



Ensemble perception includes information from multiple spatial scales

Timothy D. Sweeny¹ · Andrew Bates¹ · Elric Elias²

© The Psychonomic Society, Inc. 2020

Abstract

Most visual scenes contain information at different spatial scales, including the local and global, or the detail and gist. Global processes have become increasingly implicated in research examining summary statistical perception, initially as the output of ensemble coding, and more recently as a gating mechanism for selecting which information is included in the averaging process itself. Yet local and global processing are known to be rapidly integrated by the visual system, and it is plausible that global-level information, like spatial organization, may be included as an input during ensemble coding. We tested this hypothesis using an ensemble shape-perception task in which observers evaluated the mean aspect ratios of sets of ellipses. In addition to varying the aspect ratios of the individual shapes, we independently varied the spatial arrangements of the sets so that they had either flat or tall organizations at the global level. We found that observers made precise summary judgments about the average aspect ratios of the sets by integrating information from multiple shapes. More importantly, global flat and tall organizations were incorporated into ensemble judgments about the sets; summary judgments were biased in the directions of the global spatial arrangements on each trial. This global-to-local integration even occurred when the global organizations were masked. Our results demonstrate that the process of summary representation can include information from both the local and global scales. The gist is not just an output of ensemble representation – it can be included as an input to the mechanism itself.

Keywords Ensemble coding · Summary statistical perception · Shape perception · Global · Local

Introduction

Most visual scenes contain information at different spatial scales (Palmer, 1977), including the local and global, or the detail and gist. Curiosity about how information is processed and perceived at these levels has been central to the study of visual perception for just about as long as the field has existed. This is evident in the early work of the Gestaltists (Köhler, 1930; Wertheimer, 1923), as well as in more modern research on perceptual organization (Kimchi, 1994; Wagemans et al., 2012), visual search (Wolfe et al., 2011), scene perception (Brady & Shafer-Skelton, 2017; Oliva & Torralba, 2006), and visual awareness (Hochstein & Ahissar, 2002). Local

and global information appear to be processed in parallel (Gerlach & Poirel, 2020), and by distinct neural mechanisms (Bijanzadeh et al., 2018; Liu & Luo, 2019) that operate at different timescales. According to Reverse-Hierarchy Theory, global information is made available to awareness before local information (Campana et al., 2016; Hochstein et al., 2015; Hochstein & Ahissar, 2002). And yet global and local processing interact, with global information altering more localized processing in early cortical areas (Altmann et al., 2003), presumably via feedback connectivity (Angelucci et al., 2017).

Recently, local and global processes have been implicated in research examining summary statistical perception (Cohen et al., 2016; Whitney et al., 2014; Whitney & Yamanashi Leib, 2018). This visual mechanism, also known as ensemble coding, enables perceivers to extract summary statistical information about sets of simple and complex objects (see Whitney & Yamanashi Leib, 2018, for a review), in a fraction of a second (Haberma & Whitney, 2009), with remarkable precision (Alvarez, 2011; Baek & Chong, 2020; Sun & Chong, 2019; Sweeny et al., 2013) and with limited demands

✉ Timothy D. Sweeny
timothy.sweeny@du.edu

¹ Department of Psychology, University of Denver, 2155 S Race St, Frontier Hall, Denver, CO 80210, USA

² Department of Psychology, University of Colorado Denver, Denver, CO, USA

on attention (Ji et al., 2018). In the context of ensemble coding, it is often the case that local processes or analyses are described as occurring first, pertaining to the encoding of individual set members. These local representations then presumably feed into a summary representation of the set via a pooling mechanism, at which point local information is then lost in favor of a percept or judgment at the gist level (Haberma & Whitney, 2009, 2011). This characterization of the ensemble mechanism is consistent with the feed-forward architecture of the visual system, and simulations that feature this type of approach are able to approximate human perception quite well (Allik et al., 2013; Baek & Chong, 2020; Ji et al., 2020; Sweeny & Whitney, 2014; Sweeny, et al., 2015). In this characterization, global-level information is the *outcome* of the ensemble process, not an input.

It is clear, however, that information at the global level can have a profound impact on summary representation. Grouping cues like color (Brady & Alvarez, 2011), similarity, proximity, and common region (Corbett, 2017), sharing a 2D surface (Cha & Chong, 2018), and category membership (Elias & Sweeny, 2020) appear to influence the computation of ensemble codes. Indeed, summary judgments of emotional crowds are known to be more accurate when their members express emotion synchronously, as a collective (Elias et al., 2017). In some cases, the average of a set can even bias the perception (Ross & Burr, 2008) or memory of its constituents (Utochkin & Brady, 2020). These studies demonstrate that information at the global, grouped level, can act as a sort of gating mechanism, influencing which features or objects are integrated into summary representations. They also demonstrate that summary representations can influence how individuals within a set are perceived. However, they do not necessarily indicate that unique global-level information, like the spatial layout of a set, can be included in summary representations.

All sets of objects must have some sort of spatial structure, hierarchy, or global organization. In fact, any set of objects that might be represented as an ensemble should carry some information at the global level. Can information about a set's global organization be included in the summary representation of that set? Might computations of ensemble properties include multiple spatial scales? There are several reasons to predict that they should. First, information at the global scale is known to take precedence in perception (Campana et al., 2016; Kimchi, 2015; Navon, 1977; Nie et al., 2017), potentially being processed more quickly than local information (Gerlach & Poirel, 2020). Second, even though local and global information are known to be processed separately (Bijanzadeh et al., 2018; Hübner & Volberg, 2005; Liu & Luo, 2019), their integration has been proposed to occur automatically, or pre-attentively (Gerlach & Poirel, 2020). Third, information from both the local and the global scales appear to be stored in visual working memory, simultaneously (Brady & Alvarez, 2011), with the global scale being

prioritized (Nie et al., 2017). It therefore follows that when individual objects appear in sets, or groups, their ongoing visual representations should include information about both their local properties *and* the global spatial organization in which they appear. Consequently, we tested the hypothesis that local *and* global information are available to the pooling mechanism at the heart of ensemble coding, and that both can be used, simultaneously, to form summary representations.

We focused on the perception of shapes, and the computation of the ensemble aspect ratio in particular, for a few reasons. First, all global organizations have aspect ratios. For example, a set of square windows might be stacked into a tall organization on a skyscraper. Second, ensemble coding is known to operate for perception of aspect ratio (Elias & Sweeny, 2020). Third, global-to-local interactions occur during the perception of shape (e.g., the spatial organization of a set of objects can bias perception of individual aspect ratios in the set) (Sweeny et al., 2011a; Sweeny et al., 2017). Fourth, interesting local distortions emerge during the perception of aspect ratio. In particular, during brief viewing, the perceived aspect ratios of individual shapes tend to be exaggerated away from the null-point (i.e., a circle or square) toward extreme values (Dickinson et al., 2017; Elias & Sweeny, 2020; Suzuki & Cavanagh, 1998; Sweeny, Grabowecy, Kim, et al., 2011b), such that flat shapes are reported as being flatter than they are, and vice versa. Examining judgments of shape, or aspect ratio, thus provided us with a means to (1) test our hypothesis that information about a set's global organization can be included in an ensemble code, and (2) determine whether this occurs in conjunction with computations known to occur at the local or individual-object level.

Our investigation featured two identical experiments, both of which featured a design in which observers viewed sets of six briefly presented shapes, and on each trial used the method-of-adjustment to report the mean aspect ratio of the entire set. The local aspect ratios of individual shapes within each set varied on each trial, but on average each set had a relatively flat or tall aspect ratio. The six shapes in each set were always spatially distributed such that, overall, the set had either a flat or a tall spatial arrangement. This global organization was always independent of the local aspect ratios of the individual shapes within each set (e.g., a set of tall shapes could have been arranged in a globally flat arrangement). This allowed us to separately measure the extent to which observers incorporated local and global information about aspect ratio into their ensemble judgments.

We also included a masking condition so that we could examine whether global organizations could permeate ensemble representations even when those global patterns were difficult or impossible to see. In this condition, we used four-dot masks to disrupt perception of the two most peripheral shapes in each set. The placement of these two peripheral shapes provided our sets with global flat or tall organizations. By

masking these, but not the other four shapes, we aimed to disrupt perception of global organizations, more generally, and then determine whether these hidden global organizations could nevertheless influence local estimates of ensemble aspect ratio. This question is worth asking – information about global form is known to be processed even when it is suppressed from awareness (Chung & Khoo, 2014; Mudrik et al., 2011), masked global organizations can bias perception of individual shapes (Sweeny et al., 2017), and ensemble representations may include visual information about which a perceiver is unaware (Fischer & Whitney, 2011; Parkes et al., 2001). We note that this question was of secondary interest to us, since answering it would only be of value if we first found an effect of global integration *without* masking (our primary aim).

We made the following predictions: First, observers' estimates of average aspect ratio should be tightly correlated with the actual means of the sets. These estimates of local mean shape should reflect the operation of ensemble representation, including information from multiple set members. Second, estimates of set means should be exaggerated away from the null (circular) value, reflecting a perceptual effect of local repulsion from the category boundary. Such a finding for sets of shapes would dovetail with previous work on judgments of individual shapes (Dickinson et al., 2017; Elias & Sweeny, 2020; Suzuki & Cavanagh, 1998; Sweeny, Grabowecky, Kim, et al., 2011a). Third, we predicted that estimates of set averages should be biased toward the aspect ratio at the global level on each trial. For example, the mean of a set of tall shapes should be reported as even taller when seen in a globally tall spatial configuration. Fourth, we speculated that sets of tall shapes might be most susceptible to influence from the global spatial organizations. This may seem like a surprising prediction, but in fact we found in a previous investigation that global organizations only distorted perception of individual shapes when those shapes had tall aspect ratios (Sweeny, Grabowecky, & Suzuki, 2011b). Finally, in our previous work we found that global organizations biased perception of individual shapes even when they were not visible (Sweeny et al., 2017). We thus predicted that, here, global aspect ratio would still bias perception of a set's average shape in the masked condition, but potentially with reduced strength compared to the unmasked condition.

Experiments 1 and 2

We conducted two identical experiments, both of which featured the same design and analysis, to test our predictions and then examine replicability. Rather than report each experiment separately, we instead present one common Methods section, and then one Results section with analyses from Experiments 1 and 2 presented side-by-side. Our intention was to facilitate

comparisons across the Experiments and focus the narrative only on the findings that replicated.

Materials and methods

Observers We selected the sample size for Experiment 1 based on results of our previous investigation examining effects of global organization on the perception of an individual shape's aspect ratio (Sweeny, Grabowecky, & Suzuki, 2011b). In this previous work, we replicated an effect whereby global organization was assimilated into the perception of tall but not flat shapes, using a sample size of eight in two experiments. The effect size for this result was quite large in both experiments ($\eta_p^2 = 0.403$ and 0.272), but in the current investigation we aimed to examine potential interactions with masking, and our analytical approach was different as well. We thus took a conservative approach and ran 50 observers in Experiment 1 and then ran 50 new observers in Experiment 2 (we had to drop one observer from each experiment due to failure to follow instructions). Observers were undergraduates at the University of Denver and participated for course credit.

This study was approved by the Institutional Review Board at the University of Denver, and all participants gave informed consent before participating in the study.

Stimuli The stimulus set included 27 ellipses (0.2° thick lines) drawn in Adobe Photoshop CS6 v. 13.0 x64, each rendered in dark gray (luminance: 19 cd/m^2) (Fig. 1A). Circular shapes subtended a visual angle of 1.77° . The aspect ratios were symmetrically distributed (in log scale) around the circular value ranging from flat to tall ($-.602, -.556, -.510, -.463, -.417, -.371, -.324, -.278, -.232, -.185, -.139, -.093, -.046, 0.00$ (circle), $0.046, 0.093, 0.139, 0.185, 0.232, 0.278, 0.324, 0.371, 0.417, 0.463, 0.510, 0.556, 0.602$). Note that the appearance of unequal changes in aspect ratio across the stimulus range is due to rounding error. The incremental change between adjacent aspect ratios across the stimulus set was equated, in log units, past the tenth decimal. The areas of all ellipses were equated to the second decimal, and the edges of each ellipse were blurred in Adobe Photoshop using the Gaussian blur tool with a 2-pixel radius.

Flat ellipses present in set displays included the following aspect ratios: $-.463, -.417, -.371, -.324, -.278, -.232, -.185, -.139, -.093, -.046$, and 0.00 (circle). Additionally, at the response stage only, three extremely flat ellipses ($-.602, -.556$, and $-.510$) were available as response options in addition to the rest of the flat ellipses. Tall ellipses present in set displays included the following aspect ratios: 0.00 (circle), $0.046, 0.093, 0.139, 0.185, 0.232, 0.278, 0.324, 0.371, 0.417$, and 0.463 . Again, at the response stage only, three extremely tall ellipses ($0.510, 0.556$, and 0.602) were available as response options in addition to the rest of the

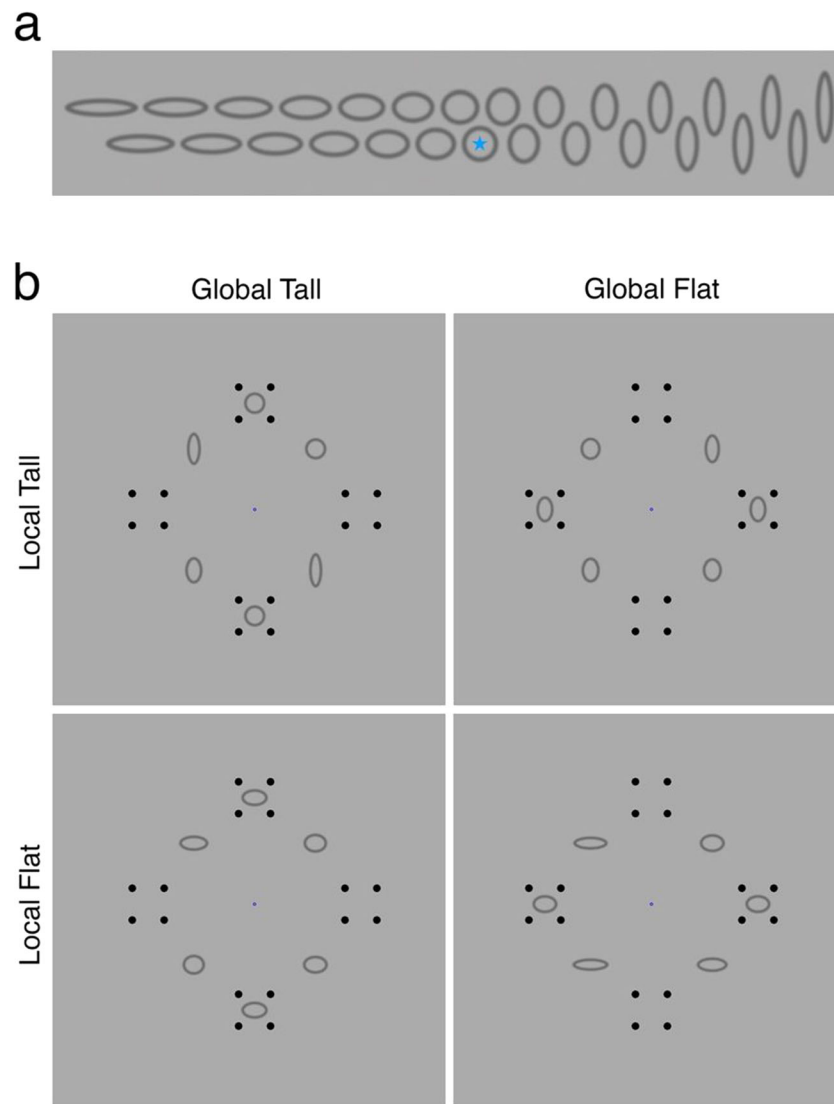


Fig. 1 (A) The full stimulus set of twenty-seven ellipses. (B) Sets of six ellipses depicting the four combinations of local flat or tall average aspect ratios and global flat or tall organizations. For example, in the bottom left display, the individual shapes have flat aspect ratios (although they vary in the extent of their flatness) and they are organized in a global-tall

organization. Note that the arrangement of the shapes produces the global organization; all four quartets of masking dots were presented on each trial, always in a diamond configuration, meaning they never provided information about the global organization or location of the shapes. Thus, only the shapes produced global tall and flat organizations

tall ellipses. This prevented compression in the response stage (see Procedure).

Procedure Each observer was seated in a dimly lit cubicle after providing consent. A researcher then demonstrated a few example trials to the observer in order to illustrate the experimental design. Next, observers were allowed to complete an unlimited number of practice trials until they felt comfortable with the task. The instructions were to “Estimate the average shape. Maintain your gaze on fixation at all times. Move the mouse L or R to adjust response.”

There were nine trial types. Some trials featured the presentation of flat ellipses from the flat range of aspect ratios and the circular value, but not the most extreme values (-0.463 ,

-0.417 , -0.371 , -0.324 , -0.278 , -0.232 , -0.185 , -0.139 , -0.093 , -0.046 , 0.00). We refer to these as local-flat trials. Other trials included only ellipses from the tall range of aspect ratios and the circular value, but not the most extreme values (0.00 , $.046$, 0.093 , 0.139 , 0.185 , 0.232 , 0.278 , 0.324 , 0.371 , 0.417 , 0.463). We refer to these as local-tall trials. For each of these trials (i.e., local-flat or local-tall), the aspect ratios of the six ellipses in the set were randomly selected from the ranges listed above (i.e., six aspect ratios were randomly selected from the flat and circular values for a given local-flat trial).

These local-flat and local-tall trials were fully crossed with flat and tall global organizations, producing four trial types. Local-tall/Global-tall trials included six ellipses with tall aspect ratios arranged in a globally-tall spatial organization (see

the top-left panel of Fig. 1B). Local-flat/Global-tall trials included six ellipses with flat aspect ratios arranged in a globally-tall spatial organization (bottom-left panel of Fig. 1B). Local-flat/Global-flat trials included six ellipses with flat aspect ratios arranged in a globally flat spatial organization (bottom-right panel of Fig. 1B). Local-tall/Global-flat trials included six ellipses with tall aspect ratios arranged in a globally flat organization (top-right panel of Fig. 1B). Each of these four trial types was crossed with our masking manipulation (which included masked and unmasked conditions), producing eight trial types in what we refer to as the *multiple-shape condition*.

We also included a *single-shape condition*, which featured the presentation of a single ellipse. On these trials, a set of six ellipses was generated as if for a multiple-shape trial, but only a single ellipse was randomly selected from this set and then displayed on the screen at a random location. This single-shape condition served as a control condition that allowed us to examine whether observers used ensemble coding, averaging information from more than one shape, to make estimates of the sets, or if they based estimates on multiple-shape trials from one shape in a given set (see Results). Each observer completed 50 trials from each of the nine trial types, and 450 trials overall.

Each trial began with the presentation of a blue fixation circle (0.31° visual angle) at the center of the screen. Observers were instructed to keep their eyes fixed at this point, but to let their attention spread across the entire screen. On each trial from the multiple-shape condition, six ellipses appeared on the screen for 60 ms. Four shapes were presented around the fixation circle, with locations to the upper left, upper right, bottom right, and bottom left, with the centroid of each ellipse 7.15° from the fixation circle (Fig. 1B). The centroids of each of these four shapes were 9.9° from each other. The fifth and sixth shapes in a set could appear at two of four peripheral locations, each of which was 8.98° from fixation. These locations included positions directly above, to the right, below, and left of fixation, with the centroid of each shape 12.1° from each other, and 6.12° from the shapes in the central locations. Crucially, the fifth and sixth shapes always appeared in the top and bottom locations or the left and right locations. In this way, when combined with the four central shapes, the six shapes formed a globally flat or tall organization (Fig. 1B). On trials from the single-shape condition, the one visible shape appeared randomly at any of the locations.

All trials from the multiple-shape and single-shape conditions included the presentation of quartets of black masking dots surrounding the four peripheral locations, regardless of whether shapes appeared at those locations (see Fig. 2). We elected to use object-substitution masking (i.e., OSM; Enns, 2004; Enns & Di Lollo, 1997; Goodhew et al., 2013) because it was useful for disrupting visual awareness of our shape

stimuli in a previous investigation (Braun & Sweeny, 2019; Elias et al., 2018), but we acknowledge that other forms of masking (e.g., metacontrast or backward masking) could have met our needs as well. We note that this was not an investigation of OSM, so we limit our discussion of its mechanisms and simply note that we selected dot sizes, distances, and timing parameters based on values that produced effective masking in our previous work (Braun & Sweeny, 2019; Elias et al., 2018).

By placing masking dots at all four peripheral locations, we ensured that these dots did not contribute to a globally flat or globally tall organization on any trial (only the shapes produced these organizations). On unmasked trials in the multiple-shape condition, these masking dots were displayed on the screen for the same amount of time as the ellipses, on-setting and off-setting simultaneously. On masked trials in the multiple-shape condition, the masking dots onset with the ellipses, but then remained on the screen for an additional 100 ms after the ellipses disappeared. All trials from the single-shape condition were unmasked. Each masking dot subtended a visual angle of 0.63° and appeared 1.82° from the centroid of each shape. Masking dots were 2.6° apart from each other.

At the end of each trial, participants used the method-of-adjustment to adjust a response ellipse presented at the center of the screen to report the average aspect ratio of the entire set of ellipses on multiple-shape trials (which we refer to as the local mean), or the aspect ratio of the single ellipse on single-shape trials. Observers moved a mouse leftward and rightward to adjust the aspect ratio of the response ellipse. The starting aspect ratio of the response ellipse was randomly selected from the 27 values in the stimulus set on each trial. The aspect ratio of the response ellipse was free to cycle across the entire range of 27 aspect ratios, and it stopped adjusting once each observer reached either the lower or the upper limit of the range. The range of response aspect ratios was greater than the range of actual shape values on any trial so that observers would be free to overestimate perceived shape values, and thus avoid compression (artificial clumping of responses away from the endpoints of the range) in the response stage. After the observer reported the average aspect ratio of the set or single shape by clicking a button on the mouse, the response ellipse disappeared and was replaced by a backward mask, which was an image of the circular shape divided into a 54×54 grid, scrambled, and shown for 250 ms. A blank screen then appeared for a duration between 800 ms and 1,200 ms, randomly selected from a uniform distribution.

Experiments were conducted on a CRT monitor with a refresh rate of 100 Hz at a viewing distance of 55 cm. Observers were given two breaks at the one-third and two-thirds marks of the experiment. Stimuli were presented against a uniform gray background (RGB value = 170, 170, 170; luminance = 41.5 cd/m^2). Experiments were coded and run

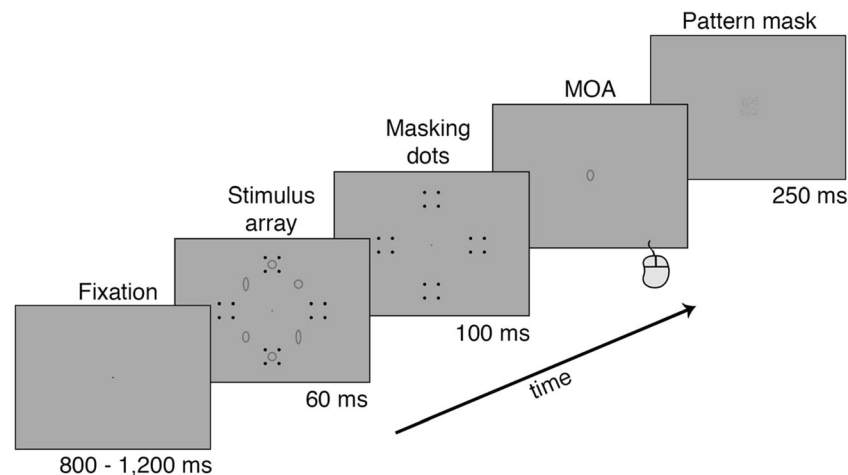


Fig. 2 A typical trial sequence. In this example, a set of six shapes is shown in the context of a masking trial, but the stimulus array could have included just one shape on single-shape trials, and on no-masking trials the masking dots would not have been shown after the offset of the stimulus array

using MATLAB (Release 2014b; The MathWorks, Natick, MA, USA) with the Psychophysics Toolbox (Brainard, 1997).

Results

Multiple regression analysis Our primary analysis featured one multiple regression (conducted using R) on data from trials with sets of flat shapes and another multiple regression on data from trials with sets of tall shapes (see rationale for separating these analyses below). Each regression equation predicted perceived-aspect-ratio-of-the-set with fixed effects of intercept, local-mean-aspect-ratio-of-the-set, global organization, and masking ($y = \text{int} + \text{mean} * \text{global} * \text{mask}$).

The means of the local aspect ratios of the sets in our design were not evenly distributed across the full range of aspect ratios available in our stimulus range (i.e., there were no trials in which the means were at or close to the circular value; see the top density panels in Fig. 3). Rather, the means of the local aspect ratios of the sets were clearly bimodally distributed. This was intentional, because recent evidence indicates that the precision of ensemble coding is lower for sets of shapes with aspect ratios that cross the flat-tall category boundary than for sets that do not, and the perception of variance of such sets is also greater (Elias & Sweeny, 2020). However, this meant that it would have been inappropriate to run a single multiple regression on the full dataset. If we had done so, for example, a single linear fit would have had an intercept close to zero, running through the middle of the data and clearly missing the offsetting intercepts evident in Fig. 3, obtained by running regressions separately for trials with sets of flat and tall shapes. We thus conducted separate multiple regressions for trials with sets of flat shapes and trials with sets of tall shapes.

Mostly for illustration purposes, we depicted every data point from all observers from the multiple-shape condition

in Fig. 3, with reported aspect ratio of the set (our dependent variable) shown as a function of mean of the local aspect ratio of the set. A few things are notable even from a quick visual inspection. First, observers were clearly able to perform the task (the two variables were positively correlated). Second, the distribution of reported aspect ratios was bimodal (although observers did report some sets with means around zero). Third, reported average aspect ratios appear to have been exaggerated, especially for trials in which the local mean aspect ratio of the set was close to the flat/tall category boundary.

The regression weights (or β values) for our model fits indicated the extent to which the set's actual local mean aspect ratio, the set's global organization, the presence of masking, or their interactions, influenced judgments of the set's average aspect ratio. Values near zero would indicate no influence, whereas positive (or negative) values would indicate a positive (or negative) relationship between any variable and reported aspect ratio of the set. Tables 1 and 2 report the β values, p -values, and 95% confidence intervals for these estimates, for each variable and interaction, for both Experiment 1 and Experiment 2, separately for flat sets (Table 1) and tall sets (Table 2). In the interest of simplicity and avoiding over-explaining our data, we now describe in more detail *only* the β values that were statistically significant in both experiments. For flat sets, we found significant effects of intercept and local-mean-aspect-ratio-of-the-set across both experiments (see Table 1). For tall sets, we found significant effects of intercept, local-mean-aspect-ratio-of-the-set, as well as global organization across both experiments (see Table 2).

The effects of intercept (also evident in Fig. 3) reflect a phenomenon of exaggeration. That is, the average aspect ratios of flat sets were reported to be flatter than they actually were, and the average aspect ratios of tall sets were reported to be taller than they actually were. The effects of mean (local

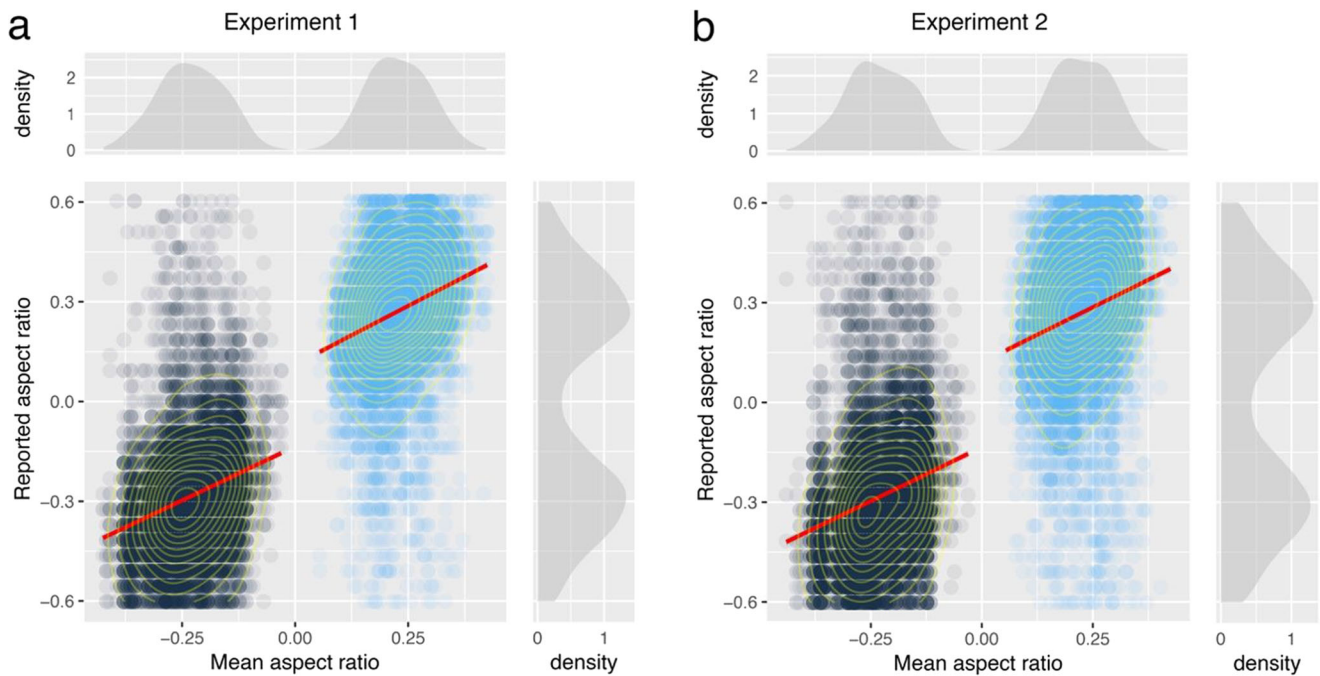


Fig. 3 Raw data from Experiment 1 (A) and Experiment 2 (B) depicting the relationship between the mean of the local aspect ratio of each set (x-axis) and reported average aspect ratio (y-axis). Data are from the multiple-shape condition (trials from the single-shape condition are not included here). All trials from all observers are shown at once, as if from a single observer. Trials from sets with flat shapes are depicted in dark blue, and trials from sets with tall shapes are depicted in light blue. Red lines

depict separate linear fits to the data from trials with flat and tall shapes. The density panels above and to the right of each scatterplot depict the distributions of displayed and reported aspect ratios, respectively. The offsetting intercepts of the linear fits clearly illustrate the fact that the reported average aspect ratios of the sets tended to be exaggerated from their actual mean aspect ratios

mean of the aspect ratios of the set) are reassuring, although not surprising, and simply reflect the fact that as the average aspect ratios of sets became flatter (or taller), so too did judgments of those set's aspect ratios. Flat and tall global organizations biased judgments of set means in these same directions, but only when the individual shapes within those sets had tall aspect ratios. Although the specificity of this effect to tall shapes may seem surprising, we did, in fact, suspect that it

would occur based on our previous work, and it replicated across two experiments here. Finally, we note that we did not expect a main effect of mask in our multiple regression because the presence (or absence) of masking should not have made shapes appear to be flatter or taller. We included this variable in the multiple regression mainly to determine if it consistently interacted with the effect of global organization, which it did not.

Table 1 Regression weights (i.e., β values), p-values, and 95% confidence intervals for fixed effects and their interactions from our multiple regression examining data from trials with sets of flat shapes, from both Experiment 1 and Experiment 2

Experiment 1				Experiment 2			
IV	β	p-value	95% CI	β	p-value	95% CI	
Intercept	-.131	*<.001	-.147	-.114	*<.001	-.135	-.099
Mean	.619	*<.001	.551	.687	*<.001	.589	.737
Global	-.023	*.004	-.04	-.007	.812	-.015	.02
Mask	-.0009	.42	-.03	-.01	*.008	-.05	-.008
Mean \times Global	-.1	*.003	-.17	-.03	.901	-.07	.069
Mean \times Mask	.05	.25	-.03	.15	.543	-.13	.07
Global \times Mask	.001	.9	-.02	.02	*.019	.05	-.004
Mean \times Mask \times Global	.01	.76	-.08	.11	*.043	-.21	-.003

Note: Significant effects are noted by an asterisk (*) next to the corresponding p-value. Effects that were significant in both experiments are in bold type

Table 2 Regression weights (i.e., β values), p -values, and 95% confidence intervals for fixed effects and their interactions from our multiple regression examining data from trials with sets of tall shapes, from both Experiment 1 and Experiment 2

Tall sets								
Experiment 1					Experiment 2			
IV	β	p -value	95% CI		β	p -value	95% CI	
Intercept	.096	*<.001	.07	.11	.123	*<.001	.103	.142
Mean	.713	*<.001	.64	.78	.597	*<.001	.51	.67
Global	.026	*.003	.008	.04	.019	*.04	.002	.039
Mask	.029	*.018	.005	.05	-.005	.71	-.032	.02
Mean \times Global	-.105	*.004	-.17	-.03	-.053	.204	-.13	.02
Mean \times Mask	-.01	.77	-.11	.08	-.012	*.03	-.012	.24
Global \times Mask	-.004	.69	-.02	.01	-.003	.82	-.024	.03
Mean \times Mask \times Global	.05	.33	-.05	.15	.026	.649	-.14	.08

Note: Significant effects are noted by an asterisk (*) next to the corresponding p -value. Effects that were significant in both experiments are in bold type

Alternative analyses – local and global distortion The β weights listed above describe how strongly, and in which direction, each variable influenced perception of set means. Yet these sorts of results can sometimes feel opaque, or less accessible than the sorts of results one finds when collapsing across conditions and running simple contrasts. We now present a subset of our data using this latter approach in order to provide a different (yet consistent) perspective on some of our findings. Specifically, we focused on the signed error (i.e., too flat or too tall) on multiple-shape trials (each observer's judgments of the set mean as a function of the combination of local aspect ratio and the global configuration of the set; the

combinations are depicted in Fig. 1B, like local-flat/global-flat, etc.). These difference scores are depicted in Fig. 4. First, we examined the extent to which perception of the set was exaggerated away from the category boundary of null aspect ratio (e.g., a circle). For each combination of local and global aspect ratios, the average error relative to the true set mean was significantly different from zero (Experiment 1: all t -values > 4.63 , all p -values for one-sample t -tests $< .001$, all Cohen's d values $> .66$; Experiment 2: all t -values > 3.29 , all p -values for one-sample t -tests $< .002$, all Cohen's d values > 0.47). More important, perceived aspect ratio was always distorted in the direction of the local aspect ratios in the set.

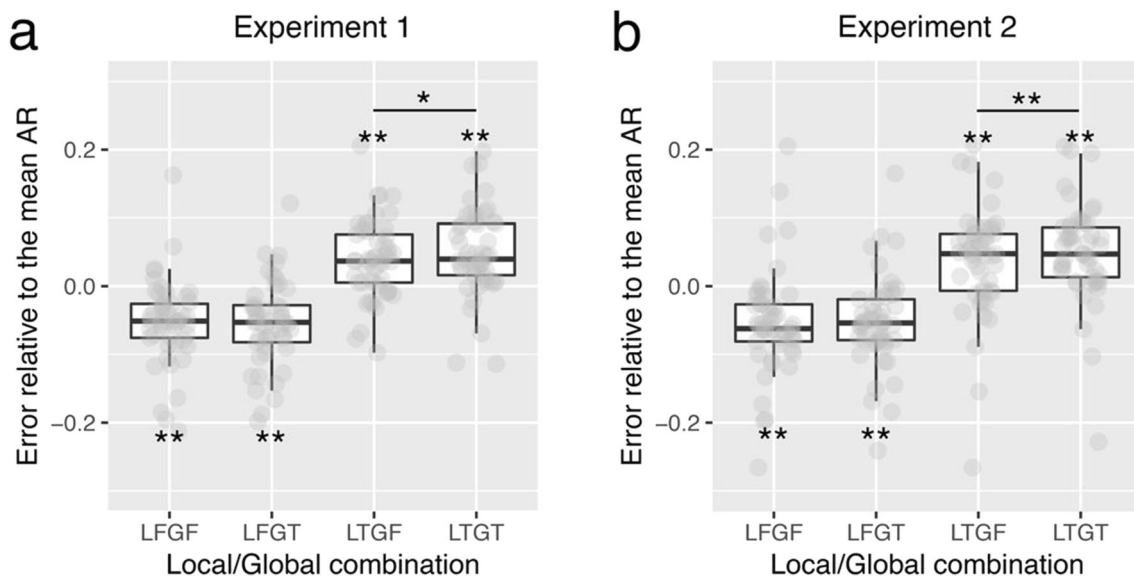


Fig. 4 Average error of judgments relative to the mean aspect ratio as a function of local and global aspect ratio combinations from Experiment 1 (**A**) and Experiment 2 (**B**). Values represent data collapsed across masking conditions. Each gray dot in each condition represents average signed error from one observer. Each boxplot shown behind the dots

depicts the median and interquartile range. We elected to present our data in this format in order to be as transparent as possible, even though their appearance may not necessarily convey the significance of the within-observer comparisons. * indicates $p < .05$, ** indicates $p < .01$

So, for example, if a set contained flat shapes, regardless of the global organization (flat or tall), the perceived mean aspect ratio of the set was perceived to be flatter than it was (e.g., the LFGF and LFGT conditions in Fig. 4A). This effect occurred in both Experiment 1 and Experiment 2, and it occurred in addition to the effect of global organization. These effects reflect the same underlying mechanism behind the effects of intercept in the multiple regressions described above.

Next, we examined the effect of global organization. Signed errors in the local-tall/global-tall condition (LTGT) were greater than those in the local-tall/global-flat condition (LTGF) both in Experiment 1, $t(48) = 2.46$, $p = .01$, $d = .35$, and in Experiment 2, $t(48) = 3.36$, $p < .01$, $d = 0.47$ (Fig. 4). Comparisons between the LFGF and LFGT conditions were non-significant in Experiment 1, $t(48) = 0.72$, $p = .46$, $d = .1$, and in Experiment 2, $t(48) = 0.34$, $p = .74$, $d = 0.04$. These effects reflect the same underlying mechanism behind the global effects in the multiple regressions described above. These effects may appear subtle in the overlapping distributions in Fig. 4, but the differences within observers were nonetheless real and reliable.

It is worth considering that the effects of exaggeration away from the category boundary (described at the beginning of this section) may not have reflected a true perceptual distortion. After all, if observers had simply noted whether shapes were flat or tall in a given set and then correctly responded with an aspect ratio from the *middle* of the flat or tall response range (which was in fact more extreme than the average flat or tall set, because the response range was extended), then, artifactually, errors relative to the true set mean could have appeared exaggerated, like the patterns in Fig. 4. This would have *not* been the case, however, for trials in which the mean aspect ratios of the sets were very flat or very tall. On these trials, responding from the middle of the flat or tall range would have produced a pattern of data consistent with perceptual attraction, with positive errors for flat trials and negative errors for tall trials. We thus re-examined error-relative-to-the-mean only for trials in which the mean aspect ratio was very flat (less than $-.324$) or very tall (greater than $.324$). We found the same pattern of results for both Experiment 1 and Experiment 2 (Fig. 5), whereby aspect ratios were numerically exaggerated from the circular value, albeit not significantly (Experiment 1; flat trials, $t(48) = -1.08$, $p = .284$, $d = .15$, and tall trials, $t(48) = 1.75$, $p = .08$, $d = .25$, Experiment 2; flat trials, $t(48) = -1.28$, $p = .2$, $d = 0.18$, and tall trials, $t(48) = 1.87$, $p = .067$, $d = 0.27$). Thus, the effects of perceptual exaggeration described above appear not to have been due to a response artifact.

Mean versus median We now describe the results of planned comparisons designed to reveal insights about what kind of summary information observers used to make their judgments about average aspect ratio. First, we examined whether responses more closely reflected the mean or median aspect

ratio of the sets. For each observer, we recorded the signed error of their estimate relative to the mean and median of the set on each trial, and then recorded the standard deviation of each distribution across all trials. We then compared the average standard deviation of these error distributions, across observers, when made relative to the mean or the median. Figure 6 illustrates that errors calculated relative to the mean were lower than those calculated relative to the median, both for Experiment 1, $t(48) = -14.91$, $p < .001$, $d = 2.13$, and for Experiment 2, $t(48) = -17.11$, $p < .001$, $d = 2.44$.

Ensemble coding Next, we examined whether responses about the set means were arrived at by considering the aspect ratios of multiple shapes at once (i.e., ensemble coding) or if instead they simply reflected a process of randomly selecting and reporting the aspect ratio of one shape from each set. Recall that we included a control condition – the *single-shape condition*. We included this condition specifically for this analysis because it allowed us to determine what performance would have looked like had observers evaluated the sets based on a single randomly selected shape. On these trials, we generated sets of six shapes just as in the multiple-shape condition (and recorded the actual mean aspect ratio of the set), but then displayed only one randomly selected shape (and recorded that single shape's aspect ratio as well).

We analyzed the data from trials in the single-shape condition in two ways. In the *crowd-via-subset analysis (CvS)*, we recorded the difference between each observer's response (which could only have been based on the single visible shape) and the mean of the set of six shapes (even though observers could see only one shape from the set) on that trial. Then for each observer, we calculated the standard deviation of their distribution of errors across all trials in the single-shape condition. This calculation simulated what performance would have looked like in the multiple-shape condition if observers had based their responses on a single shape from each set. Of course, we could not have expected observers to make judgments about sets that they could not see. Rather, this analysis was analogous to an empirical simulation of what performance in the multiple-shape condition would have looked like had observers not engaged ensemble coding, and instead made their judgment based on one random shape per set.

In the *single-via-single analysis (SvS)*, we recorded the difference between each observer's response and the aspect ratio of the single visible shape on every trial. We then calculated the standard deviation of their distribution of errors across all trials. This calculation allowed us to measure baseline sensitivity for estimating aspect ratios of individual shapes.¹

¹ Scatterplots depicting the relationship between actual aspect ratio and perceived aspect ratio for the single-shape trials, from Experiments 1 and 2, are available in the Online Supplemental Materials.

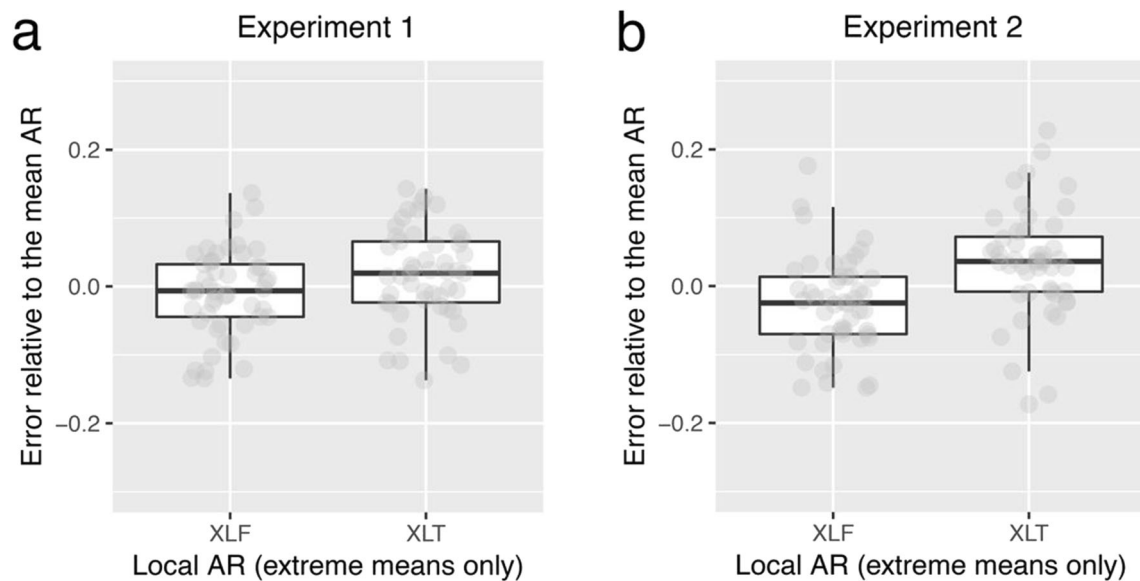


Fig. 5 Average error of judgments relative to the mean aspect ratio calculated only using trials with extremely flat (XLF) or extremely tall (XLT) sets from Experiment 1 (A) and Experiment 2 (B). Each gray dot in each condition represents average error from one observer. Each boxplot shown behind the dots depicts the median and interquartile range.

Finally, we performed a *crowd-via-crowd analysis (CvC)* using data from the multiple-shape condition. Here, we recorded the difference between each observer's response and the mean aspect ratio of the entire set (which was visible, in this case) on every trial. We then calculated the standard deviation of each observer's distribution of errors across all trials.

For illustration purposes, distributions of errors from these three analyses (using data from all observers pooled into one

If observers had simply responded using the middle of the flat and tall response ranges on flat and tall trials, respectively, then errors on XLF trials would have been positive, and errors on XLT trials would have been negative. Instead, observers produced the same overall pattern of exaggeration away from the actual aspect ratio of each set

distribution per condition) are shown in the top panels of Fig. 7. Recall that these distributions were built from the errors observers produced on each trial – each value reflected the difference between each observer's response and the mean of the set of six shapes (or the single shape's aspect ratio), with negative values indicating a response that was too flat, and positive values indicating a response that was too tall. Narrow error distributions, of course, indicate sensitive

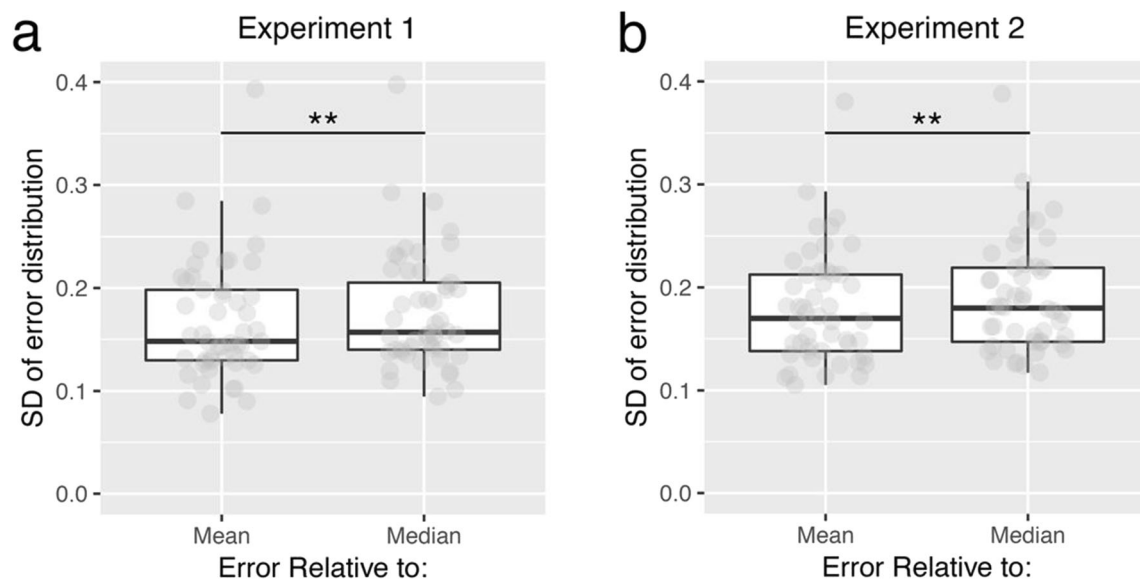


Fig. 6 Standard deviations (SDs) of distributions of error of judgments calculated relative to the mean or median of sets from Experiment 1 (A) and Experiment 2 (B). Each gray dot in each condition represents the

average error from one observer. Each boxplot shown behind the dots depicts the median and interquartile range. ** indicates $p < .01$

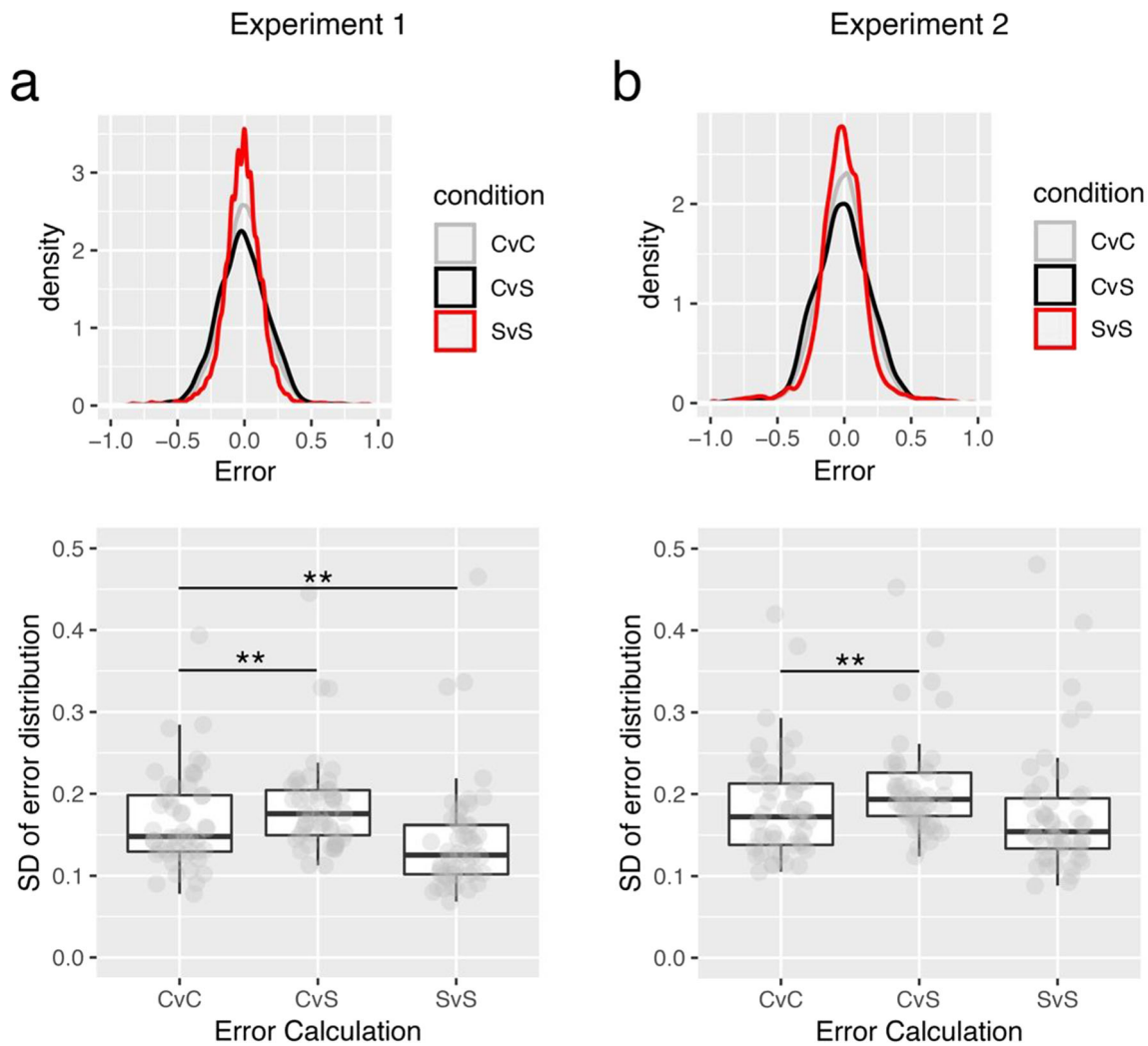


Fig. 7 Standard deviations (SDs) of distributions of error of judgments from Experiment 1 (**A**) and Experiment 2 (**B**). In the CvC analysis, errors were calculated by comparing estimates of the set of six shapes' mean aspect ratios with the actual means of those sets. In the CvS analysis, errors were calculated by comparing estimates of single visible shapes' aspect ratios with the actual means of the full (albeit hidden) sets of six. In the SvS analysis, errors were calculated by comparing estimates of the

single shape's aspect ratio with the actual aspect ratio of that single shape. The histograms at the top of each panel illustrate the distributions of errors that resulted from these analyses collapsed across all observers, for each condition. The boxplots at the bottom of each panel illustrate the SDs that resulted from these analyses with each gray dot representing an SD calculated for each observer. Each boxplot shown behind the dots depicts the median and interquartile range. ** indicates $p < .01$

perception or shape, and distributions centered on zero indicate lack of bias in reporting flat or tall aspect ratios. Most important, if observers utilized ensemble coding, their distributions of errors from the multiple-shape trials (the CvC analysis) should have been narrower than their distributions from the crowd-via-subset analysis from the single-shape trials (the CvS analysis). This should have occurred despite high baseline sensitivity for perceiving aspect ratio, and very narrow error distributions, in the single-via-single (SvS) analysis. And indeed, this is exactly what we found.

Paired-samples t -tests confirmed that, on average, the SDs of error distributions from the crowd-via-crowd (CvC) analysis were narrower than those from the crowd-via-subset (CvS) analysis. This was true for both Experiment 1, $t(48) = -3.97$, p

$< .001$, $d = 0.56$, and Experiment 2, $t(48) = -4.06$, $p < .001$, $d = 0.57$. This suggests that observers used the aspect ratios of multiple shapes to make evaluations about the means of the sets in the multiple-shape condition. Furthermore, performance in the single-shape condition was quite good, with observers producing narrower error distributions in the single-via-single (SvS) analysis than in the crowd-via-crowd (CvC) analysis in Experiment 1, $t(48) = -4.33$, $p < .001$, $d = 0.61$. The same pattern emerged in Experiment 2, but it did not reach statistical significance, $t(48) = -1.56$, $p = .12$, $d = 0.23$.

Scope of integration and masking We examined two questions in our final analysis. First, having now confirmed that observers used multiple shapes to estimate the set means, we

asked: did observers arrive at these summary representations by integrating the aspect ratios of all six shapes in each set, or did they instead base their judgments exclusively on the central four shapes? Second, if observers were able to integrate information from all six shapes, did this depend on whether the fifth and sixth shapes in the set were masked? We addressed these questions directly, and simultaneously, by conducting paired-samples *t*-tests among four conditions, with means defined by the following approach. For each observer and for each trial from the multiple-shape condition, we calculated the error of their response relative to the mean of the set as determined by the central four shapes or the mean as determined by all six shapes. We did this separately for trials from the masked and unmasked conditions. As in our previous analyses, we then calculated the standard deviation of the distribution of errors from each of these four types of analyses. We thus obtained a single value of error distribution *SD* for each observer for the following four conditions: error-versus-central-four/peripheral-masked (4-M), error-versus-central-four/peripheral-unmasked (4-UM), error-versus-all-six/peripheral-masked (6-M), and error-versus-all-six/peripheral-unmasked (6-UM).

We found that, regardless of masking, and in both experiments, errors were smaller (i.e., *SDs* of error distributions were lower) when calculated relative to the mean of all six shapes than when calculated relative to the mean of the central four shapes (Fig. 8). In Experiment 1, *SDs* from the 6-M condition were lower than the *SDs* from the 4-M condition, $t(48) = -6.18$, $p < .001$, $d = 0.88$, and *SDs* from the 6-UM condition were significantly lower than *SDs* from the 4-UM condition, $t(48) = -4.72$, $p < .001$, $d = 0.67$. Likewise, in Experiment 2, *SDs* from the 6-M condition were lower than the *SDs* from the 4-M condition, $t(48) = -5.86$, $p < .001$, $d = 0.84$, and again, *SDs* from the 6-UM condition were significantly lower than *SDs* from the 4-UM condition, $t(48) = -3.58$, $p < .001$, $d = 0.51$. These data suggest that observers used information from all six shapes to estimate the mean of a set, and that this occurred even though the fifth and sixth shapes in each set were in the visual periphery, and in some cases masked.

Discussion

Local analyses of individual objects and global analyses of spatial organization co-occur during the perception of sets and groups. Here, we have shown that these local and global analyses are incorporated into summary representations about those sets. Replicating our recent work (Elias & Sweeny, 2020), we found that observers were adept at summarizing the average aspect ratios of sets of shapes. These estimates followed the means of the sets more closely than the medians, and they reflected information from multiple shapes in each set. Again, replicating our recent work, estimates of mean aspect ratio were distorted away from the category boundary,

making tall sets appear taller than they actually were, and vice versa. Most important, though, was our novel finding that estimates of average aspect ratio were biased toward the global spatial organizations of the sets. This effect of global integration did not depend on whether the spatial organization of the set as a whole was masked, or unmasked.

We have shown that ensemble codes can include information from multiple spatial levels of analysis. This finding is important, but not because global organizations always carry meaningful information for making summary judgments (the shape of a crowd of faces is unlikely to have any relevance for a judgment about their average emotion, for example). Rather, our findings pertain more broadly to the ensemble mechanism itself – they clarify what kinds of visual information can be included in ensemble codes, and they reposition the mechanism more comfortably with decades of work indicating that local and global processing interact, with global- or gist-level information taking precedence (Kimchi, 2015; Navon, 1977; Nie et al., 2017) or being available to awareness first (Gerlach & Poirel, 2020). Our results suggest that summary representations are formed at a timepoint *after* the parallel and distinct processing of local and global information is complete (Flevaris & Robertson, 2016; Hübner & Volberg, 2005). Finally, our work adds to recent findings indicating that information at the global level, or the “gist,” is not just the output of ensemble coding. Grouping appears to gate the process of selecting which objects contribute to summary representations (Brady & Alvarez, 2011; Cha & Chong, 2018; Corbett, 2017; Elias & Sweeny, 2020), or the precision of those representations (Elias et al., 2017). We have demonstrated something novel – holistic information at the global level can serve as an input for, and be included within, the ensemble computation as well.

How might global organizations be included in summary computations? One possibility is that spatial organizations are incorporated into the sensory representations of individual shapes, subsequently distorting their perception, but only after information from each spatial scale is processed separately. Indeed, this type of global-to-local distortion was recently demonstrated for the perception of orientation (Campana et al., 2016), and we previously verified that this can occur during perception of shape (Sweeny, Grabowecky, & Suzuki, 2011b), presumably via the operation of feedback connectivity from higher-to-lower visual areas. Individual cells' responses can be driven by stimulation outside their classical receptive fields (Allman et al., 1985), and high-level areas like LOC, which are sensitive to global shapes, appear to provide information about spatial organization to retinotopic areas like V1-V4 via feedback connectivity (Altmann et al., 2003). It may be the case, for example, that separate encoding of the global and local properties of our sets occurred initially via analysis of lower- and higher-spatial frequency-tuned channels, prior to integration at a later point in time (e.g.,

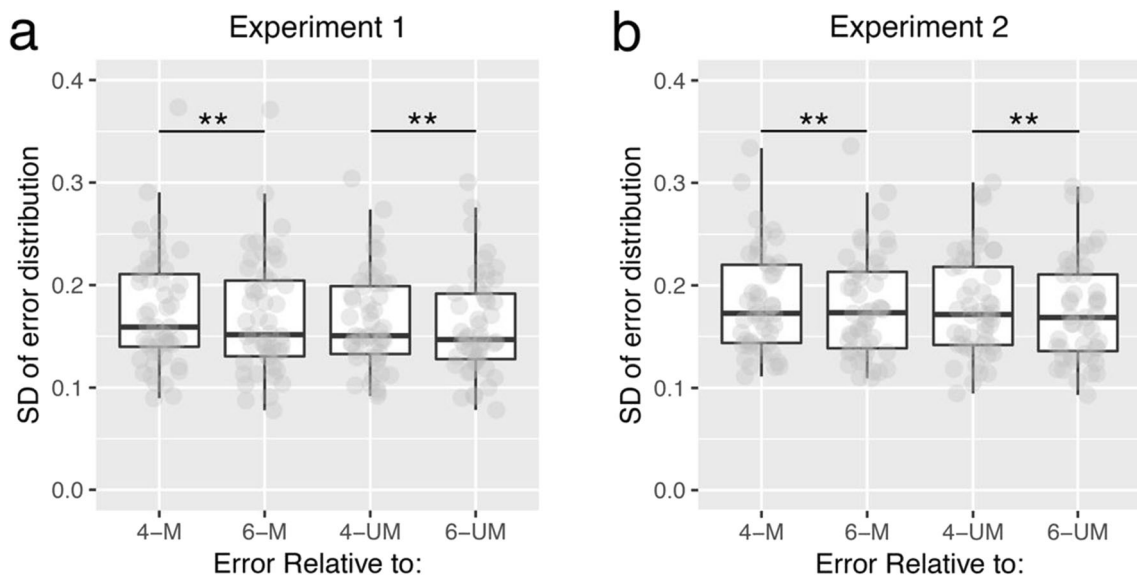


Fig. 8 Standard deviations (SDs) of error distributions of judgments calculated relative to the mean of the central four shapes (4) or all six shapes (6) in each set as a function of whether the fifth and sixth shapes were masked (M) or unmasked (UM). Data are shown separately from

Experiment 1 (**A**) and Experiment 2 (**B**). Each gray dot in each condition represents average error from one observer. Each boxplot shown behind the dots depicts the median and interquartile range. ** indicates $p < .01$

Flevaris & Robertson, 2016). A second possibility is that information about the local aspect ratios and the global organizations in our sets were encoded simultaneously, perhaps even by the same population of neurons. Individual cells in the inferotemporal cortex can be tuned to respond to particular aspect ratios, especially extremely flat and tall aspect ratios (Kayaert et al., 2005; Op de Beeck et al., 2003; Stankiewicz, 2002). At the neural population level, aspect ratio may be represented by an opponent-coding scheme (Regan & Hamstra, 1992; Suzuki, 2005), although recent work suggests a multi-channel approach may be more appropriate (Dickinson et al., 2017; Storrs & Arnold, 2017). Crucially, aspect ratio-tuned cells are also relatively invariant to an object's size (Regan & Hamstra, 1992). So theoretically, a cell tuned to taller aspect ratios could respond to *both* the tall items in a set and the tallness of the set at the global level, at the same time, obviating the need for feedback. Finally, global organizations may have biased the local representations of individual shapes in visual working memory (Brady & Alvarez, 2011). Indeed, local and global-level information from hierarchical stimuli have been shown to be stored in visual working memory, simultaneously, with a bias for global features (Nie et al., 2017). None of these explanations are mutually exclusive, although they carry different implications for when global information is incorporated into summary representation. It may be the case that when perceivers are asked to make summary judgments about sets of objects, they base these judgments on a single ensemble computation produced after the set has disappeared, drawing from lingering representations in visual short-term memory. Or, they may produce multiple ensemble computations (Yashiro et al.,

2020), with some occurring closer to initial sensory encoding, and some including more emphasis on the global properties of the set.

We found no evidence that the integration of global organizations into ensemble representations depended on whether those global organizations were masked. This is consistent with recent findings that information about global form can be processed even when it is suppressed from awareness (Chung & Khuu, 2014; Mudrik et al., 2011), as well as our previous work in which we found that global organization biased perception of individual shapes, even when they were masked out of awareness (Sweeny et al., 2017). Ensemble representations have been shown to sometimes include visual information about which a perceiver is unaware (Fischer & Whitney, 2011; Parkes et al., 2001). If global information is indeed processed more quickly than local information, and then integrated with local information automatically, or pre-attentively (Gerlach & Poirel, 2020), then global organizations should penetrate ensemble representations quickly and easily, as we found here, and when awareness is impoverished or disrupted. However, we want to point out that we cannot be certain that our masking manipulation prevented observers from becoming subjectively aware of the global organizations of the full sets, at least not on every trial. When designing our task we elected not to ask observers to report on their awareness of the peripheral shapes because this could have changed the way observers distributed their attention on each trial, potentially disrupting their attention to global organization and integration of global and local cues (Flevaris & Robertson, 2016). Based on our previous work with similar stimuli and a nearly identical masking procedure (Braun &

Sweeny, 2019), it is in fact likely that observers were sometimes aware of the masked shapes. However, a post hoc examination of our results suggested that the global effect we report here truly appears to owe little to visual awareness.² It is also notable that we never asked observers to make judgments about the global organizations or even to pay attention to them. This suggests that, like distribution shape (Chetverikov et al., 2016), global organizations may influence ensemble judgments quite easily and without explicit knowledge about them.

Our investigation did feature a few limitations. First, we produced a peculiar effect whereby global organizations only biased the perception of sets with tall shapes. This was not unexpected; in a previous investigation, we showed that when a pair of ellipses was seen side-by-side (producing a globally horizontal organization), or one-above-the-other (producing a globally vertical organization), the perception of the individual shapes in each pair was biased toward these global aspect ratios, but only when the individual shapes were tall (Sweeny, Grabowecky, & Suzuki, 2011b). Yet the mechanisms of this effect are just as unclear now as they were in our previous investigation. Cells tuned to aspect ratio do provide the basic input for visual representation of faces (Tsao et al., 2006; Young & Yamane, 1992), which tend to have tall aspect ratios. It may be that expertise discriminating faces facilitates integration of global-to-local information, but only for tall shapes. This is, of course, speculation. Second, we only examined perception of aspect ratio. It is unclear if the pattern of results we found here would occur for other visual features that are likewise ensemble coded and capable of producing conflicting information at the local and global levels, like orientation (Campana et al., 2016). Examining how the current findings compare to those with other visual features could shed additional light on mechanisms. Finally, even the local elements in our sets had global shapes. That is, the individual shapes in each set were closed contours, and thus has global organizations. It would thus be appropriate to say that we examined global information at two levels of organization, with the more global of the two levels obtaining its holistic aspect ratio via grouping. The same critique can of course be made about classic hierarchical stimuli, and in any case, this should not be a concern. Perceptual organization is hierarchical (Palmer, 1977), and we have demonstrated that so too is ensemble representation.

² If effective masking on some trials actually reduced the strength of global integration, the effect of global organization seen in Fig. 4 (LTGT-LTGF) should have at least been numerically reduced when computed separately for masking and no-masking trials. Instead, the effect of global organization was numerically stronger for masking trials ($M = .017$, $SD = .039$) than for non-masking trials ($M = .004$, $SD = .039$) in Experiment 1, bordering on significance, $t(48) = 1.98$, $p = .052$, $d = .284$, and there was no meaningful difference in Experiment 2. Thus, the global effect we report here truly appears to owe little to visual awareness.

Local and global information can be found in almost any visual scene. Integration across these levels of analysis provides perceivers with information about individual objects as well as the contexts in which they appear. We thus speculate that the biases we demonstrated here may serve to normalize or correct the perception of objects to account for the three-dimensional contexts in which they appear. More generally, we have shown that the process of summary representation is inclusive of local and global information, consistent with the visual system's goal of constructing integrated and cohesive percepts. The gist is not just an output of ensemble representation – it can be included as an input to the mechanism itself.

Acknowledgements We thank Julie Campbell and Madison Sleyster for their assistance with data collection.

Open practices statement Data and analysis scripts from this work are available at <https://osf.io/tyn62/>. The experiments in this investigation were not preregistered.

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. <https://doi.org/10.1016/j.visres.2013.02.018>
- Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience*, 8, 407–430. <https://doi.org/10.1021/ac60020a001>
- Altmann, C. F., Bühlhoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, 13(4), 342–349. [https://doi.org/10.1016/S0960-9822\(03\)00052-6](https://doi.org/10.1016/S0960-9822(03)00052-6)
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <https://doi.org/10.1016/j.tics.2011.01.003>
- Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., & Bressloff, P. C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Annual Review of Neuroscience*, 40(1), 425–451. <https://doi.org/10.1146/annurev-neuro-072116-031418>
- Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, and Psychophysics*, 82(1), 63–79. <https://doi.org/10.3758/s13414-019-01827-z>
- Bijanzadeh, M., Nurminen, L., Merlin, S., Clark, A. M., & Angelucci, A. (2018). Distinct laminar processing of local and global context in primate primary visual cortex. *Neuron*, 100(1), 259–274. <https://doi.org/10.1016/j.neuron.2018.08.020>
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384–392. <https://doi.org/10.1177/0956797610397956>
- Brady, T. F., & Shafer-Skelton, A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1160–1176. <https://doi.org/10.1037/xhp0000399>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.

- Braun, A., & Sweeny, T. D. (2019). Anisotropic visual awareness of shapes. *Vision Research*, 156, 17–27. <https://doi.org/10.1016/j.visres.2019.01.002>
- Campana, F., Rebollo, I., Urai, A., Wyart, V., & Tallon-Baudry, C. (2016). Conscious vision proceeds from global to local content in goal-directed tasks and spontaneous vision. *Journal of Neuroscience*, 36(19), 5200–5213. <https://doi.org/10.1523/JNEUROSCI.3619-15.2016>
- Cha, O., & Chong, S. C. (2018). Perceived average orientation reflects effective gist of the surface. *Psychological Science*, 29(3), 319–327. <https://doi.org/10.1177/0956797617735533>
- Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, 153, 196–210. <https://doi.org/10.1016/j.cognition.2016.04.018>
- Chung, C. Y. L., & Khuu, S. K. (2014). The processing of coherent global form and motion patterns without visual awareness. *Frontiers in Psychology*, 5(MAR), 1–11. <https://doi.org/10.3389/fpsyg.2014.00195>
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, 20(5), 324–335. <https://doi.org/10.1016/j.tics.2016.03.006>
- Corbett, J. E. (2017). The whole warps the sum of its parts: Gestalt-defined-group mean size biases memory for individual objects. *Psychological Science*, 28(1), 12–22.
- Dickinson, J. E., Morgan, S. K., Tang, M. F., & Badcock, D. R. (2017). Separate banks of information channels encode size and aspect ratio. *Journal of Vision*, 17(2017), 1–20. <https://doi.org/10.1167/17.3.27>
- Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble perception of dynamic emotional groups. *Psychological Science*, 28(2), 193–203. <https://doi.org/10.1177/0956797616678188>
- Elias, E., Padama, L., & Sweeny, T. D. (2018). Perceptual averaging of facial expressions requires visual awareness and attention. *Consciousness and Cognition*, 62, 110–126.
- Elias, E., & Sweeny, T. D. (2020). Integration and segmentation conflict during ensemble coding of shape. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 593–609. <https://doi.org/10.1037/xhp0000733>
- Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, 44(12), 1321–1331. <https://doi.org/10.1016/j.visres.2003.10.024>
- Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science*, 8(2), 135–139.
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, 106, 1389–1398.
- Flevaris, A. V., & Robertson, L. C. (2016). Spatial frequency selection and integration of global and local information in visual processing: A selective review and tribute to Shlomo Bentin. *Neuropsychologia*, 83, 192–200. <https://doi.org/10.1016/j.neuropsychologia.2015.10.024>
- Gerlach, C., & Poirel, N. (2020). Who's got the global advantage? Visual field differences in processing of global and local shape. *Cognition*, 195.
- Goodhew, S. C., Pratt, J., & Dux, P. E. (2013). Substituting objects from consciousness: A review of object substitution masking. *Psychonomic Bulletin & Review*, 20, 859–877. <https://doi.org/10.3758/s13423-013-0400-9>
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734. <https://doi.org/10.1037/a0013899>
- Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, 18, 855–859.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804. <file:///Users/cpt3270/Documents/Library.papers3/Articles/2002/Unknown/2002-17.pdf%5Cnpapers3://publication/uuid/C8A367DF-7FCF-4D29-9A27-0FD9B3EC8B16>
- Hochstein, S., Pavlovskaya, M., Bonne, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision*, 15(4). <https://doi.org/10.1167/15.4.7>
- Hübner, R., & Volberg, G. (2005). The integration of object levels and their content: A theory of global/local processing and related hemispheric differences. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3), 520–541. <https://doi.org/10.1037/0096-1523.31.3.520>
- Ji, L., Pourtois, G., & Sweeny, T. D. (2020). Averaging multiple facial expressions through subsampling. *Visual Cognition*, 28(1), 41–58. <https://doi.org/10.1080/13506285.2020.1717706>
- Ji, L., Rossi, V., & Pourtois, G. (2018). Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs. *Neuropsychologia*, 111(December 2017), 92–102. <https://doi.org/10.1016/j.neuropsychologia.2018.01.022>
- Kayaert, G., Biederman, I., Beeck, H. P. Op De, & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, 22, 212–224. <https://doi.org/10.1111/j.1460-9568.2005.04202.x>
- Kimchi, R. (1994). The role of holistic/configural properties versus global properties in visual form perception. *Perception*, 23, 489–504.
- Kimchi, R. (2015). The perception of hierarchical structure. *Oxford Handbook of Perceptual Organization, March 2018*, 129–149. <https://doi.org/10.1093/oxfordhb/9780199686858.013.025>
- Köhler, W. (1930). Human perception. In *The selected paper of Wolfgang Köhler* (pp. 142–167).
- Liu, L., & Luo, H. (2019). Behavioral oscillation in global/local processing: Global alpha oscillations mediate global precedence effect. *Journal of Vision*, 19(5), 1–12. <https://doi.org/10.1167/19.5.12>
- Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, 22(6), 764–770. <https://doi.org/10.1177/0956797611408736>
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Nie, Q. Y., Müller, H. J., & Conci, M. (2017). Hierarchical organization in visual working memory: From global ensemble to individual object structure. *Cognition*, 159, 85–96. <https://doi.org/10.1016/j.cognition.2016.11.009>
- 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. [https://doi.org/10.1016/S0079-6123\(06\)55002-2](https://doi.org/10.1016/S0079-6123(06)55002-2)
- Beeck, H. Op De, Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General*, 132(4), 491–511. <https://doi.org/10.1037/0096-3445.132.4.491>
- Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9(4), 441–474.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744. <https://doi.org/10.1038/89532>
- Regan, D., & Hamstra, S. J. (1992). Shape discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research*, 32(10), 1845–1864.

- Ross, J., & Burr, D. (2008). The knowing visual self. *Trends in Cognitive Sciences*, 12(10), 363–364.
- Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, 28(4), 913–932. <https://doi.org/10.1037//0096-1523.28.4.913>
- Storrs, K. R., & Arnold, D. H. (2017). Shape adaptation exaggerates shape differences. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 181–191.
- Sun, J., & Chong, S. C. (2019). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology: General*, 149(3), 550–563. <https://doi.org/10.1037/xge0000667>
- Suzuki, S. (2005). High-level pattern coding revealed by brief shape aftereffects. In C. Clifford. & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and aftereffects in high-level vision. Advances in visual cognition series* (Vol. 2, pp. 135–172). Oxford University Press.
- Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), 1315–1341. <http://www.ncbi.nlm.nih.gov/pubmed/9778826>
- Sweeny, T. D., D'Abreu, L. C., Elias, E., & Padama, L. (2017). Object-substitution masking weakens but does not eliminate shape interactions. *Attention, Perception, & Psychophysics*, 79(7), 2179–2189. <https://doi.org/10.3758/s13414-017-1381-y>
- Sweeny, T. D., Grabowecky, M., Kim, Y. J., & Suzuki, S. (2011a). Internal curvature signal and noise in low- and high-level vision. *Journal of Neurophysiology*, 105(3), 1236–1257. <https://doi.org/10.1152/jn.00061.2010>
- Sweeny, T. D., Grabowecky, M., & Suzuki, S. (2011b). Simultaneous shape repulsion and global assimilation in the perception of aspect ratio. *Journal of Vision*, 11(1), 1–16. <https://doi.org/10.1167/11.1.16>
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 329–337.
- Sweeny, T. D., Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, (25)10, 1903–1914.
- Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4–5 year-old children. *Developmental Science*, (18)4, 556–568.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311(February), 670–675.
- Utochkin, I. S., & Brady, T. F. (2020). Individual representation in visual working memory inherit ensemble properties. *Journal of Experimental Psychology: Human Perception and Performance*, 46(5), 458–473.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217. <https://doi.org/10.1037/a0029333>
- Wertheimer, M. (1923). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of gestalt psychology* (pp. 71–88). Routledge and Kegan Paul, Ltd.
- Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 695–710). MIT Press.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105–129. <https://doi.org/10.1146/annurev-psych-010416-044232>
- Wolfe, J. M., Vo, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences*, 15(2), 77–84.
- Yashiro, R., Sato, H., Oide, T., & Motoyoshi, I. (2020). Perception and decision mechanisms involved in average estimation of spatiotemporal ensembles. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-58112-5>
- Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256(5061), 1327–1331.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.