Annual Review of Vision Science

Rethinking Space: A Review of Perception, Attention, and Memory in Scene Processing

Monica S. Castelhano and Karolina Krzyś

Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada; email: monica.castelhano@queensu.ca

Annu. Rev. Vis. Sci. 2020. 6:563–86

First published as a Review in Advance on June 3, 2020

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-121219-081745

Keywords

scene perception, scene gist, spatial processing, visual search, attention, memory

Abstract

Scene processing is fundamentally influenced and constrained by spatial layout and spatial associations with objects. However, semantic information has played a vital role in propelling our understanding of real-world scene perception forward. In this article, we review recent advances in assessing how spatial layout and spatial relations influence scene processing. We examine the organization of the larger environment and how we take full advantage of spatial configurations independently of semantic information. We demonstrate that a clear differentiation of spatial from semantic information is necessary to advance research in the field of scene processing.

1. INTRODUCTION

Complex real-world scenes are most often described along spatial and semantic dimensions, which allow for efficient processing and interactions with the environment. Scenes are typically defined as a view of the natural world, which is made up of space-defining surfaces and smaller objects (Castelhano & Henderson 2008, Hollingworth & Henderson 1999, Oliva 2005, Williams & Castelhano 2019). In this review, we present recent advances in assessing how spatial layout and spatial relations influence scene processing. By considering the spatial knowledge of real-world scenes as information that can be exploited independently of the scene's semantic information, we can develop new approaches for examining real-world scene perception.

For the past 60 years, semantic information has played a vital role in propelling our understanding of real-world scene perception forward. Beginning with Palmer's (1975) demonstration of the effect of context on object recognition and Biederman et al.'s (1982) examination of semantic violations, the semantic link between individual objects and the larger scene context has been well established (Biederman et al. 1982, Castelhano & Heaven 2010, Castelhano & Pereira 2018, Friedman 1979, Greene 2013, Loftus & Mackworth 1978, Mackworth & Morandi 1967, Palmer 1975, Wu et al. 2014). For instance, semantic context has an effect on how objects are initially processed (Bar 2004, Boyce et al. 1989, Davenport & Potter 2004, De Graef et al. 1992, Martinez-Conde et al. 2006, Stein & Peelen 2017), how objects are remembered (Friedman 1979, Gronau & Shachar 2015, Joubert et al. 2007, Pezdek et al. 1989), and how attention is allocated to objects within the scene (Castelhano & Heaven 2010, Foulsham & Underwood 2007, Wu et al. 2014). Based on these early studies, objects were long the focus of research, as they were assumed to be the basic unit of understanding within an environment. This notion led to studying objects on blank backgrounds and randomly arranged. However, just as a word takes on different connotations in different sentences, objects can be thought of differently in different scenes. Objects can interact with the structure of the scene and produce representations that are either expected or unexpected. Given that, it is important to evaluate the influence of the scene structure on understanding.

In contrast to the results of earlier studies, the role of spatial layout in scene understanding has been strongly supported by cognitive neuroscience research into the scene processing network (Epstein & Kanwisher 1998, Epstein et al. 1999, Ferrara & Park 2016, Kravitz et al. 2011). One of the main differences between scene images and other types of stimuli, such as objects and faces, is that scene images depict an encompassing space, rather than an entity. This conceptualization of scenes was supported by early studies of the parahippocampal cortex, also referred to as the parahippocampal place area (PPA), which showed a strong response to images that conveyed a spatial layout, even when devoid of discrete objects (e.g., an empty room consisting of floor, four walls, and a ceiling) (Epstein & Kanwisher 1998, Epstein et al. 1999). Later studies have found that changes to the spatial extent of the space modified responses (Henderson et al. 2008, 2011; Park et al. 2014). In addition, other studies have found a similar pattern of effects in response to the space of a scene in the retrosplenial cortex (RSC) (Wolbers et al. 2011) and occipital place area (OPA) (Kamps et al. 2016), which, together with the PPA, are thought to make up the scene processing network (Epstein & Baker 2019). Thus, spatial information has been found to have strong influence on scene processing across several tasks in both behavioral and neuroscience studies.

We argue that new insights about scene processing will need to overcome some basic assumptions about how scene semantics and space are processed, as well as how they relate. For example, a prevailing assumption in the literature is that the spatial information is nested within the semantic context information. This assumption becomes apparent when comparing the types of

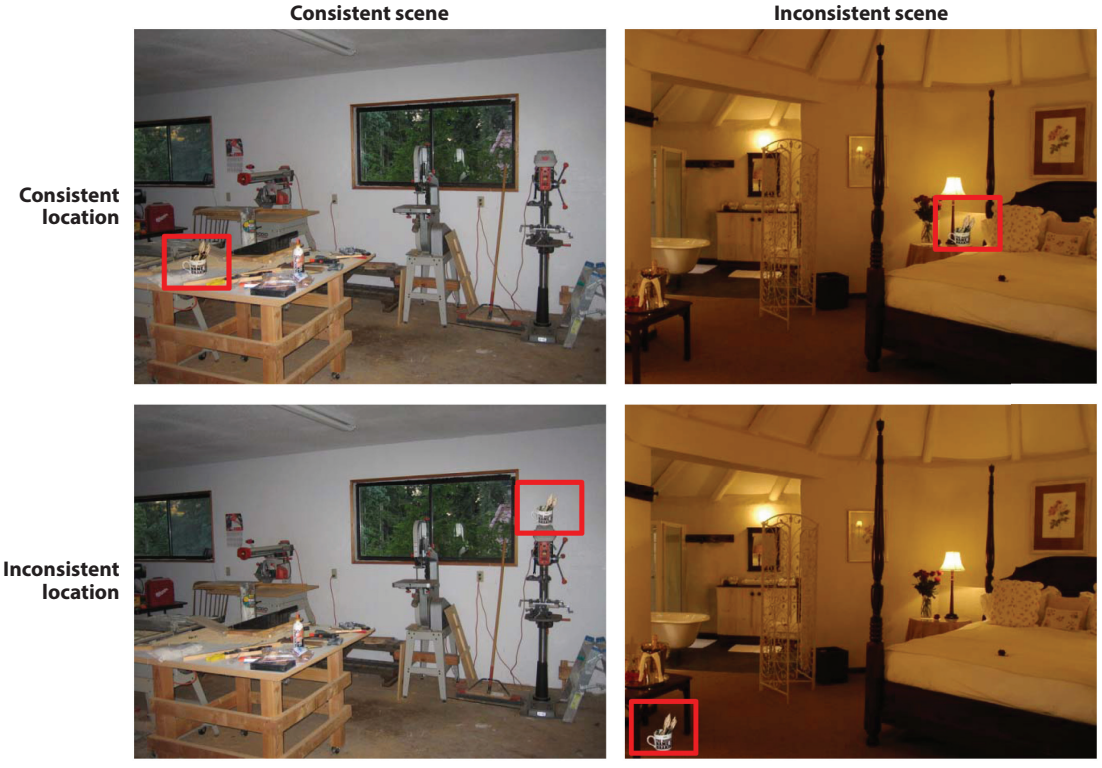

Figure 1

Example stimuli for a visual search task that manipulated the semantic and spatial relationship between the target object and the scene context. In this example, the target is a mug of paint brushes (*red box*) typically found in the workshop scenes. Figure reproduced with permission from Castelhano & Heaven (2011).

manipulations used to examine semantic versus spatial violations in scenes. When examining semantic violations, the target object is typically placed in a semantically inconsistent context (e.g., a toaster in the bathroom) (De Graef et al. 1990, Henderson et al. 1999, Võ & Henderson 2011). Conversely, when examining spatial violations, the target object is always semantically congruent with the larger scene context but placed in an unusual location (e.g., a toaster on the kitchen floor) (Eckstein et al. 2006, Hillstrom et al. 2017, Malcolm & Henderson 2010, Neider & Zelinsky 2006, Pereira & Castelhano 2014). Thus, when comparing semantic violations to spatial ones, past studies confounded the two sources due to the nested assumption, and their relationship was not made clear.

To examine the nested assumption directly, Castelhano & Heaven (2011) manipulated both the semantic association and spatial positioning of the target object orthogonally (see **Figure 1**). They found that learned spatial associations improved search performance even when there was no semantic link between the target and the scene. These findings suggest that a scene's spatial layout and its association with various objects play an important role in attentional guidance independently of whether the target object is semantically associated with that scene. These results point to an independent role of spatial information in scene processing and suggest that a different theoretical perspective could lead to new insights.

Although research on semantic relationships has greatly advanced our understanding of scene processing, in this review, we posit that scene processing is fundamentally driven by spatial

information. Why put scene space first? As put so succinctly by Epstein & Baker (2019, p. 383), "Scenes are—by definition—spaces." The spatial extent, layout, and arrangement of various subcomponents play an essential role in our understanding of, processing of, and interactions with our environment. In this review, we examine the larger organization of the environment and how we take full advantage of this organization independently of semantic information. To substantiate this point, we review behavioral and eye movement findings demonstrating a link between space and category, as well as the influence of this link on attention, memory, and navigation, both through associations with objects and across depth. We organize the analysis of spatial influences (contribution) into five processing types: early scene perception (Section 2), spatial constraints of scene–object associations (Section 3), attention and spatial processing across depth (Section 4), scene perception and navigation in space (Section 5), and scene representations and spatial information in memory (Section 6).

We review these five processes related to spatial information and demonstrate that the space of the environment plays a crucial role in understanding scene processing, as has been shown in both older and more recent research. The clearest indicator that space plays a critical role in scene perception is in how scene space and layout is linked to initial scene perception, which we turn to in the following section.

2. EARLY SCENE PERCEPTION

Our first hint that space plays a central role in scene perception lies in the way in which scene space and layout are linked to semantic category (Baldassano et al. 2013, Greene & Oliva 2010, Murray et al. 2002, Oliva & Torralba 2001). Early on, Gibson (1979, p. 34) delineated the difference between detached and attached objects within a scene. He defines detached objects as being small moveable items and attached object as the layout of a surface that is continuous with the substance of another surface, such as the ground. The rationale for this distinction is that these attached objects are not perceived as entities separate from the scene, but rather are perceived as convexities within the environment. This is similar to the scene definition discussed above and was the first description of scenes as a series of convex surfaces.

Oliva & Torralba (2001) were the first to demonstrate that scene category could be determined from the shape of the space. They proposed the spatial envelope theory, which used a sparse collection of indicators to capture different spatial properties of an environment, rather than determining scene category with piecemeal identification of a few objects. These included properties focused on the extent of the space (openness and expansion), as well as on the texture and nature of components (naturalness, roughness, and ruggedness). Together, these properties were used to distinguish and group different categories of scenes without the need to identify individual objects. Importantly, this computational work demonstrates that basic-level scene categories cluster in this descriptive space along global-property dimensions describing the spatial information of scene images.

Additionally, Greene & Oliva (2009a,b, 2010) were able to show different effects of spatial information on scene understanding. First, they demonstrated that the nature of the space depicted in an image precedes the extraction of the scene category. For instance, observers may perceive that there is a large space but not yet know if it is a field or lake (Greene & Oliva 2009a,b). Second, they demonstrated that two scene images were more likely to be confused when they had similar spatial properties than if they shared similar objects. Conversely, they also demonstrated that two images are less likely to be confused when the scenes had distinct spatial properties but shared similar objects (Greene & Oliva 2009b, 2010). Thus, it would appear that the spatial information is extracted early and used to determine the scene category, rather than the other way around.

Scene categorization studies suggest that processing follows a spatial-to-semantic direction of progression; however the results of these studies stand in contrast to how the relationship is thought of in visual search studies, where spatial information is thought to be nested within semantic information, as outlined above (Castelhano & Heaven 2011; Pereira & Castelhano 2014, 2019; Williams & Castelhano 2019). A possible source of the disconnection between these two views may lie in the roles of objects and the tasks that define those roles. In visual search, the goal of the task revolves around detecting objects, while in scene understanding, the goal revolves around deriving semantic information. While the typical location or spatial associations of objects can be inferred from the sematic information (Gronau & Shachar 2015, Võ et al. 2019), we posit that spatial associations can be used independently to guide attention during visual search, a point that we turn to in the next two sections.

3. SPATIAL CONSTRAINTS OF SCENE–OBJECT ASSOCIATIONS

The relationship between an object and a particular scene context can be understood by identifying pre-existing spatial associations of that object. The effect of those associations on processing has been known for some time (Biederman et al. 1973, 1982; Mandler & Johnson 1976; Mandler & Ritchey 1977). Also referred to as positional regularities, the spatial associations between an object and a scene play a large role in predictions of where objects should be placed (Kaiser et al. 2019, Oliva & Torralba 2007, Torralba et al. 2006, Wolfe et al. 2011). These regularities were cleverly illustrated by Oliva & Torralba (2007). As can be seen in **Figure 2**, they observed that objects' spatial associations emerged as regularities in the background of these images. If objects were placed randomly within the scene, then the background would look uniformly gray. However, when the objects were aligned, then the surfaces on which they were typically located could be seen.

The importance of spatial associations becomes most evident when one considers how information is processed in a task such as visual search. The goal in visual search is to locate and

Figure 2

Images representing the average of hundreds of images of specific objects, where objects were centered and scaled and produced statistical depictions of their typical surroundings. The surfaces on which objects are typically viewed are highlighted in these images. Figure reproduced from Oliva & Torralba (2007) with permission from Elsevier.

identify a target object as quickly as possible. For instance, when objects are out of place in a scene (i.e., inconsistent spatial position), search performance is hampered (Biederman et al. 1982, Castelhano & Heaven 2011, Castelhano & Witherspoon 2016, De Graef et al. 1990, Malcolm & Henderson 2010). Violations of the spatial associations can also violate expectations of physics, i.e., when objects hang mid-air or precariously on a surface that realistically would not be supportive (Biederman et al. 1982, Hillstrom et al. 2017, Võ & Henderson 2009). Both of these types of violations lead to effects on processing during visual search. For example, Hillstrom et al. (2017) found that, when target objects' locations switched from being spatially consistent to being either improbable (mug on floor) or impossible (mug in the air), performance worsened significantly. In another study, object function was found to underlie probable spatial associations. Castelhano & Witherspoon (2016) found a strong link between the target object's function and its spatial location in the larger scene context. They found that, when the functions of novel objects were learned, participants were able to locate them much more quickly than when only the visual features of the target object were known. Together, these and other studies have shown that object properties have a direct link to spatial location, without an intermediate step of scene semantics.

One question that remains is the degree to which attentional deployment is informed by spatial associations between objects and scenes during visual search. For a specific target, it is necessary to operationalize regions of the scene that would be target relevant (where a target object is likely to appear) or target irrelevant (where a target object is unlikely to appear). Based on the output of the contextual guidance model, Torralba and colleagues (2006) found that an object's position is captured by its likely vertical position in the scene, whereas the horizontal position was found to be less informed by the larger context, as an object is equally likely found to the left or right of a spatial region. This was especially clear when considering Torralba et al.'s experiment 2, in which, for the same set of scenes, the model highlighted the mid-level regions when the target object was a mug and the upper regions when the target object was a painting (see **Figure 3**).

More recently, Castelhano and colleagues (Castelhano & Heaven 2011; Pereira & Castelhano 2014, 2019; M.S. Castelhano and E.J. Pereira, unpublished manuscript) have argued for the importance of scene surfaces in guiding attention during visual search. The surface guidance framework model posits that attention is directed to surfaces in the scene most associated with the target object. To operationalize target-relevant scene regions, scenes are divided into three horizontal surfaces: (*a*) upper (e.g., ceiling, upper walls), (*b*) middle (e.g., countertops, tabletops, desktops, stovetops), and (*c*) lower (e.g., floor, lower walls). Target-relevant regions are then identified in association with the target object: (*a*) upper (e.g., painting, hook), (*b*) middle (e.g., toaster, alarm clock), and (*c*) lower (e.g., garbage bin, shoes). Using a method of defining a target's association with a spatial region, each scene region and target object combination could be divided into target-relevant and target-irrelevant regions (see **Figure 4**).

Dividing the scene into relevant and irrelevant surfaces that vary in their vertical positioning allows for the operationalizing of target-relevant and target-irrelevant regions. This in turn allows for examination of processing differences between relevant and irrelevant scene regions based on spatial expectations of the target object. In one recent study, Pereira & Castelhano (2019) had participants search for a target object in photographic images. After the first fixation was detected, a distractor object would suddenly appear in half of the trials. Importantly, the distractors could suddenly onset in either a relevant or an irrelevant region. They found that distractors were more likely to capture attention when they appeared in a target-relevant than in a target-irrelevant region. Thus, how attention is deployed is closely tied to the scene structure, where surfaces can act as larger object-based regions across which attention is allocated (Malcolm & Shomstein 2015, Vatterott & Vecera 2015).

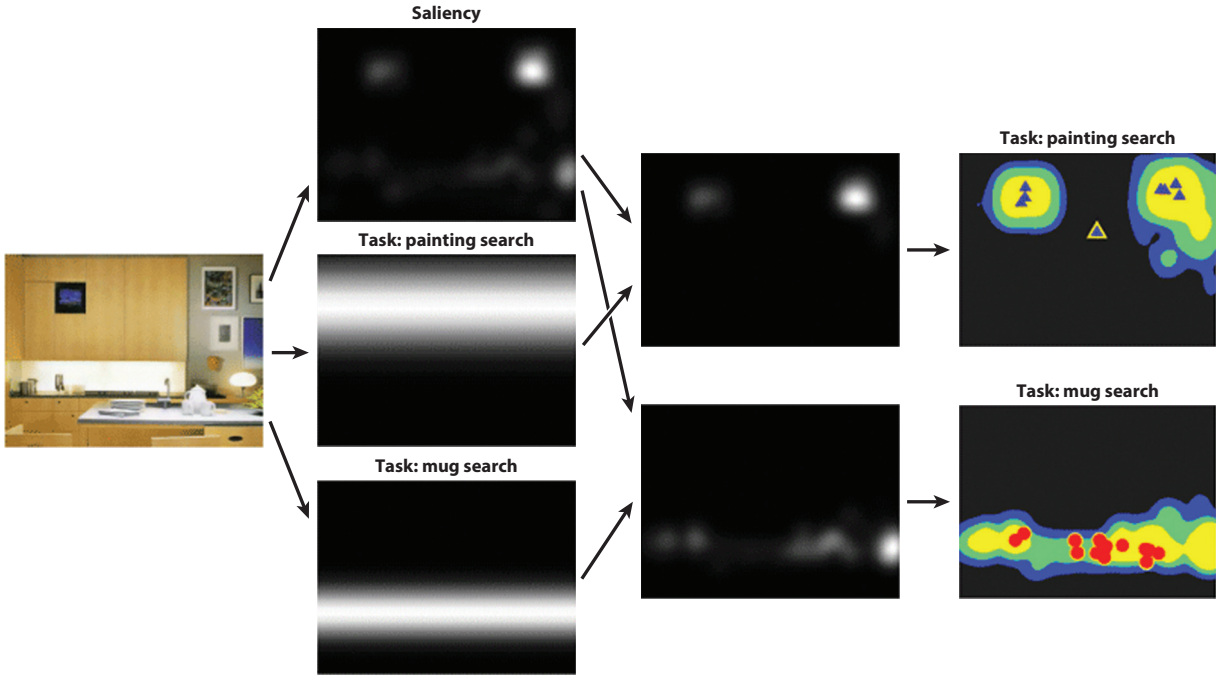

Figure 3

The contextual guidance model for two types of visual search tasks (mug versus painting). The original image properties are used to identify not only the salient regions of the image, but also the most likely locations of the target object. These are combined, and as the last panel shows, the predicted regions of fixations align quite well to the fixation patterns produced by participants searching for either a painting (*blue triangles*) or a mug (*red circles*). Figure adapted with permission from Torralba et al. (2006).

In another study, M.S. Castelhano & E.J. Pereira (unpublished manuscript) manipulated the set size in target-relevant and target-irrelevant regions (where the number of distractors was 4, 8, or 16 in each region). The logic was simply that, if scene context serves to guide search to the most relevant regions, and attentional deployment is limited to target-relevant regions, then an increase of the set size within the target-relevant region should decrease search efficiency, whereas increases of the set size in the target-irrelevant region should minimally affect search efficiency. In contrast, if attentional deployment is spread across the whole scene, then search efficiency should decrease as set size increases, regardless of where in the scene items are added. Castelhano & Pereira's results support the claim that attention and processing is limited to the target-relevant regions. Together, these studies demonstrate that the surface guidance framework allows the division of scene processing into relevant and irrelevant regions based on spatial associations of a target object within a scene context. Thus, this theoretical framework is a powerful tool for understanding complex visual information processing in real-world environments.

Central to the surface guidance framework model is that scene context knowledge is applied not as semantic informativeness or belongingness of an object in a scene, but rather as knowledge of placement within a scene. This is supported by previous studies that showed that the scene's semantic information alone did not support improvement of search performance. In one study, Castelhano & Henderson (2007) showed participants a preview image that matched the category of the search scene but that had a different layout. When compared to an identical preview of the scene, search performance was slowed. In addition, there was no benefit to performance for a

Figure 4

Scene regions being highlighted across scene and different types of targets in each region. (*a*) Examples of search scenes. (*b*) Highlighted surface regions (upper: *red*; middle: *yellow*; lower: *blue*), per the surface guidance framework. Figure reproduced with permission from Pereira & Castelhano (2019).

category-matched preview when compared to a completely different scene that did not overlap in layout, category, or content. In another study, Castelhano & Heaven (2010) examined whether a preview of the scene's semantic category name would affect search performance but found across two studies that there was no effect of knowing the scene category ahead of time. Thus, spatial associations lead to greater efficiency in search, but general scene semantics do not.

Researchers have also examined how spatial location affects object recognition (Castelhano & Heaven 2011, Kaiser et al. 2014, Katti et al. 2016, Malcolm & Henderson 2010, Munneke et al. 2013, Stein & Peelen 2017). Typically, when objects are placed in an inconsistent scene location, there is a detrimental effect on object recognition. In eye movement studies, objects are fixated for longer when placed in an inconsistent scene region, which is taken to indicate that processing is more difficult (Castelhano & Heaven 2011, Malcolm & Henderson 2010).

Relatedly, researchers have also examined how spatial associations between different objects in scenes affect attention (Draschkow & Võ 2017, Gronau & Shachar 2015, Summerfield et al. 2006, Võ et al. 2019). In a recent study, Draschkow & Võ (2017) had participants create scenes (in a virtual environment) and found that the placement of larger objects (i.e., anchor objects) preceded that of smaller objects. Later memory tests on the created scenes found that anchor objects also provided a cue for recalling information about smaller, spatially associated objects. Furthermore, cognitive neuroscience studies have also found that spatial associations benefit object processing. Researchers have examined these effects at two levels: (*a*) expectations in relation to visual space (i.e., airplanes tend to occur in the upper visual field) and (*b*) interobject relations (i.e., lamp on a desk). In one study, Kaiser & Cichy (2018) examined the effect of location within visual space (upper or lower visual field). Using a multivoxel pattern analysis, they found that, when the object's placement corresponded to their expected location, information was more accurate in

object-selective regions. In another study, Gronau & Shachar (2015) had participants memorize pairs of objects. For our purposes, they found that the spatially related object pairs (e.g., a desk lamp on a desk), regardless of whether they were functionally related, improved memory for visual detail. Although these studies did not examine objects within a scene context, the positioning of the objects reflected how they are positioned in the real world. Thus, the observed benefits are attributable to the general knowledge of the object's spatial associations. Together, these studies demonstrate that the spatial associations between an object and the environment, as well as between the object and other objects in the environment, led to a benefit to object processing.

Recently, researchers have also made connections between the action performed on an object and spatial organization (Castelhano & Witherspoon 2016, Greene et al. 2016). In one study, Castelhano & Witherspoon (2016) created a set of novel objects and had participants learn either about their features or about their function. When participants searched for those novel objects in a scene, they did better if they knew the object's putative function. Thus, knowing the function constrained where the observer assumed that the object should appear in the scene, and observers made use of that constraint even though they had never been explicitly shown the likely locations. The connections among action, scene perception, and space have recently become of interest in the literature. In the next section, we turn to how these concepts are connected to attention and memory processes.

4. ATTENTION AND SPATIAL PROCESSING ACROSS DEPTH

In recent years, interest has grown in how information processing across depth impacts attention and memory of scene representations (Bonner & Epstein 2017, 2018; Fernandes & Castelhano 2019; Josephs & Konkle 2019; Man et al. 2019). Traditionally, studies of depth perception have examined observers' estimates of distance (Cutting & Vishton 1995, Nagata 1993) and how information at different distances, from peripersonal to vista space, is processed (Costantini et al. 2011, Cutting & Vishton 1995, Previc 1998). Researchers have found that there are functionally different types of information available across different categories of depth. Peripersonal space (or personal space) is typically defined as the zone immediately surrounding the observer, generally within arm's reach and slightly beyond (Cutting & Vishton 1995) (see **Figure 5**). This space is thought to be more accurately represented, reflecting higher sensitivity to details and providing a richer source of space information from the local environment.

The immediate impact of objects and agents in this area is also thought to lead to increased sensitivity to space, distance, and visual information more generally. For instance, Costantini et al. (2011) examined whether objects would evoke action-related information depending on their apparent distance from the participant. Participants indicated whether the object presented corresponded to a subsequently presented verb. They found that, when objects were within peripersonal space, verbs associated with those objects were responded to faster than if the objects were farther away.

Recent studies on scene processing have also shown qualitatively different processing of spaces closer to the observer (Bonner & Epstein 2017, 2018; Fernandes & Castelhano 2019; Josephs & Konkle 2019; Man et al. 2019). Most recently, Fernandes & Castelhano (2019) found a foreground bias when examining rapid scene perception for images that had mismatched scene categories in the foreground and background (i.e., Chimera scenes) (see **Figure 6**). That is, foreground information (from the center of the total scene depth to the position of the observer within the scene) had a greater influence on initial scene perception than did background information.

Given the qualitative differences in processing across depth, it stands to reason that information closer in depth may have different utility than information farther away and thus may differently

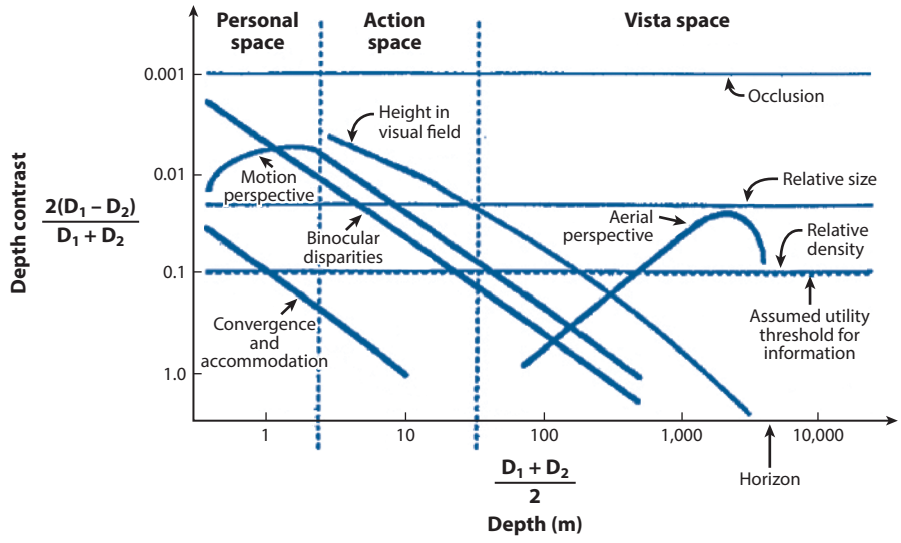

Figure 5

Different categories of space as functions of distance from the observer are linked to different perceptual sensitivities to visual information and depth cues. Figure adapted with permission Cutting & Vishton (1995).

impact eye movement guidance during visual search in a scene. Indeed, in a recent study, Man et al. (2019) found a similar foreground bias when examining visual search in these Chimera scenes. Targets were placed either closer to the observer (in the foreground of the scene) or further away (in the background of the scene) and were semantically consistent with only the region in which they were placed. Results showed that, even though targets were controlled for size, foreground targets were found more quickly and with fewer fixations than background targets. Thus, it seems reasonable to suggest that the foreground bias is driven by the enhanced processing of visual information physically closer to the observer.

The allocation of attention to a specific point in depth has been shown across numerous studies using different stimuli and tasks (Burgess et al. 2004, Costantini et al. 2011, Downing & Pinker 1985, Park & Park 2018, Previc 1998, Song et al. 2017). Using different cueing paradigms, many early studies found an increase in reaction time when invalid cues indicated a different depth than the target (Downing & Pinker 1985, de Gonzaga Gawryszewski et al. 1987). For example, Downing & Pinker (1985) found that search was slower when the target was at a different depth than the current fixation compared to when it was at the same depth. Other studies demonstrated that a unique positioning of a target in a depth plane improved search efficiency (de la Rosa et al. 2008, Finlayson & Grove 2015, Marrara & Moore 2000). Researchers also demonstrated that depth information could be used to improve search performance. In a recent study, Finlayson & Grove (2015) demonstrated that the efficiency of search is the highest for targets located in the nearest plane and that it declines as the target depth increases. Taken together, these studies established not only that attention can be allocated to a specific location along the *z* axis, but also that distractor or irrelevant information is most disruptive when present at the same or near the depth of the target.

Across numerous studies using different stimuli and tasks, researchers have also shown that the processing of information presented closer to an observer is qualitatively different. How depth affects allocation of attention has been examined more extensively within the context of driving.

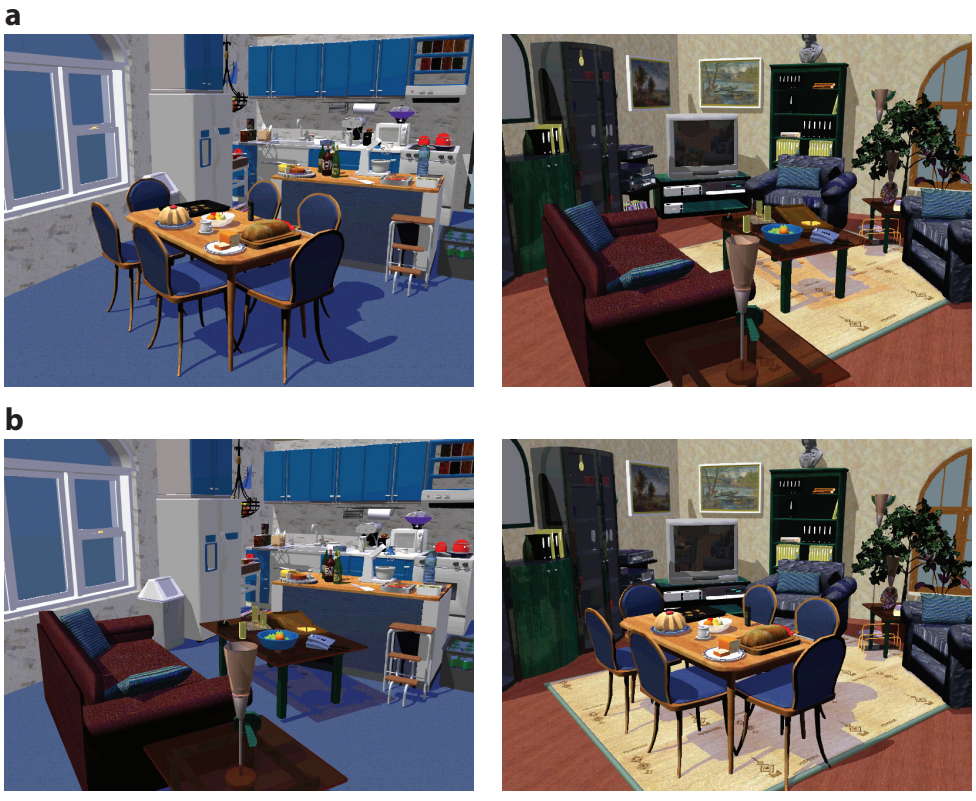

Figure 6

(*a*) Example search scenes. (*b*) Chimera scenes created by switching the foreground regions of the two normal scenes depicted in panel *a*. Figure reproduced with permission from Fernandes & Castelhano (2019).

Andersen et al. (2011) used a standard dot probe paradigm to examine attention. They asked participants to follow a lead vehicle while also monitoring and responding to light changes that were presented at different depths above the roadway. They found that reaction time to these changes depended both on the horizontal position of the light and the distance from the participant. This is consistent with other studies that have used various type of probes to examine allocation of attention while driving (Gaspar et al. 2016, Rogé et al. 2004).

There is also recent evidence in scene processing research that the area closer to the observer and within which actions can potentially occur is processed qualitatively differently than other areas. Josephs & Konkle (2019) examined the spaces that encompass workspaces within reachable distances (**Figure 7** shows the stimuli of objects, reachable spaces, and scenes). This can include images depicting a kitchen countertop, a desktop, or a dining table. The surfaces are sized and the objects are arranged such that the whole space is actionable or reachable. They found that these spaces are processed in distinct brain regions that differ from both individual objects and larger scene spaces. These findings reinforce the notion that scene information across depths is differently prioritized.

Taken together across various stimuli and tasks, these results suggest that there is a consistent pattern of prioritization of information closer to an observer, which Fernandes & Castelhano (2019) termed the foreground bias. It seems reasonable that the depth at which the information

Figure 7

The different types of stimuli with different depths. In the middle group, scene images are configured such that the whole space is depicted as within reach of an observer. Figure adapted with permission from Josephs & Konkle (2019).

occurs would play a role when searching in a real environment. The role of information that is present in closer spatial proximity introduces an interesting framework from which to consider the nexus of scene processing, navigation, and action.

5. SCENE PERCEPTION AND NAVIGATION IN SPACE

The significant role of spatial information is not in dispute when considering navigation. The very act of walking through a scene or wayfinding from point A to point B requires the simultaneous understanding of the immediate vista space (visible field of view) and the larger environment (which may be represented, but not currently in view) (Epstein et al. 1999, Gibson 1979, Hassabis & Maguire 2007, He et al. 2013, Maguire et al. 2006). What prompts more discussion is how this information is perceived and how it is represented in memory. We examine each in turn.

Researchers have found that the initial processing of the scene shape also informs the observer about the affordances of that space (Bonner & Epstein 2017, 2018; Gibson 1950, 1979). Gibson (1979) first proposed affordances of scene spaces, which he defined as perceptual properties that indicate potential for action. For instance, as with objects, if a surface is free of obstacles, and if the surface texture and structure support walking (a dirt path versus a river), then we instantaneously perceive that the environment affords us a navigational path. Recently, Bonner & Epstein (2017) found that activity in the OPA is linked to encoding of navigational affordances in a local environment. Interestingly, the patterns of activation allowed for training of a linear decoder to predict navigational affordances of previously unseen scenes, despite the fact that the participants' task was not related to navigation. This suggests that spatial properties and scene structure relevant to navigation are automatically encoded. The study of how spatial properties are encoded is relatively straightforward in static images, but motion is a powerful source of information about the structure of the scene.

When navigating through an environment, the representation of the scene and space must be updated as one progresses. Early on, Gibson (1979) proposed that moving through an environment involves changing vistas: With movement, a new vista opens in front as a former vista closes behind you. Inherent in this notion of evolving representations through space and time is the notion of anticipatory spatial representations (Intraub 2010, Intraub & Richardson 1989). The anticipatory spatial representations have an implied continuation of the scene that extends beyond the boundaries of the current view of the environment. Intraub & Richardson (1989) showed that participants were poor at remembering the absolute boundaries of the background of an image. When they were shown a close-up image, participants systematically added more background in their recall and recognition of the image. Intraub (2010) proposed that anticipatory spatial

representations are integral to relating individual views of the environment to a representation of the larger environment, facilitating the integration of successive views, as well as helping to draw attention to unexpected features that occur in an upcoming view. Interestingly, the sequential processing and integration of the current vista into a representation of the larger environment fits with the notion of a foreground bias discussed above, where information that is closer to the observer is subject to a processing preference or prioritization over information that is further away (Fernandes & Castelhano 2019, Man et al. 2019).

Studies examining scene representations in the brain have also examined representations centered on the anticipatory nature of the 3D structure of scenes (Epstein et al. 2007, Ferrara & Park 2016, Park et al. 2007). For instance, Park et al. (2007) had participants view a close-up image followed by a wide-angle picture. They found evidence for spatial extrapolation through a scene-selective attenuation in the PPA and RSC, but no such pattern of extrapolation in the lateral occipital complex. These findings demonstrated that scene layout representations are extrapolated beyond the perceptual input in areas previously found to correspond to scene processing.

In addition, Konkle & Olivia (2007) had participants view scene images with both extreme close ups and long shots of scene images. Afterwards, participants adjusted the size of the scene image boundaries to match the size viewed earlier. They found that the remembered scenes showed a systematic bias toward the normative size: The close-up shots were remembered as further away, and long shots were remembered as closer. Interestingly, when selecting the right position within the scene, Konkle & Oliva found that there was a tendency to zoom out first to assess the entire scene and then to zoom in to a comfortable distance (Konkle & Olivia 2007, Oliva et al. 2010). This study revealed that the scene representation is not only extrapolated, but also tied to canonical views based on distance, and there is a balance between the expected knowledge of the space and the spatial representations of the scene itself.

Relatedly, researchers have also examined how space is associated with the representations of objects and their surrounding space. Researchers have found a connection between an object and the space around it regardless of whether the object is imagined or directly perceived (Collegio et al. 2019; Mullally & Maguire 2011, 2013). In one study, Mullally & Maguire (2013) demonstrated that certain objects (referred to as space-defining objects) evoked depictions of the surrounding three-dimensional space (e.g., oak bed) when they were either viewed or imagined in isolation. This was in contrast to background items (e.g., a floor or a wall) and space-ambiguous objects (e.g., a laundry basket) that evoked no such associated spatial representation. The automatic processing of space in and around objects fits with the notion of attached objects [as described by Gibson (1979)] or larger objects that are spatially associated with smaller objects (e.g., anchor objects) (Võ et al. 2019).

Finally, in addition to how successive views are related across time, navigation studies have also examined how different levels of the environment are spatially related. Environments have been shown to exhibit a hierarchical organization, in which smaller local environments are represented as nested within a larger environment, which can be either visible or not (Hirtle & Jonides 1985, McNamara 1986, McNamara et al. 1989). Analogously, studies in scene perception have also shown that subregions of a scene image can be functionally dissociated from the larger scene representation (Brockmole et al. 2006, Brooks et al. 2010, Castelhano et al. 2019). For instance, Brooks et al. (2010) had participants search repeatedly through scene images with a game board (an array of dumbbell shapes), which served as a functional subregion. They found that, for observers to retrieve information about the subregion, they first needed to recognize the larger scene context. In another study, Castelhano et al. (2019) found that the subregion could be retrieved independently of the larger scene context in memory when it was spatially and semantically distinct from the larger context. Thus, it seems reasonable to posit that the way in which information across the

scene and within subregions is stored is flexible and depends on task constraints. However, even without active exploration of a place, how spatial representations of scenes are stored in memory has been a topic of much debate over the past few decades. We turn next to scene representations in memory and the role of spatial information in these representations.

6. SCENE REPRESENTATIONS AND SPATIAL INFORMATION IN MEMORY

The role of spatial information in memory has been researched extensively for decades but remains poorly understood. In this section, we examine how spatial information is stored, how it affects the retrieval of individual objects from memory, and how spatial information of scenes is represented across viewpoints.

Past studies have demonstrated that spatial information can be used as an effective cue for retrieval. For instance, early research demonstrated that associating various types of information with a specific location improved recall (Lea 1975, Roediger 1980, Schulman 1973). In addition, many theories of memory have incorporated space as a unique identifier of different representations (e.g., object file theory) (Kahneman et al. 1992, Treisman & Kahneman 1984). More pertinently, researchers have examined memory for objects within a scene and have shown that location information is both incidentally encoded (Castelhano & Henderson 2005, Tatler & Land 2011, Zelinsky & Loschky 2005) and used as a retrieval cue (Hollingworth 2005, 2006; Hollingworth & Rasmussen 2010; Mandler & Ritchey 1977). For instance, Hollingworth (2006) tested memory for objects viewed within scene images and found better memory performance when the test object was shown in the original position compared to a different position. Likewise, he also found that breaking spatial relations in the scene (scrambled scenes) disrupted the same-location benefit, but the benefit was preserved if the spatial relations were changed but not broken, as with a translation through space. This maintenance of scene context through translation and viewpoint changes suggests a robust and flexible representation of the scene context. Thus, we turn to the memory of the overall scene across changes in viewpoint.

In the past, the way in which information from multiple viewpoints of a scene is represented in memory was a matter of great debate. As mentioned above, Gibson (1979) proposed that navigating through an environment led to a changing vista, with a continual update to overall representation. The nature of place representations is thus contrasted with the representation of the immediate vista and the fluid interchange between these types of representations. An important notion of Gibson's theory was that, as one transverses through an environment, one encounters both visible and hidden surfaces, each of which needs to be represented in order to understand and navigate through an environment successfully. The notions of hidden and visible surfaces have been investigated with regard to viewpoint changes of individual objects (e.g., Biederman & Gerhardstein 1993). Past research on scene processing has largely mirrored research on object recognition, and examining the theoretical approaches to object representations may shed some light on theoretical approaches to scene representations.

Traditionally, theoretical approaches to object recognition across viewpoints fall into two main camps: viewpoint dependent and viewpoint invariant. Researchers who posited a viewpoint-dependent approach argued that object representations are largely image based, i.e., recognition of a new view of an object is based on previous experience (Bülthoff & Bülthoff 2003, Edelman 1999, Marr & Poggio 1979, Tarr & Pinker 1989, Ullman 1989). Alternatively, researchers supporting a viewpoint-invariant approach have proposed that the visual system creates a viewpoint-invariant representation of objects (Biederman 1987; Biederman & Gerhardstein 1993, 1995; Marr 1982; Marr & Nishihara 1978). For instance, Biederman (1987) posited that objects are represented as

structural descriptions of spatial relations among simple, volumetric 3D parts. For many years, these views represented two interpretations of visual representations.

Several past studies have also examined which of the two camps better explained spatial translations and viewpoint changes in scene representations in memory (Castelhano & Pollatsek 2010, Castelhano et al. 2009, Epstein et al. 2007, Waller et al. 2009). In one study, Castelhano et al. (2009) had participants study two images prior to an immediate memory test in which they were asked to discriminate between old and new views of the same scene. The two study images were always from two different viewpoints, 40° apart. For the memory test, the distractor images were either an interpolated viewpoint that was 20° from each of the study images or an extrapolated viewpoint that differed by 20° from one of the study images and by 60° from the other. Participants were less accurate at rejecting interpolated test images than at rejecting the extrapolated ones, even accounting for view similarity. Conversely, Waller, Friedman & Waller (2008) and Waller et al. (2009) found that, with extensive training on scene images from multiple viewpoints, a novel view that was taken between the trained views but not shown during training was more easily identified as the same scene than were the trained viewpoints. Taken together, these studies suggest that, under certain conditions, extrapolation of scene information outside the known viewpoints can be observed. However, it is not clear from these studies what role spatial information in the scenes (especially depth information) played in the generalization, as opposed to memory for specific objects in the scene and their relative locations in the 2D images.

Researchers have also examined how different viewpoints are integrated into the spatial representation as you move through the space (Christou & Bülthoff 1999, Epstein et al. 2007, Waller et al. 2009). In one study, Christou & Bülthoff (1999) used a navigation task in a virtual-reality setting in which participants explored an attic (consisting of multiple rooms) from certain viewpoints. They found that, when participants were asked to recognize still images taken from this environment, scene recognition was highly viewpoint dependent. Nonetheless, recent developments in object recognition have found that, depending on the experimental conditions and which parts of the brain are examined, one can obtain data supporting both view-invariant and view-based representations (Gauthier & Tarr 2016). Thus, what the relationship is between the performance and the type of representation is not the right question. Rather, it is more pertinent to ask what type of information is used when.

7. OUTSTANDING QUESTIONS AND FUTURE DIRECTIONS

The main aim of this review is to demonstrate that spatial information is a crucial element that drives processing and associations with targets. The goal is to go beyond the current state of scene processing theories. By putting spatial information at the forefront, we will be able to develop and discover different means of examining processing within scenes.

Traditionally, mechanisms of attention and memory have been examined with simplified stimuli, such as simple shapes, lines, and alphanumeric characters that often vary in color, orientation, and contrast. With these limitations, spatial information of items stands in contrast to scene space, as it is either irrelevant (i.e., objects can appear anywhere on the display) or precise (i.e., the expected effects are linked to specific x–y coordinates). Thus, although one of the main motivations for using simplified stimuli is to avoid potentially confounding effects of semantics, an overlooked question is how spatial information may have an effect. We are interested in how focusing the theory on the spatial dimension can lead to new concepts and approaches to information processing in real-world scenes.

One area that is ripe for new developments is how attention is deployed. There is much discussion in the literature pitting object-based and space-based attention against one another (Egly

et al. 1994, Malcolm & Shomstein 2015, Posner 1980). However, in scenes, these concepts do not quite capture how the object–scene spatial associations discussed above influence attention. Rather, it seems that attention could be deployed not only to specific features and specific *x*–*y* coordinates, but also to larger spatial regions. These regions could be a surface of the environment (as with the surface guidance framework; Pereira & Castelhano 2019) or could be more general (e.g., Kaakinen et al. 2011). Two outstanding questions to arise from this reframing are how to conceptualize attentional deployment across scenes and whether a new theory that is somewhere in between object-based and space-based attention, but informed by spatial associations, would be a useful tool to explore further.

The flip side to the question of how attention is deployed is what happens to areas ignored or devoid of attention. While, in the attention literature, there has recently been an increasing focus on how information is inhibited (e.g., Gaspelin & Luck 2019, Wang & Theeuwes 2018), this question has yet to be addressed in scenes. An interesting aspect in scenes is how information outside relevant regions is processed: whether it is incidentally encoded or completely disregarded. Alternatively, in considering different types of attention, information outside relevant regions may still receive a low-level, monitoring attention in which unexpected items can still capture attention. Future studies addressing these questions will be crucial to moving beyond methodologies of cueing and abrupt onsets and developing new approaches for examining the spread of attention and spatial inattention.

In terms of the spatial associations between objects and scenes, there is also an inherent naturally occurring variance in object placement. For instance, when target objects are selected for a visual search task, some objects are inherently fixed in particular positions and therefore make poor targets, as their predictable placements inevitably lead to ceiling effects. For instance, light switches, doorknobs, and faucets all have very fixed placements within a scene and are therefore quite easy to locate. This is in contrast to objects that have no fixed spatial placement, such as plants, lights, and—for those who are cat owners—cats. Somewhere between these two extremes are the objects typically used in visual search studies, such as coffee mugs, keys, and vases, which have some spatial variation but are limited to specific surfaces in the scene. Thus, there remains a question of how spatial variation is shaped and represented in the brain. Recent advancements in decoding visual information in the brain could point to new ways in which object information could be organized, as has been found with a gradient organization of object categories in the ventral cortex (Gauthier & Tarr 2016, Güçlü & van Gerven 2015).

In addition to questions of spatial information alone, there is also a question of how semantic and spatial information interact. Although this is risking oversimplification, if we examine recent findings, we can see a pattern in which different types of information are associated with different axes. For instance, scene surfaces provide direction on the *y* axis (Pereira & Castelhano 2019, Williams & Castelhano 2019), larger objects and object clusters provide direction on the *x* axis (Mack & Eckstein 2011, Pereira & Castelhano 2014, Võ et al. 2019), and the relation to peripersonal space provides direction on the *z* axis (Fernandes & Castelhano 2019, Man et al. 2019). Relating different scene properties to directions across different axes of space could provide one means of understanding interactions between semantic and spatial information given the task constraints. Putting aside the deterministic nature of assigning different types of information to specific axes, it would be helpful to be able to narrow attention to particular regions of the scene along one continuum.

Another outstanding question is how spatial information may affect how semantic information is learned. Researchers often point to the ways in which statistical regularities govern learning and updating scene knowledge (Chun 2000; Karklin & Lewicki 2005, 2009). However, in addition to passive viewing of spatial associations, the functions of objects also affect the spatial relations

with the larger scene context (Gibson 1979, Land & Hayhoe 2001) and may be a governing principle for how scenes are organized. Further investigations into how scenes are spatially organized and the connections between this organization and function and actions could prove fruitful (Castelhano & Witherspoon 2016, Greene et al. 2016).

Perhaps among the most anticipated and discussed upcoming questions in the scene perception literature are whether and how cognitive experiments will move from computer screens to virtual or augmented environments. When examining how information is processed, a few considerations may have to be taken into account. One consideration is that information is present in all 360°, with both visible and invisible parts of the environments immediately available, which hearkens back to the ideas put forth by Gibson (1979). In addition, there is the information from the current space, as well as neighboring spaces (e.g., the other room in a two room apartment) (Hayhoe & Matthis 2018). One must also consider the effect of participant movement (both head and body movements) and how these extra costs and potentially different cognitive loads will affect task strategies over the scene space. For instance, with a visual search task in a 3D space, as opposed to searching an image on a monitor, one must consider when to turn around and when to move to the next room (this is similar to foraging theories; Cain et al. 2012, Wolfe 2013).

Relatedly, there is also a question of how virtual and augmented reality methodologies will affect our understanding of dynamic scene perception. This can include moving through an environment (Kit et al. 2014, Li et al. 2018), moving and placing objects (Draschkow & Võ 2017), and merging imagined spaces and anticipating or predicting spaces (Denison et al. 2019, Fernández et al. 2019, Freyd 1987). In relation to moving through an environment while completing a task, there is also a question of how the scene properties affect the balance of costs and strategy across space. For instance, object density on a surface (increasing clutter; Rosenholtz 2016, Rosenholtz et al. 2007) presents an interesting challenge. On the one hand, during a search task, participants may avoid cluttered surfaces, as it would be more difficult to detect the target object. On the other hand, participants may prefer the clutter, as the target would have a higher probability of being present on that surface. It remains to be seen how the properties of the scenes across space affect task strategy, how they alter prioritization of information, and what their implications are for cognitive theory.

However, it should be noted that these outstanding questions should not be taken to imply that spatial information should be considered at the exclusion of all else, or that spatial information alone is sufficient to support scene processing. For instance, in the case of search performance, Pereira & Castelhano (2014) had participants search a scene using a gaze-contingent moving window, with the full search scene available within the foveated window (of 2° radius) and the peripheral manipulated to reveal different types of scene information. In one condition, the array of objects was placed on a blank (gray) background, and the positions of the objects were consistent with their placement in the scene. Results showed that there was no benefit to knowing the placement of objects in the periphery, compared to having no information (blank control condition). Based on these results, the limitations of spatial information and its interactions with and relation to semantic information will also have to be taken into consideration.

8. CONCLUSIONS

As mentioned above, we believe that the adoption of spatial information and the leading theoretical framework of scene processing will become more pertinent as experimental paradigms change and as we look forward to examining how to further explain and conceptualize cognitive processing within real-world scenes. The types of questions to be asked are in some ways yet to be uncovered. As new techniques within interactive spaces are developed, new theoretical approaches

and paradigms will lead to new questions that have yet to be asked. In many ways, these questions are not new, but the theoretical approach is; thus, the old is new again.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like to thank Carrick Williams for comments on an earlier version of this review. This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada, Canadian Foundation for Innovation, and Ontario Ministry of Research and Innovation through the Early Researcher Award to M.S.C.

LITERATURE CITED

Andersen GJ, Ni R, Bian Z, Kang J. 2011. Limits of spatial attention in three-dimensional space and dual-task driving performance. *Accid. Anal. Prev.* 43(1):381–90

Baldassano C, Beck DDM, Fei-Fei L. 2013. Differential connectivity within the parahippocampal place area. *NeuroImage* 75:228–37

Bar M. 2004. Visual objects in context. *Nat. Rev. Neurosci.* 5(8):617–29

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94(2):115–47

Biederman I, Gerhardstein PC. 1993. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J. Exp. Psychol. Hum. Percept. Perform.* 19:1162–82

Biederman I, Gerhardstein PC. 1995. Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bülthoff 1995. *J. Exp. Psychol. Hum. Percept. Perform.* 21(6):1506–14

Biederman I, Glass AL, Stacy EW. 1973. Searching for objects in real-world scenes. *J. Exp. Psychol.* 97(1):22–27

Biederman I, Mezzanotte RJ, Rabinowitz JC. 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 14(2):143–77

Bonner MF, Epstein RA. 2017. Coding of navigational affordances in the human visual system. *PNAS* 114(18):4793–98

Bonner MF, Epstein RA. 2018. Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLOS Comput. Biol.* 14(4):e1006111

Boyce SJ, Pollatsek A, Rayner K. 1989. Effect of background information on object identification. *J. Exp. Psychol. Hum. Percept. Perform.* 15(3):556–66

Brockmole JR, Castelhano MS, Henderson JM. 2006. Contextual cueing in naturalistic scenes: global and local contexts. *J. Exp. Psychol. Learn. Mem. Cogn.* 32(4):699–706

Brooks DI, Rasmussen IP, Hollingworth A. 2010. The nesting of search contexts within natural scenes: evidence from contextual cuing. *J. Exp. Psychol. Hum. Percept. Perform.* 36(6):1406–18

Bülthoff I, Bülthoff HH. 2003. Image-based recognition of biological motion, scenes, and objects. In *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes*, ed. MA Peterson, G Rhodes, pp. 146–72. Oxford, UK: Oxford Univ. Press

Burgess N, Spiers HJ, Paleologou E. 2004. Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition* 94(2):149–66

Cain MS, Vul E, Clark K, Mitroff SR. 2012. A Bayesian optimal foraging model of human visual search. *Psychol. Sci.* 23(9):1047–54

Castelhano MS, Fernandes S, Theriault J. 2019. Examining the hierarchical nature of scene representations in memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 45(9):1619–33

Castelhano MS, Heaven C. 2010. The relative contribution of scene context and target features to visual search in scenes. *Atten. Percept. Psychophys.* 72(5):1283–97
Castelhano MS, Heaven C. 2011. Scene context influences without scene gist: eye movements guided by spatial associations in visual search. *Psychon. Bull. Rev.* 18(5):890–96
Castelhano MS, Henderson JM. 2005. Incidental visual memory for objects in scenes. *Vis. Cogn.* 12(6):1017–40
Castelhano MS, Henderson JM. 2007. Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 33(4):753–63
Castelhano MS, Henderson JM. 2008. The influence of color on the perception of scene gist. *J. Exp. Psychol. Hum. Percept. Perform.* 34(3):660–75
Castelhano MS, Pereira EJ. 2018. The influence of scene context on parafoveal processing of objects. *Q. J. Exp. Psychol.* 71(1):229–40
Castelhano MS, Pollatsek A. 2010. Extrapolating spatial layout in scene representations. *Mem. Cogn.* 38(8):1018–25
Castelhano MS, Pollatsek A, Rayner K. 2009. Integration of multiple views of scenes. *Atten. Percept. Psychophys.* 71(3):490–502
Castelhano MS, Witherspoon RL. 2016. How you use it matters: object function guides attention during visual search in scenes. *Psychol. Sci.* 27(5):606–21
Christou CG, Bülthoff HH. 1999. View dependence in scene recognition after active learning. *Mem. Cogn.* 27(6):996–1007
Chun MM. 2000. Contextual cueing of visual attention. *Trends Cogn. Sci.* 4(5):170–78
Collegio AJ, Nah JC, Scotti PS, Shomstein S. 2019. Attention scales according to inferred real-world object size. *Nat. Hum. Behav.* 3(1):40–47
Costantini M, Ambrosini E, Scorolli C, Borghi AM. 2011. When objects are close to me: affordances in the peripersonal space. *Psychon. Bull. Rev.* 18(2):302–8
Cutting JE, Vishton PM. 1995. Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In *Handbook of Perception and Cognition*, ed. W Epstein, SJ Rogers, pp. 69–117. Cambridge, MA: Academic. 2nd ed.
Davenport JL, Potter MC. 2004. Scene consistency in object and background perception. *Psychol. Sci.* 15(8):559–64
De Graef P, Christiaens D, D'Ydewalle G. 1990. Perceptual effects of scene context on object identification. *Psychol. Res.* 52(4):317–29
De Graef P, De Troy A, D'Ydewalle G. 1992. Local and global contextual constraints on the identification of objects in scenes. *Can. J. Psychol. Rev. Can. Psychol.* 46(3):489–508
de la Rosa S, Moraglia G, Schneider BA. 2008. The magnitude of binocular disparity modulates search time for targets defined by a conjunction of depth and colour. *Can. J. Exp. Psychol. Rev. Can. Psychol. Exp.* 62(3):150–55
Denison RN, Yuval-Greenberg S, Carrasco M. 2019. Directing voluntary temporal attention increases fixational stability. *J. Neurosci.* 39(2):353–63
Downing CJ, Pinker S. 1985. The spatial structure of visual attention. In *Attention and Performance*, Vol. XI: *Mechanisms of Attention and Visual Search*, ed. MI Posner, OSM Martin, pp. 171–87. Hillsdale, NJ: Erlbaum
Draschkow D, Võ ML-H. 2017. Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Sci. Rep.* 7(1):16471
Eckstein MP, Drescher BA, Shimozaki SS. 2006. Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychol. Sci.* 17(11):973–80
Edelman S. 1999. *Representation and Recognition in Vision*. Cambridge, MA: MIT Press
Egly R, Rafal R, Driver J, Starrveveld Y. 1994. Covert orienting in the split brain reveals hemispheric specialization for object-based attention. *Psychol. Sci.* 5(6):380–83
Epstein RA, Baker CI. 2019. Scene perception in the human brain. *Annu. Rev. Vis. Sci.* 5:373–97
Epstein RA, Harris A, Stanley D, Kanwisher N. 1999. The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23(1):115–25
Epstein RA, Higgins JS, Jablonski K, Feiler AM. 2007. Visual scene processing in familiar and unfamiliar environments. *J. Neurophysiol.* 97(5):3670–83

Epstein RA, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* 392(6676):598–601

Fernandes S, Castelhano MS. 2019. *The Foreground Bias: Initial Scene Representations Across the Depth Plane*. PsyArXiv. **https://doi.org/10.31234/OSF.IO/S32WZ**

Fernández A, Denison RN, Carrasco M. 2019. Temporal attention improves perception similarly at foveal and parafoveal locations. *J. Vis.* 19(1):12

Ferrara K, Park S. 2016. Neural representation of scene boundaries. *Neuropsychologia* 89:180–90

Finlayson NJ, Grove PM. 2015. Visual search is influenced by 3D spatial layout. *Atten. Percept. Psychophys.* 77(7):2322–30

Foulsham T, Underwood G. 2007. How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception* 36(8):1123–38

Freyd JJ. 1987. Dynamic mental representations. *Psychol. Rev.* 94(4):427–38

Friedman A. 1979. Framing pictures: the role of knowledge in automatized encoding and memory for gist. *J. Exp. Psychol. Gen.* 108(3):316–55

Friedman A, Waller D. 2008. View combination in scene recognition. *Mem. Cogn.* 36(3):467–78

Gaspar JG, Ward N, Neider MB, Crowell J, Carbonari R, et al. 2016. Measuring the useful field of view during simulated driving with gaze-contingent displays. *Hum. Factors* 58(4):630–41

Gaspelin N, Luck SJ. 2019. Inhibition as a potential resolution to the attentional capture debate. *Curr. Opin. Psychol.* 29:12–18

Gauthier I, Tarr MJ. 2016. Visual object recognition: Do we (finally) know more now than we did? *Annu. Rev. Vis. Sci.* 2:377–96

de Gonzaga Gawryszewski L, Riggio L, Rizzolatti G, Umiltá C. 1987. Movements of attention in the three spatial dimensions and the meaning of "neutral" cues. *Neuropsychologia* 25(1):19–29

Gibson J. 1950. *The Perception of the Visual World*. Boston: Houghton Mifflin

Gibson JJ. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin

Greene MR. 2013. Statistics of high-level scene context. *Front. Psychol.* 4:777

Greene MR, Baldassano C, Esteva A, Beck DM, Fei-Fei L. 2016. Visual scenes are categorized by function. *J. Exp. Psychol. Gen.* 145(1):82–94

Greene MR, Oliva A. 2009a. The briefest of glances: the time course of natural scene understanding. *Psychol. Sci.* 20(4):464–72

Greene MR, Oliva A. 2009b. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 58(2):137–76

Greene MR, Oliva A. 2010. High-level aftereffects to global scene properties. *J. Exp. Psychol. Hum. Percept. Perform.* 36(6):1430–42

Gronau N, Shachar M. 2015. Contextual consistency facilitates long-term memory of perceptual detail in barely seen images. *J. Exp. Psychol. Hum. Percept. Perform.* 41(4):1095–111

Güçlü U, van Gerven MAJ. 2015. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* 35(27):10005–14

Hassabis D, Maguire EA. 2007. Deconstructing episodic memory with construction. *Trends Cogn. Sci.* 11(7):299–306

Hayhoe MM, Matthis JS. 2018. Control of gaze in natural environments: effects of rewards and costs, uncertainty and memory in target selection. *Interface Focus* 8(4):20180009

He C, Peelen MV, Han Z, Lin N, Caramazza A, Bi Y. 2013. Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *NeuroImage* 79:1–9

Henderson JM, Larson CL, Zhu DC. 2008. Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: an fMRI study. *Brain Cogn.* 66(1):40–49

Henderson JM, Weeks PAJ, Hollingworth A. 1999. The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25(1):210–28

Henderson JM, Zhu DC, Larson CL. 2011. Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: an fMRI study. *Vis. Cogn.* 19(7):910–27

Hillstrom AP, Segabinazi JD, Godwin HJ, Liversedge SP, Benson V. 2017. Cat and mouse search: the influence of scene and object analysis on eye movements when targets change locations during search. *Phil. Trans. R. Soc. Lond. B* 372(1714):20160106

Hirtle SC, Jonides J. 1985. Evidence of hierarchies in cognitive maps. *Mem. Cogn.* 13(3):208–17

Hollingworth A. 2005. Memory for object position in natural scenes. *Vis. Cogn.* 12(6):1003–16

Hollingworth A. 2006. Scene and position specificity in visual memory for objects. *J. Exp. Psychol. Learn. Mem. Cogn.* 32(1):58–69

Hollingworth A, Henderson JM. 1999. Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychol.* 102(2–3):319–43

Hollingworth A, Rasmussen IP. 2010. Binding objects to locations: the relationship between object files and visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 36(3):543–64

Intraub H. 2010. Rethinking scene perception: a multisource model. *Psychol. Learn. Motiv.* 52:231–64

Intraub H, Richardson M. 1989. Wide-angle memories of close-up scenes. *J. Exp. Psychol. Learn. Mem. Cogn.* 15(2):179–87

Josephs EL, Konkle T. 2019. Perceptual dissociations among views of objects, scenes, and reachable spaces. *J. Exp. Psychol. Hum. Percept. Perform.* 45(6):715–28

Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M. 2007. Processing scene context: fast categorization and object interference. *Vis. Res.* 47(26):3286–97

Kaakinen JK, Hyönä J, Viljanen M. 2011. Influence of a psychological perspective on scene viewing and memory for scenes. *Q. J. Exp. Psychol.* 64(7):1372–87

Kahneman D, Treisman A, Gibbs BJ. 1992. The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24(2):175–219

Kaiser D, Cichy RM. 2018. Typical visual-field locations facilitate access to awareness for everyday objects. *Cognition* 180:118–22

Kaiser D, Quek GL, Cichy RM, Peelen MV. 2019. Object vision in a structured world. *Trends Cogn. Sci.* 23(8):672–85

Kaiser D, Stein T, Peelen MV. 2014. Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *PNAS* 111(30):11217–22

Kamps FS, Lall V, Dilks DD. 2016. The occipital place area represents first-person perspective motion information through scenes. *Cortex* 83:17–26

Karklin Y, Lewicki MS. 2005. A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Comput.* 17(2):397–423

Karklin Y, Lewicki MS. 2009. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457(7225):83–86

Katti H, Peelen MV, Arun SP. 2016. Deep neural networks can be improved using human-derived contextual expectations. arXiv:1611.07218 [cs.CV]

Kit D, Katz L, Sullivan B, Snyder K, Ballard D, Hayhoe M. 2014. Eye movements, visual search and scene memory, in an immersive virtual environment. *PLOS ONE* 9(4):e94362

Konkle T, Olivia A. 2007. Normative representation of objects: evidence for an ecological bias in object perception and memory. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 29, pp. 407–12. Austin, TX: Cogn. Sci. Soc.

Kravitz DJ, Peng CS, Baker CI. 2011. Real-world scene representations in high-level visual cortex: It's the spaces more than the places. *J. Neurosci.* 31(20):7322–33

Land MF, Hayhoe M. 2001. In what ways do eye movements contribute to everyday activities? *Vis. Res.* 41(25–26):3559–65

Lea G. 1975. Chronometric analysis of the method of loci. *J. Exp. Psychol. Hum. Percept. Perform.* 1(2):95–104

Li C-L, Aivar MP, Tong MH, Hayhoe MM. 2018. Memory shapes visual search strategies in large-scale environments. *Sci. Rep.* 8(1):4324

Loftus GR, Mackworth NH. 1978. Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 4(4):565–72

Mack SC, Eckstein MP. 2011. Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *J. Vis.* 11(9):9

Mackworth NH, Morandi AJ. 1967. The gaze selects informative details within pictures. *Percept. Psychophys.* 2(11):547–52
Maguire EA, Nannery R, Spiers HJ. 2006. Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain* 129(11):2894–907
Malcolm GL, Henderson JM. 2010. Combining top-down processes to guide eye movements during real-world scene search. *J. Vis.* 10(2):4
Malcolm GL, Shomstein S. 2015. Object-based attention in real-world scenes. *J. Exp. Psychol. Gen.* 144(2):257–63
Man L, Krzys K, Castelhano M. 2019. The foreground bias: differing impacts across depth on visual search in scenes. PsyArXiv. **https://doi.org/10.31234/OSF.IO/W6J4A**
Mandler JM, Johnson NS. 1976. Some of the thousand words a picture is worth. *J. Exp. Psychol. Hum. Learn. Mem.* 2(5):529–40
Mandler JM, Ritchey GH. 1977. Long-term memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* 3(4):386–96
Marr D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman
Marr D, Nishihara HK. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B* 200(1140):269–94
Marr D, Poggio BT. 1979. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B* 204(1156):301–28
Marrara MT, Moore CM. 2000. Role of perceptual organization while attending in depth. *Percept. Psychophys.* 62(4):786–99
Martinez-Conde S, Macknik SL, Martinez LM, Alonso J-M, Tse PU, et al. 2006. Top-down facilitation of visual object recognition: object-based and context-based contributions. *Prog. Brain Res.* 155:3–21
McNamara TP. 1986. Mental representations of spatial relations. *Cogn. Psychol.* 18(1):87–121
McNamara TP, Hardy JK, Hirtle SC. 1989. Subjective hierarchies in spatial memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 15(2):211–27
Mullally SL, Maguire EA. 2011. A new role for the parahippocampal cortex in representing space. *J. Neurosci.* 31(20):7441–49
Mullally SL, Maguire EA. 2013. Exploring the role of space-defining objects in constructing and maintaining imagined scenes. *Brain Cogn.* 82(1):100–7
Munneke J, Brentari V, Peelen MV. 2013. The influence of scene context on object recognition is independent of attentional focus. *Front. Psychol.* 4:552
Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL. 2002. Shape perception reduces activity in human primary visual cortex. *PNAS* 99(23):15164–69
Nagata S. 1993. How to reinforce perception of depth in single two-dimensional pictures. In *Pictorial Communication in Virtual and Real Environments*, ed. SR Ellis, pp. 527–45. Philadelphia, PA: Taylor & Francis
Neider MB, Zelinsky GJ. 2006. Scene context guides eye movements during visual search. *Vis. Res.* 46(5):614–21
Oliva A. 2005. Gist of the scene. In *Neurobiology of Attention*, ed. L Itti, G Rees, JK Tsotsos, pp. 251–56. Cambridge, MA: Academic
Oliva A, Park S, Konkle T. 2010. Representing, perceiving, and remembering the shape of visual space. In *Vision in 3D Environments*, ed. LR Harris, MRM Jenkin, pp. 308–40. Cambridge, UK: Cambridge Univ. Press
Oliva A, Torralba A. 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42(3):145–75
Oliva A, Torralba A. 2007. The role of context in object recognition. *Trends Cogn. Sci.* 11(12):520–27
Palmer SE. 1975. The effects of contextual scenes on the identification of objects. *Mem. Cogn.* 3(5):519–26
Park J, Park S. 2018. Coding of navigational distance in the visual scene-selective cortex. *J. Vis.* 18(10):739
Park S, Intraub H, Yi D-J, Widders D, Chun MM. 2007. Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron* 54(2):335–42

Park S, Konkle T, Oliva A. 2014. Parametric coding of the size and clutter of natural scenes in the human brain. *Cereb. Cortex* 25(7):1792–805
Pereira EJ, Castelhano MS. 2014. Peripheral guidance in scenes: the interaction of scene context and object content. *J. Exp. Psychol. Hum. Percept. Perform.* 40(5):2056–72
Pereira EJ, Castelhano MS. 2019. Attentional capture is contingent on scene region: using surface guidance framework to explore attentional mechanisms during search. *Psychon. Bull. Rev.* 26:1273–81
Pezdek K, Whetstone T, Reynolds K, Askari N, Dougherty T. 1989. Memory for real-world scenes: the role of consistency with schema expectation. *J. Exp. Psychol. Learn. Mem. Cogn.* 15(4):587–95
Posner MI. 1980. Orienting of attention. *Q. J. Exp. Psychol.* 32(1):3–25
Previc FH. 1998. The neuropsychology of 3-D space. *Psychol. Bull.* 124(2):123–64
Roediger HL. 1980. The effectiveness of four mnemonics in ordering recall. *J. Exp. Psychol. Hum. Learn. Mem.* 6(5):558–67
Rogé J, Pébayle T, Lambilliotte E, Spitzenstetter F, Giselbrecht D, Muzet A. 2004. Influence of age, speed and duration of monotonous driving task in traffic on the driver's useful visual field. *Vis. Res.* 44(23):2737–44
Rosenholtz R. 2016. Capabilities and limitations of peripheral vision. *Annu. Rev. Vis. Sci.* 2:437–57
Rosenholtz R, Li Y, Nakano L. 2007. Measuring visual clutter. *J. Vis.* 7(2):17
Schulman AI. 1973. Recognition memory and the recall of spatial location. *Mem. Cogn.* 1(3):256–60
Song J, Bennett P, Sekuler A, Sun H-J. 2017. Effect of apparent depth in peripheral target detection in driving under focused and divided attention. *J. Vis.* 17(10):388
Stein T, Peelen MV. 2017. Object detection in natural scenes: independent effects of spatial and category-based attention. *Atten. Percept. Psychophys.* 79(3):738–52
Summerfield JJ, Lepsien J, Gitelman DR, Mesulam MM, Nobre AC. 2006. Orienting attention based on long-term memory experience. *Neuron* 49(6):905–16
Tarr MJ, Pinker S. 1989. Mental rotation and orientation-dependence in shape recognition. *Cogn. Psychol.* 21(2):233–82
Tatler BW, Land MF. 2011. Vision and the representation of the surroundings in spatial memory. *Phil. Trans. R. Soc. B* 366(1564):596–610
Torralba A, Oliva A, Castelhano MS, Henderson JM. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113(4):766–86
Treisman A, Kahneman D. 1984. Changing views of attention and automaticity. In *Varieties of Attention*, ed. R Parasuraman, DR Davies, pp. 29–61. Cambridge, MA: Academic
Ullman S. 1989. Aligning pictorial descriptions: an approach to object recognition. *Cognition* 32(3):193–254
Vatterott DB, Vecera SP. 2015. The attentional window configures to object and surface boundaries. *Vis. Cogn.* 23(5):561–76
Võ ML-H, Boettcher SE, Draschkow D. 2019. Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Curr. Opin. Psychol.* 29:205–10
Võ ML-H, Henderson JM. 2009. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J. Vis.* 9(3):24
Võ ML-H, Henderson JM. 2011. Object-scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Atten. Percept. Psychophys.* 73(6):1742–53
Waller D, Friedman A, Hodgson E, Greenauer N. 2009. Learning scenes from multiple views: Novel views can be recognized more efficiently than learned views. *Mem. Cogn.* 37(1):90–99
Wang B, Theeuwes J. 2018. How to inhibit a distractor location? Statistical learning versus active, top-down suppression. *Atten. Percept. Psychophys.* 80(4):860–70
Williams CC, Castelhano MS. 2019. The changing landscape: high-level influences on eye movement guidance in scenes. *Vision* 3(3):33
Wolbers T, Klatzky RL, Loomis JM, Wutte MG, Giudice NA. 2011. Modality-independent coding of spatial layout in the human brain. *Curr. Biol.* 21(11):984–89
Wolfe JM. 2013. When is it time to move to the next raspberry bush? Foraging rules in human visual search. *J. Vis.* 13(3):10

Wolfe JM, Võ ML, Evans KK, Greene MR. 2011. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn. Sci.* 15(2):77–84

Wu C-C, Wick FA, Pomplun M. 2014. Guidance of visual attention by semantic information in real-world scenes. *Front. Psychol.* 5:54

Zelinsky GJ, Loschky LC. 2005. Eye movements serialize memory for objects in scenes. *Percept. Psychophys.* 67(4):676–90