values from two dimensions at once. But does selection of multiple dimensions actually help in this context—can it help when inspecting a collection of objects?

One surprising recent result from the information visualization literature suggests that, in the context of a simulated real-world task, redundant encoding offers no advantage whatsoever (Gleicher, Correll, Nothelfer, & Franconeri, 2013). Participants were shown scatterplots containing two "point clouds" of data (similar to Figure 1a, but with 50 points per collection), and were asked to judge which data group had the higher average. Performance was no different when judging collections that differed by color alone (orange vs. purple), or shape alone (circles vs. triangles), compared to judging collections that were redundantly encoded (orange circles vs. purple triangles). However, there are a number of reasons for why this study may have failed to find a redundancy benefit (see Conclusion).

In summary, no existing study has demonstrated a benefit from redundant encoding of a collection of objects, as is often the case in real-world displays. There is evidence from the dimensional categorization and visual search literatures that redundancy can be helpful in some visual tasks, but those tasks differ from the present ones in critical ways. While some work shows that the visual system is capable of selecting two values in two dimensions at once, one recent study found no benefit for redundant encoding in a simulated data display task. In Experiments 1 and 2, we tested for a benefit of redundant encoding in a new type of real-world display meant to simulate the requirements of interpreting a large class of visual representations of data. In Experiment 3, we tested for a similar benefit in an established test of visual grouping strength.

Experiment 1

Data visualizations often require the observer to judge the shape of the distribution of a collection, whether they are points in a graph, values in a chart, or glyphs on a map (see Figure 1). Where are the outliers, clumps, and regions of greater or lower concentration? We constructed an abstracted task designed to emulate such judgments, requiring the participant to select a collection of objects holistically in order to judge its shape envelope of the collection (see Figure 2a), by reporting the quadrant of the display that was missing elements of a given color or shape. Three different sets of redundant visual features (Experiment 1a: blue or asterisk; Experiment 1b: blue or circle; Experiment 1c: red or triangle) were used to test for generalizability.

Because we were interested in whether redundant encoding improves performance for "in-a-glance" decisions—as opposed to slow and serial inspection over the course of several seconds—we use a brief presentation time (around 90 ms on average; see below). Despite using a brief presentation, we simulated the experience a viewer should have from previous experience with a specific display (including knowledge of the relevant and irrelevant features within it) by showing a preview screen depicting the objects to be judged, and ignored, before every trial, which should improve overall performance (Wang, Cavanagh, & Green, 1994).

Method

Participants. We recruited 44 Northwestern University students and community members (ages 18 to 28; Experiment 1a: 12 subjects (one was author C.N.); Experiment 1b: 16 subjects; Experiment 1c: same 16 subjects from Experiment 1b) in exchange for course credit or payment.

Stimuli. Stimuli were created using MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997), and presented on a 32 × 24 cm CRT monitor (75-Hz refresh rate, 1,024 × 768 resolution). All visual angles were

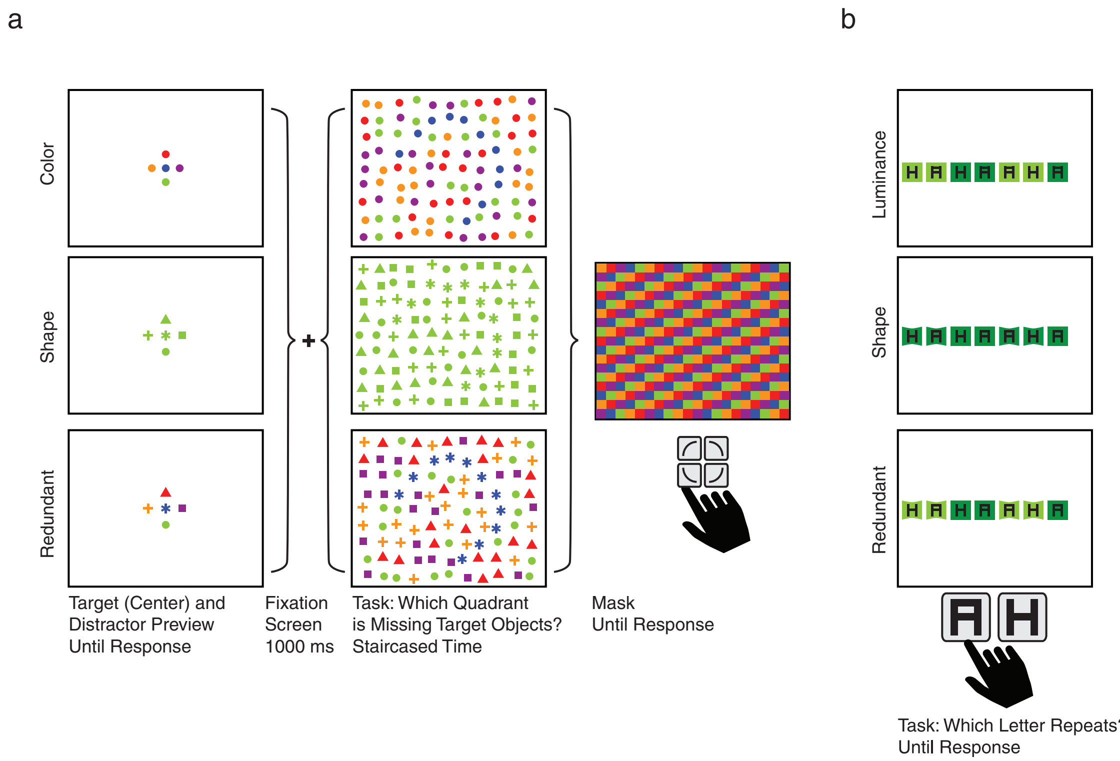

Figure 2. (a) Experiment 1's design. Stimuli shown here are from Experiment 1a and are not drawn to scale. Participants saw a preview screen (left column; until response), followed by a fixation cross (1,000 ms), and test display (center column; staircased display time). Trials concluded with a mask screen (right column) until participants indicated which quadrant was missing from the ring of target objects (all trials shown here; correct answer: bottom left). Target objects differed from distractors either by color (top), shape (center), or color and shape redundantly (bottom). (b) In Experiment 3, participants first viewed a fixation screen, followed by the test display until response. Displays contained objects pairs which differed by luminance (top), shape (center), or both luminance and shape (*redundant*; bottom). See the online article for the color version of this figure.

calculated assuming a typical distance of 40 cm from the monitor. Ninety-nine objects were arranged across a medium gray screen in a 9×11 grid with square cells 3 visual degrees in diameter, centered on a black fixation cross (Figure 2a). Each object spanned 1.0–1.5 visual degrees in diameter. Each object's x and y coordinates were jittered by ± 0.6 visual degree for each trial. Eleven targets formed a partial ring embedded among 88 distractor objects. The ring was always missing five adjacent target elements, restricted to one quadrant of the screen and replaced with randomly picked (without replacement) objects from the set of available distractors. Target objects were always presented in the same location (prior to jittering) for a given missing quadrant trial type (e.g., the location of the targets in the top-left-quadrant-missing *color* trials was the same as that in *redundant* trials where the same quadrant is missing). Objects in Experiments 1a–1c were orange, red, purple, blue, and green. Colors were approximately perceptually equiluminant, as determined by a separate experiment (see the online supplemental materials). Experiment 1a used plus signs, triangles, squares, asterisks, and circles, whereas Experiments 1b and 1c used only triangles, squares, and circles.

Target objects were identical to each other, and differed from distractors in color only (*color* trials), shape only (*shape* trials), or in both color and shape dimensions (*redundant* trials). Targets were always the same color, shape, or both through the entire experiment (Experiment 1a: blue or asterisk; Experiment 1b: blue or circle; Experiment 1c: red or triangle; e.g., targets were always blue in *color* trials, asterisks in *shape* trials, and blue asterisks in *redundant* trials in Experiment 1a). Distractors in *color* and *shape* trials consisted of every remaining feature value in the relevant feature dimension, and were identical in the irrelevant feature (e.g., a *color* trial with blue circle targets would have orange, red, purple, and green circle distractors; a *shape* trial with red asterisk targets would have red triangle, square, circle, and plus sign distractors). *Redundant* trials in Experiment 1a used unique color–shape pairs for all distractors (blue asterisk targets among orange plus sign, red triangle, purple square, and green circle distractors). Because Experiments 1b and 1c used fewer shapes than colors, shape and color were randomly and independently assigned to each distractor in *redundant* trials (e.g., Experiment 1b presented blue circle targets among orange triangles, orange squares, red triangles, red squares, purple triangles, purple squares, green triangles, and green squares), but the total number of shapes and colors used on these trials remained consistent. Experiments 1b and 1c used fewer shapes because pilot experiments revealed that if participants attend to circular and triangular targets, respectively, there need to be fewer distractor shapes (less heterogeneity) in order for performance to be above chance.

Preview screens featured the target object for the given trial at the center of a medium gray screen, beneath a black fixation cross, and surrounded by the subsequent distractor objects on an imaginary circle. The mask screen was a grid of 52 × 45 adjacent repeating orange, red, purple, green, blue, and green rectangles that filled the screen. The fixation screen consisted of a black cross at the center of a medium gray screen.

Procedure. Participants viewed the preview screen and responded with the space bar to continue after viewing that trial's target object. A fixation screen appeared for 1,000 ms, followed by the stimulus display for a variable amount of time, and the mask screen until response (the *1*, *2*, *4*, and *5* keys on a number pad covered with stickers showing the appropriate portion of a circle in the corresponding key location—e.g., the bottom left quadrant of a circle was placed on the bottom left [*1*] key). Participants were instructed to indicate the quadrant of the screen where the target object ring was missing elements. To encourage simultaneous visual selection of target objects, participants were asked to attend to all of the targets at once rather than attempting to check each quadrant serially for missing targets, because the stimulus display would flash only briefly. Participants were told to fixate through the entire trial after studying the preview screen until they saw the mask. The trial concluded with a blank medium gray screen presented for 200 ms after response.

Factors in the fully crossed design included: feature condition (*color, shape, redundant*), irrelevant features for *color* and *shape* trials (*color* trials used objects of all the same shape from the set of five [Experiment 1a] or three [Experiments 1b and 1c] possible shapes; *shape* trials used objects of all the same color from a set of five possible colors), and gap condition (gap in the target ring appeared in the top left, top right, bottom left, or bottom right quadrant). Because *color* and *shape* trials needed to display every possible irrelevant feature, these conditions had 5 times more types of unique trials than *redundant* trials in Experiment 1a. In light of this, *redundant* trials were repeated more often to maintain the number of trials within each feature condition (120 trials each). All possible *color* and *shape* trials were repeated six times (five possible irrelevant features, four gap conditions, six repetitions, yielding 120 trials per condition) while *redundant* trials were repeated 30 times (four gap conditions, 30 repetitions, yielding 120 trials), resulting in a total of 360 trials. The results were the same when examining the first half of the trials within each feature condition, so Experiments 1b and 1c each had a total of 180 experiment trials. Because these experiments had only three possible shapes, *color* trials were repeated five times (three possible irrelevant shapes, four gap conditions, and five repetitions yielded 60 trials), *shape* trials were repeated three times (five possible irrelevant colors, four gap conditions, and three repetitions yielded 60 trials), and *redundant* trials were repeated 15 times (four gap conditions and 15 repetitions yielded 60 trials).

Participants first completed 12 unrecorded practice trials in which the stimulus was presented for 200 ms. This was followed by 36 calibration trials (extra trials, randomly selected from the set of test trials) in which the display time of the stimulus (starting at 200 ms) was increased by 8 ms after incorrect answers or decreased by 4 ms for correct answers. This ratio allowed display time to staircase, automatically producing performance halfway between chance (25%) and ceiling (100%). Calibration trials were excluded from analysis unless otherwise noted. For the remaining test trials, display time was instead increased by 2 ms after incorrect answers, or decreased by 1 ms for correct answers. Averaged across the three experiments, mean display time was 89 ms (SD = 32 ms), measuring from the last 50 trials of each participant. Trials were randomly ordered within each block (practice, calibration, test trials).

Results and Discussion

Some participants were removed from the analysis due to an average display time (including calibration trials) greater than 200 ms (the starting staircase time) or because the standard deviation of the final 100 trials' display time exceeded 20 ms (Experiment 1a: one removed; Experiment 1b: three removed; Experiment 1c: two removed). One additional participant was removed from Experiments 1b and 1c due to an inability to remain alert throughout the experiment.

Figure 3 shows accuracy results for Experiments 1a–1c. If attending to objects encoded by multiple dimensions yields better visual selection and subsequent global shape detection, then participants should be most accurate in the *redundant* condition. Indeed, accuracy was highest for *redundant* trials (Experiment 1a: $M = 92.3\%$, $SD = 4.7\%$; Experiment 1b: $M = 86.5\%$, $SD = 4.3\%$; Experiment 1c: $M = 84.5\%$, $SD = 6.4\%$). Accuracy values were submitted to a repeated-measures analysis of variance (ANOVA; degrees of freedom were Greenhouse-Geisser corrected for sphericity violations), revealing a main effect of feature condition, Experiment 1a: $F(1.12, 11.16) = 26.64$, $p < .001$, $\eta_p^2 = 0.73$; Experiment 1b: $F(1.31, 14.42) = 40.88$, $p < .001$, $\eta_p^2 = 0.79$; Experiment 1c: $F(1.08, 13.00) = 12.54$, $p = .003$, $\eta_p^2 = 0.51$. *Redundant* accuracy was significantly higher than whichever condition—*color* or *shape*—was better for each participant (average accuracy for participants' best condition [*color* or *shape*]— Experiment 1a: $M = 71.4\%$, $SD = 2.2\%$; Experiment 1b: $M = 71.7\%$, $SD = 7.3\%$; Experiment 1c: $M = 71.4\%$, $SD = 8.9\%$), as confirmed by two-tailed t tests, Experiment 1a: $t(10) = 11.74$, $p < .001$, $d = 3.54$; Experiment 1b: $t(11) = 5.50$, $p < .001$, $d = 1.59$; Experiment 1c: $t(12) = 5.50$, $p < .001$, $d = 1.52$. Thus, visual selection benefited from objects encoded by multiple, redundant features than by either feature alone. See the online supplemental materials for additional analyses.

There are two possible models for how the present redundancy benefit operates. According to a combination model, information from both color and shape dimensions of the redundant targets contribute activation toward a participant's response. Alternatively, a race model specifies that color and shape dimensions of redundant targets provide independent sources of information that are never to be combined, and whichever is detected first contributes toward the participant's response on any given trial. Related redundancy gain work has discussed this issue extensively, particularly examining response time distributions rather than response time means (e.g., Miller, 1982; Mordkoff & Yantis, 1993; see Townsend, 1990, for a review of approaches disentangling the two models).

Within a race model, if our participants are unaware of which feature is more reliable, they should make their decisions based on an arbitrarily chosen feature. If this is the case, accuracy in the

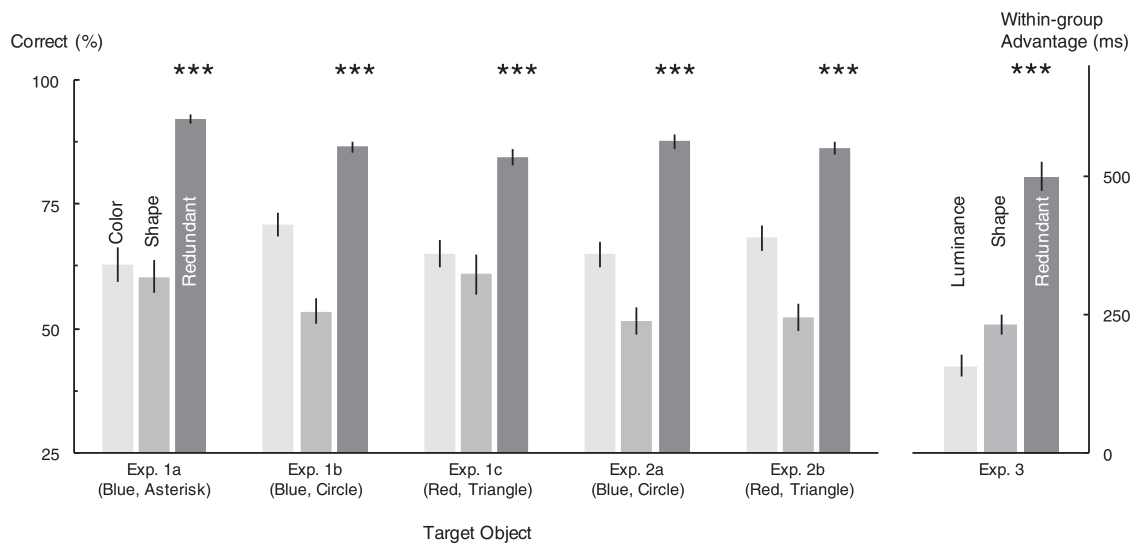

Figure 3. Experiments 1–3 results. The graph shows accuracy in each of three conditions across five experiments (Experiments 1a–2b), and *within-group* response time advantage (*between-groups* response times minus *within-group* response times) for Experiment 3). Error bars indicate within-subject standard errors of the mean, and symbols above the error bars refer to the p value of two-tailed t tests across observers: $^{***}\ p < .001$.

redundant condition —if the two features are not integrated—should range from p(s) (if the participant always chooses shape), [p(s) + p(c)]/2 (if they randomly choose either feature from trial to trial), to p(c) (if the participant always chooses color). Conversely, if participants know which feature is more reliable, they should always choose that feature in making their decisions. In this case, the accuracy in the *redundant* condition should be equivalent to the greater accuracy of the two features, that is, whichever is greater between p(s) and p(c). Because the actual accuracy in the *redundant* condition is significantly larger than either of these estimates based on separate processing of the two features, the result suggests that shape and color are integrated.

Consistent with this result, Grubert, Krummenacher, & Eimer (2011) have shown that the redundancy advantage arises even at early stages of attentional allocation, as demonstrated by an earlier N2pc onset to redundant versus single dimension trials in a pop-out visual search task. Furthermore, Krummenacher et al. (2001) showed that response times for redundant pop-out targets support a combination model (though only when trials are blocked by pop-out feature, attributing this to single-feature trials (e.g., a *color* trial) attracting weight away from the feature map of the other single-feature (e.g., *orientation*) on a subsequent *redundant* trial, resulting in a weaker synergy effect of the two features).

Experiment 2

Experiment 1 provided participants with a preview of the target object and distractor objects, which allowed participants to infer how many features would distinguish targets from distractors in the subsequent screen. For example, in the *color* trial preview depicted in Figure 2a, participants could have determined that they only need to attend to the color blue, since both target and distractor objects are circles. Thus, participants could have prepared to attend to only one feature in *color* and *shape* trials, while preparing to attend to both color and shape in *redundant* trials. To ensure that differences in these preparation strategies cannot account for the redundant encoding advantage, the preview specified the color and shape of the target, omitting descriptions of the distractor objects for the upcoming trial so that participants would not know if an upcoming display would contain a redundant encoding of the target. In addition, to test whether redundant encoding can control feature-based attention in a rapid presentation where the feature had not already been visually primed (see Zhang & Luck, 2009, for a test of this idea using single-feature dimensions), the target was described not by an image, but by printed text that identified a single shape and color (e.g., "blue circle").

Method

Participants. We recruited 31 Northwestern University students and community members (ages 19 to 31; Experiment 2a: 15 subjects; Experiment 2b: 14 subjects from Experiment 2a, plus 2 more subjects) in exchange for course credit or payment.

Stimuli. Stimuli in Experiments 2a and 2b were the exact same as those in Experiment 1b (target object: blue/circle) and Experiment 1c (target object: red/triangle), respectively, except for the preview screen. Preview screens featured the color and shape of target object, written out, for the given trial (e.g., "blue circle"). The black, lowercase text appeared on a single line at the center of a medium gray screen, spanning 4.3–6.8 visual degrees wide and 1.1 visual degrees tall.

Procedure. The procedure was the exact same as in Experiment 1.

Results and Discussion

One participant from Experiment 2a was removed from the analysis due to an average display time (including calibration trials) greater than 200 ms (the starting staircase time). No participant's standard deviation of the final 100 trials' display time exceeded 20 ms.

Figure 3 shows accuracy results for Experiments 2a–2b. As with Experiment 1, if attending to objects encoded by multiple dimensions yields better visual selection and subsequent global shape detection, then participants should be most accurate in the *redundant* condition. Indeed, accuracy was highest for *redundant* trials (Experiment 2a: $M = 87.6\%$, $SD = 6.1\%$; Experiment 2b: $M = 86.3\%$, $SD = 6.3\%$). Accuracy values were submitted to an ANOVA, revealing a main effect of feature condition, Experiment 2a: $F(1.43,18.56) = 40.99$, $p < .001$, $\eta_p^2 = 0.76$; Experiment 2b: $F(1.35,20.24) = 37.99$, $p < .001$, $\eta_p^2 = 0.72$. Specifically, *redundant* accuracy was significantly higher than whichever condition—*color* or *shape*—was better for each participant (average accuracy for participants' best condition (*color* or *shape*)—Experiment 2a: $M = 67.1\%$, $SD = 8.0\%$; Experiment 2b: $M = 71.0\%$, $SD = 7.8\%$), as indicated by a two-tailed t test, Experiment 2a: $t(13) = 8.65$, $p < .001$, $d = 2.31$; Experiment 2b: $t(15) = 6.01$, $p < .001$, $d = 1.50$. See the online supplemental materials for additional analyses.

Visual selection benefited from attending to objects encoded by multiple, redundant features than by either feature alone. Participants approached each trial with the same type of knowledge due to the nature of the preview, and thus should have approached *color*, *shape*, and *redundant* trials with the same strategy. Because a redundancy benefit arose even when participants did not know whether which feature—or both—would be useful in distinguishing the target collection from the distractors, these data rule out any pre-trial differences in strategy as the root cause of the redundancy advantage. This is also apparent because performance on *color* and *shape* trials would have been lower if participants attended to only one of the features presented in the preview (e.g., attending to only to only color would lead to worse performance on *shape* trials); single-dimension trial accuracies here are similar to those in Experiment 1. Participants' average display time of the last 50 trials (Experiment 2a: $M = 104$ ms, $SD = 33$ ms; Experiment 2b: $M = 97$ ms, $SD = 31$ ms) were similar to those in Experiment 1 (Experiment 1b: $M = 96$ ms, $SD = 23$ ms; Experiment 1c: $M = 78$ ms, $SD = 28$ ms).

Experiment 3

Experiments 1 and 2 showed that participants were better able to perceive objects' global shape when encoded by redundant features than either feature alone. Experiment 3 explored whether this benefit generalizes to other tasks. We used the repetition discrimination task (Palmer & Beck, 2007), which assesses the strength of a grouping cue. Specifically, we tested grouping by luminance similarity, shape similarity, and luminance combined with shape similarity. This task differed from that in Experiments 1 and 2 in several key ways: (a) a longer time scale (hundreds of milliseconds), (b) a different dependent measure (response time), (c) continuously, versus discretely, different features, and (d) a target that is not directly tied to the presence of the redundant features. We expected that grouping features would combine such that redundant encoding would produce a stronger effect than either grouping feature alone in this task, whether demonstrated by interference (*between-groups* trials) or stronger grouping (*within-group* trials).

Method

Participants. We recruited 17 Northwestern University students and community members (ages 18 to 29) in exchange for course credit or payment.

Stimuli. Each display consisted of seven objects, each 2.9 visual degrees wide, arranged horizontally with 0.7 visual degree between each object on a black screen (Figure 2b). There were always three or four objects to the left of the center of the screen and four or three objects to the right, respectively. *Luminance* trial objects were the same shape but one of two luminances (light green and dark green), *shape* trial objects were the same luminance but one of two shapes (square and square with a concave top and bottom, which will be referred to as *curved squares*), and *redundant* trial objects were one of two luminance–shape combinations (light green square and dark green curved square, or dark green square and light green curved square). Pairs of adjacent objects were identical in the variable feature (e.g., luminance in *luminance* trials), which alternated between object pairs (e.g., two light green squares, followed by two dark green squares, followed by two light green squares). The outermost object on the side of the screen with four objects was always unpaired. On top of each shape was a black *H* or *A*, always white-outlined to maintain contrast despite changes in object luminance across trials. *H*s and *A*s differed only in that *A*s had an additional connecting line across the top. The letters alternated, except there was always one repetition of one of the letters (e.g., H A H H A H A), which occurred in one of any six possible locations within a display. Critically, as shown in Figure 4a, the repeated letters were either on a pair of matching features (e.g., *H* on a dark green square next to another) or between adjacent pairs (e.g., *H* on a dark green square next to an *H* on a light green square). The fixation display consisted of a white cross at the center of a black screen.

Procedure. The factors included similarity grouping cue (*luminance*, *shape*, or *redundant* combination of luminance and shape), object arrangement (three objects left, four objects right of screen center; four objects left, three objects right of screen center), irrelevant dimension for all objects (squares or curved squares for *luminance* trials; dark green or light green for *shape* trials; not applicable for *redundant* trials), relevant feature of leftmost object (light green or dark green for *luminance* trials; square or curved square for *shape* trials; light green square or dark green curved square or dark green square or light green curved square for *redundant* trials), repeating letter (*A* or *H*), and location of repeating letter (object positions 1–2, 2–3, 3–4, 4–5, 5–6, or 6–7). There were 288 trials, half of which contained the letter repetition within a pair of matching objects (*within-group* trials), half between adjacent pairs (*between-groups* trials).

Participants indicated which letter, *A* or *H*, repeated, by using the *V* and *B* keys covered with *A* and *H* stickers, respectively. They were asked to respond as quickly as possible while keeping their error rate under 5%. After participants viewed a fixation screen for 500 ms, the test display was presented until response. After 24 unrecorded practice trials, test trials were done in eight blocks of 36 trials, lasting 10–15 min.

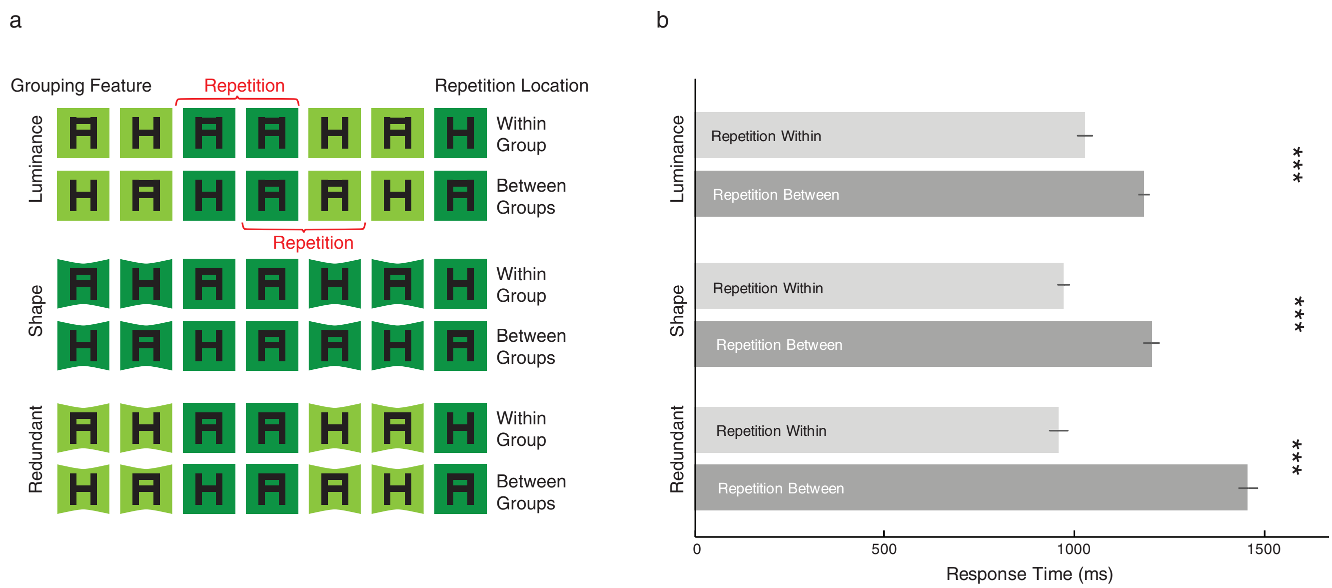

Figure 4. (a) Example of stimuli for Experiment 3 (not drawn to scale). Each display contained objects pairs which differed by luminance, shape, or both luminance and shape (*redundant*). Participants indicated which letter, *A* or *H*, repeated in the display, unpredictably appearing either within or between object groupings. Performance was expected to be worse for *between-groups* trials. If redundant grouping cues can be combined, participants should be slowest on *between-groups redundant* trials (sixth row). (b) Results for Experiment 3. The graph shows response time for each of three similarity grouping cues (*luminance*, *shape*, *redundant*), depending on whether the letter repetition occurred *within* or *between* object pairs. Note that the difference between the last two bars (*redundant* condition) is larger than the difference between either of the first two sets of bars (these differences are explicitly plotted in Figure 3). Error bars represent within-subject standard errors of the mean. See the online article for the color version of this figure.

Results and Discussion

Following Palmer and Beck (2007), two participants were removed from the analysis for overall accuracies less than 95%. Figure 4b shows participants' median response times to *within-group* and *between-groups* letter repetitions in objects grouped by *luminance*, *shape*, and both luminance and shape (*redundant*). Figure 3 shows the same data plotted as the *within-group* response time advantage (*between-groups* response times minus *within-group* response times) for Experiment 3. In the repetition discrimination task, the difference in response time between the *between-groups* and *within-group* trials is interpreted as the strength of the grouping effect. If similarity grouping is indeed stronger when using redundant features rather than individual features, then *redundant* trials should show a greater response time difference between *between-groups* and *within-group* trials than *luminance* and *shape* trials.

Replicating Palmer and Beck (2007), an ANOVA on participants' median response times revealed a main effect of letter repetition within/between-groups location, $F(1, 14) = 101.93$, $p < .001$, $\eta_p^2 = 0.88$. Specifically, participants were significantly slower when the letter repetition was between groups (M = 1,280 ms, SE = 42 ms) than within groups (M = 985 ms, SE = 26 ms), all $ts(14) > 5.00$, $ps < 0.001$, $ds > 1.30$, for each similarity grouping cue (*luminance*: *between-groups* M = 1,182 ms, SD = 148 ms, *within-group* M = 1,025 ms, SD = 132 ms; *shape*: *between-groups* M = 1,202 ms, SD = 177 ms, *within-group* M = 971 ms, SD = 110 ms; *redundant*: *between-groups* M = 1,456 ms, SD = 190 ms, *within-group* M = 958 ms, SD = 101 ms; all depicted in Figure 4b). There was also a main effect of similarity grouping cue, $F(2, 28) = 23.34$, $p < .001$, $\eta_p^2 = 0.63$, such that participants were slowest on *redundant* trials, slower than *luminance* trials, $t(14) = -5.80$, $p < .001$, $d = -1.50$, and *shape* trials, $t(14) = -5.93$, $p < .001$, $d = -1.53$, and no significant difference between *luminance* and *shape* trials, $t(14) = 0.88$, $p > .250$, $d = 0.23$. While participants were slowest on *redundant* trials, this main effect was driven by slow *between-groups* performance—participants were significantly slower on *redundant between-groups* trials than *luminance between-groups* trials, $t(14) = 10.80$, $p < .001$, $d = 2.79$ and *shape between-groups* trials, $t(14) = 7.90$, $p < .001$, $d = 2.04$, whereas performance on *redundant within-group* trials was faster than that of *luminance within-group* trials, $t(14) = -2.31$, $p = .037$, $d = -0.60$, and no different from performance on *shape within-group* trials, $t(14) = -0.59$, $p = .566$, $d = -0.15$.

Critically, we found a significant interaction between letter repetition within/between-groups location and similarity grouping cue, $F(2, 28) = 50.07$, $p < .001$, $\eta_p^2 = 0.78$. Specifically, as shown in Figure 3, the *within-group* response time advantage was greater for *redundant* trials (M = 498 ms, SD = 177 ms) than for *luminance* trials (M = 156 ms, SD = 114 ms), $t(14) = -8.26$, $p < .001$, $d = -2.13$, and *shape* trials (M = 231 ms, SD = 116 ms), $t(14) = -7.04$, $p < .001$, $d = -1.82$. The *within-group* response time advantage was also greater for *shape* than *luminance* trials, $t(14) = -2.79$, $p = .015$, $d = -0.72$. The *within-group* response time advantage was also significantly greater for *redundant* trials than whichever single grouping cue (luminance or shape) produced the greatest response time difference for each subject (M = 245 ms, SD = 109 ms), $t(14) = -6.89$, $p < .001$, $d = -1.78$. Thus, similarity grouping is stronger when objects are similar on redundant features than when similar by only a single feature.

We had no a priori prediction for whether redundant encoding would produce a stronger grouping effect via interference (slower performance on *between-groups* trials), stronger grouping (faster performance on *within-group* trials), or both. Our results suggest that redundancy could produce increases in interference, but only marginal strengthening of grouping. It is possible that redundant encodings might not make objects easier to attend to, but instead harder to ignore. However, it is also very possible that performance in the *within-group* condition was already at effective ceiling.

Conclusion

The present data provide the first empirical demonstration that redundant encoding of objects can be beneficial to viewers of visual data displays. Experiments 1 and 2 presented participants with a brief display designed to mimic a dense data visualization, and asked them to report the display quadrant that was missing objects of a specified color and shape. Performance was substantially better when a collection was redundantly specified by both color and shape, regardless of whether participants knew the collection's encoding type before each trial. This advantage was echoed in response times (see the online supplemental materials), and was present even when comparing redundant encoding with each participant's best single dimension. Experiment 3 replicated a similar benefit by showing that redundant encoding of visual groups created stronger effects within a measure of visual grouping. These results apply directly to the redundant encoding design technique used in data visualization, but also more broadly to how we attend to objects in our daily environment—it is much more typical that an object will differ from its surrounding objects in multiple feature dimensions than a single feature dimension.

While the present results show that redundancy can be beneficial, there are many open questions surrounding when it will have a benefit. Our displays balanced the relative signal strength of each dimension, but when one dimension, such as color, is more easily discriminable, adding shape differences will likely not yield a redundancy advantage. Our dimensions were also generally easily discriminable, but it is unclear whether a redundancy benefit would be even stronger in cases with less perceptually salient differences between data groups (e.g., such as graphs with many data groups, which use increasingly less perceptually salient feature differences as more data groups are added). In addition, while our displays contained well-spaced points, some real-world displays may contain objects that are more densely spaced and even overlapping points—it is unclear whether redundant coding would still help here. If redundancy operates by reducing noise within the attended collection, then the type of task may also matter: The display segmentation task in Experiment 1 may be especially affected by noise, but other tasks may not. In fact, we suspect that the lack of redundancy benefit in past work (Gleicher et al., 2013, which also carefully balanced the signal strength between dimensions) is due to the use of a noise-resistant task: estimation of the center of a cloud of points. Adding substantial noise to the position representation of each point should barely affect the mean position judgment, which should average over such noise. But there were other differences between that study and the present one that could also explain the differing results, including longer display durations (several seconds), a different population and testing environment (online experiments using Mechanical Turk), and displays with less heterogeneity of distracting colors and shapes.

The differences in results between the past and present study light a path to a broader set of studies exploring how people interpret these types of visual displays. What are the processing bottlenecks on different types of visual decisions, and what stages of processing describe how people inspect such complex displays over time? Candidates include determining the possible features to select by inspecting a first "statistical snapshot" of the features available (Szafir et al., 2016; Haberman & Whitney, 2012), using top-down control of feature- and location-based attention to select collections of interest, computing one of many potential properties of that collection, and comparing that property (e.g., height, size, heterogeneity) to that of another collection. Are these stages serial (Huang & Pashler, 2007; Levinthal & Franconeri, 2011), or can they progress simultaneously for multiple collections at once (Halberda et al., 2006)? Making progress on such questions would plant a researcher firmly in Pasteur's Quadrant (Stokes, 1997), simultaneously informing our broad understanding of how visual attention works, while also having immediate translational importance to the outside world. There are several excellent examples in the domain of data visualization (e.g., Healey, Booth, & Enns, 1996; Lam, Rensink, & Munzner, 2006), which is ripe for further cross-disciplinary work.

References

Biederman, I., & Checkosky, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology, 83*(3, Pt. 1), 486–490. http://dx.doi.org/10.1037/h0028841

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Callaghan, T. C. (1984). Dimensional interaction of hue and brightness in preattentive field segregation. *Perception & Psychophysics, 36,* 25–34. http://dx.doi.org/10.3758/BF03206351

Callaghan, T. C. (1989). Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & Psychophysics, 46,* 299–311. http://dx.doi.org/10.3758/BF03204984

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. http://dx.doi.org/10.1037/0033-295X.96.3.433

Egeth, H., & Pachella, R. (1969). Multidimensional stimulus identification. *Perception & Psychophysics, 5,* 341–346. http://dx.doi.org/10.3758/BF03210654

Eriksen, C. W., & Hake, H. W. (1955). Multidimensional stimulus differences and accuracy of discrimination. *Journal of Experimental Psychology, 50,* 153–160. http://dx.doi.org/10.1037/h0047863

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Oakland, CA: Analytics Press.

Garner, W. R. (1969). Speed of discrimination with redundant stimulus attributes. *Perception & Psychophysics, 6,* 221–224. http://dx.doi.org/10.3758/BF03207021

Gleicher, M., Correll, M., Nothelfer, C., & Franconeri, S. (2013). Perception of average value in multiclass scatterplots. *IEEE Transactions on*

Visualization and Computer Graphics. IEEE Transactions, *19,* 2316–2325.

Gleick, J. (2011). *The information: A history, a theory, a flood*. London, UK: Fourth Estate.

Grubert, A., Krummenacher, J., & Eimer, M. (2011). Redundancy gains in pop-out visual search are determined by top-down task set: Behavioral and electrophysiological evidence. *Journal of Vision, 11,* 10. http://dx.doi.org/10.1167/11.14.10

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:osobl/9780199734337.003.0030

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17,* 572–576. http://dx.doi.org/10.1111/j.1467-9280.2006.01746.x

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science, 24,* 2409–2419.

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. [TOCHI]. *ACM Transactions on Computer-Human Interaction, 3,* 107–135. http://dx.doi.org/10.1145/230562.230563

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

Krummenacher, J., Müller, H. J., & Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics, 63,* 901–917. http://dx.doi.org/10.3758/BF03194446

Lam, H., Rensink, R. A., & Munzner, T. (2006, July). Effects of 2D geometric transformations on visual memory. In R. W. Fleming & S. Kim (Eds.), *Proceedings of the 3rd symposium on applied perception in graphics and visualization* (pp. 119–126). New York, NY: ACM.

Leonard, C. J., Balestreri, A., & Luck, S. J. (2015). Interactions between space-based and feature-based attention. *Journal of Experimental Psychology: Human Perception and Performance, 41,* 11–16. http://dx.doi.org/10.1037/xhp0000011

Levinthal, B. R., & Franconeri, S. L. (2011). Common-fate grouping as feature selection. *Psychological Science, 22,* 1132–1137. http://dx.doi.org/10.1177/0956797611418346

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology, 72,* 95–104. http://dx.doi.org/10.1037/h0023319

Lustig, A. G., & Beck, D. M. (2012). Task-relevant and task-irrelevant dimensions are modulated independently at a task-irrelevant location. *Journal of Cognitive Neuroscience, 24,* 1884–1895. http://dx.doi.org/10.1162/jocn_a_00249

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science, 25,* 1663–1673.

Microsoft. (2014). *Microsoft by the numbers*. Retrieved from http://www.microsoft.com/en-us/news/bythenumbers/ms_numbers.pdf

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279. http://dx.doi.org/10.1016/0010-0285(82)90010-X

Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence of coactivation. *Perception & Psychophysics, 53,* 357–366. http://dx.doi.org/10.3758/BF03206778

Morton, J. (1969). The use of correlated stimulus information in card sorting. *Perception & Psychophysics, 5,* 374–376. http://dx.doi.org/10.3758/BF03210661

Munzner, T. (2014). *Visualization analysis and design*. Boca Raton, FL: CRC Press.

Palmer, S. E., & Beck, D. M. (2007). The repetition discrimination task: An objective method for studying perceptual grouping. *Perception & Psychophysics, 69,* 68–78. http://dx.doi.org/10.3758/BF03194454

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

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

Slatcher, R. B., Selcuk, E., & Ong, A. D. (2015). Perceived partner responsiveness predicts diurnal cortisol profiles 10 years later. *Psychological Science, 26,* 972–982.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of Vision, 16,* 11–11. http://dx.doi.org/10.1167/16.5.11

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1,* 46–54. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00067.x

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48. http://dx.doi.org/10.1037/0033-295X.95.1.15

Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1983). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics Press.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics, 56,* 495–500. http://dx.doi.org/10.3758/BF03206946

Ware, C. (2013). *Information visualization: Perception for design*. Amsterdam, The Netherlands: Elsevier.

Williams, R. (2014). *The non-designer's design Book* (4th ed.). San Francisco, CA: Peachpit Press.

Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are *not* the same thing. *Vision Research, 32,* 757–763. http://dx.doi.org/10.1016/0042-6989(92)90190-T

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433. http://dx.doi.org/10.1037/0096-1523.15.3.419

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12,* 24–25. http://dx.doi.org/10.1038/nn.2223

Received February 17, 2016
Revision received August 17, 2016
Accepted August 28, 2016 ■