Shared cognitive mechanisms involved in the processing of scene texture and scene shape

Vignash Tharmaratnam

Graduate Program in Psychology, University of Toronto, Toronto, ON, Canada ✉

Mihilkumar Patel

Faculty of Medicine, University of Ottawa, Ontario, Canada ✉

Matthew X. Lowe

Graduate Program in Psychology, University of Toronto, Toronto, ON, Canada ✉

Jonathan S. Cant

Graduate Program in Psychology, University of Toronto, Toronto, ON, Canada
Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada ✉

Recent research has demonstrated that the parahippocampal place area represents both the shape and texture features of scenes, with the importance of each feature varying according to perceived scene category. Namely, shape features are predominately more diagnostic to the processing of artificial human–made scenes, while shape and texture are equally diagnostic in natural scene processing. However, to date little is known regarding the degree of interactivity or independence observed in the processing of these scene features. Furthermore, manipulating the scope of visual attention (i.e., globally vs. locally) when processing ensembles of multiple objects—stimuli that share a functional neuroanatomical link with scenes—has been shown to affect their cognitive visual representation. It remains unknown whether manipulating the scope of attention impacts scene processing in a similar manner. Using the well-established Garner speeded-classification behavioral paradigm, we investigated the influence of both feature diagnosticity and the scope of visual attention on potential interactivity or independence in the shape and texture processing of artificial human–made scenes. The results revealed asymmetric interference between scene shape and texture processing, with the more diagnostic feature (i.e., shape) interfering with the less diagnostic feature (i.e., texture), but not vice versa. Furthermore, this interference was attenuated and enhanced with more local and global visual processing strategies, respectively. These findings suggest that the scene shape and texture processing are mediated by shared cognitive mechanisms and that, although these representations are governed primarily via feature diagnosticity, they can nevertheless be influenced by the scope of visual attention.

Introduction

We frequently encounter a wide and complex variety of visual stimuli daily, such as objects, faces, animals, trees, and buildings. Each of these types of visual stimuli has been thoroughly investigated in both the behavioral psychophysical and neuroimaging literature, thus contributing to our understanding of their cognitive and neural representations, respectively. Although these stimuli are typically studied in isolation, one commonality that unites our experiences with these stimuli is their ubiquitous presence within scenes, a point that underscores the need to better understand how entire scenes are represented. Critically, it is essential to investigate the shared and independent cognitive mechanisms responsible for representing visual feature information within a scene. Using a behavioral paradigm, we will be investigating the functional interactivity and independence of scene feature processing (specifically scene shape and scene texture) and comparing our findings to what is known about scene representation from the functional magnetic resonance imaging (fMRI) literature.

Within the past 20 years, fMRI studies have demonstrated that different regions of the brain participate in different aspects of scene representation. For example, the retrosplenial complex is involved in

Received August 28, 2020; published July 16, 2021

navigation using knowledge retrieved from memory (Malcolm, Groen, & Baker, 2016; Silson, Steel, & Baker, 2016), the lateral occipital complex (LOC) in processing local objects within scenes (MacEvoy & Epstein, 2011), and the occipital place area in processing local features of scenes (e.g., boundaries, navigational affordances) for navigation through the immediate environment (Bonner & Epstein, 2017; Julian, Ryan, Hamilton, & Epstein, 2016; Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016). Notably, the parahippocampal place area (PPA) is involved in global scene perception (Epstein & Kanwisher, 1998).

PPA represents and identifies scenes on the basis of a wide range of global scene features. For example, PPA is sensitive to processing the layout or geometry of space (Epstein, Graham, & Downing, 2003), the spatial boundary (open/closed) of a scene (Park, Brady, Greene, & Oliva, 2011), the object content of a scene (Harel, Kravitz, & Baker, 2013), the category a scene belongs to (artificial/natural; Walther, Caddigan, Fei-Fei, & Beck, 2009), the contour junction statistics of scenes (Choo & Walther, 2016), and the spatial frequency content of scenes (Berman, Golomb, & Walther, 2017). One explanation for the finding that PPA is sensitive to processing a wide variety of scene features lies with the principle of feature diagnosticity: visual features that are most important to perform the task at hand are preferentially used, given the available visual information (Oliva & Schyns, 1997). For instance, colored surfaces within a scene are diagnostic to the categorization of some types of scenes (e.g., forests are green, oceans are blue) but not others (e.g., urban landscapes) (Oliva & Schyns, 2000). Thus, although PPA rapidly processes and represents multiple types of scene features, the principle of feature diagnosticity suggests that some features would receive preferential focus depending on how useful the feature was to the task at hand. This was precisely shown by Lowe, Gallivan, Ferber, and Cant (2016) in a study that used multivoxel pattern analysis to examine the potential diagnosticity of different visual features (i.e., scene geometry vs. scene texture) to the representation of different categories of scenes (i.e., natural vs. manufactured) in PPA. Results indicated that scene geometry was more diagnostic than scene texture when classifying manufactured scenes (e.g., city landscapes or indoor rooms, scene categories that are typically dominated by horizontal and vertical lines; see Oliva & Torralba, 2001). In contrast, when classifying natural scenes—where textural surfaces can play a larger role in scene identification (e.g. sand in a desert, water in a lake; see Oliva & Torralba, 2001)—shape and texture were equally diagnostic. These findings suggest that the expansive list of scene features represented in PPA may be explained by the principle of feature diagnosticity.

From the literature discussed above, it is clear that both structural (i.e., shape/geometric information)

and surface-related (i.e., texture, color) visual cues are diagnostic to scene perception. This comes from human neuroimaging research demonstrating the importance of spatial layout in scene processing (for review, see Epstein & Baker, 2019), and the joint contribution of both shape and texture to scene representation in PPA (Lowe et al., 2016; Lowe, Rajsic, Gallivan, Ferber, & Cant, 2017). Beyond neuroimaging research in neurologically intact individuals, neuropsychological research with patients 1, who has a profound visual form agnosia and cannot accurately perceive structual details of objects but can perceive their surface properties (Humphrey, Goodale, Jakobson, & Servos, 1994), demonstrated that patients 1 exhibited higher activation in PPA for appropriately colored scenes compared with black-and-white versions of these scenes (the latter of which can only be recognized on the basis of structural cues; Steeves, Humphrey, Culham, Menon, Milner, & Goodale, 2004). Moreover, recent neuroimaging and neurophysiological research has demonstrated the importance of both shape and texture to scene-selective regions of the macaque brain, with potentially greater representational weighting for the latter (Kornblith, Cheng, Ohayon, & Tsao, 2013). Importantly, these neuroimaging, neuropsychological, and neurophysiological findings are consistent with behavioral scene research. For many years, much of this research focused on shape-related visual cues, owing in part to the findings that both rats and human infants reorient themselves in their environment solely on the basis of geometric cues (Cheng, 1986; Hermer & Spelke, 1994). However, subsequent research revealed the additional role of surface-based visual cues (such as color and texture) to scene perception and recognition. For example, natural scenes are recognized more quickly when presented in color, as opposed to in black-and-white (Gegenfurtner & Rieger, 2000), a finding that mirrors the neuropsychological research discussed above (Steeves et al., 2004). Moreover, surface-based cues can be used to categorize scenes (i.e., scene "gist") without the need to identify particular objects in those scenes (Biederman, Mezzanotte, & Rabinowitz, 1982; Møller & Hurlbert, 1996; Oliva & Schyns, 1997, 2000; Oliva & Torralba, 2001; Schyns & Oliva, 1994, 1997; Vailaya, Jain, & Zhang, 1998). Finally, models that use textural information have successfully demonstrated how the visual system represents scenes, particularly beyond the fovea where visual information is degraded (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, 2011). Taken together, the number of lines of evidence discussed above demonstrate the importance of both shape and surface-properties to scene representation, but it remains unclear how the processing of these features are represented in scene-selective cortex (i.e., interactively or independently), and how their representation relates to the processing of global scene properties in PPA. Gaining a deeper understanding of

these issues is important, because it will not only inform future computational and neural-network models of scene representation, but may also resolve ambiguities in the neuropsychological literature relating to the relative weighting of shape- versus surface-related cues in everyday scene perception and recognition (e.g., Robin, Lowe, Pishdadian, Rivest, Cant, & Moscovitch, 2017).

Recently, studies have demonstrated that the representation in PPA extends beyond the processing of global scene properties, to global features of a variety of stimuli. For instance, Cant and Goodale (2007, 2011) demonstrated that attention to different features of single objects presented in isolation (i.e., not within the context of a scene) activates different regions of occipitotemporal cortex, with object shape preferentially activating LOC, and object texture preferentially activating a region of the collateral sulcus overlapping PPA. These results suggest that engaging global processing in PPA may depend upon the stimulus attended to (i.e., within a scene, object shape and texture are processed more locally, whereas when focusing on a single object, texture may be processed more globally than shape). Moreover, Cant and Xu (2012) found that in addition to processing texture, PPA is sensitive to processing the shape and texture features from ensembles of multiple objects (e.g., leaves on a tree, grapes on a vine, etc.), likely because scene, texture, and ensemble perception share similar underlying computational processes (i.e., the extraction of global statistical features from repeating and redundant visual information). This again demonstrates that the degree of global processing of a visual feature in PPA may be stimulus dependent (e.g., shape is processed more locally in single objects but more globally in object ensembles). Furthermore, changes in the global ratio of different objects comprising a heterogeneous ensemble are encoded in PPA, whereas LOC encodes changes in the spatial arrangement of objects within an ensemble, a change that requires processing of the local visual elements of the group (Cant & Xu, 2015). These findings suggest that the representation in PPA is biased more toward global feature processing of visual stimuli, whereas the representation in LOC may be biased more toward local feature processing of visual stimuli, which is consistent with findings in the scene perception literature (e.g., MacEvoy & Epstein, 2011). Overall, these findings demonstrate that, in addition to feature diagnosticity, another important factor governing the cognitive representation in PPA (and LOC) is the scope of visual attention, with the PPA involved in the global processing of multiple types of stimuli.

Previous studies have focused on fMRI neuroimaging to help disentangle the heterogeneous nature of visual feature representation in PPA. Investigating potential functional sub-divisions within PPA, Baldassano and colleagues (2013) revealed a posterior/anterior functional split within this scene-selective region, with the posterior region being more involved in visual perception, and the anterior being more involved in memory. However, it remains unclear how this varied functional sensitivity seen across different subdivisions of PPA translates into everyday scene perception. Moreover, we do not know whether various scene features (e.g., texture, geometry, spatial frequency, spatial boundary, etc.) are processed and subsequently represented in a more independent or interactive manner. Finally, it remains unclear whether any potential feature independence or interactivity is influenced by a feature's diagnosticity to scene perception or the scope of visual attention (i.e., local vs. global).

As an alternative approach to neuroimaging research, Garner's speeded-classification task (Garner, 1974) is a sensitive behavioral paradigm that can be used to investigate potential interactivity or independence in visual feature processing. Participants attend and classify different values of a relevant feature in baseline trials, where only the relevant feature varies (e.g., scene spatial geometry). This performance is then compared against that in filtering trials, where participants still classify variations in the relevant feature, but this time in the presence of random variation in a second, irrelevant feature (e.g., scene texture). Importantly, the type of stimuli presented and the number of trials are the same for baseline and filtering blocks. If participants can easily ignore changes in the irrelevant feature (i.e., no changes in response latency and classification accuracy between baseline and filtering trials) when classifying the relevant one, then these features are said to be processed independently and are labeled as separable dimensions. Examples of known separable dimensions include the position of lines and their luminance contrast (Shechter & Hochstein, 1992), the color and texture of objects (Cant, Large, McCall, & Goodale, 2008), and the sex and emotion of bodies (Gandolfo & Downing, 2020). If, however, participants have worse performance in the filtering trials (i.e., increased response latency or a greater number of classification errors) compared with the baseline, then these features are not processed independently and are said to be integral feature dimensions, which demonstrate Garner interference. Examples of known integral features that interfere with each other are the length of lines and their orientation (Dick & Hochstein, 1988) and the length and width of objects (Cant & Goodale, 2009; Cant et al., 2008; Dykes & Cooper, 1978; Felfoldy, 1974; Ganel & Goodale, 2003). Garner interference can also be observed asymmetrically, where one feature interferes with the processing of another feature, but not vice versa. Examples of this include identity/sex interfering with emotion processing in face perception (Atkinson & Burt, 2005; Schweinberger, Burton, & Kelly, 1999; Schweinberger, Soukup, & Konstanz, 1998), sex interfering with weight processing in body

perception (Johnstone & Downing, 2017), and body posture interfering with facial identity processing (Reed, Bukach, Garber, & McIntosh, 2018). Thus Garner's task has the potential to elucidate similarities or differences in cognitive representation for different types of visual feature processing, including perceptual processing of high-level stimulus classes (e.g., faces and bodies), and this can inform novel predictions about the underlying neural representation of these features. Of course, this relationship is reciprocal, as neuroimaging findings can lead to the testing of novel hypotheses regarding cognitive processing in behavioral paradigms.

Garner's speeded-classification task has been used extensively to link brain activity with behavior. As mentioned previously, Cant and Goodale (2007, 2011) demonstrated that attending to object shape selectively activates LOC whereas attending to object texture selectively activates PPA. The fact that different regions processed different object features, combined with neuropsychological findings that surface-property perception is preserved in cases of extreme disruptions to shape perception (Humphreys, Romani, Olson, Riddoch, & Duncan, 1994; Milner, Perrett, Johnston, Benson, Jordan, Heeley, Bettucci D, Mortara F, Mutani R, Terazzi E, Davidson, 1991), suggested that object shape and texture may be processed independently of each other. Using Garner's speeded-classification task, Cant et al. (2008) confirmed this prediction, demonstrating functional independence in the processing of object shape and texture, which underscored the utility of using neuroimaging results to generate testable behavioral hypotheses.

Furthermore, Cant and colleagues used both Garner's behavioral task and a neuroimaging approach to investigate global feature processing in object-ensemble perception. Using Garner's task, Cant, Sun, and Xu (2015) demonstrated that, unlike single-object perception, global processing of the shape and texture of object ensembles interfere with each other, suggesting that the two features may share a similar underlying neural substrate. Using fMRI-adaptation, Cant and Xu (2017) found that this was indeed the case, because similar results were found within PPA when participants selectively attended globally to either the shape or texture of object ensembles. Interestingly, Cant et al. (2015) also found that the scope of attention influenced the presence of Garner interference between ensemble shape and texture. Specifically, a task requiring a global-processing strategy when perceiving ensemble shape and texture—effectively expanding the scope of visual attention—resulted in interference, while implementing a local-processing strategy resulted in independence. Taken together, these studies demonstrate a link between neuroimaging and behavioral data and, importantly, reveal that Garner's task has the potential to inform our understanding of

the cognitive representation of different visual features, specifically within PPA.

In the current study, across 3 experiments (and 2 control experiments; See Supplementary Materials and Figures), we used Garner's speeded-classification task with artificial human-made scene stimuli (i.e., indoor rooms) to investigate the processing of scene shape and texture, two features that are represented in PPA (Lowe et al., 2016; Lowe et al., 2017). Specifically, we investigated whether these scene features are processed interactively or independently of each other (for a more detailed treatment of this issue, see "Experiment 1"), and whether feature diagnosticity and the scope of attention can influence this relationship.

General methods

Participants

Thirty-one participants (eight males, 23 females; 28 right-handed, three left-handed; mean age 18.2 years; age range, 17–23 years) participated in Experiment 1 of this study, 32 participants (12 males, 20 females; 28 right-handed, four left-handed; mean age 18.8 years; age range, 17–25 years) participated in Experiment 2 of this study, and 31 participants (10 males, 21 females; 30 right-handed, one left-handed; mean age, 19.9 years; age range, 18–32 years) participated in Experiment 3 of this study. The Supplementary Materials include a description of two additional experiments, S1 and S2, which provide important controls to, and replications of, the findings in Experiments 2 and 3, respectively. The participants were chosen from undergraduate students taking introductory psychology courses at the University of Toronto Scarborough. Participants had normal or corrected-to-normal vision and received course credit for participation. All participants provided informed consent, and the study was approved by the University of Toronto Research Ethics Review Board.

Stimuli and apparatus

Stimuli were computer rendered using Adobe Photoshop CC 2015 (Adobe Systems Incorporated, San Jose, CA, USA) and Blender 2.77 software (Stichting Blender Foundation, Entrepotdok, Amsterdam, the Netherlands). The mean and standard deviation of luminance across the two stimulus sets used within each experiment (see below) was matched using the SHINE toolbox within MATLAB software.

In each experiment, two types of stimuli were presented to participants, and each type of stimulus contained variations in both shape and texture. In Experiment 1, participants viewed single isolated

objects (i.e., not presented within the context of a scene), rendered in one of two different shapes (square vs. rectangular) and in one of two different textures (mesh vs. spotted patterns), and artificial scenes (i.e., indoor rooms) devoid of objects and containing variations in both features (shape: triangular vs. pentagonal; texture: mesh vs. spotted patterns; henceforth referred to simply as "original scene stimuli"; see Figure 2). In Experiments 2 and 3, only scenes were presented, and the same variations in shape (i.e., triangular vs. pentagonal) and texture (i.e., mesh vs. spotted) were used in both experiments. However, additional textural manipulations, distinct from those used in Experiment 1, were applied to the scenes used in Experiments 2 and 3. In Experiment 2, one set of scene stimuli was constructed that contained bold line segments at the junction of adjacent surfaces (henceforth referred to as "bolded-edge" scene stimuli), and another set was constructed that contained a mixture of both types of textures (i.e., one texture was applied to the lateral walls/surfaces, and the other texture was applied to the center/background wall/surface; henceforth referred to as "two-textured" scene stimuli; see Figure 3). The scene stimuli used in Experiment 3 were identical to those used in Experiment 2, with the exception that the floors of the scene stimuli were textured with the same texture as their lateral surfaces/walls (see Figure 4).

Participants were tested in a darkened room and had their head mounted on a headrest, elevated 32 cm from the surface of the table and 40 cm from a CRT monitor (1920 × 1080 pixels, screen refresh rate: 60 Hz). Object stimuli subtended 43.3° vertically and 43.3° horizontally, and scene stimuli subtended 95.7° vertically and 65.6° horizontally. Participants made classifications of shape or texture (for both objects and scenes) by pressing the "1" or "3" button on the number pad of the keyboard in front of them, with their right index finger on the "1" key and their right middle finger on the "3" key. Stimulus presentation, as well as the collection of response latency and accuracy data were controlled using E-prime 2.0 software (Psychology Software Tools, Inc., Sharpsburg, PA, USA), and data analysis was conducted using MATLAB (ver. R2018b) and R software (2020).

Procedure

The procedure of this experiment was adapted from Cant and colleagues (2008). The order of stimulus set presentation (e.g., objects vs. unmodified scenes in Experiment 1) was counterbalanced between participants within each experiment. Within each stimulus set, or task, there were two conditions involving attention to a specific feature: the classification of stimulus shape and texture. The order of these conditions within each task was counterbalanced across participants. Furthermore, within each attended feature condition, participants completed two baseline blocks (where only the relevant, or attended, feature varied; e.g., shape) and two filtering blocks (where both relevant and irrelevant features varied; e.g., shape and texture), with 32 trials within each block (block order was pseudorandom for each combination of attended feature and task). This design yielded a total of eight blocks of trials per task (two baseline and two filtering blocks for both the shape and texture conditions) and thus 16 total blocks per experiment (2 tasks × 2 attended features/task × 4 blocks/attended feature × 32 trials/block = 512 trials/participant).

For each attended feature, the participants were asked to attend to the center of the screen and classify as quickly and accurately as possible either the shape of the stimulus in the shape condition, or the texture of the stimulus in the texture condition. Stimuli were presented in a pseudorandom order within each block of trials and participants responded by pressing either the "1" or "3" key for a specific shape or texture (e.g., in Experiment 1, pressing "1" for a triangular room and "3" for a pentagonal room in the scene shape condition, and pressing "1" for a spotted textured room and "3" for a mesh textured room in the scene texture condition). For two-textured scenes in Experiments 2 and 3, participants were asked to classify the texture seen only in the center of the scene on the back wall/surface of the room. The assignment of a particular shape or texture to a particular number key on the keyboard was counterbalanced across participants. Each block began with instructions to the participant about the task they were about to complete. The first trial in each block began with a central fixation cross for 2000 ms, was followed by the presentation of a stimulus that remained onscreen until the participant's response, and was then followed by an interstimulus interval indicated by a fixation point at the center of the screen that lasted for 2000 ms (see Figure 1). This trial structure repeated until all stimuli within a block were presented.

At the beginning of each of the four unique combinations of task and attended feature (i.e., shape and texture conditions for both stimulus sets) within each experiment, participants were shown examples of the stimuli and were given instructions on how to complete the classification task. Next, before commencing the experimental blocks of trials for a particular attended feature (e.g., object shape in Experiment 1), participants completed 20 practice trials that had the same stimuli, presentation parameters, and response key assignment as the ensuing experimental blocks of trials. In these practice trials participants received feedback as to the accuracy of their responses (i.e., "correct" or "incorrect" was presented on the screen after each response). Verbal feedback was also provided as necessary. This feedback was specific to the practice trials and was not provided in the experimental

Garner Speeded-Classification Task
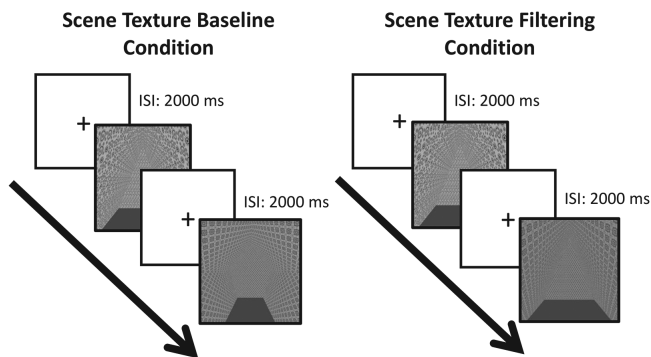**Example of Task: Attend only to scene texture**

Figure 1. Schematic of experimental design. In all blocks, the stimuli are presented in a pseudorandom order and remained on the screen until a response is made. Once a response is made an inter-stimulus interval of 2000 milliseconds (ms) commences and is followed by the presentation of another stimulus on the subsequent trial. In the baseline blocks only the attended feature varies between stimuli (in the example above, scene texture), while in the filtering blocks both the attended feature (e.g., scene texture) and the unattended feature (e.g., scene shape) can vary between stimuli.

trials. Once the participants completed the practice trials, they proceeded to the experimental blocks of trials, and were provided with written instructions onscreen between each block of trials reminding them of the task, attended feature, and number key assignment in the ensuing block of trials. Between the two tasks of each experiment, participants had up to one minute to take a break and relax their eyes before being allowed to proceed to the second half of the experiment.

Data analysis

In all experiments, we focused our data analysis on participants' response latencies, rather than error rates, since previous studies using an equivalent paradigm demonstrated better reliability for the former measure when using Garner's speeded-classification task (Cant & Goodale, 2009; Cant et al., 2008; Cant et al., 2015). Nevertheless, for completeness we report the error data in the supplementary material. Only correct responses were entered into the response latency analysis, and an outlier analysis on correct responses was conducted. For each participant separately and for each experiment as a whole, responses that were more than 1.5 times the interquartile range above or below the third and first quartile mean response latency, respectively, for a given block type were excluded. For accuracy, the mean and standard error (SEM) were calculated for

each block type and participant separately, and then mean accuracies that were 1.5 interquartile range above or below the third and first quartile mean accuracy, respectively, for a given block type were excluded. Finally, if a participant's mean response latency or accuracy was deemed an outlier (using the criteria above for each combination of stimulus, attended feature, and block type), both dependent measures were excluded from all analyses.

These outlier analyses resulted in the removal of different numbers of participants across the four conditions used in each experiment. Note that the type of statistical analysis we used (i.e., repeated measures mixed-effects model; see below) is robust to differences in sample size across experimental conditions in a repeated-measures design (Quené & Van Den Bergh, 2004). Below we describe how many participants were removed from each condition in each experiment (based on the outlier exclusion criteria for both response latency and accuracy outlined above). In Experiment 1, different numbers of participants were excluded from the object shape, object texture, scene shape, and scene texture conditions (six, two, three, and six participants, respectively), and the same was true for the bolded-edge scene shape, bolded-edge scene texture, two-textured scene shape, and two-textured scene texture conditions in Experiments 2 and 3 (Experiment 2: six, three, five, and two participants, respectively; Experiment 3: one, six, three, and three participants, respectively).

The data within each experiment was analyzed using a $2 \times 2 \times 2$ repeated measures mixed-effects model ($\alpha = 0.05$; type III Wald F tests with Kenward-Roger df). Main effects included stimulus (Experiment 1: objects vs. original scenes; Experiments 2 and 3: bolded-edge vs. two-textured scenes), attended feature (attending to shape vs. texture), and block type (baseline vs. filtering), with a random intercept given for each subject (Barr, Levy, R., Scheepers, C., & Tily, 2013). Significant effects were investigated using planned pairwise comparisons ($\alpha = 0.05$, two-tailed), with corresponding Bayes factors (BF_{10}) reported using JASP software (JASP Team, 2016) to verify nonsignificant effects where appropriate (i.e., the absence of interference in object or scene processing). According to standard convention (Biel & Friedrich, 2018; Jeffreys, 1998), a $BF_{10} < 1$ constitutes evidence in favor of the null hypothesis over the alternative hypothesis. Specifically, Garner interference or independence was evaluated by comparing response latencies in the baseline and filtering blocks separately for each combination of attended feature and stimulus set (i.e., only two comparisons were made for each stimulus set).

The data across experiments was analyzed using a 6×2 mixed-effects model ($\alpha = 0.05$; type III Wald F tests with Kenward-Roger df), collapsing across block type. Main effects included stimulus (objects, original scenes, bolded-edge scenes, two-textured scenes, bolded-edge

scenes with textured floors, two-textured scenes with textured floors), and attended feature (attending to shape vs. texture), with a random intercept given for each subject. Planned comparisons were conducted between the object stimuli used in Experiment 1 and the various scene stimuli used in Experiments 2 and 3, and also between the original scenes used in Experiment 1 and the scene stimuli used in Experiments 2 and 3 ($\alpha = 0.05/8 = 0.00625$, two-tailed), separately for each scene feature attended to.

In addition, to evaluate the potential impact of baseline differences in the processing of shape and texture on resultant Garner interference, for each stimulus within each experiment, differences in response latency at baseline (i.e., texture baseline response latency – shape baseline response latency) across participants were correlated with differences in response latency for the Garner interference effect (i.e. filtering – baseline), separately for each stimulus, for either shape and texture (Pearson's correlation coefficient, $\alpha = 0.05/2 = 0.025$, two-tailed; see Supplementary Materials and Supplementary Figure S4).

Experiment 1

The main purpose of Experiment 1 was to assess whether scene shape and scene texture are processed independently or interactively. Based on previous fMRI results, there are at least three different possibilities that we may observe when using Garner's task. First, given that both of these global scene features are processed within PPA, the processing of these scene features may mutually interfere with each other (i.e., full or reciprocal Garner interference), similar to what is observed with the processing of ensemble shape and texture (i.e., interference between both features using Garner's task and sensitivity to processing both features in PPA; Cant et al., 2015; Cant & Xu, 2017). Second, the finding of different populations of PPA neurons involved in processing the shape versus the texture of artificial scenes may translate into behavioral independence when processing these features. In other words, this would mirror the same type of independence observed when different cortical regions are involved in processing different object features (i.e., independence of object shape vs. texture in LOC and PPA, respectively; Cant & Goodale, 2007; Cant & Goodale, 2011; Cant et al., 2008). A third possibility stems from the finding of feature diagnosticity in the representation of scenes in PPA. Although multiple scene features are processed within the PPA, certain features may receive preferential processing depending on the task at hand. As mentioned previously, Lowe et al. (2016) demonstrated that shape was more diagnostic

than texture when processing artificial human–made scenes. Given that we are using indoor rooms as our scene category in this study, we may observe asymmetric interference between the processing of scene shape and texture. That is, the more diagnostic feature (i.e., shape) will interfere with the processing of the less diagnostic feature (i.e., texture), but not vice versa.

Since observing independence between scene features would rest on a null result, it was important to pair the scene task with another task that reliably produces independence between shape and texture processing. This would ensure that any independence we observe in the scene task is not likely to be explained by confounds within our experimental design. To that end, participants also classified the shape and texture of single objects (not presented within the context of a scene), and we predicted independence in the processing of these object features, consistent with previous studies (Cant et al., 2008; Cant et al., 2015).

Results and discussion

Participants performed very accurately in each task, as overall accuracy was near ceiling (mean accuracy across all blocks across participants was 97.35%, $SEM = 0.15\%$; see Supplementary Materials and Supplementary Figure S3A). For response latency, the main effect of stimulus (objects: $M = 540.28$ ms, $SEM = 6.63$ ms; original scenes: $M = 585.94$ ms, $SEM = 9.10$ ms; $F(1,189.18) = 41.24$, $p < 0.001$) was significant. In contrast, nonsignificant results were observed for the main effects of attended feature and block type, and all interactions (all $Fs(1,\sim189) < 1.88$, all $ps > 0.17$).

Based on our prediction of independence between shape and texture processing in object perception, and the three distinct predictions for the processing of these features in scene perception, we conducted pairwise comparisons between the baseline and filtering blocks separately for each combination of attended feature (i.e., shape and texture) and stimulus (i.e., objects and scenes). As predicted, we found nonsignificant differences between baseline and filtering blocks when participants attended both object shape (baseline: $M = 531.34$ ms, $SEM = 15.28$ ms; filtering: $M = 529.91$ ms, $SEM = 13.53$ ms; $t(24) < 0.14$, $p = 0.89$, $BF_{10} = 0.368$, $p = 0.269$) and object texture (baseline: $M = 547.22$ ms, $SEM = 12.75$ ms; filtering: $M = 548.82$ ms, $SEM = 13.59$ ms; $t(28) < 0.19$, $p = 0.85$, $BF_{10} = 0.299$, $p = 0.165$). In the scene shape task we did not observe a significant difference between baseline and filtering blocks (baseline: $M = 593.20$ ms, $SEM = 17.01$ ms; filtering: $M = 582.74$ ms, $SEM = 18.67$ ms; $t(27) = 1.03$, $p = 0.31$, $BF_{10} = 0.474$, $p = 0.321$) but did find a significant difference when participants attended scene texture (baseline: $M = 568.15$ ms, $SEM = 17.99$ ms; filtering: $M = 598.58$ ms,
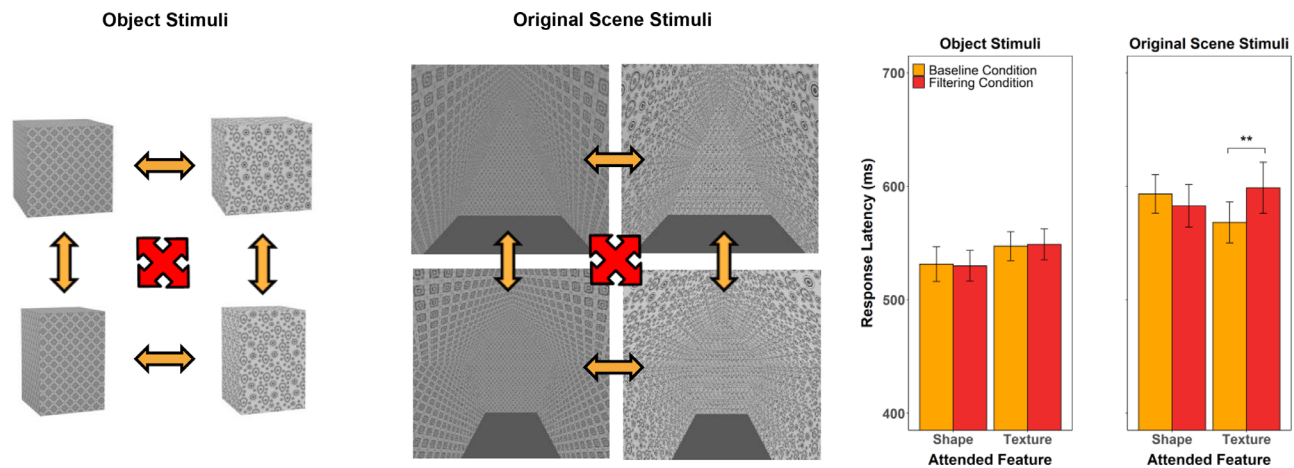
Figure 2. Results from Experiment 1 (n = 31), using object stimuli and original scene stimuli. Orange arrows/bars represent baseline blocks (where only the attended feature changes across trials), and red arrows/bars represent filtering blocks (where both the attended and unattended features change across trials). These (and all subsequent) stimuli displayed are for illustrative purposes only and thus do not reflect the actual size of stimuli used in the experiments. **$p < 0.01$; ms = milliseconds.

$SEM = 22.55$ ms; $t(24) = 2.81$, $p < 0.01$) (see Figure 2). This result is consistent with our third prediction, namely, asymmetric interference between the processing of scene shape and scene texture. That is, the more diagnostic feature when processing artificial scenes (scene shape) interferes with the processing of the less diagnostic feature (scene texture), but not vice versa. Taken together, these results replicate previous and well-established findings of independence between shape and texture processing in object perception (e.g., Cant & Goodale, 2009; Cant et al., 2008; Cant et al., 2015), but importantly, demonstrate, for the first time, behavioral interference between shape and texture processing in scene perception.

Experiment 2

Having established asymmetric interference between the processing of scene shape and texture in Experiment 1, the purpose of Experiment 2 was to examine whether this asymmetric interference effect, which was governed by feature diagnosticity, is affected by manipulations to the scope of visual attention. Using object ensembles (which, like scenes, are processed globally within PPA; see Cant & Xu, 2017) in a behavioral paradigm, Cant et al. (2015) demonstrated that implementing a global-processing style accentuated Garner interference between ensemble shape and texture, whereas attending locally eliminated interference. With this in mind, we made two manipulations to the scene stimuli in this experiment to influence the participants' scope of attention.

Previous research has demonstrated that contour junctions are particularly important visual cues in scene categorization (Walther & Shen, 2014; Wilder, Dickinson, Jepson, & Walther, 2018), and the neural representation of global scene features in high-level scene-selective cortex (e.g., PPA) is particularly sensitive to contour junctions (Choo & Walther, 2016). On the basis of these findings, we created one type of stimulus (bolded-edge scenes) whereby the contour junctions of adjacent surfaces were demarcated with a bold line, thus enhancing the saliency of the scene's global shape (see Figure 3).

One way that we parse objects within scenes is based on texture segmentation, where local textural differences across surfaces demarcate foreground (i.e., "object") from background (i.e., "scene") (Wagemans, Elder, Kubovy, Palmer, Peterson, Singh, & von der Heydt, 2012). Such texture segmentation processes have been proposed to be localized in early visual cortex (i.e., V2), where neurons process local textural discontinuities (Schmid & Victor, 2014). With this in mind, we created a second type of stimulus (two-textured scenes), in which two different textures were used in each scene, with one applied to the lateral walls/surfaces and another applied to the central wall/surface. This effectively created a perceptual "pop-out" effect based on the segmentation of dissimilar textures across surfaces and increased the saliency of the central wall's shape and texture (as if an object had been placed in the center of the scene).

Given that Parkhurst, Law, and Niebur, (2002) have shown that increasing the saliency of features (through regional differences in color, luminance, and orientation) within a scene increases the allocation of attention toward them, we predicted that our stimulus manipulations would promote global attention
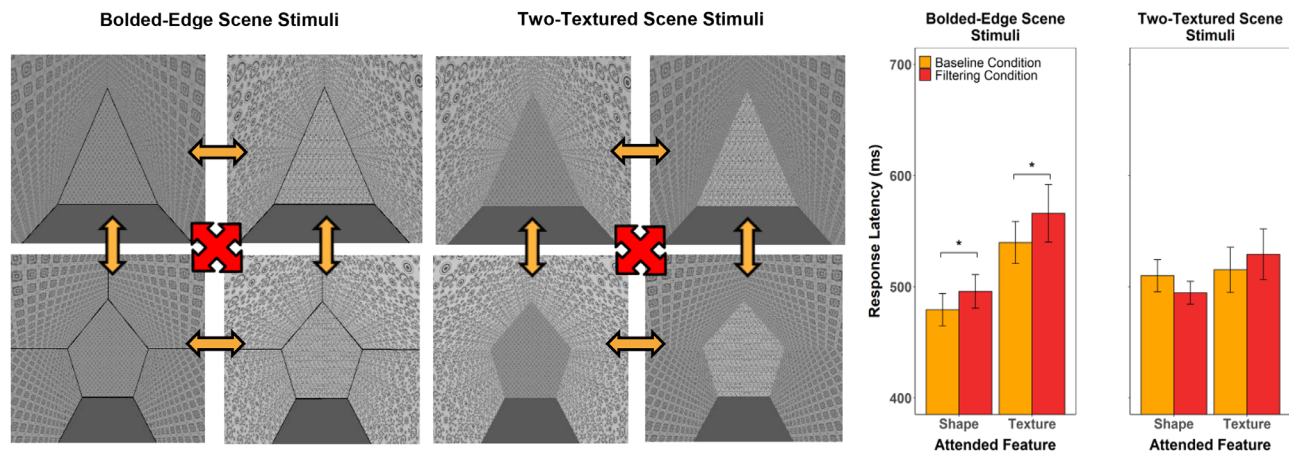
Figure 3. Results from Experiment 2 (n = 32), using bolded-edge and two-textured scene stimuli. Orange arrows/bars represent baseline blocks (where only the attended feature changes across trials), and red arrows/bars represent filtering blocks (where both the attended and unattended features change across trials). *$p < 0.05$; ms = milliseconds.

in bolded-edge scenes, and local attention to the central wall in two-textured scenes. Of particular interest in this experiment is the comparison of the importance of feature diagnosticity and attentional scope in modulating Garner-interference effects in scene perception. If scene feature diagnosticity largely governs the asymmetric interference effect, then asymmetric interference between shape and texture should persist in both types of stimuli, even when attention is drawn to local elements in two-textured scenes. If, however, attentional scope plays a larger role, then we should see stronger interference (compared with Experiment 1) when attention is manipulated globally, and the elimination of interference with a local attentional manipulation. A finding in between these possibilities (e.g., similar interference in bolded-edge scenes and the elimination of interference in two-textured scenes) would point to the dual importance of both cognitive processes.

Results and discussion

Participants performed very accurately in each task in Experiment 2, as overall accuracy was again near ceiling (mean accuracy across all blocks was 97.22%, $SEM = 0.23\%$; see Supplementary Materials and Supplementary Figure S3B). For response latency, the main effect of attended feature (shape: $M = 505.60$ ms, $SEM = 7.82$ ms; texture: $M = 540.46$ ms, $SEM = 10.87$ ms; $F(1,197.75) = 13.20$, $p < 0.001$) and the stimulus-by-attended feature interaction ($F(1,197.51) = 10.41$, $p < 0.01$) were both significant, but all other main effects and interactions were not significant (all $Fs(1,\sim198) < 2.61$, all $ps > 0.11$).

Based on our predictions, we investigated potential Garner interference by conducting pairwise

comparisons between the baseline and filtering blocks separately for each attended feature (i.e., shape and texture) in each stimulus set (i.e., bolded-edge and two-textured scenes). For bolded-edge scenes we found that reaction times in the baseline and filtering blocks differed when attending to both scene shape (baseline: $M = 479.37$ ms, $SEM = 14.54$ ms; filtering: $M = 495.81$ ms, $SEM = 15.02$ ms; $t(25) = 2.48$, $p < 0.05$), and scene texture (baseline: $M = 539.78$ms, $SEM = 18.87$ ms; filtering: $M = 566.11$ ms, $SEM = 25.86$ ms; $t(28) = 2.13$, $p < 0.05$), indicating reciprocal interference. In contrast, there was no interference observed in the two-textured scene stimuli when attending to either shape (baseline: $M = 510.01$ ms, $SEM = 14.48$ ms; filtering: $M = 494.57$ ms, $SEM = 10.19$ ms; $t(26) = 1.69$, $p = 0.10$, $BF_{10} = 1.400$, $p = 0.58$) or texture (baseline: $M = 515.17$ ms, $SEM = 20.36$ ms; filtering: $M = 529.15$ ms, $SEM = 22.83$ ms; $t(29) = 1.59$, $p = 0.12$, $BF_{10} = 0.264$, $p = 0.208$) (see Figure 3).

These results reveal that our scene stimulus manipulations replicated (i.e., bolded-edge scenes) and eliminated (two-textured scenes) the interference of scene shape on texture processing observed in Experiment 1, by expanding and contracting the scope of attention, respectively. This demonstrates that both scene feature diagnosticity and the scope of visual attention contribute to the pattern of interference (or lack thereof) observed in the processing of scene shape and texture (see below). Surprisingly, scene texture interfered with scene shape processing in bolded-edge scenes, which we did not observe with the original scene stimuli used in Experiment 1. However, as shown in subsequent experiments (see Experiments 3, and S1 and S2 in the Supplementary Materials), this result does not replicate, whereas the interference of scene shape on scene texture processing does in each instance. This is further supported by the lack of a significant

effect of block type in the mixed-effects model. Thus the interference of scene texture on shape processing in this experiment is not consistently reliable across experiments and should not be interpreted further. As such, the more reliable finding, as will be seen in subsequent experiments, is asymmetric interference between scene shape and scene texture processing (as seen in Experiment 1).

The replication of asymmetric interference by bolding the edges of adjacent surfaces in the scene highlights the importance of feature diagnosticity in scene perception, since a global attentional manipulation did not strengthen the asymmetric interference effect observed in Experiment 1. However, it is difficult to conclude that feature diagnosticity plays a stronger role than attentional scope in governing scene feature interference effects based on these results alone. Indeed, it is still possible that the global spread of attention contributed to the asymmetric interference observed in bolded-edge scenes. To further investigate the effect of manipulating attention, we compared the response latencies when responding to bolded-edge scenes(and two-textured scenes) with the responses to the original scenes in Experiment 1 (see "Comparisons Across Experiments" for details).

In contrast to the results seen with bolded-edge scenes, using two-textured stimuli to localize attentional processing eliminated Garner interference between scene shape and texture processing. Since independence between the processing of shape and texture is routinely observed in object perception (Cant et al., 2008; Cant et al., 2015), this finding provides support for the notion that participants may have been using more local, "object-like" processing resources to perceive the two-textured scenes, based on texture segmentation. This result is also consistent with those in (Cant et al., 2015), which examined global and local processing of shape and texture in object-ensemble stimuli using a Garner interference paradigm. This validates the use of our scene stimulus manipulations since such effects would be expected given that both scenes and ensembles are functionally related and share similar underlying neural substrates (Cant & Xu, 2012; Cant & Xu, 2015; Cant & Xu, 2017). Together, the results of Experiment 2 reveal that interference in the perception of scene features depends partly upon the diagnosticity of the scene feature in question, and partly upon the scope of visual attention. This latter effect is further explored in Experiment 3, where additional scene manipulations were applied to accentuate global processing styles.

Experiment 3

The purpose of Experiment 3 was to further investigate how feature diagnosticity and the scope of visual attention contribute to interference effects observed in scene processing. In Experiment 2 we manipulated the focus of attention globally in bolded-edge scenes and more locally in two-textured scenes. In Experiment 3 we examine the processing of scene shape and texture further by manipulating both types of scene stimuli to potentially accentuate global-processing styles. Specifically, we textured the floor of each type of scene with the same texture that was applied to the lateral walls, where previously all scene stimuli contained a gray-colored textureless floor (see Figure 4). For bolded-edge scenes, texturing the floor to match the lateral and central walls would increase the saliency of the scene's global shape and texture, spreading attention globally, based on Gestalt grouping cues (i.e., texture similarity). For the same reason, we believe that attention will spread globally when viewing the two-textured scenes. Indeed, given that Gestalt grouping cues have been found to modulate the automatic spread of attention (Wannig, Stanisor, & Roelfsema, 2011), we predict that we will find asymmetric Garner interference for bolded-edge and two-textured scenes with textured floors. Alternatively, for two-textured scenes specifically, adding textured floors may accentuate a local-processing style by reinforcing the strong object "pop out" effect, also explained by Gestalt principles (i.e., segmentation based on texture dissimilarity). In this case, we would expect to see independence in the processing of shape versus texture in the two-textured scenes (as seen in Experiment 2). However, we believe the stronger effect will result from texture similarity, and thus predict asymmetric Garner interference (i.e., shape interferes with texture processing but not vice versa) for both types of scene stimuli.

Results and discussion

As in Experiments 1 and 2, accuracy was near ceiling in Experiment 3 (mean accuracy across all blocks was 97.72%, $SEM = 0.15\%$; see Supplementary Materials and Supplementary Figure S3C). For the response latency data, the stimulus-by-attended feature ($F(1,193.24) = 56.64$, $p < 0.001$) and the attended feature-by-block type ($F(1,193.50) = 6.43$, $p < 0.05$) interactions were significant. In contrast, the main effects, stimulus-by-block type interaction, and three-way interaction were all non-significant (all $Fs(1,\sim193) < 3.44$, all ps > 0.065).

Based on our predictions, we conducted pairwise comparisons between the baseline and filtering blocks separately for each combination of attended feature (i.e., shape vs. texture) and stimulus set (i.e., bolded-edge vs. two-textured scenes). These comparisons revealed significantly faster responses in the baseline compared with the filtering blocks when attending to texture in bolded-edge scenes with textured floors
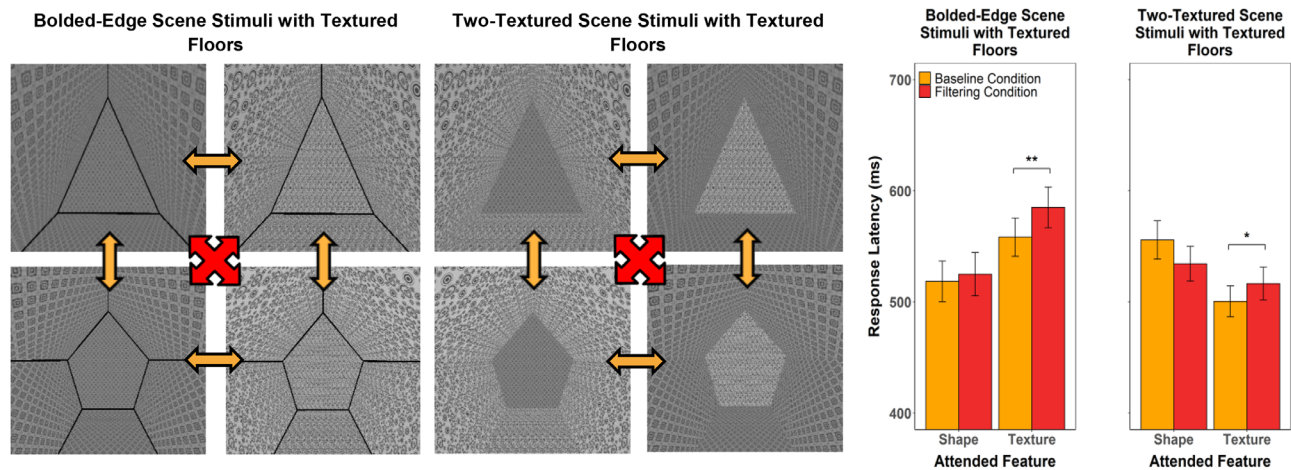
Figure 4. Results from Experiment 3 (n = 31), using bolded-edge and two-textured scene stimuli with textured floors. Orange arrows/bars represent baseline blocks (where only the attended feature changes across trials), and red arrows/bars represent filtering blocks (where both the attended and unattended features change across trials). *$p < 0.05$; **$p < 0.01$; ms = milliseconds.

(baseline: $M = 558.34$ ms, $SEM = 17.15$ ms; filtering: $M = 585.04$ ms, $SEM = 18.30$ ms; $t(24) = 3.06$, $p < 0.01$), but no difference across blocks when attending to shape (baseline: $M = 518.52$ ms, $SEM = 18.30$ ms; filtering: $M = 524.97$ ms, $SEM = 19.49$ ms; $t(29) = 0.73$, $p = 0.47$, $BF_{10} = 0.204$, $p = 0.246$).

Similarly, there were significantly faster response latencies in the baseline blocks, compared with the filtering blocks, when attending to texture in the two-textured scenes with textured floors (baseline: $M = 500.59$ ms, $SEM = 13.90$ ms; filtering: $M = 516.52$ ms, $SEM = 14.89$ ms; $t(27) = 2.47$, $p < 0.05$), but no difference between blocks when attending to shape (baseline: $M = 555.78$ ms, $SEM = 17.21$ ms; filtering: $M = 534.43$ ms, $SEM = 15.70$ ms; $t(27) = 1.77$, $p = 0.09$, $BF_{10} = 0.398$, $p = 0.556$) (see Figure 4).

These results replicate the asymmetric interference effect between texture and shape in bolded-edge scene stimuli (i.e., shape interfered with texture processing but not vice versa), and interestingly, we observe the same asymmetric interference effect in two-textured scenes (which was not observed in Experiment 2). Together, this demonstrates the importance of feature diagnosticity in scene perception, since the more diagnostic feature (i.e., shape) interfered with the less diagnostic feature (i.e., texture) in two different types of manufactured scenes. But these findings, and those in Experiment 2, also point to the contribution of the scope of visual attention, since using a local attentional manipulation eliminated interference (i.e., two-textured scenes in Experiment 2), and adding a global manipulation led to interference (i.e., two-textured scenes with textured floors in this experiment).

We contend that adding textured floors to both types of stimuli effectively spread attention globally, resulting

in asymmetric Garner interference by overriding the object "pop-out" effect present in two-textured scenes. To more comprehensively explore the impact that each of our scene stimulus manipulations had, we conducted an analysis where the response latency data was compared across all three experiments (for the results of two control experiments that completely replicated the results of Experiments 2 and 3 and demonstrated that the interference effects observed could not be explained by differences in luminance across the textures used, see experiments S1 and S2, respectively, in the Supplementary Materials).

Comparison across experiments

The results of Experiment 1 demonstrated that feature diagnosticity likely explained the asymmetric interference observed between the processing of scene shape and texture. The results of Experiments 2 and 3 reinforced the importance of feature diagnosticity, but also revealed the contribution of the scope of visual attention to scene feature processing. The goal of this cross-experiment analysis was to understand the relative contributions of feature diagnosticity and the scope of attention to Garner interference. To do so, we compared the speed of processing scene features in Experiments 2 and 3 against the speed of processing features of the original scenes in Experiment 1, to gauge how manipulating the scope of visual attention affected overall processing speed (i.e., comparing RTs for stimuli across experiments, separately for each attended feature, and collapsed across block type). To preview, we show that asymmetric interference between the processing of scene shape and scene texture is governed primarily
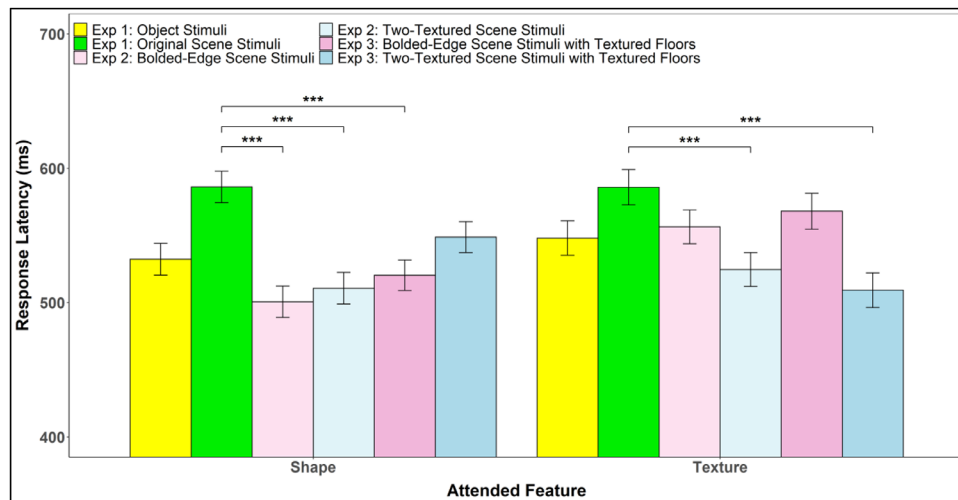
Figure 5. Results of the analysis comparing response latencies for the various stimuli across experiments. Planned comparisons were conducted between the original scene stimuli used in Experiment 1 with each of the scene stimuli used in Experiments 2 and 3, and between the object stimuli used in Experiment 1 with the scene stimuli used in Experiments 2 and 3 ($\alpha = 0.05/8 = 0.00625$). Each bar represents data collapsed across baseline and filtering blocks. ***$p < .001$; ms = milliseconds.

via feature diagnosticity, but it can nevertheless be influenced by the scope of visual attention.

If a global-processing strategy is being employed on manufactured scenes, we expect asymmetric interference because of feature diagnosticity, with the more diagnostic feature (i.e., scene shape) interfering with the less diagnostic feature (i.e., scene texture). We would also expect feature diagnosticity to largely dictate the speed of feature processing under a global processing strategy. Because rapid global scene perception prioritizes the processing of more diagnostic features over lesser ones, if a global attentional manipulation in Experiments 2 or 3 made a scene feature more salient (compared to Experiment 1), we would expect to see larger decreases in response latency for more diagnostic features (i.e., scene shape) than less diagnostic ones (i.e., scene texture).

In contrast, if a local-processing strategy is employed, we expect independence between the processing of scene features. Furthermore, we expect that the saliency of the local feature attended to would dictate processing speed, irrespective of the diagnosticity of that feature. Given that increased attention has been shown to speed up visual processing of localized attended features (Tünnermann, Petersen, & Scharlau, 2015) and the participants' task in all three experiments was to attend to the back wall of the scene when classifying scene features, if a scene manipulation promoted a more local processing strategy of shape or texture (i.e., two-textured scenes in Experiment 2), we expect response latencies for the locally attended scene feature to decrease (compared with the response latencies in Experiment 1).

Results and discussion

The main effects of stimulus ($F(5,189.30) = 9.49$, $p < 0.001$), attended feature ($F(1,591.67) = 11.73$, $p < 0.001$), and their interaction ($F(5,591.57) = 12.77$, $p < 0.001$) were all significant.

In Experiment 2, adding bolded edges to the original scenes was meant to promote a global processing strategy, and enhance Garner interference. However, because the interference observed was similar to the original scenes used in Experiment 1, it was difficult to ascertain whether expanding attention globally contributed to this interference effect, above and beyond the contribution of feature diagnosticity. When comparing response latencies across experiments, we found that bolded-edge scenes had significantly faster response latencies compared with the original scenes used in Experiment 1 when attending to shape (original scenes: $M = 586.07$ ms, $SEM = 11.71$ ms; bolded-edge scenes: $M = 500.56$ ms, $SEM = 11.71$ ms; $t(341) = 5.16$, $p < 0.001$), but not texture (original scenes: $M = 585.81$ ms, $SEM = 13.19$ ms; bolded-edge scenes: $M = 556.35$ ms, $SEM = 12.53$ ms; $t(341) = 1.62$, $p = 0.11$) (see Figure 5). The reduction of response latency for the more diagnostic feature (scene shape) but not the less diagnostic one (scene texture) suggests that adding salient contours in bolded-edge scenes did in fact spread attention globally and impacted feature processing.

It should be noted that because participants always focus on the center wall, and the saliency of the contours surrounding the center wall are enhanced, a similar reduction in response latency for scene shape (but not scene texture) would also be expected if using

a local processing strategy. However, because a local processing strategy would also likely eliminate Garner interference (Cant et al., 2015; results with two-textured scenes in Experiment 2), and we instead observed interference between shape and texture, the use of a global processing strategy in bolded-edge scenes appears to be the more consistent conclusion.

Two-textured scenes in Experiment 2 had significantly faster response latencies when attending to both shape (original scenes: $M = 586.07$ ms, $SEM = 11.71$ ms; two-textured scenes: $M = 510.64$ ms, $SEM = 11.71$ ms; $t(340) = 4.55$, $p < 0.001$) and texture (original scenes: $M = 585.81$ ms, $SEM = 13.19$ ms; two-textured scenes: $M = 524.57$ ms, $SEM = 12.53$ ms; $t(340) = 3.37$, $p < 0.001$). Combined with the finding that shape and texture were processed independently in two-textured scenes, these results suggest that the use of texture dissimilarity in two-textured scenes did in fact facilitate a local-processing strategy compared with the original scenes used in Experiment 1.

The response latencies for bolded-edge scenes with textured floors in Experiment 3 were significantly faster than those for the original scenes in Experiment 1 when attending scene shape (original scenes: $M = 586.07$ ms, $SEM = 11.71$ ms; bolded-edge scenes with textured floors: $M = 520.29$ ms, $SEM = 11.32$ ms; $t(341) = 4.04$, $p < 0.001$) but not scene texture (original scenes: $M = 585.81$ ms, $SEM = 13.19$ ms; bolded-edge scenes with textured floors: $M = 568.00$ ms, $SEM = 13.32$ ms; $t(340) = 0.95$, $p = 0.34$). Combined with the finding of asymmetric Garner interference between scene shape and texture, these results suggest that, in addition to the contribution of feature diagnosticity, a global-processing strategy was indeed being implemented when processing bolded-edge scenes with textured floors.

In Experiment 3, adding a textured floor to two-textured scenes created a scenario where global (i.e., texture similarity) and local (i.e., texture dissimilarity) processing strategies potentially competed against each other. The fact that we observed asymmetric Garner interference for two-textured scenes with textured floors (compared with independence for two-textured scenes without textured floors in Experiment 2) suggests that these scenes were processed in a more global manner. However, it is unclear whether this effect is due to a complete shift from local to global processing, or whether the existing local processing strategy was simply attenuated by the spread of attention globally away from the back wall.

The results support the second scenario. When attending to shape, response latencies increased compared with the two-textured scenes in Experiment 2 and were now no longer significantly different from the original scenes in Experiment 1 (original scenes: $M = 586.07$ ms, $SEM = 11.71$ ms; two-textured scenes with textured floors: $M = 548.72$ ms, $SEM = 11.51$ ms;

$t(341) = 2.27$, $p = 0.02$; note, this is not significantly different based on the Bonferroni correction value of $\alpha = 0.00625$). In contrast, response latencies when attending texture remained significantly faster compared with the original scenes from Experiment 1 (original scenes: $M = 585.81$ ms, $SEM = 13.19$ ms; two-textured scenes with textured floors: $M = 509.22$ ms, $SEM = 12.96$ ms; $t(340) = 4.14$, $p < 0.001$).

In summary, the effect of adding textured floors to two-textured scenes had a more pronounced impact on shape, compared with texture classifications. Spreading attention globally during shape classifications does not impact accuracy, since global and local scene shape match. However, during texture classifications, attention still has to have a degree of focus on the texture of the central wall to maintain accuracy, since classifications based on the global texture of the lateral walls and floor would be incorrect (this likely occurred, given the high accuracy in this condition; see Supplemental Materials and Figures). Thus, it appears that the local-processing strategy based on texture dissimilarity was attenuated by the global spread of attention from the addition of a textured floor, which ultimately led to the reemergence of Garner interference that was absent in the two-textured scenes used in Experiment 2.

Finally, we also made all of the same pairwise comparisons between the object stimuli used in Experiment 1 and the scene stimuli used in Experiments 2 and 3, but all comparisons were non-significant (all $t(\sim341) < 2.13$, $p > 0.03$, Bonferroni correction value of $\alpha = .00625$).

When considering all of these results, an important point to make is that we do not simply see asymmetric interference when the diagnostic feature is facilitated, because shape processing was facilitated for two-textured scenes in Experiment 2 yet we found independence between the processing of scene shape and texture. Thus a stimulus manipulation that focuses attention locally can remove the interfering influence of the diagnostic feature. Furthermore, a stimulus manipulation that expanded the scope of attention globally for both the diagnostic and less diagnostic feature resulted in the reemergence of asymmetric Garner interference. Taken together, although feature diagnosticity appears to dominate scene processing and leads to asymmetric interference, the results of this across experiment response latency analysis demonstrate that it can be influenced by the scope of visual attention.

General discussion

Using Garner's speeded-classification task (Garner, 1974) along with traditional null-hypothesis significance

testing and Bayesian analyses across 3 experiments (plus two additional control experiments; see Supplementary Materials), the potential interactivity between scene shape and texture processing was examined, motivated by fMRI findings that a region of scene-selective visual cortex (i.e., PPA) is sensitive to processing both features (Lowe et al., 2016). In Experiment 1, asymmetric interference between scene shape and scene texture processing was observed, and this effect was governed by feature diagnosticity. Specifically, the more diagnostic feature for the perception of a given scene category (in this case, shape for indoor rooms; see Oliva & Torralba, 2001) interfered with the processing of the less diagnostic scene feature (i.e., scene texture), but not vice versa. Importantly, the lack of interference of scene texture on scene shape processing is not likely explained by confounds within our experimental design, because independence was observed between the processing of object shape and texture, replicating a well-established finding in the object perception literature (Cant et al., 2008; Cant et al., 2015). This behavioral finding of asymmetric interference (which persists through Experiments 2 and 3, S1, and S2) reliably supports fMRI literature indicating that feature diagnosticity significantly contributes to scene representation in PPA (Lowe et al., 2016) and demonstrates that scene shape and texture are integral feature dimensions. Given that shape and texture are equally diagnostic in natural scenes (Lowe et al., 2016), future studies should investigate whether symmetric or reciprocal interference between shape and texture would be observed when using natural scene categories. Importantly, the results of these experiments demonstrate that the processing of scene shape and scene texture are mediated by at least partially shared cognitive mechanisms.

After establishing asymmetric interference between the processing of scene shape and texture, Experiments 2 and 3 (as well Experiments S1 and S2) revealed that, in addition to feature diagnosticity, the scope of visual attention also influences Garner interference. Specifically, scene stimulus manipulations that expanded the scope of attention globally (i.e., adding bolded edges or textured floors) led to asymmetric interference, whereas manipulations that contracted the scope of attention locally (i.e., adding two textures) eliminated such interference. These findings directly mirror results from Cant et al. (2015), which used object ensembles (a type of visual stimulus functionally and anatomically linked with scene perception; Cant & Xu, 2012; Cant & Xu, 2015; Cant & Xu, 2017) and demonstrated that expanding the scope of visual attention led to Garner interference between shape and texture processing, whereas contracting attention eliminated it. Because we consistently observe asymmetric interference between the processing of scene shape and texture, it seems that feature diagnosticity is a strong and reliable contributor to scene representation. However, the fact that this

interference can be eliminated and reinstated using local and global attentional manipulations, respectively, speaks to the additional influence of the scope of visual attention. Interestingly, adding textured floors to the scenes in Experiment 3 revealed that global- and local-processing strategies can compete with each other, thereby influencing the processing and interference observed between scene features. Further research is necessary to disambiguate the effects of competing global and local processing strategies on scene feature processing. It is also important to note that the results of Experiments 2 and 3 completely replicate after correcting for differences in luminance across textures in Supplementary Experiments S1 and S2, respectively. This demonstrates that the presence or absence of asymmetric interference cannot be explained simply by appealing to low-level visual processing (see Supplementary Materials and Figures), and the reliability of these results reveal new insights into the cognitive mechanisms underlying scene representation. We contend that these results are not attributable to differences in cognitive strategy across the baseline and filtering trials (i.e., trading off speed for accuracy and vice versa) as performance in all Experiments was uniformly near ceiling. Thus differences in response latency across baseline and filtering blocks where interference is observed cannot be accounted for by differences in accuracy across those blocks. Furthermore, we do not believe our asymmetric interference results are attributable to the type of scene-processing task we used. An argument could be made that we see no interference in the scene shape condition because participants are fixating on the central wall and this strategy engages more localized processing, whereas we see interference in the scene texture condition because this condition is better suited to engaging global scene-processing mechanisms. This is not likely given that the task requirements for both attending shape and texture required the participant to fixate the central wall. If participants were explicitly sampling information from the lateral walls when making their texture discriminations, then we would expect to see lower accuracy in the two-textured scenes compared with the bolded-edge scenes, which was never the case (i.e., the main effect of stimulus was non-significant in the accuracy analysis in all experiments). Thus it is more likely that participants were complying with task requirements in all experiments, and interference in scene processing is better explained by the more diagnostic feature (i.e., shape) interfering with the less diagnostic feature (i.e., texture), but not vice versa.

The validity in interpreting asymmetric Garner interference as integral processing is dependent on an assumption that response latencies at baseline between scene shape and texture are not significantly different. If one scene feature (e.g., shape) is processed

faster than the other (e.g., texture) at baseline, then the interference observed in texture processing (but not shape processing) could be partially explained by the serial processing of these features (Algom & Fitousi, 2016; Gandolfo & Downing, 2020; Garner, 1976; Johnstone & Downing, 2017; Schweinberger et al., 1999). In a serial-processing hypothesis, the faster the first feature is perceptually represented, the more potential it has to interfere with the processing of the second feature, thereby increasing Garner interference. Thus, a larger difference between baseline shape and texture processing would translate to a stronger Garner interference effect. To investigate this possibility, a Pearson correlation coefficient was calculated across participants between baseline differences in response latency (i.e., texture baseline – shape baseline) and differences in Garner interference response latency (i.e., filtering – baseline), separately for shape and texture (see Supplementary Materials, and Supplementary Figure S4). Across all three experiments (and both control experiments), we find no significant positive correlations for the attended scene feature that demonstrated Garner interference and thus conclude that the asymmetric interference observed was likely due to integral processing and is not fully explained by serial processing differences between the scene features at baseline.

Beyond furthering our understanding of the cognitive mechanisms mediating global and local processing in scene perception, our results also relate to their underlying neural representations. Specifically, expanding or contracting attention to create or eliminate Garner interference is directly supported by fMRI studies investigating the functional properties of PPA and LOC. For example, studies suggest the role of the LOC and PPA extends beyond merely object and scene processing, respectively, as these regions are also involved in representing local and global visual aspects of the environment (which of course could be mechanisms with which to derive object and scene representations from). There have been several examples where the PPA is involved in object processing, provided the objects contribute to the perception of the global spatial layout of a scene. For instance, it was found that the PPA not only responds strongly to scenes but also to non-scene landmark objects like buildings (Bastin, Vidal, Bouvier, Perrone-Bertolotti, Bénis, Kahane, David, Lachaux, Epstein, 2013; Cate, Goodale, & Köhler, 2011). Furthermore, objects are represented in brain regions based on their real-world size, with smaller objects showing increased activation in the occipital temporal sulcus and LOC, and larger objects showing increased activation in the PPA (Konkle & Oliva, 2012). In addition, simple rectangles of identical size that were perceived as close and far away within a scene increased activation in the LOC and the PPA, respectively (Cate et al., 2011). Conversely, LOC plays a

role in scene processing whereby its greater sensitivity to scene-specific local object details within scenes aids in scene categorization (MacEvoy & Epstein, 2011; Walther et al., 2009), and errors in scene categorization are correlated with LOC activity when the object content between scene stimuli are more similar (Park et al., 2011). Combined with these fMRI findings, our behavioral findings reinforce the idea that the scope of attention plays a key role in determining how the brain processes visual features.

Our scene perception results relate particularly well to ensemble perception. Ensemble perception refers to the visual system's ability to extract global summary statistical information (e.g., average orientation of a group of gabors: Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; average facial expression of a crowd of faces: (Haberman & Whitney, 2007; Roberts, Cant, & Nestor, 2019; Sama, Nestor, & Cant, 2019) from large groups of objects, at the expense of insensitivity to local features of individual objects (for reviews, see Alvarez, 2011; Whitney & Yamanashi Leib, 2017). A region in parahippocampal cortex along the collateral sulcus and overlapping the PPA has been shown to be sensitive to processing both scenes and object ensembles (Cant & Xu, 2012; Cant & Xu, 2015; Cant & Xu, 2017; Cant & Xu, 2020). This common neuroanatomical architecture is supported by similar findings regarding the functional information processing of visual features in ensemble and scene perception. Namely, the processing of shape and texture in both stimulus domains is mediated by shared cognitive mechanisms (ensembles: Cant et al., 2015; scenes: results of the present study), which are co-localized within anterior-medial ventral visual cortex (i.e., PPA; ensembles: Cant & Xu, 2017; scenes: Lowe et al., 2016; Lowe et al., 2017). Moreover, Brady, Shafer-Skelton and Alvarez (2017) have provided compelling evidence for a correlation between scene, texture, and ensemble processing and have proposed that scene recognition ability can be explained via the processing of global ensemble textures (i.e., spatial patterns of orientation). Thus the shared neural substrates for scene and ensemble processing are likely explained by a reliance on similar underlying computational processes (i.e., the extraction of statistical features from repeating and redundant visual information). Given the wide range of global visual features processed by PPA, future studies could use Garner's task (1974) to investigate the global (and local) processing of objects and object ensembles within scenes, to explore the pattern of interference and independence observed across multiple combinations of different stimulus types and visual features.

In summary, across three experiments (and two control experiments) using Garner's speeded-classification task, we demonstrate asymmetric interference between the processing of scene shape and scene texture. This asymmetry can be explained by

the more diagnostic feature (scene shape) interfering with the processing of the less diagnostic feature (scene texture), but not vice versa. Moreover, we demonstrated that stimulus manipulations that promote more global-based processing can lead to interference between shape and texture, whereas manipulations that promote more local-based processing can lead to the elimination of interference (i.e., independence), possibly because of a greater reliance on local object-based processing mechanisms. Together, these novel behavioral results consistently and reliably demonstrate that the processing of scene shape and texture are mediated by shared cognitive resources and thus provide important constraints on neuroimaging and neurocomputational models of scene representation. Moreover, these results further our understanding of other cognitive processes that are functionally related to scene perception, such as texture and ensemble perception, and reveal the interactive nature of visual feature processing in everyday global perception.

Keywords: attention, feature diagnosticity, garner interference, global vs. local processing, lateral occipital cortex, parahippocampal place area, scene perception, shape perception, texture perception

Acknowledgments

Commercial relationships: none.
Corresponding author: Vignash Tharmaratnam.
Email: vignash.tharmaratnam@mail.utoronto.ca.
Address: 1265 Military Trail, Science Wing Room 411, Toronto, ON, Canada, M1C 1A4.

References

Algom, D., & Fitousi, D. (2016). Half a century of research on garner interference and the separability-integrality distinction. *Psychological Bulletin, 142*(12), 1352–1383, https://doi.org/10.1037/bul0000072.

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.

Atkinson, A. P., & Burt, D. M. (2005). Asymmetric interference between sex and emotion in face perception. *Perception & Psychophysics, 67*(7), 1199–1213, https://doi.org/https://doi.org/10.3758/BF03207617.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12), 1–18, https://doi.org/10.1167/9.12.1.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the parahippocampal place area. *NeuroImage, 75*, 236–245, https://doi.org/10.1016/j.neuroimage.2013.02.073.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bastin, J., Vidal, J. R., Bouvier, S., Perrone-Bertolotti, M., Benis, D., Kahane, P., . . . Epstein, R. A. (2013). Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *Journal of Neuroscience, 33*(24), 10123–10131, https://doi.org/10.1523/JNEUROSCI.4646-12.2013.

Berman, D., Golomb, J. D., & Walther, D. B. (2017). Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *PLoS ONE, 12*(12), 1–16, https://doi.org/10.1371/journal.pone.0189828.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177, https://doi.org/10.1016/0010-0285(82)90007-X.

Biel, A. L., & Friedrich, E. V. (2018). Why you should report bayes factors in your transcranial brain stimulation studies. *Frontiers in Psychology, 9*, 1125.

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences, 114*(18), 4793–4798, https://doi.org/10.1073/pnas.1618228114.

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 and Performance, 43*(6), 1160–1176, https://doi.org/10.1037/xhp0000399.

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex, 17*(3), 713–731, https://doi.org/10.1093/cercor/bhk022.

Cant, J. S., & Goodale, M. A. (2009). Asymmetric interference between the perception of shape and the perception of surface properties. *Journal of Vision, 9*(5), 13–13, https://doi.org/10.1167/9.5.13.

Cant, J. S., & Goodale, M. A. (2011). Scratching beneath the surface: New insights into the functional properties of the lateral occipital area and parahippocampal place area. *Journal of Neuroscience, 31*(22), 8248–8258, https://doi.org/10.1523/JNEUROSCI.6113-10.2011.

Cant, J. S., Large, M.-E., McCall, L., & Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception, 37*(1), 57–78, https://doi.org/10.1068/p5727.

Cant, J. S., Sun, S. Z., & Xu, Y. (2015). Distinct cognitive mechanisms involved in the processing of single objects and object ensembles. *Journal of Vision, 15*(4), 12, https://doi.org/10.1167/15.4.12.

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience, 32*(22), 7685–7700, https://doi.org/10.1523/JNEUROSCI.3325-11.2012.

Cant, J. S., & Xu, Y. (2015). The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. *Cerebral Cortex, 25*(November), 4226–4239, https://doi.org/10.1093/cercor/bhu145.

Cant, J. S., & Xu, Y. (2017). The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. *Journal of Cognitive Neuroscience, 29*(2), 398–412, https://doi.org/10.1162/jocn_a_01050.

Cant, J. S., & Xu, Y. (2020). One bad apple spoils the whole bushel: The neural basis of outlier processing. *NeuroImage, 211*(October 2019), 116629, https://doi.org/10.1016/j.neuroimage.2020.116629.

Cate, A. D., Goodale, M. A., & Köhler, S. (2011). The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Research, 1388*, 109–122, https://doi.org/10.1016/j.brainres.2011.02.022.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*(2), 149–178, https://doi.org/10.1016/0010-0277(86)90041-7.

Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex: Contour junctions underlie neural representations of scenes. *NeuroImage, 135*, 32–44, https://doi.org/10.1016/j.neuroimage.2016.04.021.

Dick, M., & Hochstein, S. (1988). Interactions in the discrimination and absolute judgement of orientation and length. *Perception, 17*(2), 177–189, https://doi.org/10.1068/p170177.

Dykes, J. R., & Cooper, R. G. (1978). An investigation of the perceptual basis of redundancy gain and orthogonal interference for integral dimensions. *Perception & Psychophysics, 23*(1), 36–42, https://doi.org/10.3758/BF03214292.

Epstein, R., & Baker, C. I. (2019). Scene perception in the human brain. *Annual Review of Vision Science, 5*, 373–397, https://doi.org/10.1146/annurev-vision-091718-014809.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37*(5), 865–876, https://doi.org/10.1016/S0896-6273(03)00117-X.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598–601, https://doi.org/10.1038/33402.

Felfoldy, G. L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics, 15*(3), 453–459, https://doi.org/10.3758/BF03199285.

Gandolfo, M., & Downing, P. E. (2020). Perceiving emotion and sex from the body: evidence from the Garner task for independent processes. *Cognition and Emotion, 34*(3), 427–437, https://doi.org/10.1080/02699931.2019.1634003.

Ganel, T., & Goodale, M. A. (2003). Visual control of action but not perception requires analytical processing of object shape. *Nature, 426*(6967), 664–667, https://doi.org/10.1038/nature02156.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: L. Erlbaum Associates.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*(1), 98–123, https://doi.org/10.1016/0010-0285(76)90006-2.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10*(13), 805–808, https://doi.org/10.1016/S0960-9822(00)00563-7.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751–R753, https://doi.org/10.1016/j.cub.2007.06.039.

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cerebral Cortex, 23*(4), 947–957, https://doi.org/10.1093/cercor/bhs091.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature, 370*(6484), 57–59, https://doi.org/10.1038/370057a0.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23*(12), 1457–1481, https://doi.org/10.1068/p231457.

Humphreys, G. W., Romani, C., Olson, A., Riddoch, M. J., & Duncan, J. (1994). Non-spatial extinction following lesions of the parietal lobe in humans. *Nature, 372*(6504), 357–359, https://doi.org/10.1038/372357a0.

Jeffreys, H. (1998). *The theory of probability*. Oxford: OUP.

Johnstone, L. T., & Downing, P. E. (2017). Dissecting the visual perception of body shape with the Garner selective attention paradigm. *Visual Cognition, 25*(4–6), 507–523, https://doi.org/10.1080/13506285.2017.1334733.

Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology, 26*(8), 1104–1109, https://doi.org/10.1016/j.cub.2016.02.066.

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage, 132*(3), 417–424, https://doi.org/10.1016/j.neuroimage.2016.02.062.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74*(6), 1114–1124, https://doi.org/10.1016/j.neuron.2012.04.036.

Kornblith, S., Cheng, X., Ohayon, S., & Tsao, D. Y. (2013). A network for scene processing in the macaque temporal lobe. *Neuron, 79*(4), 766–781, https://doi.org/10.1016/j.neuron.2013.06.015.

Lowe, M. X., Gallivan, J. P., Ferber, S., & Cant, J. S. (2016). Feature diagnosticity and task context shape activity in human scene-selective cortex. *NeuroImage, 125*, 681–692, https://doi.org/10.1016/j.neuroimage.2015.10.089.

Lowe, M. X., Rajsic, J., Gallivan, J. P., Ferber, S., & Cant, J. S. (2017). Neural representation of geometry and surface properties in object and scene perception. *NeuroImage, 157*, 586–597, https://doi.org/10.1016/j.neuroimage.2017.06.043.

MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience, 14*(10), 1323–1329, https://doi.org/10.1038/nn.2903.

Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*(11), 843–856, https://doi.org/10.1016/j.tics.2016.09.003.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., . . . Davidson, D. L. W. (1991). Perception and action in "visual form agnosia." *Brain, 114*(1), 405–428, https://doi.org/10.1093/brain/114.1.405.

Møller, P., & Hurlbert, A. C. (1996). Psychophysical evidence for fast region-based segmentation processes in motion and color. *Proceedings of the National Academy of Sciences of the United States of America, 93*(14), 7421–7426, https://doi.org/10.1073/pnas.93.14.7421.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*(1), 72–107, https://doi.org/10.1006/cogp.1997.0667.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210, https://doi.org/10.1006/cogp.1999.0728.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience, 31*(4), 1333–1340, https://doi.org/10.1523/JNEUROSCI.3885-10.2011.

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.

Parkhurst, D., Law, K., & Niebur, E. (2002). *Modeling the role of salience in the allocation of overt visual attention. 42*, 107–123.

Quené, H., & Van Den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*(1–2), 103–121, https://doi.org/10.1016/j.specom.2004.02.004.

Reed, C. L., Bukach, C. M., Garber, M., & McIntosh, D. N. (2018). It's not all about the face: Variability reveals asymmetric obligatory processing of faces and bodies in whole-body contexts. *Perception, 47*(6), 626–646, https://doi.org/10.1177/0301006618771270.

Roberts, T., Cant, J. S., & Nestor, A. (2019). Elucidating the neural representation and the processing dynamics of face ensembles. *The Journal of Neuroscience, 39*(39), 7737–7747, https://doi.org/10.1523/JNEUROSCI.0471-19.2019.

Robin, J., Lowe, M. X., Pishdadian, S., Rivest, J., Cant, J. S., & Moscovitch, M. (2017). Selective scene perception deficits in a case of topographical disorientation. *Cortex, 92*, 70–80, https://doi.org/10.1016/j.cortex.2017.03.014.

Rosenholtz, R. (2011). What your visual system sees where you are not looking. *Human Vision and Electronic Imaging XVI, 7865*, 786510, https://doi.org/10.1117/12.876659.

Sama, M. A., Nestor, A., & Cant, J. S. (2019). Independence of viewpoint and identity in face ensemble processing. *Journal of Vision, 19*(5), 1–17, https://doi.org/10.1167/19.5.2.

Schmid, A. M., & Victor, J. D. (2014). Possible functions of contextual modulations and receptive field nonlinearities: Pop-out and texture segmentation. *Vision Research, 104*, 57–67, https://doi.org/10.1016/j.visres.2014.07.002.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception and Psychophysics, 61*(6), 1102–1115, https://doi.org/10.3758/BF03207617.

Schweinberger, S. R., Soukup, G. R., & Konstanz, U. (1998). *Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. Journal of Experimental Psychology: Human Perception and Performance 24*(6), 1748–1765.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-sependent scene recognition. *Psychological Science, 5*(4), 195–201.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26*(8), 1027–1038, https://doi.org/10.1068/p261027.

Shechter, S., & Hochstein, S. (1992). Asymmetric interactions in the processing of the visual dimensions of position, width, and contrast of bar stimuli. *Perception, 21*(3), 297–312, https://doi.org/10.1068/p210297.

Silson, E. H., Steel, A. D., & Baker, C. I. (2016). Scene-selectivity and retinotopy in medial parietal cortex. *Frontiers in Human Neuroscience, 10*(August), 17, https://doi.org/10.3389/fnhum.2016.00412.

Steeves, J. K. E., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patients with visual form agnosia. *Journal of Cognitive Neuroscience, 16*(6), 955–965, https://doi.org/10.1162/0898929041502715.

Tünnermann, J., Petersen, A., & Scharlau, I. (2015). Does attention speed up processing? Decreases and increases of processing rates in visual prior entry. *Journal of Vision, 15*(3), 1–27, https://doi.org/10.1167/15.3.1.

Vailaya, A., Jain, A., & Zhang, H. J. (1998). On image classification: city images vs. Landscapes. *Pattern Recognition, 31*(12), 1921–1935, https://doi.org/10.1016/S0031-3203(98)00079-X.

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: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*(6), 1172–1217, https://doi.org/10.1037/a0029333.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience, 29*(34), 10573–10581, https://doi.org/10.1523/JNEUROSCI.0559-09.2009.

Walther, D. B., & Shen, D. (2014). Nonaccidental Properties Underlie Human Categorization of Complex Natural Scenes. *Psychological Science, 25*(4), 851–860, https://doi.org/10.1177/0956797613512662.

Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience, 14*(10), 1243–1244, https://doi.org/10.1038/nn.2910.

Whitney, D., & Yamanashi Leib, A. (2017). Ensemble Perception. *Annual Review of Psychology, 12*(16), annurev-psych-010416-044232, https://doi.org/10.1146/annurev-psych-010416-044232.

Wilder, J., Dickinson, S., Jepson, A., & Walther, D. B. (2018). Spatial relationships between contours impact rapid scene classification. *Journal of Vision, 18*(8), 1–15, https://doi.org/10.1167/18.8.1.