Ensemble summary statistics as a basis for rapid visual categorization

Igor S. Utochkin

National Research University Higher School of Economics (HSE), Russian Federation, Moscow, Russia

✉

Ensemble summary statistics represent multiple objects on the high level of abstraction—that is, without representing individual features and ignoring spatial organization. This makes them especially useful for the rapid visual categorization of multiple objects of different types that are intermixed in space. Rapid categorization implies our ability to judge at one brief glance whether all visible objects represent different types or just variants of one type. A framework presented here states that processes resembling statistical tests can underlie that categorization. At an early stage (primary categorization), when independent ensemble properties are distributed along a single sensory dimension, the shape of that distribution is tested in order to establish whether all features can be represented by a single or multiple peaks. When primary categories are separated, the visual system either reiterates the shape test to recognize subcategories (in-depth processing) or implements mean comparison tests to match several primary categories along a new dimension. Rapid categorization is not free from processing limitations; the role of selective attention in categorization is discussed in light of these limitations.

Introduction

The power of ensemble summary statistics

Almost every article on ensemble summary statistics starts with establishing their potential to surmount the severe limitations imposed by attention or working memory in visual cognition of individual objects (e.g., Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988; Treisman & Gelade, 1980). Indeed, at every moment when our eyes are open, we see much more than just a few (probably three to four) objects. Ensemble summary statistics allow us to compress hundreds and thousands of visible properties of objects into compact descriptions, such as approximate number (Chong & Evans, 2011; Feigenson, Dehaene, &

Spelke, 2004; Halberda, Sires, & Feigenson, 2006), average across multiple dimensions (Alvarez & Oliva, 2008; Ariely, 2001; Bauer, 2009; Chong & Treisman, 2003; Dakin & Watt, 1997; Haberman & Whitney, 2007, 2009), or variance (Morgan, Chubb, & Solomon, 2008; Solomon, 2010). The rapid ascribing of those statistics to all visible objects provides a surprisingly precise global representation (or gist) of a visible scene (Alvarez, 2011) with little or no conscious access to individuals (Alvarez & Oliva, 2008; Ariely, 2001; Corbett & Oriet, 2011; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), even when attention is occupied by other objects (Alvarez & Oliva, 2009; Burr, Turi, & Anobile, 2010). Another important property of ensemble summary statistics (that is critical in the context of the present article) is their *high level of abstraction*. That is, statistical descriptions can be built as "pure" global features regardless of the spatial arrangement of individual items in the visual field (Cant & Xu, 2012; Chong, Joo, Emmanouil, & Treisman, 2008; Utochkin, 2013).

The rapid categorization of multiple objects

Imagine that you are picking berries. Once you look at a new bush, you need to understand how many ripe red berries there are on that bush. This problem can be interpreted in terms of ensemble summary statistics: You should estimate the average redness of multiple berries whose shade can vary in a wide range from completely green to saturated red. However, this standard averaging task is complicated by the fact that berries are interspersed with leaves on the same bush. Note that the leaves are more numerous than the berries and are green. If we estimated the average redness of the entire ensemble, then our judgments about the berries would have been wrong because the leaves would have shifted the estimate toward green. Somehow we can easily recognize berries among leaves and judge their average color independently.

This example illustrates a striking ability that I term *rapid visual categorization*. In complex visual scenes where different types of objects are present in numerous copies and often intermixed in space, observers are able to recognize those objects as representing the same or different types (categories) without paying attention to individuals. When we look at a huge, dense crowd of birds on a square, just a brief glance is enough to see that there are pigeons and sparrows among those birds. Although pigeons have quite variable features (different sizes, feather color, turned differently to the observer) they still look quite similar within the category compared with sparrows, which represent a different category.

The notion that the visual system can split multiple objects into clearly distinguishable global subsets is not novel. Much work on this notion has been conducted within the studies of texture segmentation and visual search. In those studies, the roles of various spatial factors are widely discussed as principal determinants of subset formation: local proximity and local contrasts (Bacon & Egeth, 1991; Bravo & Nakayama, 1992; Itti & Koch, 2001; Treisman, 1988; Wolfe, 1994) as well as more global factors, such as abrupt violation of spatial statistics over a region (Nothdurft, 1992, 1993). Indeed, these spatial principles work well for perceiving textures because they are spatially linked to surfaces. However, multiple objects, unlike textures, are organized and processed a bit differently (Cant & Xu, 2012). In the physical world, objects have more independence than textural elements, so their spatial organization is not always that regular: Objects of the same type can be widely disseminated, and objects of different types can be adjacent. In his analysis, Wolfe (1992) showed that processing textures can sometimes be different from processing individual objects to be searched for in those textures, even if both the texture and the objects are defined by the same sensory features.

We can conclude from the previous paragraph that rapid visual categorization of multiple objects cannot rely solely on the spatial organization of elements. It appears that some more abstract, nonspatial representations should drive that process. As mentioned above, ensemble summary statistics have the proper level of abstraction, so they can be considered to be a candidate representation for understanding the mechanisms of rapid categorization.

From the standard idea of representing the features of multiple objects in the form of average, variance, and number, there is just a short step to the central statement of the view of rapid visual categorization presented in this article: If the visual system is good at computing those descriptive statistics, then it probably can use them for testing statistical hypotheses. In regular statistics, primary statistical parameters can be used to check out differences between distributions. At the end, both statistical tests and categorization aim to establish whether compared items have same or different properties. The main goal of this article is to show how the visual system performs statistical tests and what tests it can use for rapid visual categorization.

Primary categorization

To begin to understand how the visual system categorizes multiple objects, we must turn to the simplest case of primary statistical processing. This requires a statistical decision along a single perceptual dimension, while other dimensions are invariable or neglected.

A "reverse statistical inference" problem of visual perception

Although it seems very simple, the primary categorization of ensembles is not a trivial problem for the visual system. The main difficulty is that the only informative input is the distribution of pooled activity along a certain dimension and that there is no prior information about an item's category. Based solely on this continuous distribution, the visual system must derive discrete (categorical) entities. To illustrate how problematic this type of decision can be, I turn to an example that can be termed *reverse statistical inference*. In normal research practice, when conducting an experiment, the experimenter manipulates conditions between trials. Each trial has a *prior* mark by its condition. This prior marking allows further unambiguous grouping of experimental data and statistical comparisons between those groups (samples). If sample characteristics are significantly different, then the experimenter can decide that the manipulated factor had an effect on the dependent variable. This is the logic behind direct statistical inference. However, now imagine that the prior conditional marks are lost for some reason, only dependent variable data are left, and all trials are pooled together. Can one judge the experimental effect knowing how various these data are? Of course, this logic is unacceptable in normal statistical reasoning. However, it looks as if reverse statistical inference is the only way to estimate the differences between ensemble members because the prior marks exist only in physical reality and are thus unavailable for internal processing.

The shape test as a solution to the reverse inference problem

Whereas a correct reverse inference is hardly possible in regular statistics, it appears that the visual system
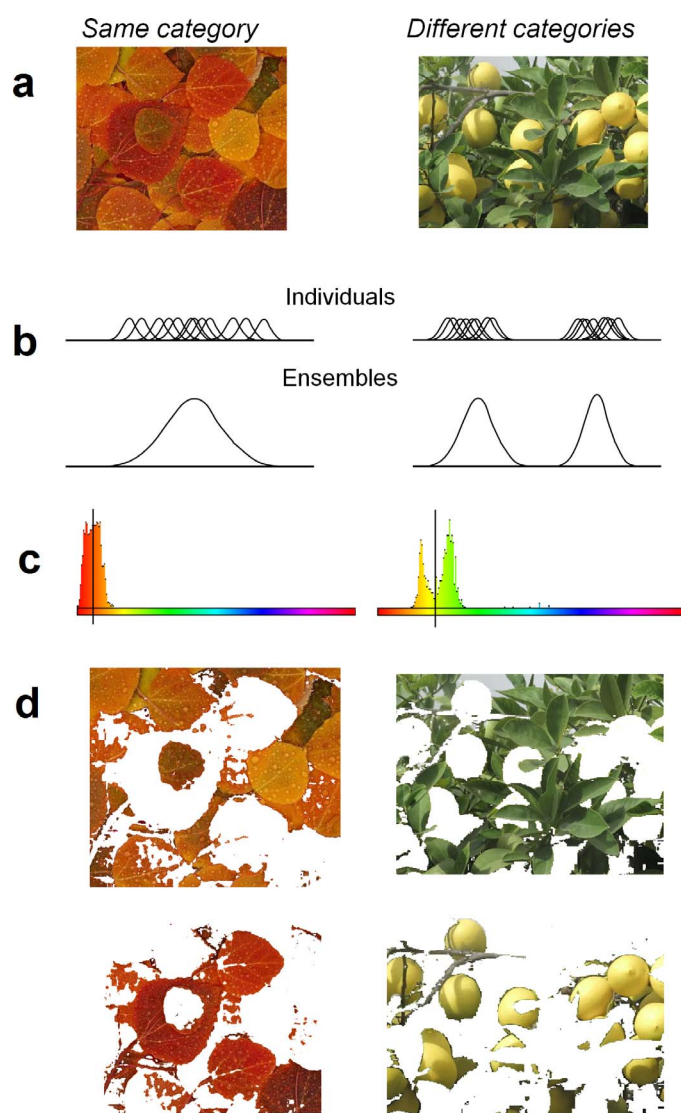
Figure 1. The example representation of natural ensembles along the color dimension for categorically identical (left panel) and categorically different objects. (a) Original images; (b) hypothetical internal representations of individual items and ensembles; (c) physical hue distributions in the HSB (hue-saturation-brightness) color space, with the vertical line depicting the half-split threshold; and (d) the images after filtering the upper half or the lower half of the hue distribution. Hue histograms (c) and processed images (d) were obtained via ImageJ image analysis software (Schneider, Rasband, & Eliceiri, 2012).

can use a heuristic that helps in the categorization of ensembles. This heuristic is statistical by nature, as is ensemble representation itself, and is based on the shape of distribution.

Testing a pooled distribution for shape can be an efficient way to estimate the categorical unity or separability of an ensemble. The usefulness of such a test is justified by the statistical structure of the natural world, as our visual system has evolved for its

perception. The physical properties of most natural objects (both animate and inanimate) are distributed in accordance with the single-peak law (e.g., Gaussian distribution), with the majority of values grouped around the average and deviants showing some natural variability of that property within the type. In contrast, when objects of different types are placed together, their properties are not likely to form a single peak. Rather, they should form a multiple-peak distribution, with each peak corresponding to a local average property of each presented type. This is illustrated by an example in Figure 1. In the left panel of Figure 1a, an ensemble of fall leaves is presented. Here, the leaves are substantially variable in both color and size, but both those dimensions have a quite gradual variation between colors or sizes, with maximum at some intermediate points (orange shades or medium size). In the right panel of Figure 1a, two types of objects—leaves and lemons—are intermixed in space. Although their colors cover approximately the same range as in the left panel, it is obvious that there is no single peak describing overall color distribution. There is no shared peak at intermediate shades, but there are two local peaks at the extremes instead—one for each type of object. Figure 1b shows hypothetical internal color representations of those two images in terms of individual items end ensembles. Figure 1c shows the physical hue distributions of those images. It is easy to see that the shape of the physical distribution for the fall leaves tends to have a single peak, while the distribution for lemons and leaves is more like a two-peak one. Finally, in Figure 1d, two versions of the original images are presented; each version was processed using a half-split threshold on the hue axis. This threshold was approximately established at the middle point of each distribution (see Figure 1c). When filtering away a left-most or right-most half of a hue distribution, the remaining part represents a subsample of items showing how the physical distribution of the shape is related to categories. Indeed, for the leaves image, this filtering removes a subsample of more reddish or more yellowish leaves, which do not look categorically different in the original image. In the lemons and leaves image, the threshold provides almost perfect separation between the lemons and the leaves. (Note that this separation is made based on hue only, but we can visually verify the quality of separation using an independent dimension, such as the shape.)

An idea that the visual system is testing ensemble distributions for their shapes is also supported by the principles of feature representation across multiple levels of neural processing. Since Hubel and Wiesel's (1959) pioneering work, growing evidence is accumulating from numerous sensory domains that each individual feature is encoded as a single-peak Gaussian distribution of firing activity among feature-selective

cells (Yantis, 2014). Relatively small receptive fields of those feature-selective cells in early cortical visual fields (e.g., V1 and V2) are then pooled into larger receptive fields (e.g., V3, V4, MT), which are appropriate for the processing of more substantial portions of visual information. This poses a plausible substrate for the integration of feature information from individual locations into a global ensemble percept. Alvarez (2011) and Haberman and Whitney (2012) proposed a simple model predicting that pooled local neural responses are then represented by an averaged Gaussian at higher levels of analysis. The peak of that Gaussian represents the mean (Alvarez, 2011), and its standard deviation affects the precision of averaging (Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Im & Halberda, 2013; Utochkin & Tiurina, 2014).

It is easy to see that the mechanism proposed by Alvarez (2011) and Haberman and Whitney (2012) is effective mainly for those objects whose features are quite similar en masse (as are the physical properties of same-type objects). Those features statistically benefit from the single-peak pooling because they gain greater representation in the pooled response due to the following rules: (a) The higher between-elements similarity, the narrower the standard deviation of the pooled response, and (b) the more numerous those elements, the higher the global peak. However, the flip side is that averaging should wash out elements whose features are rare and highly deviant from the mean. Elsewhere, Haberman and Whitney (2010) showed that those highly deviant elements are devaluated during ensemble perception and do not contribute to the resulting average estimation. Of course, devaluation does not mean that the deviant gets unseen (pop-out effects in visual search show us the opposite). Rather, it means that those deviants are not represented under the same peak of a single Gaussian.

So how does the visual system use the shape of the distribution to decide whether ensemble members belong to same or different categories? Perhaps the most plausible mechanism relies on peak separation in a pooled ensemble representation. In other words, the system verifies whether the overall pattern of local neural activities can be summarized as a single-peak distribution. If the distribution satisfies this condition, then all ensemble members are recognized as belonging to the same category; otherwise, they are recognized as belonging to different categories.

How does the visual system distinguish between the single-peak and multiple-peak distributions? It most likely depends on a critical distance between neighboring features in the feature space. To understand how it works, we must consider again the process of pooling that transforms local spatial feature representations into the ensemble percept. As was mentioned above, each particular feature is encoded by the single-peak distribution of activity in feature-selective cells, and that makes each individual representation noisy (Alvarez, 2011). When local neural responses are pooled together, their distinctiveness is the matter of overlap. Overlap is large for highly similar features, and this makes those features hardly discriminable, so all such features will contribute to the same global peak corresponding to one category. In contrast, the representations of dissimilar features have a negligible overlap that produces some discontinuity in pooled activity. Indeed, when such discontinuity takes place between two clusters of pooled activities, it is impossible to build a peak between them. Consequently, no single peak can be built, but each cluster generates its own peak. The number of peaks defines the number of categories that one can perceive in the ensemble.

A good quantitative measure of critical feature distance to produce categorical separation can be obtained in the visual search task. This measure was termed the *preattentive just noticeable difference* (PJND; Wolfe, 1994). A paradigm for measuring PJND involves searching for a feature singleton among homogeneous nontargets. The target–nontarget difference is systematically manipulated (Foster & Ward, 1991). The PJND is a critical difference where a switch takes place between two search patterns—efficient (the search time does not change with set size) and inefficient (the search time increases with set size). Elsewhere, Wolfe, Friedman-Hill, Stewart, and O'Connell (1992) suggested considering such a shift as a result of categorical separation. In terms of the peak separation mechanism introduced in the previous paragraph, if a target and nontargets differ more than a PJND, then they are likely to produce a two-peak pooled distribution. This explains the efficient search pattern for the above-PJND targets as an act of rapid categorization: Here, one peak is ascribed to the target and another is ascribed to the rest of the items that allows their momentary rejection as a single unit. In contrast, if a target–nontarget difference is below the PJND, their local responses fall under the same pooled peak and, therefore, are recognized as being categorically the same. To complete the search among these categorically identical items, the visual system will require the deployment of focused attention that serially moves from item to item and provides better discrimination. This is what predicts the inefficient search pattern for categorically identical stimuli.

Treue, Hol, and Rauber (2000) gave a prominent example of ensemble categorical separation based on testing the shape of pooled neural activity. They registered single-cell responses from neurons selective for the direction of motion in the macaque MT area when presenting moving-dot patterns. The dots were split into two spatially overlapping subsets, each moving in its own direction. The angle of separation

varied from 0° to 120°. Treue et al. arrived at two important findings, both highly consistent with the model of shape-based categorization. First, they found that in small angular separations (up to 60°), the maximum activity was shown by neurons preferring an average direction between two physically presented directions, so overall neural activity can be represented by a single peak. However, in large angular separations (90°–120°, which probably exceeded the PJND), the single central peak tended to be divided into two peaks, each moving to one physical direction, while neural responses to the average direction decreased. Treue et al. interpreted this pattern as two peaks instead of one. Presumably, this neural separation facilitates the phenomenal perception of bidirectionally moving ensembles (or textures), although Treue et al. noted that original directions can be recovered from the bandwidth of the single-peak distribution as well.

The idea of a single-peak versus multiple-peak distribution of pooled activity has another important implication. It turns out that two equally distinct objects (or sets of objects) can be recognized as belonging to the same or different categories depending on other ensemble members. When two features are present, their difference exceeds the PJND, and no other features are present, then they should be represented as two peaks and perceived as categorically different. However, if more features are present and their values lie between the initial ones—I term those new values *transition features*—all pairwise differences may happen to be below the PJND. In that case, the visual system would probably collect all features under the same peak, as there is no gap within the entire cluster of features. In line with this notion, Chong and Treisman (2003) found a slight decrease in the accuracy of mean size judgments when only extremely large and small items were presented (the two-peak distribution) compared with other displays containing both extreme and intermediate sizes (distributed either normally or uniformly). Although this effect was present only in naïve observers, Chong and Treisman made a claim about its nature that the two-peak distribution was not represented by a single peak, whereas others were. A quite similar effect and interpretation were obtained by Utochkin and Tiurina (2014). They compared the accuracy of averaging in ensembles consisting of only extreme-size circles (very small and very large) and ensembles including the variety of more intermediate sizes within the same range. Like in Chong and Treisman (2003), averaging performance was poorer under the former condition in Utochkin and Tiurina's (2014) study.

Simultaneously, the results of Chong and Treisman (2003) and Utochkin and Tiurina (2014) can be explained without reference to peak separation. Better averaging performance with transition features could also arise from variance reduction: It is easy to see that the two-peak distribution provides maximum variance as containing only extremes, and transition features add some moderate values, reducing the overall variance within the same range. According to this view, the effect can be explained by transformations within the unitary peak rather than its split. A critical experiment that distinguished between those two explanations was conducted by Yurevich and Utochkin (2014). Their participants performed a visual search for a singleton orientation among homogeneous or heterogeneous nontargets of different orientations. Heterogeneous nontarget orientations always varied within the same range of 45°, so the distance between extremes was always the same. A critical factor manipulated in those displays was the number of intermediate orientations, which defined the smoothness of transition from one extreme to another. In the distinct condition there were only the extremes and no transition, in the sharp transition condition there was one transition orientation halfway between the extremes (transition step was approximately 22.5° on average), and in the smooth transition conditions eight transition orientations (step 5°) were added. Yurevich and Utochkin found a nonmonotonous effect of transition on the speed of the visual search: The sharp transition condition yielded the slowest search performance, while the smooth transition condition yielded the fastest search performance among all heterogeneous displays. This cannot be explained by variance reduction because any intermediate nontarget feature decreases overall variance and thus should predict a faster visual search (Rosenholtz, 1999). Therefore, the sharp transition condition should facilitate rather than inhibit performance. The only plausible explanation of such a divergence is based on peak separation. In large steps of transition, any additional feature is represented as a separate Gaussian that complicates performance. In contrast, small steps of transition cause the visual system to represent the entire range as a single-peak though high-variance distribution, which is easy to treat as a single set. The aforementioned example with leaves and lemons (Figure 1) directly demonstrates how such duality of categorizing extreme features applies to real-world perception. With a smooth transition between colors or sizes and relatively stable shapes, items in Figure 1a are perceived as categorically the same. With a sharp transition between colors, sizes, or shapes, items in Figure 1b are perceived as representing two categorically different ensembles.

Within-category similarity and between-categories contrast

An important aspect of categorical perception is diminishing differences between objects belonging to

the same category and exaggerating differences between objects belonging to different categories (Goldstone & Hendrickson, 2010). The principles of ensemble summary statistics can help us understand how these categorical effects arise when we are categorizing multiple objects.

This explanation is based on the previously mentioned ideas of Alvarez (2011) about precision-weighted averaging and those of Haberman and Whitney (2012) about ensemble representation as the average of local responses. Both those ideas imply that the visual system favors more numerous and similar features and devaluates deviants when building an ensemble percept. This inevitably leads to strengthening perceived homogeneity among ensemble members under the same peak. Consequently, all same-category items should be seen more similarly than they would be seen in isolation. Recent studies showed that ensemble summary statistics systematically affect the perception of individual members and retention in working memory: They cause a strong bias toward the mean (e.g., Brady & Alvarez, 2011; Olkkonen, McCarthy, & Allred, 2014). In other words, individual items tend to be seen in the ensemble as being more averaged than they actually are. Furthermore, as Brady and Alvarez (2011) showed, this effect is category specific, as they dissociated independent biases in color-segmented ensemble subsets from the general bias toward the total ensemble mean. Elsewhere, Solomon (2009; Morgan et al., 2008) noted that presenting items in multielement textures (or ensembles) allows the visual system to discount the individual noisiness of each such item. Morgan et al. (2008) suggested that the visual system computes the summary variance of those elements, and this makes the texture seem relatively uniform and stable.

Obviously, forces responsible for increasing within-category similarity should also increase between-categories contrast. If individual features tend to be seen and memorized as more averaged under their common peak, then they probably should move away from other features represented by another peak. This predicts an effect for differently categorized items opposite to that found for same-categorized items. The Ebbinghaus illusion gives support for this prediction for one who looks at that illusion in terms of ensemble summary statistics (Ariely, 2001; Im & Chong, 2009). When surrounded by homogeneous small circles, a medium central circle is seen as larger than when it is surrounded by large circles. It is important here that the central circle and surroundings are sufficiently different to produce a gap between their distributions; therefore, they are likely to be categorically different as well. The experiments of Im and Chong (2009) convincingly show that the Ebbinghaus illusion acts upon ensembles as well as individual objects. When

they presented sets of differently sized central circles, each with either small or large surrounds, they found systematic biases in judging the mean size depending on the surround size. Observers tended to underestimate the mean size of circles with a large surround and vice versa. In terms of categorical effects in perception, this can be interpreted as an increasing between-categories contrast in spatial groups. Certainly, because of its strong dependence on spatial layout, the Ebbinghaus illusion is not a pure example of rapid categorization based on abstract ensemble statistics. It is likely that modified versions of Im and Chong's (2009) experiments with random layouts can give more convincing evidence for between-categories effects in ensemble perception.

Secondary categorization

When primary categorization is performed and an ensemble is divided into several subsets along one dimension, the visual system can further elaborate the processing of those subsets along other dimensions. I term this process *secondary categorization*. Again, this process can be described as testing statistical hypotheses. However, at this stage, the variety of testable hypotheses is wider because the reverse statistical inference problem is already solved—at least for one dimension. Categories divided in course of primary processing (if any) now serve as labels marking each individual item at an input, much like fixed conditions of an independent variable mark each individual trial at an input to analysis of variance. Following this analogy, a target dimension for secondary processing can be considered to be a dependent variable.

Secondary categorization can progress in two modes depending on an ongoing task. The first mode can be termed *in-depth categorization*. It implies that the observer selects one of the primary categories and applies the same algorithm to a new dimension to figure out whether some subcategories can be recognized within the initial one. The second mode is *in-breadth categorization*, when the observer compares whether several primary categories are same or different in terms of the other dimension. Figure 2 illustrates both of these strategies of secondary categorization, and a detailed explanation is given below.

In-depth categorization

There is no need to describe the probable mechanism of in-depth categorization in detail because it seems to be very similar to that of primary categorization. In summary, it can be based on the shape test described
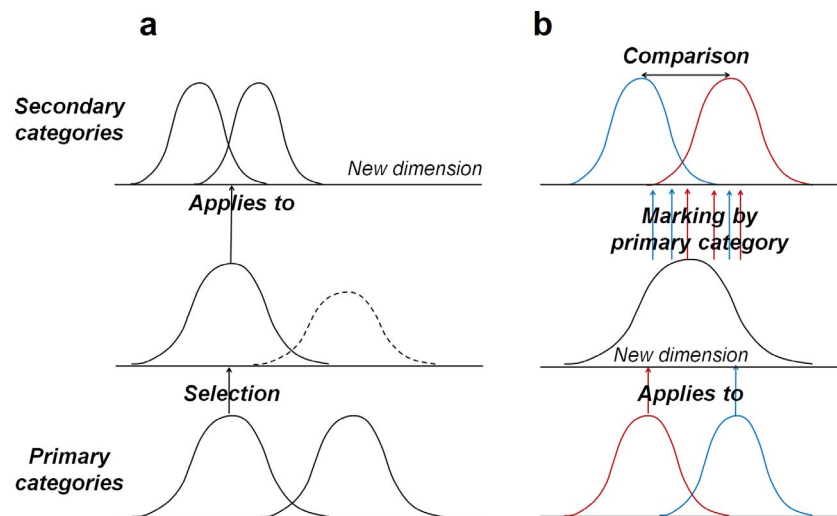
Figure 2. Two modes (strategies) of secondary categorization along a new dimension given a good peak separation of primary categories. (a) In-depth categorization. (b) In-breadth categorization.

above. One important addition is that performing in-depth categorization requires the prior selection of a relevant subset that involves attentional processes to some degree. This process is shown in Figure 2a. Once we have access to a set of primary categories along one dimension, we can selectively attend to only one such category (e.g., when picking berries from a bush, we can attend to round items that look like berries while ignoring oblong items that look like leaves). Within that attended category, we switch to a new dimension trying to apply the peak separation algorithm and see whether the items are categorically the same or different along that dimension. For example, keeping attention on the berries, we can estimate their ripeness using color. Some statistical information can be still available for unattended primary categories (shown by a dotted line in Figure 2a), but the attended category would prevail (De Fockert & Marchant, 2008; Pavlovskaya, Soroker, Bonneh, & Hochstein, 2015).

In-depth categorization has an important implication for understanding rather efficient visual searches for some feature conjunctions, which is commonly predicted to be inefficient (Treisman & Gelade, 1980)—that is, producing the increased search time with the number of items in the display. Indeed, in a number of classical experiments, it was found that some conjunction searches, such as for color × depth, motion × depth (Nakayama & Silverman, 1986), and color × orientation (Friedman-Hill & Wolfe, 1995), appear to be rather efficient. This is hard to explain with the item-by-item mode of focused attention. In Friedman-Hill and Wolfe's (1995) experiments, participants searched for a certain color × orientation target among distractors sharing either color or orientation with that target. A rather effective strategy for such a search, as described by Friedman-Hill and Wolfe, includes two

steps. In the first step, the observer attends to one color at a time. In the second step, an efficient search for an odd orientation is carried out within this selected subset. Later on, Nakayama and Martini (2011) reported that Nakayama and Silverman's (1986) observers seemed to attend to different depth planes in order to rapidly detect a second feature. In terms of in-depth categorization, those findings indicate observers' ability to split the set into categories along one dimension (primary categorization) and then reiterate the same along another dimension within a selected category (secondary categorization). As the target had a unique secondary feature within the primary category, it could be found almost as efficiently as a standard feature-defined target. This sort of reiterative categorization is useful. Being able to attend to one feature-defined category and filter out other such categories, the observer does not need to focus attention on every individual item to bind two features. Instead, the visual system operates with more substantial portions of the input, making the deployment of attention more efficient (Wolfe, Cave, & Franzel, 1989).

In-breadth categorization

In many studies of ensemble summary statistics, researchers measure the thresholds of mean or variance discrimination of two subsets presented simultaneously or serially. These two subsets are either spatially separated (e.g., Attarha, Moore, & Vecera, 2014; Chong et al., 2008; Chong & Treisman, 2003; Corbett et al., 2012; Im & Chong, 2009; Morgan et al., 2008) or intermixed but marked by well-discriminable colors (Chong & Treisman, 2005; Im & Chong, 2014). Both methods of presentation represent the in-breadth
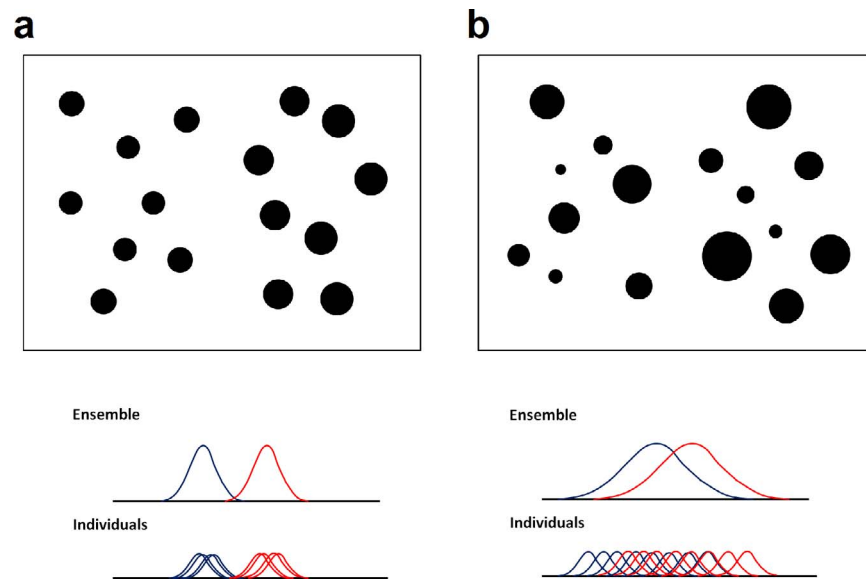
Figure 3. Illustration of the mean discrimination between (a) low-variance sets and (b) high-variance sets. Top panels depict example stimulus displays requiring one to determine which of two sides (left or right) has a larger mean size. Bottom panels represent the corresponding hypothetical individual and ensemble representations (blue Gaussians depict left side; red Gaussians depict right side). As variances in panel b are larger, the mean discrimination is more difficult than that in panel a.

categorization task, where primary categories are well defined by spatial statistics (centroids) or color statistics and other dimensions are statistically compared between those categories.

It appears that in-breadth processing requires statistical procedures other than those used for primary or in-depth categorization. As the primary categories are well defined and represented as clearly distinguished peaks along their dimension, further statistical processing can be performed as testing mean or variance equality. Figure 2b illustrates this idea. Once the visual system has access to primary peaks it can mark each secondary-dimension value by a certain primary value. Thus, even if the secondary dimension originally has no clear peak separation (as in Figure 2b), it is still possible to compare summary statistics along this dimension due to primary marking. Experimental data show that visual comparisons of this sort do behave like regular statistical tests intended for similar purposes.

Figure 3 demonstrates two displays for a typical size-averaging task. The displays are divided into two halves, each containing an ensemble with its own mean size. An observer is asked to determine which side has larger average size. Although both examples have exactly the same mean difference (right circles are 30% larger on average than left ones), the task gets progressively more difficult from panel a to panel b. When a naïve observer is asked why the task in panel b is harder than that in panel a, he or she says that it is because panel a contains very similar sizes on each size, whereas in panel b the sizes are much more variable.

Thus, this naïve explanation (sometimes from people not very familiar with ensemble summary statistics and even regular statistical tests) reproduces the basic logic of mean comparison tests, such as the t test or the F test underlying analysis of variance. In brief, those tests estimate how much between-groups variability defined by mean difference exceeds within-group variability. In both panels of Figure 2, between-groups variability is the same but within-group variance increases from panel a to panel b, thus reducing the overall between-groups:within-group ratio and making the right and left ensembles less discriminable in terms of mean size.

Apart from Figure 2, there is established experimental evidence that the visual comparison of means between ensembles behaves like the t test or the F test. Im and Halberda (2013) manipulated the size variance of two successively presented ensembles and measured the thresholds of mean size discrimination. They found that the threshold rose with variance. Corbett et al. (2012) exhibited a similar finding using a different paradigm. They adapted their observers to ensembles with large and small mean sizes and then measured an adaptation aftereffect on ensembles with slightly different mean sizes. In one of the experiments, they varied the size variance of adapting ensembles. They found that the aftereffect was attenuated by high variance. The findings of both Im and Halberda (2013) and Corbett et al. (2012) are consistent with the logic of the t test and the F test. In Im and Halberda's experiment, as variance increased, the mean difference between ensembles increased as well to achieve a critical between-groups:within-group ratio to reject the

null hypothesis. In Corbett et al.'s experiment, increasing variance reduced the between-groups:within-group ratio and, as a consequence, the visible contrast between adapting ensembles and the magnitude of the aftereffect.

In line with those findings, Fouriezos, Rubenfeld, and Capstick (2008) asked observers to rate their confidence while comparing the mean height of two clusters of vertical bars. Height variance and the number of bars were manipulated. Fouriezos et al. found that accuracy and confidence are positive functions of numerosity and negative functions of variance. They directly related their results to statistical decisions, which are akin to regular mean comparison tests. Variance had the same effect as described by Im and Halberda (2013). As for the effect of numerosity, it can be explained by the statistical power of a test—its ability to confirm the H_1 hypothesis if the hypothesis is correct—which is known to increase with sample size.

The limitations of rapid categorization

Ensemble perception is often considered to oppose the limited-capacity processing of individual objects (e.g., Alvarez, 2011; Ariely, 2001; Chong & Evans, 2011; Robitaille & Harris, 2011; Treisman, 2006). Other researchers, in contrast, argue that this ability is provided by the same limited-capacity processes (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Marchant, Simons, & De Fockert, 2013; Myczek & Simons, 2008; Simons & Myczek, 2008). Establishing the locus and the nature of the processing bottleneck is a fundamental challenge for vision theory. Viewing the problem from the perspective of rapid visual categorization can contribute to this debate.

How many ensembles can be recognized and stored?

Numerous studies show that our ability to attentionally track and store objects is severely limited by three to five items on average, depending on task, stimulus type, and so on (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Pylyshyn & Storm, 1988). When the visual system expands its analysis to ensembles, it meets even more stringent limitations. There is growing evidence that different ensemble tasks can be performed without substantial loss when no more than two subsets are presented at one time (Attarha et al., 2014; Halberda et al., 2006; Im & Chong, 2014; Poltoratski & Xu, 2013; Zosh, Halberda, & Feigenson, 2011; but see Treisman, 2006, who did

not find any decrease in judging the proportion among three concurrently presented feature-based ensembles). An important question in light of the present framework of rapid categorization concerns the locus of this limitation. Does it arise at primary processing when the number of peaks is being roughly determined? Or does categorization encounter the bottleneck at the stage of secondary categorization? A partial answer can be found in the experiments by Poltoratski and Xu (2013) and Watson, Maylor, and Bruce (2005), who estimated the critical number of concurrently presented and spatially intermixed colors that can be reported without a loss in accuracy or speed. As they did not require the reporting of any specific properties of colored subsets, their tests can be admitted measuring primary processing. Both studies came to the conclusion that no more than two ensembles can be seen and stored at once. Given those results, it is likely that ensemble categorization has very early limitations. Moreover, Poltoratski and Xu (2013) suggest that the working memory limit for ensembles (primary categories) is a critical determinant of the limitations of further processing (secondary).

The binding problem and attentional control for rapid categorization

Another source of limitations typically associated with the processing bottleneck is the binding problem. The statement of the problem is based on a critical finding that basic features of multiple objects are likely to be processed in parallel and separately, so that at any given moment we have full information about feature distribution but no knowledge of how individual basic features are bound in objects (Wolfe & Cave, 1999). Perhaps the most influential attempt to describe a possible solution to this problem is feature integration theory (Treisman, 2006; Treisman & Gelade, 1980), where a limited-capacity attentional mechanism is supposed to move from one location to another and bind corresponding features to an object percept. For correct perception, attention should select one object at one time.

As ensemble perception requires the simultaneous processing of many objects, the binding problem predicts that the visual system should face insuperable difficulties when seeing ensembles filled with varieties of feature conjunctions. It is easy to see that the problem is associated with secondary ensemble categorization, which implies the processing of one dimension given another. However, the aforementioned severe limit of primary categorization heightens the problem.

In a series of studies, Treisman and her colleagues tried to discover how the visual system works with conjunctive ensembles. In one such study, Treisman

(2006) briefly presented observers with sets of colored letters and asked them to report the percentage of a particular color (e.g., all green letters), letter (e.g., all Ts), or conjunction (e.g., green Ts). She found that the observers were good at judging proportions of features but were much worse at judging proportions of conjunctions. In another study, Emmanouil and Treisman (2008) tested observers' accuracy in estimating ensemble summary statistics along two dimensions at one time. Their participants were presented with sets of differently sized and moving (or differently oriented in one of experiments) objects and had to estimate average size and average speed. The relevant dimension was either precued or postcued, making observers either focus on one dimension or divide attention between two. In almost all cases, Emmanouil and Treisman found a significant (though not dramatic) cost of dividing attention between two dimensions. They also separated relevant dimensions between shape-defined subsets (differently sized Os and moving Xs) and found that the cost tended to increase under that condition. In summary, both studies showed that the visual system indeed is imperfect when it is trying to combine different ensemble properties at one time.

However, the previous analysis in this article shows us that the visual system is quite good at performing both primary and secondary categorization. How is it possible if ensemble perception suffers from the binding problem? First, we must keep in mind that ensemble processing does not require binding features of all objects. As ensemble properties are represented in the form of summary statistics, then the only thing needed is binding those summary statistics. This manner of binding is definitely more economic and can probably explain why Emmanouil and Treisman (2008) found the cost of dividing attention to be poor.

Second and more important, if secondary categorization is exposed to binding limitations, then it appears that its effectiveness may be provided by the strategies of attentional selection. One promising candidate for such a strategy is that used for guided search. Although it was originally described to explain attentional phenomena in visual searches (Wolfe, 1994; Wolfe et al., 1989), this strategy is also supposed to be appropriate for secondary ensemble processing (Chong & Treisman, 2005). A core similarity between visual search and ensemble categorization is that both can be guided, on one hand, by bottom-up processes considered to be parallel and, on the other hand, by top-down processes controlling the selection of task-relevant information.

It appears that processes underlying primary categorization resemble the bottom-up and preattentive mechanisms of attentional guidance (Wolfe, 1994; Wolfe et al., 1989). These processes permit one to split the visual field into several categorically different subsets along a number of independent dimensions, as preattentive processes build separate feature maps within each basic dimension. It appears that, like the preattentive processes, primary categorization is automatic (Alvarez & Oliva, 2009), at least within its limitations. Under top-down attentional control, the primary categories can be used for secondary categorization. This means that the observer can selectively attend to one or two (and probably no more at one time given the aforementioned limitations) primary categories and launch secondary processing (either in depth or in breadth) within them.

The top-down character of attentional selection also implies that we can flexibly attend to those primary categories, which have ensemble features similar with a certain template of the relevant object type. Returning to the example at the beginning of this article, when estimating whether berries on a bush are sufficiently ripe, one can selectively attend to a subset of only relatively round items and judge the average redness of those items. Indeed, experimental data show that when relevant primary categories are precued, observers have no problem with reporting secondary summary statistics (Chong & Treisman, 2005; Halberda et al., 2006).

Relation to scene categorization

As stated at the beginning of this article, ensemble perception is an important part of the rapid perception and comprehension of scenes. Previous work has shown that our striking ability to categorize the gist of natural scenes does not require the full recognition of objects in those scenes (this process is time consuming and limited by the relatively narrow foveal area of the visual field). Instead, rough image statistics (e.g., spatial frequencies, line, luminance, or color distributions) can be used to distinguish between the variety of landscapes (Oliva & Torralba, 2001, 2006; Torralba & Oliva, 2002) or between animals and vehicles (Rosenholtz, Huang, & Ehinger, 2012). The approach presented in this article continues the previous work, as it implies that rapid ensemble categorization requires access to overall summary statistics (e.g., the mean, the variance, and the shape of distribution) instead of individual object properties. At the same time, it seems to require more elaborate and fine processing than just rough image statistics—this is supported by a finding that extracting ensemble summary statistics requires substantial time (Whiting & Oriet, 2011). First, ensemble summary statistics are built on entities representing objects (or maybe proto-objects; see Rensink, 2000) rather than raw elements of images. Second, ensemble categorization implies that more than one category can be recognized within the same image, while scene catego-

rization requires only one category per image (but that can be just a lack of the corresponding paradigm). Certainly, more detailed theoretical and empirical analyses are needed in the future to establish an overlap between the mechanisms of rapid categorization based on ensemble and image summary statistics.

Summary and conclusions

The approach presented in this article is based on a simple statement that the visual system can use primary ensemble summary statistics (the mean, the variance, and the numerosity) to test hypotheses that the features of multiple visible objects are statistically identical or different. This type of statistical testing is supposed to be a possible mechanism of rapid visual categorization. Additionally, two levels (stages) of categorization were identified. Primary categorization is presumed to run as a shape test of pooled activity across a dimension. Using this test, the visual system recognizes whether this pooled activity can be approximated by a single-peak distribution or splits into multiple peaks, each corresponding to a separate category. In secondary categorization, the visual system operates primary categories as separate units. This stage can be performed in depth when subcategories are identified within one primary category via the shape test or can be performed in breadth when several (probably no more than two at one time) primary categories are compared along other dimensions by tests resembling the t test or the F test. Finally, it turns out that rapid categorization is more or less susceptible to fundamental limitations that work in the perception of individual objects—attention and working memory capacities and the binding problem. However, the flexible allocation of attention to relevant categories supports the effectiveness of that process despite the limitations. The efficiency of ensemble categorization is higher than that of object identification (which is required for visual search) because the former exploit summary statistics.

Certainly, statistical tests in visual perception are a metaphor. This metaphor is based somewhere on neurally grounded models of ensemble summary statistics (Haberman & Whitney, 2012) and incorporates strong evidence from the neuroscience of vision (Treue et al., 2000). However, in many cases, we do not know which computations actually underlie ensemble summary statistics and statistical tests. What is important is that both ensemble summary statistics and statistical tests capture the real observer's performance and give new insights into the nature of the underlying internal processes.

Acknowledgments

References

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83,* 25–39.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Science, 15,* 122–131.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15,* 106–111.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble features outside the focus of attention. *Psychological Science, 19,* 392–398.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, USA, 106,* 7345–7350.

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

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

Bacon, W. F., & Egeth, H. E. (1991). Local processes in preattentive feature detection. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 77–90.

Bauer, B. (2009). Does Stevens's power law for

brightness extend to perceptual brightness averaging? *Psychological Record, 59,* 171–186.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22,* 384–392.

Bravo, M., & Nakayama, K. (1992). The role of attention in different visual search tasks. *Perception & Psychophysics, 51,* 465–472.

Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision, 10*(6):20, 1–10, doi:10.1167/10.6.20. [PubMed] [Article]

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience, 32,* 7685–7700.

Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count versus estimate). *WIREs Cognitive Science, 2,* 634–638.

Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics, 70,* 1327–1334.

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

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

Corbett, J., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition, 20,* 211–231.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138,* 289–301.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37,* 3181–3192.

De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70,* 789–794.

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70,* 946–954.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science, 8,* 307–314.

Foster, D. H., & Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society London: Series B, 243,* 75–81.

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics, 70,* 456–464.

Friedman-Hill, S. R., & Wolfe, J. M. (1995). Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 531–551.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science, 1,* 65–78.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17,* R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 718–734.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72,* 1825–1838.

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). Oxford, United Kingdom: Oxford University Press.

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

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology, 148,* 574–591.

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics, 71,* 375–384.

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

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics, 75,* 278–286.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Luck, S. J., & Vogel, E. K. (1997). The capacity of

visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Marchant, A. P., Simons, D. J., & De Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142,* 245–250.

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A "dipper" function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11):9, 1–8, doi:10.1167/8.11.9. [PubMed] [Article]

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70,* 772–788.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51,* 1526–1537.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264–265.

Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion, and color cues. *Vision Research, 33,* 1937–1958.

Nothdurft, H. T. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics, 52,* 355–375.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36.

Olkkonen, M., McCarthy, P., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision, 14*(11):5, 1–15, doi:10.1167/14.11.5. [PubMed] [Article]

Pavlovskaya, M., Soroker, N., Bonneh, Y. S., & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience, 27,* 1397–1411.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. J. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739–744.

Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision, 13*(8): 6, 1–14, doi:10.1167/13.8.6. [PubMed] [Article]

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

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7,* 17–42.

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11*(12):18, 1–8, doi:10.1167/11.12.18. [PubMed] [Article]

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39,* 3157–3163.

Rosenholtz, R., Huang, J., & Ehinger, K. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology.* doi:10.3389/fpsyg.2012.00013.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods, 9,* 671–675.

Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics, 70,* 1335–1336.

Solomon, J. A. (2009). The history of dipper functions. *Attention, Perception, & Psychophysics, 71,* 435–443.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10*(14):19, 1–16, doi:10.1167/10.14.19. [PubMed] [Article]

Torralba, A., & Oliva, A. (2002). Depth estimation from image structure. *IEEE Pattern Analysis and Machine Intelligence, 24,* 1226–1238.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40A,* 201–237.

Treisman, A. M. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14,* 411–443.

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

Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion—Physiology and psychophysics. *Nature Neuroscience, 3,* 270–276.

Utochkin, I. S. (2013). Visual search with negative slopes: The statistical power of numerosity guides attention. *Journal of Vision, 13*(3):18, 1–14, doi:10.1167/13.3.18. [PubMed] [Article]

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica, 146,* 7–18.

Watson, D. G., Maylor, E. A., & Bruce, L. A. (2005). The efficiency of feature-based subitization and counting. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1449–1462.

Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast. *Psychonomic Bulletin and Review, 18,* 484–489.

Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are *not* the same thing. *Vision Research, 32,* 757–763.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1,* 202–238.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysics of the binding problem. *Neuron, 24,* 11–17.

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.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology, 18,* 34–49.

Yantis, S. (2014). *Sensation and perception.* New York, NY: Worth.

Yurevich, M. A., & Utochkin, I. S. (2014). Distractor heterogeneity effects in visual search are mediated by "segmentability". *Journal of Vision, 14*(10): 921, doi:10.1167/14.10.921. [Abstract]

Zosh, J. M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General, 140,* 141–158.