



# Ensemble perception without attention depends upon attentional control settings

Zhimin Chen<sup>1</sup> · Ran Zhuang<sup>1</sup> · Xiaolin Wang<sup>1</sup> · Yanju Ren<sup>1</sup> · Richard A. Abrams<sup>2</sup>

© The Psychonomic Society, Inc. 2020

## Abstract

People are able to rapidly extract summary statistical information about common patterns, or ensembles, that may exist in a scene, such as repeated textures or colors. Here we examined the extent to which such an ensemble perception can occur in the absence of focal visual attention using a method that has some advantages over methods previously used to study the issue. In particular, we assessed the extent to which ensembles can be processed without attention by measuring the indirect effect of a to-be-ignored ensemble on judgments of an attended ensemble. The results show that ensembles outside the focus of attention do influence judgments of attended ensembles when the to-be-ignored ensemble contains summary statistics that match a sought-for target category. Thus, an attentional control setting for specific summary statistical information permits the processing of ensembles outside of focal attention, facilitating the rapid perception of visual scenes.

**Keywords** Ensemble perception · Visual attention · Unattended processing · Attentional control settings

## Introduction

When we first view a scene, the visual system rapidly extracts information about common patterns that may exist in the scene such as repeated colors or textures. The representation of such patterns is referred to as “summary statistical information” or “ensemble perception,” and is thought to play a critical role in the perception of natural scenes (Brady, Shafer-Skelton, & Alvarez, 2017). In particular, the ability to rapidly extract summary statistical information from a scene may reduce the burden of processing that would be otherwise needed by limited-capacity attentional and cognitive mechanisms. Nevertheless, despite evidence of rapid extraction of statistical information (e.g., Ariely, 2001; Chong & Treisman, 2003; Haberman & Whitney, 2007; also see review by Whitney & Yamanashi, 2018), it is still unclear whether ensemble perception can occur in the absence of visual attention. Resolving the issue has important implications for the understanding of perception more generally, and is the focus of the present paper.

Several recent studies have examined the extent to which ensemble perception could be carried out without attention, and they have yielded mixed results. On the one hand are studies that suggest that ensemble information can be processed without attention, or with only minimal attention. For example, Alvarez and Oliva (2008) asked participants to track a set of moving objects while ignoring a set of moving distractors. Although the to-be-ignored distractors were well outside the focus of attention, participants were still able to extract accurate summary statistics specifying their center of mass. Similarly, participants in Alvarez and Oliva (2009) were able to detect changes in an unattended background pattern more effectively when the change produced a different ensemble structure (compared to equivalent local changes that did not alter the summary statistics), suggesting reduced attentional demands for ensemble perception. In another study, Bronfman, Brezis, Jacobson, and Usher (2014) found that participants could report the diversity of colors contained in objects outside of focal attention with no cost to the performance of their primary task, which required focal attention elsewhere in the display, showing that color diversity, even outside focal attention, could be perceived automatically (see also Ward, Bear, & Scholl, 2016). Several other studies have also reached similar conclusions involving summary statistical information of other global visual attributes, such as circle size (Chong & Treisman, 2005) and gabor patch orientation (Alvarez & Oliva, 2009).

✉ Yanju Ren  
renyanju@gmail.com

<sup>1</sup> School of Psychology, Shandong Normal University, No. 1 Daxue Road, Changqing District, Jinan 250358, People’s Republic of China

<sup>2</sup> Department of Psychological and Brain Sciences, Washington University in St. Louis, St Louis, WA, USA

On the other hand, some studies have shown that ensemble perception does incur an attentional cost. Jackson-Nielsen, Cohen, and Pitts (2017) found that participants had no information about color diversity, size diversity, or the mean size of elements outside the focus of attention. Huang (2015) had participants make judgments about either individual visual features or summary statistics of stimuli presented sometimes in unexpected locations. He found that judgments of summary statistics benefited just as much from a spatial precue (which permitted focal attention) as did judgments of an individual feature. These findings suggest that ensemble perception is indeed attention-demanding, and cannot be accomplished in the absence of attention.

One reason for the conflicting results may be that many of the studies that have provided evidence for attention-free ensemble perception have employed dual-task paradigms in which participants perform a primary task with high attentional demands and are then probed on a secondary task about summary statistical information for unattended elements in the display (e.g., Alvarez & Oliva, 2008, 2009; Bronfman et al., 2014; Ward et al., 2016). Such paradigms leave open the possibility that participants may have allocated some attentional resources to the secondary (ensemble) task, rendering these experiments imperfect tests of the attentional demands of ensemble processing. However, studies that have revealed attentional costs of ensemble perception have used different methods. For example, Jackson-Nielsen et al. (2017) employed an inattentive blindness paradigm in which participants performed a focal task for several trials and then received an unexpected query regarding unattended elements after one of the trials. In that study, the participants did not have the motivation to allocate any attention beyond what was required of the focal task, and indeed there was no evidence of ensemble perception for unattended parts of the display. In the study by Huang (2015), focal attention was contrasted with divided attention across trials – with no incentive for participants to divide their attention on the focal attention trials. The results also showed that ensemble perception relies on focal attention.

The studies by Jackson-Nielsen et al. (2017) and Huang (2015) suggest that attention may be necessary for ensemble perception; however, those studies also suffer from a potential weakness. In particular, correct responses regarding the ensemble summary statistics in those experiments required the participants to successfully remember details of the ensemble in order to correctly respond. Thus, any failures of *memory* for the ensemble might be incorrectly assumed to reflect the absence of ensemble perception itself (Chen & Wyble, 2016; Jiang, Shupe, Swallow, & Tan, 2016; but see Ward and Scholl, 2015a, b). As a result, it is still unclear whether ensemble perception does or does not require attention.

In order to address this question, here we adopted a method that permits the assessment of ensemble perception without any memory requirement and without any motivation for the participant to attend to irrelevant portions of the display. The

method is based on that used by Gronau and Izoutcheev (2017), who studied a similar question regarding the extent to which attention is required for the perception of scenes. The method also bears some similarity to methods used by others in which the processing of an unattended stimulus is assessed by examining its (indirect) effect on the processing of an attended stimulus (e.g., Du & Abrams, 2008, 2012; Eriksen & Eriksen, 1974; Gaspelin, Ruthruff, & Jung, 2014; Theeuwes, 1992, 1994, 2010). Here we use the method to examine ensemble perception. In the critical experiment reported below, participants were required to attend to one region of a display that contained the relevant stimulus and ignore another region that contained a distractor. Our goal was to determine the extent to which ensemble information about the distractor (outside of attention) was processed – but the distractors were not probed by explicit report. Instead, the processing of the unattended distractor was inferred on the basis of interference or facilitation caused by the distractor on evaluation of the relevant, attended stimulus. Thus, the method indirectly assesses the extent to which ensemble information is processed outside the focus of attention without any motivation for participants to attend to the critical (distractor) stimulus and without any requirement that features of the unattended stimulus be retained in memory.

## Overview of experiments

We report two experiments below. In the first experiment, participants were briefly exposed to two ensemble stimuli (consisting of clusters of lines) and were asked to determine the presence of an ensemble that matched a pre-specified target category (vertical, horizontal, or oblique line orientations). Our interest was to determine whether performance would be facilitated when the two stimuli were from the same category. If such facilitation does occur, that would show that shared ensemble category membership can affect ensemble judgments when both ensemble stimuli are attended. To anticipate the results, such facilitation did occur. Then, in Experiment 2 our goal was to seek evidence for the same facilitation when one of the ensembles is outside the focus of attention. Such an effect there would indicate that ensemble perception can take place in the absence of attention.

## Experiment 1

In this experiment two ensemble stimuli were briefly presented. The stimuli consisted of clusters of lines that were mostly vertical (the “vertical” category), mostly horizontal (“horizontal”), or mostly oblique (“oblique”; similar to stimuli used by others, e.g., Huang, 2015). Participants were to decide whether either ensemble was from a pre-specified target category.

Our interest was to determine whether the decision was influenced by the extent to which both ensembles did or did not share the same category. Such a result would serve as an important prerequisite for the test conducted in Experiment 2, in which some ensembles were presented outside the focus of attention.

## Method

**Participants** The sample size here and in Experiment 2 was based on the study by Gronau and Izoutcheev (2017), who used a similar paradigm. The sample of 18 participants in their Experiment 1 yielded a medium effect size (partial eta squared = 0.66) when stimuli were fully attended. In order to enhance the power of the present experiment, 24 undergraduate students (13 females, 11 males, age 19–22 years) with normal or corrected-to-normal vision participated. They were paid 15 RMB (equivalent to about \$2.14) for their participation.

**Apparatus and procedure** Stimuli were presented on a 17-in. CRT with an 85-Hz refresh rate viewed from a distance of 57 cm, on a gray background. The sequence of events on each trial is shown in Fig. 1. At the beginning of each trial, participants attended to a red fixation cross ( $.8^\circ \times .8^\circ$ ) presented at the center of the screen. After 600 ms, two ensemble stimuli were presented – one above and one below fixation – for 47 ms. The ensembles were followed by a 129-ms pseudonoise pattern mask and then a 1,082-ms blank screen. Participants were to press one key on the computer keyboard as quickly as possible if either ensemble was a member of the pre-specified target category, and another key if neither ensemble was a member. Trials without responses by the end of the blank interval were considered errors.

The ensembles were selected from one of three categories: *vertical*, *horizontal*, or *oblique*, with one category designated in advance as the target category. Each stimulus subtended

$9.3^\circ$  by  $9.3^\circ$  and consisted of 16 black line segments ( $1.2^\circ \times .3^\circ$ ) arranged in a  $4 \times 4$  grid in which a randomly selected 12 lines matched the category designation and the other four lines had orientations selected randomly from the other categories. Ensembles were centered approximately  $5^\circ$  above and below fixation. Depending on the particular line orientations, rows within an ensemble were between  $1.2^\circ$  and  $2^\circ$  apart with the space between ensembles at least  $2^\circ$ .

For both the target-present and target-absent trials, the two ensembles were on some trials from the same category, whereas on other trials, the categories differed. Figure 2 shows examples of the different trial types when the target category was horizontal. On *same-category target-present* trials, both ensembles were from the same (pre-specified target) category (horizontal in the example). On *different-category target-present* trials, one ensemble was from the target category and the other was from one of the other categories. Finally, on *target-absent* trials, the two ensembles could be either from the same category or a different category, but never included ensembles from the target category.

**Design** The experiment contained 180 target-present trials and 240 target-absent trials. For the target-present trials, one-third (60 trials) contained two ensembles from the target category (e.g., both horizontal when horizontal is the target category), and two-thirds contained one ensemble from the target category and one ensemble from one of the other two categories (60 trials for each of the possible non-target categories; e.g., one horizontal and one vertical, or one horizontal and one oblique). For the target-absent trials, one-half of them (120 trials) contained two ensembles from the same category (60 for each of the two non-target categories; e.g., both vertical or both oblique), while the other half (120 trials) contained one ensemble from each of the two non-target categories (e.g., one vertical and one oblique). Each of the three ensemble categories served as the target category for one-third of the

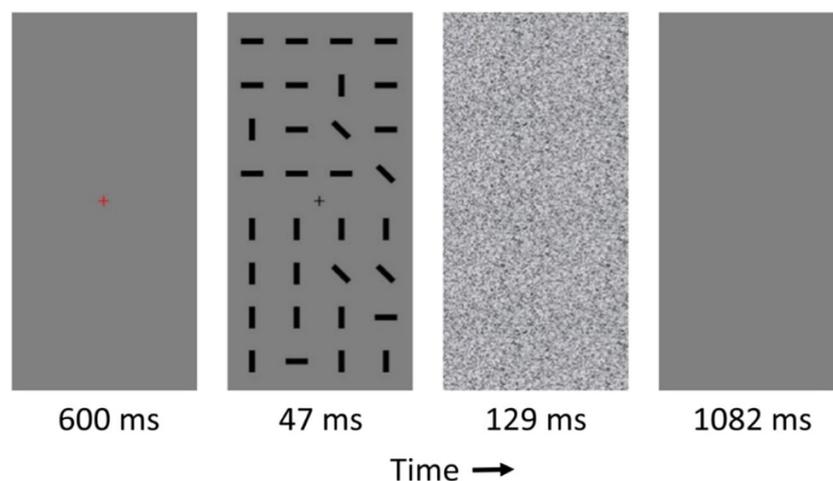
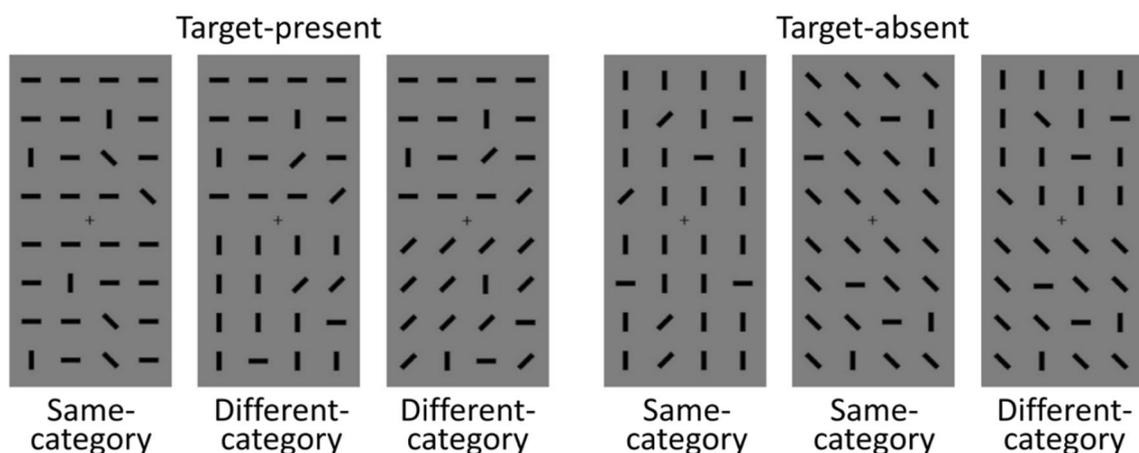


Fig. 1 Sequence of events on a trial in Experiment 1



**Fig. 2** Examples of the stimuli for each trial type in Experiment 1 when “horizontal” was the specified target category. The designations “same-

category” and “different-category” refer to the relations between the categories of the two ensembles on a trial. See text for additional details

participants. When the target category was present, it was equally likely to appear above or below fixation. When one or two oblique ensembles were in the display, all oblique lines had the same orientation. Trials were presented in a random order. At the beginning of the session, participants completed a practice block of 42 trials (with trial types in the same proportions as in the formal testing) that was not included in the analysis. Two prospective participants were replaced because they were unable to achieve 80% accuracy in the practice block.

## Results

Trials with errors and those with reaction times more than three standard deviations above or below each participant’s mean in each experimental condition were excluded from analysis. Mean reaction times are shown in Fig. 3. We conducted a target-presence (present or absent) by category relation (same-category or different-category) ANOVA. Reaction times were faster for target-present than target-absent judgments,  $F(1, 23) = 49.19, p < .001, \eta_p^2 = .68$ . Reaction times were also faster when both ensembles were from the same category,  $F(1, 23) = 104.74, p < .001, \eta_p^2 = .82$ . The effect of category relation was greater for target-present than for target-absent trials,  $F(1, 23) = 9.25, p = .006, \eta_p^2 = .29$ . Importantly, follow-up contrasts showed that the category relation effect was significant not only for target-present trials,  $F(1, 23) = 131.27, p < .001, \eta_p^2 = .85$ , but also for target-absent trials,  $F(1, 23) = 39.05, p < .001, \eta_p^2 = .63$ .

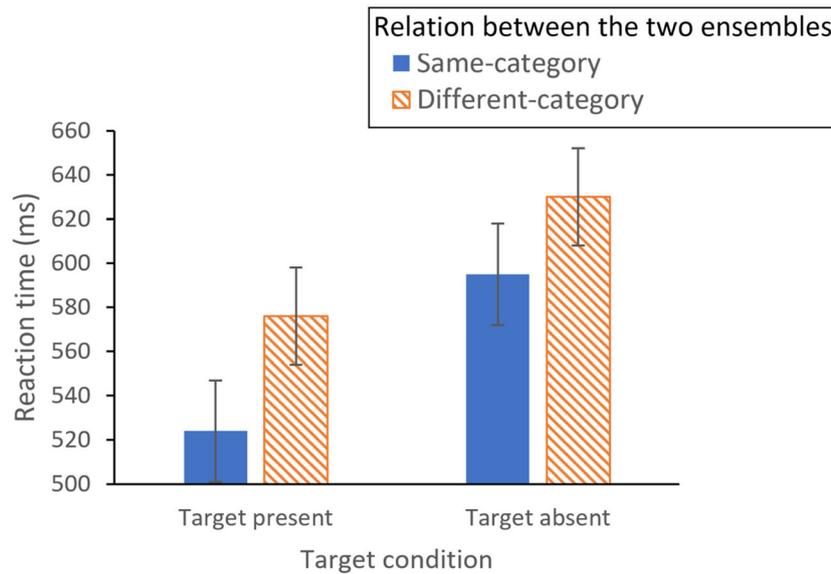
Accuracy rates are shown in Table 1. Participants were more accurate when the two ensembles were from the same category,  $F(1, 23) = 40.87, p < .001, \eta_p^2 = .64$ , matching the effect in reaction times and ruling out a speed-accuracy tradeoff. There was no overall difference between target-present and target-absent trials,  $F(1, 23) = .01, p = .90, \eta_p^2 = .001$ , but the effect of category relation was greater for the

target-present trials, as revealed by an interaction between the two factors,  $F(1, 23) = 24.22, p < .001, \eta_p^2 = .51$ . Follow-up contrasts showed that the category relation effect was significant not only for target-present trials,  $F(1, 23) = 49.53, p < .001, \eta_p^2 = .68$ , but also for target-absent trials,  $F(1, 23) = 4.76, p = .040, \eta_p^2 = .17$ .

## Discussion

In this experiment, participants attended to two ensembles of lines, searching for the presence of a pre-specified category. When the target was present in both ensembles (target-present, same-category trials) participants were faster than when the target was present in only one ensemble (target-present, different-category). This occurred presumably in part because assessment of the orientation of either ensemble would lead to a “present” response. More importantly, there was also a same-category advantage on the target-absent trials despite the fact that target-absent trials always required both stimulus ensembles to be inspected prior to an “absent” response. This result shows that when an entire scene is attended, ensemble perception is influenced by the ensemble category relations present in the scene. While the source of that result could lead to insight into ensemble perception, doing so was not our objective.<sup>1</sup> Most importantly, it serves as an important prerequisite for Experiment 2, in which we examined the possibility that ensemble category membership can influence ensemble perception outside the focus of attention.

<sup>1</sup> We can speculate briefly about what the results from Experiment 1 might reveal about ensemble perception more generally. One possibility is that summary statistical information about one part of a scene may prime perception of similar ensembles elsewhere in the scene (e.g., Bajo, 1988). Another possibility is that an initial rapid global analysis is performed that reveals whether any part of a scene contains summary statistical information that differs from that in other parts of the scene (e.g., Davenport & Potter, 2004). Either of these possibilities would predict more rapid evaluation of two ensembles when they share the same category.



**Fig. 3** Mean reaction times from Experiment 1 as a function of target presence and the relation between the two ensembles. Reaction times were faster when a target was present, and when the two ensembles

were from the same category (for both target-present and target-absent conditions). Error bars represent the standard errors of the means in each condition

### Experiment 2

Experiment 1 revealed a same-category advantage: when both stimuli in the display were from the same ensemble category, the stimuli were processed more quickly. Because that experiment required participants to indicate if a target category was present anywhere in the display, presumably both elements were attended there. Here we repeated the experiment with one important difference: we cued one of the ensemble stimulus locations in advance, and asked subjects to report only whether the stimulus in the cued location matched the pre-specified target category. As a result, the other ensemble stimulus was a distractor – outside the scope of spatial attention. If ensemble statistics are perceived automatically and without attention, then the ensemble category of the unattended distractor stimulus would be expected to influence responses here and reveal a same-category advantage for judgments, as in Experiment 1. Alternatively, if attention is required for ensemble perception then there should be no effect of the category relation between the attended and unattended ensembles on judgments of the attended ensemble. Importantly, the task measures the processing of the distractor ensemble when there is no motivation for participants to split their attention

between the two stimuli – the distractor is completely irrelevant to the task. Additionally, there is no requirement that participants remember anything about the distractor in order for us to determine that ensemble information about the distractor was processed.

### Method

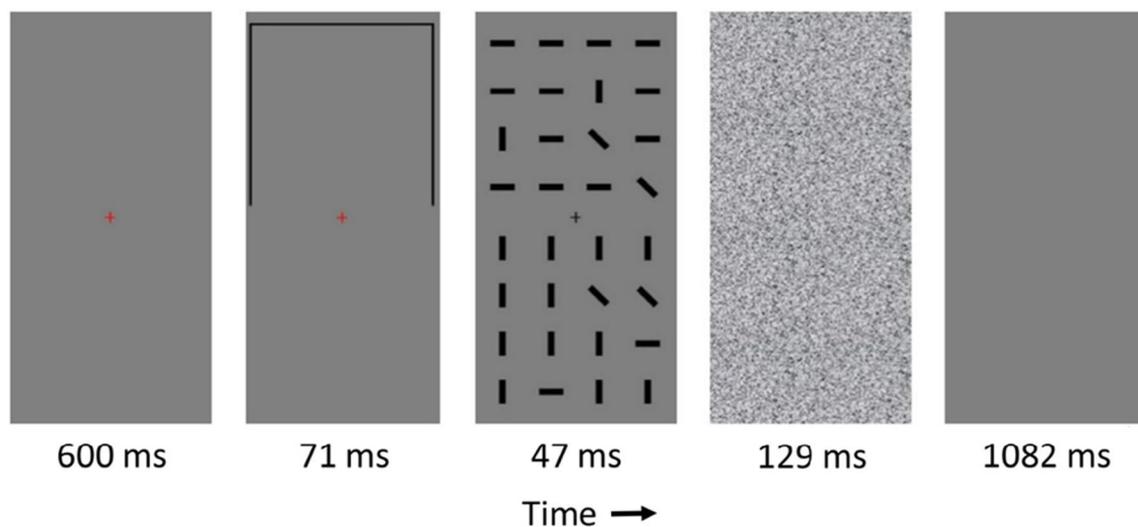
**Participants** A new group of 24 undergraduate students (13 females, 11 males, age 19–21) participated in Experiment 2. They were paid 15 RMB (equivalent to about \$2.14) for their participation. All participants had normal or corrected-to-normal vision.

**Procedure** The sequence of events on each trial is shown in Fig. 4. The procedure was identical to that used in Experiment 1 with only two exceptions. First, we inserted a 71-ms spatial cue (a three-sided frame) at one of the stimulus locations prior to presentation of the ensemble stimuli. Second, participants were instructed to report only whether the *cued* stimulus did or did not match the pre-specified target category – the uncued ensemble was a distractor that could be ignored.

Examples of the stimuli are shown in Fig. 5. When an ensemble from the target category appeared in a cued location (target-present/cued) the distractor (i.e., uncued) ensemble could either be from the same (i.e., target) category, or a different category (Fig. 5a). Similarly, when a target was absent from the display, the two ensembles could come from the same or from different categories (Fig. 5b). These conditions allowed us to assess the effect of a same-category ensemble in the unattended (i.e., uncued) location. Finally, there

**Table 1** Mean accuracy rates (proportion correct) from Experiment 1. Standard errors are shown in parentheses

Target present		Target absent	
Same-category	Different-category	Same-category	Different-category
.96 (.01)	.83 (.02)	.91 (.02)	.88 (.02)

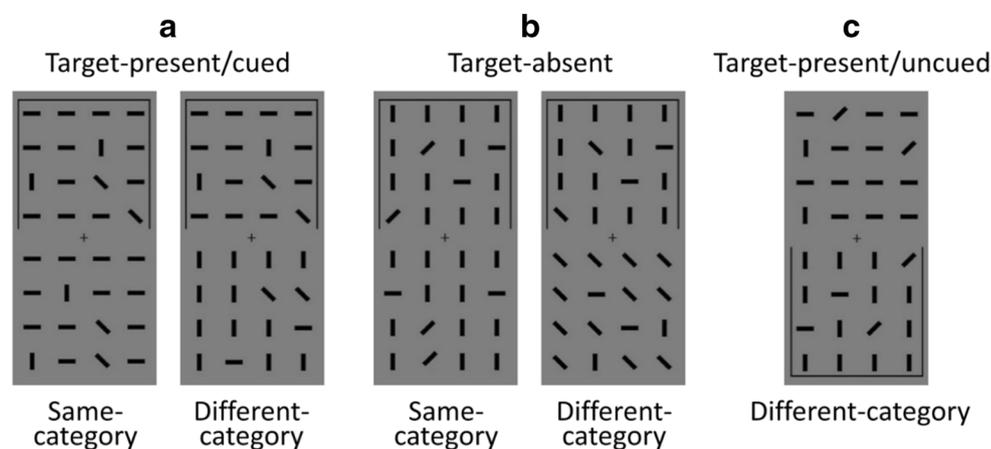


**Fig. 4** Sequence of events in Experiment 2. Participants were asked to assess only whether the cued ensemble (as cued by the three-sided frame) matched the pre-specified target category

were trials that contained one ensemble from the target category that was uncued (Fig. 5c).

**Design** The experiment included 180 trials in which the target category was present and cued (100 of which included a same-category distractor in the uncued location; 80 of which had a different-category distractor). 160 trials were target-absent trials (with half of those containing a same-category distractor, and half a different-category distractor). Finally 80 trials included a non-target category in the cued location and a target-category ensemble in the distractor location (note that this

number matched the number of trials containing a target category that was cued and a different-category distractor; a similar design was used by Gronau & Izoutcheev, 2017, who studied scene perception). The trial numbers were selected so that the ratio of “present” responses to “absent” responses here (based on only the ensemble in the cued location) matched that of Experiment 1 ( $180:240 = .75$ ). All other aspects of the design matched Experiment 1. In particular, for trials that contained non-target categories, each of the two non-target categories were represented equally often. The top and bottom locations were equally likely to be cued.



**Fig. 5** Examples of stimuli from Experiment 2 when “horizontal” was the target category. Participants here indicated only whether the cued ensemble matched the target. (a) When the target was present and cued (and hence the correct response was “present”), the distractor (the uncued ensemble) could be either from the same-category (i.e., matching the target category), or from a different category. (Note that the figure does not show one additional different-category trial type where the uncued ensemble belonged to the other non-target category [oblique in this example].) (b) When the target was absent, the distractor could either be

from the same category as the cued ensemble or from a different one. (Note that one additional combination of same-category target-absent [oblique, oblique] was tested but is not shown in the figure.) (c) Finally, there were also trials in which an ensemble matching the target category was present, but a non-target category was cued (and hence the required response would be “absent”). (Not shown in the figure are trials in which the cued non-target ensemble was from the other non-target category [oblique])

**Results**

Trials in which participants made errors or trials that had RTs that deviated from each participant’s mean RT by more than three standard deviations were excluded from the RT analysis. RTs are shown in Fig. 6. To assess the effects of the unattended (i.e., uncued) ensemble, we conducted a target-presence (present or absent) by category relation (same-category or different-category) ANOVA similar to that used in Experiment 1, excluding the target-present/uncued trials. Although the main effect of target presence was not significant,  $F(1, 23) = 2.32, p = .141, \eta_p^2 = .09$ , RTs were faster when the uncued ensemble was from the same category,  $F(1, 23) = 4.87, p = .037, \eta_p^2 = .18$ , showing that the ensemble statistics of the unattended stimulus were processed. However, as seen in the figure, this occurred only for the target-present trials, as revealed by an interaction between the two factors,  $F(1, 23) = 5.36, p = .03, \eta_p^2 = .19$ . Follow-up tests showed that the category effect was significant for the target-present trials,  $F(1, 23) = 7.46, p = .012, \eta_p^2 = .25$  (comparison “A” in Fig. 6), but not the target-absent trials,  $F(1, 23) = .03, p = .859, \eta_p^2 = .001$ .

To further examine the potential processing of the uncued (unattended) ensemble, we conducted a planned contrast comparing the target-absent different category condition with the target-present uncued condition (comparison B in Fig. 6). Both of these conditions cued a non-target category (requiring an “absent” response), yet in the former condition the distractor also came from a non-target category whereas in the latter condition the distractor was from the target category. The result of the comparison revealed a marginally significant cost to responding “absent” when the distractor was from the target category,  $t(23) =$

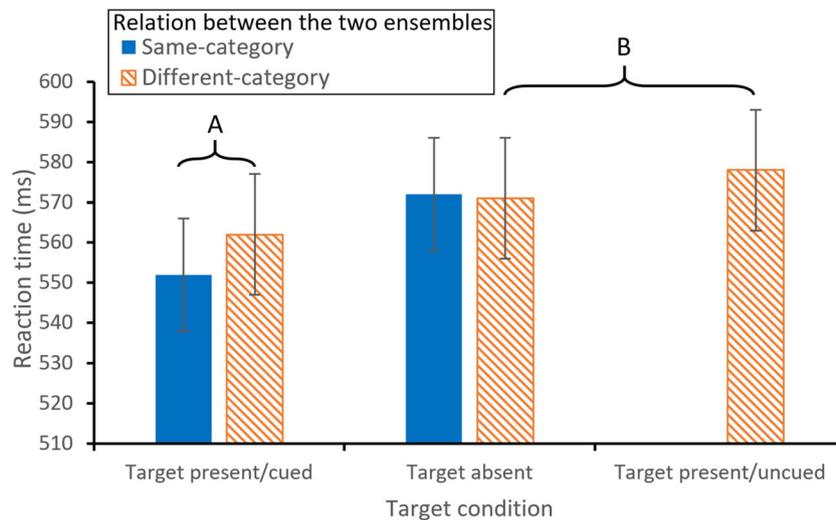
$2.05, p = .052$ , Cohen’s  $d = .85$ , suggesting that the unattended distractor’s category was indeed processed.

Additional insight into the attentional effects of the target category comes from a comparison of the different target category conditions. Recall that participants with horizontal and vertical categories defined as the target were searching for the presence of a specific ensemble feature whereas those who were assigned the oblique category were searching for oblique lines that were either tilted to the left or tilted to the right. As a result, the target category was less specific for those searching for an oblique target, and that might be expected to result in a reduced same-category advantage (the comparison marked “A” in Fig. 6). Indeed, oblique targets yielded a numerically smaller advantage when the unattended ensemble was also oblique ( $m = 8.79$  ms) compared to the same-category advantage for horizontal and vertical target categories ( $m = 10.42$  ms), but the difference was not statistically significant,  $t(22) = .21, p = .837$ , Cohen’s  $d = .09$ .

Accuracies are shown in Table 2. There was no effect of target presence,  $F(1, 23) = 2.97, p = .098, \eta_p^2 = .115$ , or of category relation,  $F(1, 23) = .04, p = .842, \eta_p^2 = .002$ , and the two factors did not interact,  $F(1, 23) = .25, p = .623, \eta_p^2 = .01$ .

**Discussion**

In this experiment participants attended to one cued ensemble of lines and ignored a second, distractor ensemble. Nevertheless, the category of the distractor ensemble influenced judgments of the cued ensemble. Participants were faster to indicate that a cued ensemble was in the target category when the distractor was in the target category. They were also (marginally) slower to indicate that the cued ensemble



**Fig. 6** Reaction times from Experiment 2. The comparison marked “A” reveals the benefit to respond “present” when the unattended ensemble was also from the target category. The comparison marked “B” shows the cost to responding “absent” when the cued ensemble was not from the

target category but the unattended ensemble did come from the target category. Error bars represent the standard errors of the means in each condition

**Table 2** Mean accuracy rates (proportion correct) from Experiment 2. Standard errors are shown in parentheses

Target-present/cued		Target-absent		Target-present/uncued
Same-category	Different-category	Same-category	Different-category	Different-category
.92 (.01)	.92 (.01)	.93 (.01)	.94 (.01)	.93 (.01)

was not a member of the target category when the distractor was a member of the category. Because the experiment did not encourage any division of attention between the two ensembles (i.e., there was never any reason for participants to assess the distractor), the results show that ensemble statistics under at least some circumstances can be perceived in the absence of focal attention.

## General discussion

The present study examined ensemble perception in the absence of focal attention. In Experiment 1, we found that when two ensembles are attended, the category membership of both ensembles is processed – yielding a same-category advantage in responding. In Experiment 2 we required that participants attend to only one of two ensemble stimuli that were presented – the uncued ensemble could be safely ignored. Nevertheless, we also found a same-category advantage there: the category membership of the unattended ensemble influenced performance. These results show that summary ensemble statistics, under at least some circumstances, can be processed in the absence of focal attention.

Importantly, our experiments did not have the same shortcomings that have influenced earlier attempts to address the same issue. In particular, we assessed perception of the unattended ensemble implicitly by examining any effects of the unattended ensemble on decisions related to the attended ensemble. Because the unattended ensemble was entirely irrelevant to the task, there was no need for participants to partially divide their attention between the two ensembles, unlike in some past studies that examined the same question (e.g., Alvarez & Oliva, 2008, 2009; Bronfman et al., 2014). Second, our experiments also did not impose any memory requirements upon the participants in order to assess perception of summary statistics outside the focus of attention. Some past studies that have examined the same question did require that participants retain and then explicitly recall certain aspects of the unattended stimuli (e.g., Jackson-Nielsen et al., 2017; Huang, 2015). Thus, the method used here has some advantages over previous methods.

## Attentional control settings and task set

One noteworthy aspect of our findings is that the category identity of the unattended ensemble in Experiment 2 only influenced performance when it matched the target category. In particular, when the to-be-ignored stimulus matched the target category, it facilitated judgments when an ensemble from the target category had been cued and impaired judgments when the cued stimulus was not a member of the target category. On the other hand, when the to-be-ignored ensemble did not match the target category, it had no influence on performance. These results suggest two possible interpretations. First, it is possible that ensemble statistics are processed outside the focus of attention – but only when the ensemble matches the participant's task set. This interpretation is consistent with late-selection theories of attention that propose that meaning can be evaluated across the visual field prior to the selectivity of attention (e.g., Deutsch & Deutsch, 1963; Duncan, 1984). Results consistent with such a possibility have been reported by Gronau, Cohen, and Ben-Shakhar (2009). They showed that some distractors outside the focus of attention are able to exert their effect on the processing of focal stimuli without themselves attracting attention (see also Eriksen & Eriksen (1974), LaBerge (1983), and Peelen, Fei-Fei, & Kastner (2009), who have results with similar interpretation). Importantly, however, in Experiment 2, such processing of the uncued ensemble only occurred when it matched the target category. Target non-matching ensembles in the uncued location were not processed (because if they had been, they would have affected RTs, as they did in Experiment 1 when both ensembles were attended). That aspect of the results suggests that the attentional control setting or task set for the target category may have acted as an early filter over the entire scene, allowing only target-matching elements to be processed more deeply because only such ensembles matched the properties of the sought-for target. Related findings have been reported by Folk, Leber, and Egeth (2002). They showed that

<sup>0</sup> Our results are consistent in many ways with the model of attention proposed by Huang and his colleagues (Huang & Pashler, 2007; Huang, Treisman, & Pashler, 2007). They proposed that people have conscious access to only one feature value per dimension in a scene at any one time. Consistent with that, our participants' responses in Experiment 2 revealed that they were able to process both ensembles after a brief exposure when the ensembles were both from the same target category. Importantly, we did not find a similar result for ensembles from non-target categories.

distractors in to-be-ignored locations were nevertheless processed when they matched the participant's task set.<sup>2</sup>

The match between target-category ensembles and the participant's task set also leads to a second interpretation of our findings. It is possible that target-matching ensembles captured attention automatically precisely because they were consistent with the participants' attentional control settings. Such contingent capture has been demonstrated in a wide range of situations (e.g., Folk, et al., 2002; Folk, Remington, & Johnson, 1992; Gaspelin et al., 2014; Reeder, van Zoest, & Peelen, 2015; Wyble, Folk, & Potter, 2013). One interpretation of contingent capture effects is that the attentional control setting enhances the salience of elements that match the task set, thus causing them to capture attention (e.g., Biggs & Gibson, 2010, 2014; Cosman & Vecera, 2010).<sup>3</sup> If this had occurred in Experiment 2 then our results would be better characterized as resulting not from pre-attentive processing of the target-matching ensemble statistics, but instead from contingent capture of attention by such ensembles.

Distinguishing between these two possibilities will be difficult because tasks that attempt to assess the locus of attention typically require presenting occasional probes at the locations being tested (e.g., Kim & Cave, 1995). The possibility of such probes could motivate the subjects to intentionally allocate attention to uncued regions, rendering such a test ineffective for ensemble perception.

Although it is not possible to distinguish between the two possibilities with the present results, both interpretations reveal the critical role of ensemble processing for perception: Ensembles matching one's task set are capable either of being processed in the absence of attention or of summoning attention to their location. In either case, perception of important summary statistics of complex scenes can be rapidly and efficiently accomplished.

## Relation to scene perception

The present findings and conclusions closely match those reported by Gronau and Izoutcheev (2017) in their study of scene perception (see also Gronau, 2020). Those researchers examined perception of the gist of scenes in attended and unattended locations using a method very similar to the one we used here. As in our study, Gronau and Izoutcheev (2017) found that scenes outside the focus of attention influenced

judgments only when they matched the sought-for scene category, suggesting that scene gist is processed without attention when the scene is consistent with the observer's goals, parallel to our results for ensemble perception (in Experiment 2). Some researchers have argued that processing of summary statistics is a fundamental part of scene gist processing (e.g., Brady et al., 2017), sharing mechanisms with visual object categorization (Khayat & Hochstein, 2019). The similarity of the present findings to the Gronau and Izoutcheev (2017) results provides support for that idea.

## Alternative explanations

It is possible that the results that we reported stem from properties of the *responses* required in our task as opposed to the properties of the stimuli, as we have suggested. In particular, in Experiment 2, when the unattended ensemble matched the target category, the response to that ensemble (if one had been required) also matched the required response when the cued stimulus was in the target category but it did not match the required response when the cued stimulus was not a target category member. Thus our results might simply derive from the response congruency associated with the attended and unattended ensembles. For the case of scene processing, Gronau (2020, Experiment 4) has shown that response congruency cannot completely account for gist processing outside of focal attention. Although we cannot rule out that possibility here, if it did occur, the present findings would still indicate that ensemble summary statistics can be processed outside the focus of attention. If, in fact, the effect was entirely due to response congruence, then that could indicate that both target-matching and target-nonmatching ensembles can be processed without attention. More work will be needed to rule out that alternative.

## Conclusion

In summary, the present results reveal robust processing of summary statistical information of ensembles outside of focal attention when the ensembles matched the properties of a sought-for target. Such ensembles (but not non-matching ones) were either processed preattentively or they caused attention to be directed to their location. The results are similar to those from studies of scene gist processing (e.g., Gronau & Izoutcheev, 2017) and help to further illuminate the way in which we rapidly assess the contents of visual scenes.

**Acknowledgements** This research was partly supported by Shandong Provincial Key Research & Development Program (No. 2018GSF118090), Natural Science Foundation of Shandong Province

<sup>3</sup> Although we have characterized our results as having arisen from the participant's task set, it is also likely that responses to target-category matching ensembles were influenced by selection history effects (cf. Awh, Belopolsky, & Theeuwes, 2012). This is because not only did participants have an attentional control setting for a specific ensemble category, but they also repeatedly searched for and responded to that specific category throughout the experiment. Thus selection history effects, in part caused by intertrial priming, may also have exerted an influence on attention. We do not distinguish here between these two influences (see also Egeth, 2018).

(No. ZR2017MC058), and a Visiting Scholar Grant from Shandong Provincial Government (No. 2017190) to Y. Ren.

**Open practices statement** The data for this study are available on the Open Science Framework project page (<https://osf.io/7btg4/>).

## Compliance with ethical standards

**Competing interests statement** The authors have declared they have no competing financial interests.

## References

- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398. doi: <https://doi.org/10.1111/j.1467-9280.2008.02098>
- 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 of the United States of America*, *106*(18), 7345–7350. doi: <https://doi.org/10.1073/pnas.0808981106>
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. doi: <https://doi.org/10.1111/1467-9280.00327>
- Awh, E., Belopolsky, A., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. doi: <https://doi.org/10.1016/j.tics.2012.06.010>
- Bajo, M. T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 579–589. doi: <https://doi.org/10.1037/0278-7393.14.4.579>
- Biggs, A. T., & Gibson, B. S. (2010). Competition between color salience and perceptual load during visual selection can be biased by top-down set. *Attention, Perception, & Psychophysics*, *72*(1), 53–64. doi: <https://doi.org/10.3758/APP.72.1.53>
- Biggs, A. T., & Gibson, B. S. (2014). Visual salience can co-exist with dilution during visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 7–14. doi: <https://doi.org/10.1037/a0033922>
- Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception & Performance*, *43*(6), 1160–1176. doi: <https://doi.org/10.1037/xhp0000399>
- Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: “Cost free” color phenomenality outside focal attention. *Psychological Science*, *25*(7), 1394–1403. doi: <https://doi.org/10.1177/0956797614532656>
- Chen, H., & Wyble, B. (2016). Attribute amnesia reflects a lack of memory consolidation for attended information. *Journal of Experimental Psychology: Human Perception & Performance*, *42*(2), 225–234. doi: <https://doi.org/10.1037/xhp0000133>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. doi: [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. doi: <https://doi.org/10.1016/j.visres.2004.10.004>
- Cosman, J. D., & Vecera, S. P. (2010). Attentional capture under high perceptual load. *Psychonomic Bulletin & Review*, *17*(6), 815–820. doi: <https://doi.org/10.3758/PBR.17.6.815>
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. doi: <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. doi: <https://doi.org/10.1037/h0039515>
- Du, F., & Abrams, R. A. (2008). Synergy of stimulus-driven salience and goal-directed prioritization: Evidence from the spatial blink. *Perception & Psychophysics*, *70*(8), 1489–1503. doi: <https://doi.org/10.3758/pp.70.8.1489>
- Du, F., & Abrams, R. A. (2012). Out of control: Attentional selection for orientation is thwarted by properties of the underlying neural mechanisms. *Cognition*, *124*, 361–366. doi: <https://doi.org/10.1016/j.cognition.2012.05.013>
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. doi: <https://doi.org/10.1037/0096-3445.113.4.501>
- Egeth, H. (2018). Comment on Theeuwes’s characterization of visual selection. *Journal of Cognition*, *1*(1), 26. doi: <https://doi.org/10.5334/joc.29>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. doi: <https://doi.org/10.3758/BF03203267>
- Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*(5), 741–753. doi: <https://doi.org/10.3758/bf03194741>
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. doi: <https://doi.org/10.1037/0096-1523.18.4.1030>
- Gaspelin, N., Ruthruff, E., & Jung, K. (2014). Slippage theory and the flanker paradigm: An early-selection account of selective attention failures. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1257–1273. doi: <https://doi.org/10.1037/a0036179>
- Gronau, N. (2020). Vision at a glance: The role of attention in processing object-to-object categorical relations. *Attention, Perception, & Psychophysics*. doi: <https://doi.org/10.3758/s13414-019-01940-z>
- Gronau, N., & Izoutcheev, A. (2017). The necessity of visual attention to scene categorization: dissociating “task-relevant” and “task-irrelevant” scene distractors. *Journal of Experimental Psychology: Human Perception & Performance*, *43*(5), 954–970. doi: <https://doi.org/10.1037/xhp0000365>
- Gronau, N., Cohen, A., & Ben-Shakhar, G. (2009). Distractor interference in focused attention tasks is not mediated by attention capture. *The Quarterly journal of experimental psychology*, *62*(9), 1685–1695. doi: <https://doi.org/10.1080/17470210902811223>
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753. doi: <https://doi.org/10.1016/j.cub.2007.06.039>
- Huang, L. (2015). Statistical properties demand as much attention as object features. *PLoS One*, *10*(8), e0131191. doi: <https://doi.org/10.1371/journal.pone.0131191>
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*(3), 599–631. doi: <https://doi.org/10.1037/0033-295X.114.3.599>
- Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317*(5839), 823–825. doi: <https://doi.org/10.1126/science.1143515>

- Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness & Cognition*, *48*, 149–160. doi: <https://doi.org/10.1016/j.concog.2016.11.007>
- Jiang, Y. V., Shupe, J. M., Swallow, K. M., & Tan, D. H. (2016). Memory for recently accessed visual attributes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1331–1337. doi: <https://doi.org/10.1037/xlm0000231>
- Khayat, N., & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception, & Psychophysics*, *81*(8), 2850–2872. doi: <https://doi.org/10.3758/s13414-019-01792-7>
- Kim, M. S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science*, *6*(6), 376–380. doi: <https://doi.org/10.1111/j.1467-9280.1995.tb00529.x>
- LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(3), 371–379. doi: <https://doi.org/10.1037/0096-1523.9.3.371>
- Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. doi: <https://doi.org/10.1038/nature08103>
- Reeder, R. R., van Zoest, W., & Peelen, M. V. (2015). Involuntary attentional capture by task-irrelevant objects that match the search template for category detection in natural scenes. *Attention, Perception, & Psychophysics*, *77*(4), 1070–1080. doi: <https://doi.org/10.3758/s13414-015-0867-8>
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & psychophysics*, *51*(6), 599–606. doi: <https://doi.org/10.3758/BF03211656>
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 799–806. doi: <https://doi.org/10.1037/0096-1523.20.4.799>
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(1), 77–99. doi: <https://doi.org/10.1016/j.actpsy.2010.02.006>
- Ward, E. J., & Scholl, B. J. (2015a). Inattention blindness reflects limitations on perception, not memory: Evidence from repeated failures of awareness. *Psychonomic Bulletin & Review*, *22*(3), 722–727. doi: <https://doi.org/10.3758/s13423-014-0745-8>
- Ward, E. J., & Scholl, B. J. (2015b). Stochastic or systematic? Seemingly random perceptual switching in bistable events triggered by transient unconscious cues. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 929–939. doi: <https://doi.org/10.1037/a0038709>
- Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals? the role of statistical perception in determining whether awareness overflows access. *Cognition*, *152*, 78–86. doi: <https://doi.org/10.1016/j.cognition.2016.01.010>
- Whitney, D., & Yamanashi, L. A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105–129. doi: <https://doi.org/10.1146/annurev-psych-010416-044232>
- Wyble, B., Folk, C., & Potter, M. C. (2013). Contingent attentional capture by conceptually relevant images. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 861–871. doi: <https://doi.org/10.1037/a0030517>

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