The Architecture of Object-Based Attention

Patrick Cavanagh[1,2] · Gideon P. Caplovitz[3] · Taissa K. Lytchenko[3] · Marvin R. Maechler[4] · Peter U. Tse[4] ·
David L. Sheinberg[5,6]

Abstract

The allocation of attention to objects raises several intriguing questions: What are objects, how does attention access them, what anatomical regions are involved? Here, we review recent progress in the field to determine the mechanisms underlying object-based attention. First, findings from unconscious priming and cueing suggest that the preattentive targets of object-based attention can be fully developed object representations that have reached the level of identity. Next, the control of object-based attention appears to come from ventral visual areas specialized in object analysis that project downward to early visual areas. How feedback from object areas can accurately target the object's specific locations and features is unknown but recent work in autoencoding has made this plausible. Finally, we suggest that the three classic modes of attention may not be as independent as is commonly considered, and instead could all rely on object-based attention. Specifically, studies show that attention can be allocated to the separated members of a group—without affecting the space between them—matching the defining property of feature-based attention. At the same time, object-based attention directed to a single small item has the properties of space-based attention. We outline the architecture of object-based attention, the novel predictions it brings, and discuss how it works in parallel with other attention pathways.

Keywords Attention · Vision · Object-based attention

Introduction

To make sense of the world, we must pick and choose what sensory information will inform our thoughts and actions. Attention lies at the heart of this selection step, allowing a flexible, transient improvement of the rate and fidelity of processing of the selected target. The nature of the selection process has been the subject of decades of debate, with many studies suggesting that attention can independently select locations, features, or objects (Kravitz & Behrmann, 2011; Posner, 1980;

✉ Patrick Cavanagh
patcav1@yorku.ca

1 Department of Psychology, Glendon College, 2275 Bayview Avenue, North York, ON M4N 3M6, Canada

2 CVR, York University, Toronto, ON, Canada

3 University of Nevada, Reno, Reno, NV, USA

4 Dartmouth College, Hanover, NH, USA

5 Department of Neuroscience, Brown University, Providence, RI, USA

6 Carney Institute for Brain Science, Brown University, Providence, RI, USA

Posner et al., 1982). Space-based attention is directed to a specific location and enhances the processing of all features at a specific location of the visual field. Feature-based attention enhances the processing of a specific feature across the entire visual field (Baylis & Driver, 1992; Driver & Baylis, 1989). Finally, object-based attention enhances only the features at the locations that define the selected object. Interestingly, it can be allocated rapidly throughout an object even if attention was initially directed to only part of the object (Egly et al., 1994). In all three cases, space-, feature-, and object-based attention, the selection can be triggered bottom-up—by a flashed cue or a sudden appearance—or top-down, by a central cue or by instructions. And in each case, the consequence is a short-lasting improvement in performance on further, typically conscious, processing of the attended location, object, or feature.

There have been several reviews of object-based attention that focused on methods and effect size (Chen, 2012; Francis & Thunell, 2022; Reppa et al., 2012). Here, we review the literature on object-based attention to determine the processes that underlie it. First, what are the targets of object-based attention? Evidence from anesthetized humans and nonhuman primates, and from unconscious priming and threat responses shows that visual input gets preattentively

parsed very deeply, often to the level of complete object representations. Object-based attention appears to operate on these fully developed, preattentive object representations that have reached the level of identity and semantic connectivity. This is in stark contrast to the claims from the visual search literature that objects cannot exist until attention arrives to bind their features together (Treisman & Gormican, 1988; Wolfe & Bennett, 1997). An analysis of task demands and stimulus configurations shows how objects can indeed exist preattentively and yet fail to act that way in visual search tasks. Second, we examine several possibilities for the allocation of attention to objects. Several studies reveal that the process must involve projections from object areas to the locations and features of the object. Finally, we find that the three classic modes of attention may not be as independent as is commonly considered—instead, all three modes of selection could rely on the mechanisms of object-based attention. Studies on object-based attention directed to groups demonstrate that object-based attention in this case shows the critical nonspatial property of feature-based attention: extending over group elements but not the space between them. The processes of object-based attention may be the source of feature-based attention as well. Even so, the object-based system cannot be the only mechanism for allocating attention. There is overwhelming evidence for a space-based system controlled by saccade centers (Awh et al., 2006).

We start by briefly reviewing the three types of attentional selection before focusing on the properties required for one of them, object-based attention. This will lead us to a number of novel conclusions.

Types of attention

Space-based attention

When visual items are selected from a predefined or cued location, many suggest that the effect is like shining a spotlight on that region. This is often represented as locations of high activity on a salience or priority map (Itti & Koch, 2001), or as we present in a more general framework in Fig. 1, a gain map that influences responses on corresponding location of cortical or subcortical maps. Across the attention literature, these maps have been called attention, salience, and priority maps but for the rest of this review, we will use the term "gain maps" as a general label for all of these. The hot spots on the gain map can be initiated by top-down (endogenous) or bottom-up signals (exogenous). Posner's (1980) original cueing experiments demonstrated both. In this proposal, the effects of exogenous and endogenous cues on a gain map are identical—both lead to increased activity at the cued location which then boosts responsiveness in early visual cortex (Brefczynski & DeYoe, 1999). The difference is whether the activation of the gain map comes directly, bottom-up, from the image (e.g., a salient flash) or top-down from another cortical area that has computed the location of interest (e.g., based on expectation or a symbolic cue like an arrow). In both cases, the effects of the downward projections to early visual cortex are extended in time and space so that targets appearing briefly after the cue are processed more rapidly (Posner, 1980).
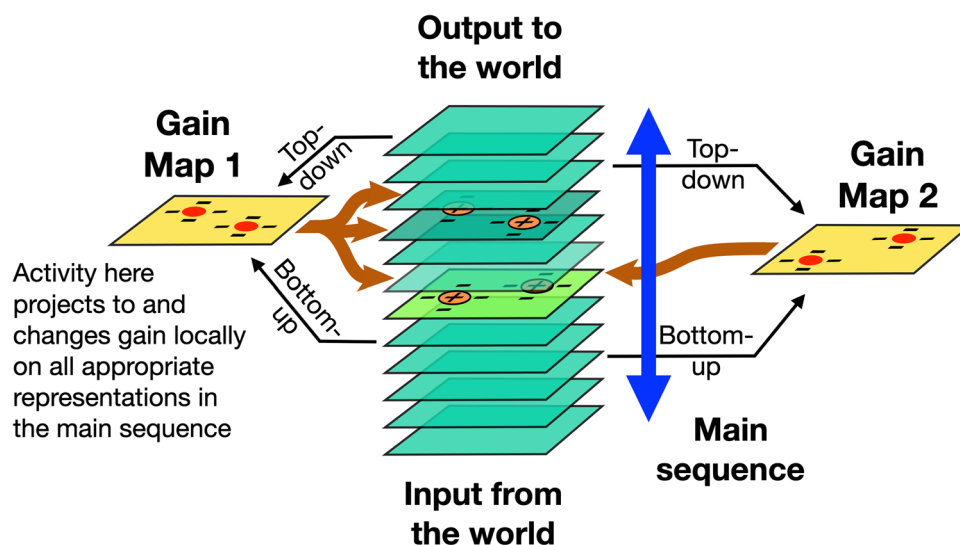

Fig. 1 Gain maps implement attention benefits. The gain map implements salience or attention control by receiving locations of interest (with inhibitory surrounds) from top-down or bottom-up projections. These activity peaks project to the corresponding locations on cortical or subcortical areas where they increase gain and suppress surrounding activity. Two gain maps are shown here, but there could be many. These two maps are in retinotopic coordinates but other maps might be in other spaces such as semantic or feature space so that the projection to retinotopic areas would require a recovery of locations. (Color figure online)

Gain is a ubiquitous factor in all connections in the brain, whether feedforward or feedback;. why would some maps, and the gain they mediate, be special? One reason is related to a second property of attention: it not only changes gain but it is limited to one or at most, a few, targets at a time (Pylyshyn & Storm, 1988). This limited capacity is most easily understood as a winner-take-all or biased competition where the target with the highest activity suppresses the others (Desimone, 1998; Lee et al., 1999; Vecera & Behrmann, 2001). This suggests that the reach of this suppression must be very wide to deal with competitors from virtually anywhere in the visual field. Consider, in contrast, activity on, say, V2 or V3. Clearly the feedforward and feedback projections from these areas change the gain elsewhere but the size of the suppressive surrounds is quite limited (Gegenfurtner et al., 1997; Shushruth et al., 2009). Local suppression may resolve local competition but cannot mediate the image-wide competition that is a critical element of attention (Franconeri et al., 2013).

A number of areas are involved in directing spatial attention (Bisley, 2011; Maunsell, 2015). Parietal areas (Corbetta & Shulman, 2002; Gazzaniga & Ladavas, 1987; Posner et al., 1984) show strong attention-linked responses and parietal lesions have devastating effects on attention (Battelli et al., 2001; Friedrich et al., 1998). Saalmann et al. (2012) have reported that the pulvinar may be part of the control of attention and Zhang et al. (2012) have argued that V1 can act as a salience (gain) map. There is also compelling evidence that saccade centers located in the frontal eye fields in the frontal cortex and the superior colliculus in the midbrain also direct spatial attention (Awh et al., 2006; Moore & Zirnsak, 2017). These findings suggest that the space-based allocation of attention can be realized in the activity patterns of these saccade/gain maps and the effects of their projections to visual cortex. Whatever the source of these projections to early visual cortex, their effect once they arrive has been described variously as an increase of gain for the stimulus at that location, a sharpening of tuning, and a suppression of responses to adjacent stimuli (Maunsell, 2015). The end result is a benefit for the stimulus at the attended location.

Feature-based attention

It is also possible to direct attention to a particular feature throughout the visual field, independently of space-based attention. Flashing a red cue, for example, leads to speeded processing of a red target at an unknown location in visual search (Lin et al., 2011). Unlike space-based attention, this processing benefit is also seen on task-irrelevant stimuli elsewhere in the visual field that share the attended feature (Saenz et al., 2002; Treue & Trujillo, 1999; White et al., 2013). Indeed, there is strong evidence that space-based and feature-based attention are operationally distinct. For example, space-based effects are seen when the location of the upcoming target is known even though its features are not (Awh et al., 2003). In contrast, the effects of feature-based attention can be observed even when an irrelevant feature is superimposed on the target feature so that spatial attention is of no use at all (Saenz et al., 2002; Scolari et al., 2014).

A defining property of feature-based attention (Fig. 2A) is that attention to one feature increases the gain of cortical neurons tuned to the attended feature, anywhere in the visual field (Saenz et al., 2002; Treue & Trujillo, 1999). Further evidence of this broadcasting of feature-specific activation is seen in an fMRI study by Serences and Boynton (2007). Participants attended to one of two overlapping sets of dots, moving either up or down. The attended direction could be decoded from the stimulus location, as expected, but also from a location in the other hemifield that had no stimulus. Two further studies demonstrated the same point by showing that attention to one of the two directions generated motion aftereffects in the direction opposite to the attended motion at locations all over the visual field (Boynton et al., 2006; Liu & Mance, 2011).

Interestingly, the attentional benefit that is directed to one feature of an object appears to extend to other features of the object. O'Craven et al. (1999) superimposed a face and a house, with one moving whereas the other was stationary. When observers directed their attention to the motion, for example, there was an enhanced activation for the cortical representation of the other attributes of the moving object but not for the other attributes of the stationary object. This result suggested that the processing of all attributes (features) of an attended object will be enhanced whether or not they are relevant to the task. Similar results were reported in an EEG study by Adamian et al. (2020) and in an MEG study by Schoenfeld et al. (2014). These studies were testing top-down activation of feature-based attention—the feature to attend to was specified on each trial and the extension of attentional benefits to task irrelevant stimuli only occurred when they shared the attended feature. Feature-based attention can also be allocated without prior specification of the feature to be attended. For example, during saccades, the attention directed to the saccade target also extends to other stimuli that have the same direction of motion as the target (White et al., 2013).

Liu (2019), in a recent review, continued Müller's (2014) earlier proposal that feature-based attention is controlled by the same network of frontoparietal regions responsible for space-based attention. There is surprisingly little physiological evidence for an area that controls the allocation of feature-based attention in the same way that saccade areas have been shown to control space-based attention (Awh et al., 2006; Moore & Zirnsak, 2017). We could imagine a feature gain map, equivalent to a spatial gain map but with dimensions of features rather than space. For example, the region of inferotemporal (IT) cortex (Fig. 2B) could act as such a feature map (Tanaka, 2003). Once a feature is activated here, downward projections to all locations with that feature would
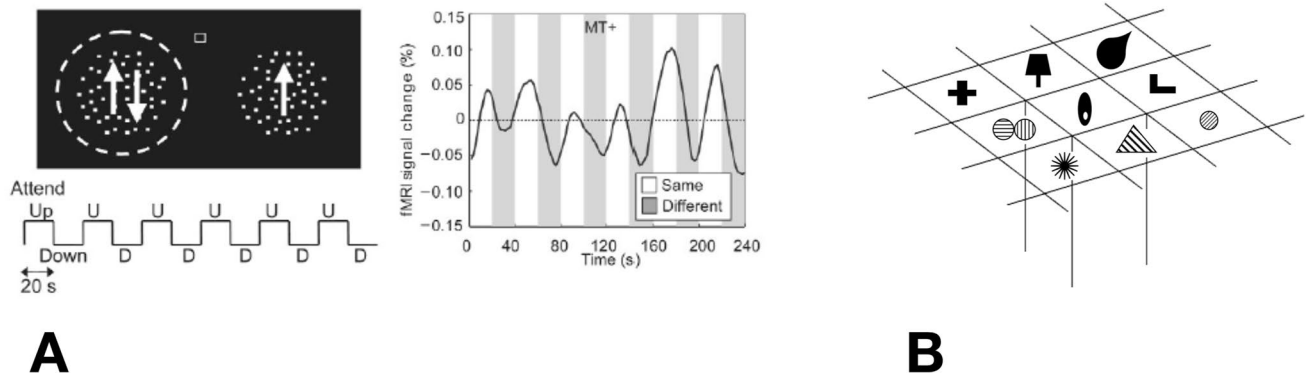
A **B**

Fig. 2 **A** Feature-based attention. Left. Two fields of dots are superimposed on the left, one moving up and the other down. Participants attended to one direction of motion on the left and made a speed discrimination. The patch of dots on the right moved in only one direction and it was irrelevant to the task. Participants switched attention between monitoring upward and downward dots every 20 s. Right. The activation from MT+ contralateral to the ignored stimulus increased whenever the attended direction on one side matched the ignored direction on the other. This demonstrated that feature-based attention is not spatially constrained. From Saenz et al. (2002), with permission. **B** Feature organization in area TE of IT cortex. Cells in this area respond selectively to complex visual object features. Those that respond to similar features cluster in a columnar organization, they are arranged by feature type not spatial location. Adapted from Tanaka (2003)

allocate attention to appropriate locations throughout the visual field. While this may be a plausible structure for the control of feature-based attention, this region of IT cortex in particular is not a likely candidate because the features here are more complex than the simple features that are the basis of pop-out visual search (Treisman & Gormican, 1988) and feature-based attention tasks (Saenz et al., 2002). However, the features used in feature-based attention tasks have not yet explored the full range of features that may be supported (see Box 1). Whatever the case, the anatomy underlying feature-based attention remains a missing piece of attentional control—a missing piece that we will uncover later in this review.

Object-based attention

Duncan (1984) was the first to note that attention could select one of two items that were at the same location (Fig. 3,

left). This extended the simple idea of space-based selection, by demonstrating that attention could selectively enhance the processing of a subset of features at the same location in the visual field. This initial evidence indicated that there was some object-specificity in selection. Blaser et al. (2000) extended this result by superimposing two Gabor patches that smoothly and independently changed their orientations, colors, and bar widths over time. Participants were able to keep track of the features of one of the two objects, accurately reporting the direction of any abrupt changes in two properties, say, the target's orientation and color. However, they could not track one property of each object at the same location even though they still had to report only the changes in two features, say, the orientation of one and the color of the other.

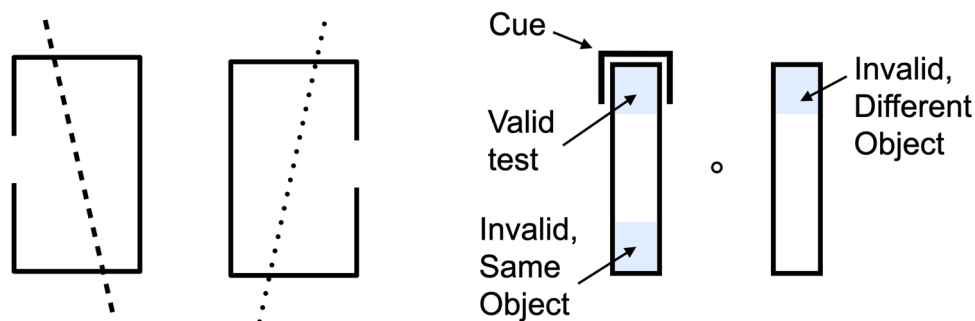Egly et al. (1994) further confirmed and extended this object-specificity by showing that attention, once directed

Fig. 3 Object-based attention. Left, When two objects are superimposed two features of one object are more accurately reported than one feature from each object. Adapted from Duncan (1984). Right.

A probe is detected more rapidly in the uncued location if it is within the cued object, compared with on a different object. Adapted from Egly et al. (1994)

to part of an object, can be allocated to locations within the object faster than to locations outside the object (Fig. 3, right). A large literature now documents numerous ways in which the spatially cueing of one part of an object can lead to attentional facilitation at uncued locations of the same object (Chen, 2012; Müller, 2014).

Many of the studies of object-based attention have used the two-rectangle style of Egly et al. (1994). Effects of object-based attention have also been found using paradigms that show observers two overlapping surfaces (Mitchell et al., 2004; Valdes-Sosa et al., 2000), or superimposed objects (houses vs. faces: O'Craven et al., 1999; Tong et al., 1998) or adjacent, real objects (Malcolm & Shomstein, 2015). Across this wide range of paradigms, outcomes have been quite consistent: when response time is the measurement, a response advantage of about 15 ms is found for tests at uncued locations within the object versus uncued location in an adjacent object.

Francis and Thunell (2022) recently reported that this modest 15 ms effect size has led to an "excess success" in reporting object-based experiments based on the original Egly et al. (1994) paradigm. Nevertheless, they did show that with the appropriate number of participants, 264 in their case, the two-rectangle measure of object attention holds up (at 14 ms). Possibly the objects typically used in the experiments are not very salient.

Malcolm and Shomstein (2015) did find stronger effects using real world objects (around 50 ms for the same object, invalid cue advantage). Future, behavioral, EEG and fMRI experiments could give a better picture of the effect size and spatial extent of object-based attention (see Box 1).

Box 1: New object-based experiments.

1. Would an inattentional blindness test reveal a strong object influence (Fig. B1A)? The cue is always valid and there is only one critical trial where a probe is presented in an unexpected location, which may or may not be in the same rectangle as the colored test that the participants are monitoring. The experiment ends at that point (Mack & Rock, 1998). The probe may be missed if it is not in the attended rectangle.

2. What is the shape of the downward projections to a target object? The experiment in Figure B1B adapts Drew et al.,'s (2009) paradigm to the two-rectangle stimulus. Participants do not respond to the flashes but the P1 and N1 components they trigger should be increased for flashes on regions (and at times) with more attention. The cue is always valid.
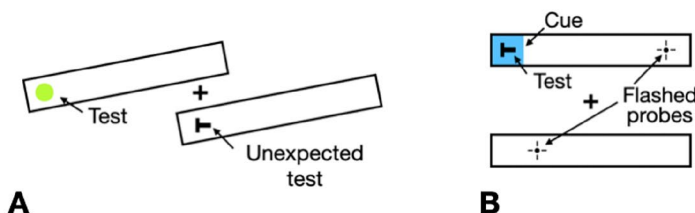

Figure B1. A. Inattentional blindness with objects shape. Participants report if the test disk is more red or more green. On the third trial, in addition to the color test, an unexpected probe is flashed either in the test rectangle or the other rectangle. There are no further trials. **B. EEG study**. Two rectangle paradigm, test always at cued location, small probe flashes test EEG response for different locations and times as in Drew et al., (2009).

In many of these studies the target is presented in advance allowing its shape to reach a high level of shape representation where it may guide the allocation of attention in a top-down manner. In the two-rectangle experiments Egly et al. (1994), the rectangles are irrelevant to the task. Nevertheless, the presence of the rectangles before the flashed cue gives the rectangles the chance to exert top-down influences. This lead time does not appear to be necessary—the same-object advantage has been found with invisible rectangles (Lin & Yeh, 2015; Zhang & Fang, 2012) suggesting that the object's involvement does not require awareness. Attentional capture by threat stimuli (Tamietto & de Gelder, 2010) is another example of object-based attention without object preview. No particular low-level feature is salient in these stimuli, instead, processing must reach the level of objects preattentively for the threat value to attract attention.

One study of multiple object tracking (Scholl et al., 2001) suggests that object processing is obligatory—paying attention to one part of an object automatically brings attention onto the whole object. This study took a standard tracking display but joined each target with a distractors linking them together with a bar. The end points of the bars moved on exactly the same trajectories as in the standard display and if an end point could be considered an object, tracking should not have been affected. However, performance plummeted, suggesting that the object is the minimum unit on which attention can operate.

Cohen and Tong (2015) and others (Al-Aidroos et al., 2012) suggest that feedback from object areas to early visual areas underlies the processing advantages seen for the attended object. This requires that the feedback to early visual areas targets the low-level features corresponding to the attended object at their actual locations. Cohen and Tong (2015) found evidence for this object-specific enhancement even when two targets (a house and a face) overlapped. They proposed that the feedback from object-specific areas (face and scene areas in particular) projects to the complex cells at the object's location. This involvement of object-selective areas is consistent with tests on patient DF who has bilateral damage to the lateral occipital area that mediates object processing (de-Wit et al., 2009). DF showed no evidence of object-based enhancement in two standard paradigms. Although DF had normal spatial orienting in Posner cueing tasks (Downing & Pinker, 1985), the cueing effects were completely uninfluenced by object structure.

While the source of the direct projections to early visual areas may arise from object-specific cortex, frontal and parietal areas may control which object representations are activated (Serences, 2004). This top-down activation would be based on task relevance or expectation.

Three systems for allocating attention

These three systems have very different modes of operation (Kravitz & Behrmann, 2011; Müller, 2014), two being spatially constrained and one nonspatial. In the discussion, we consider a variety of proposals for how the three systems may overlap (Liang & Scolari, 2020; Shomstein & Behrmann, 2006; Vecera, 1994). For example, in his 2014 review, Müller (2014) suggested that feature-based and space-based attention are guided by the same cortical networks with the difference that feature-based attention acts globally, while spatial attention is more local. However, according to Müller (2014), "very little is known concerning the direct comparison between feature- and object-based attention" (p. 133). Our next section digs deeper into the functioning of object-based attention and does uncover a strong link to feature-based attention.

How does object-based attention work?

Here we do not review the evidence for object-based attention (Chen, 2012; Francis & Thunell, 2022; Reppa et al., 2012), instead we cover the prerequisites to object-based attention. First, concerning the targets of object-based attention, we see that visual input may reach, preattentively, the level of complete object representations. Second, we examine several possibilities for the allocation of attention to objects and find that the most likely process involves projections from object areas to the locations and features of the object. This recovery of location from high-level areas has become more plausible with the recent work on autoencoders (Hinton, 2007; Hinton & Salakhutdinov, 2006; see Recovering Features and Locations section). We also find that attention can be rapidly allocated not just within an object defined by bounded contours, but also to the disconnected members of a perceptual group, matching the defining property of feature-based attention.

What is an object?

The definition of an object has been debated for a long time (Feldman, 2003; Palmeri & Gauthier, 2004; Scholl et al., 2001; Spelke, 1990). We need only think of the wide range of things that we can consider as an object (a pencil, a swimming pool, an entryway, the moon, a chair, a flame) to see the challenge. Although it is clear that no single definition will hold, we can agree that connectivity is a general property of objects: They are connected, bounded volumes with an inside and an outside. They have a shape and a three-dimensional configuration of features. These are the properties—features and their locations—that object-based attention uses to boost relevant target activity at earlier levels. Separate entities that are connected can form the parts of a larger object, like the wheels and frame of a bicycle. But disconnected entities like an orchestra, a flower arrangement or a flock of birds are groups, not objects. The

connectivity that glues together the parts of an object may be mechanical, like that for sweaters or a tractor, or simply adjacency like that for a pudding or a tornado, a flame or a cloud. While connectivity is fine as a defining property for objects in the world, objects in our perception, whether real or depicted, raise further issues because occlusions can break the object into separate parts, destroying visual connectivity. We could appeal to the connections that would have been there if the occluder were absent but to do that we need to make the best grouping of the visible parts using Gestalt principles of Good Form (known object shapes) and connectivity. But these processes just tell us that we know what objects are because we already know what objects look like. In the end, it is object knowledge that determines what is an object and this object knowledge resides in the object areas of the cortex. Interestingly, this leads us to propose a brain-based definition—objects are what object-areas learn to represent. For now, this is completely circular. Further study of responses in object-based areas may reveal actual, general properties true of all objects or may show instead that, like the categories of, say music and art, objects are a consequence of learning and the elements in the set may differ between individuals and cultures.

What are the preattentive objects that attention can select?

Whatever objects are, some neural representation of objects must exist preattentively for attention to be directed to them. According to Wolfe and Bennett (1997), preattentive processes can segment visual input into candidate objects that they called preattentive object files. Rensink (2000) proposed calling these preattentive objects, "proto-objects" and described them as "volatile units of visual information that can be bound into a coherent and stable object once accessed by focused attention" (Walther & Koch, 2006, p. 1395). Von der Heydt (2015) reported that grouping cells in area V4 are the first to represent an object. He proposed this to account for border ownership units in V2 that responded to contours based on their location relative to the object they belonged to. Franken and Reynolds (2021) recently reported columns of border ownership cells in V4 that may correspond to this object-specific representation. These early representations of grouped contours might serve as proto-objects but if so, they lack the identity and semantic information necessary for the results seen across the extensive literature of high-level unconscious processing (reviewed below). We will later propose that these early V4 units participate in the object's representation but cannot be, on their own, the preattentive representations that are required for object-based attention. Based on the unconscious processing literature we will take a very different stance and claim that the preattentive object representations can reach the level of identity and semantic connectivity and this does not necessarily change qualitatively once accessed by attention. We will then address and explain the apparently contradictory findings, consistent with Rensink's (2000) proposal, that attention is necessary to bind features together and that without attention, there are no objects.

Unconscious object representations

Research into the neural responses to unattended visual stimuli and their behavioral consequences challenge the view that attention is necessary to form object representations. Several studies have reported high-level processing in animals under anesthesia, for example, with object-specific responses of neurons in the inferotemporal cortex (Gross et al., 1972; Tanaka et al., 1991). Deeprose and Andrade (2006) reviewed studies on priming in humans during anesthesia and concluded that perceptual priming takes place in the absence of conscious awareness. There is also a broad literature in human behavior showing that preconscious and preattentive representations can reach the level of objects and words (Ansorge et al., 2014; Elgendi et al., 2018; Van den Bussche et al., 2009). Semantic priming has been demonstrated even when participants are unable to identify the prime (Yeh et al., 2012) or unaware that there has been a prime (Mack & Rock, 1998). Another piece of evidence for preattentive objects is the capture of attention by threat stimuli (Tamietto & de Gelder, 2010). For example, West et al. (2009), using temporal order judgements, found that when two faces were presented side by side, angry faces appeared first (prior entry). Low-level features cannot explain the difference in appearance order—there was no angry face advantage when the faces were inverted. The objects must reach a level of identification preattentively for the threat value to draw attention and speed its processing. Several authors also claim that the gist of a scene can be categorized rapidly without consciously identifying the objects in the scene (Greene & Oliva, 2009; Potter, 1976; review, Wolfe et al., 2011). Finally, a number of studies (Bahrami et al., 2008; Shin et al., 2009) have reported that unconscious processing can be boosted if a spatial cue directs attention to the region of the unseen stimulus. These results suggest that the downward projections of spatial attention have a generalized faciliatory effect on processing even on stimuli that never reach awareness. Although these studies, and others to be discussed below, do support extensive processing of unconscious targets, others suggest that this remains controversial as there may be contamination by weak conscious perception (Zerweck et al., 2021).

Unconscious object-based attention

If there are preattentive objects, it is reasonable to ask whether these can support classic object-based attention effects. First, a number of studies have shown that salient objects attract attention even if they are not seen (Huang

et al., 2015; Jiang et al., 2006; Lin & Murray, 2013; Lu et al., 2012). Of course, if attention is required for a stimulus to enter awareness (Cohen et al., 2012), then salient stimuli must be able to attract attention to themselves. If attention could only operate on stimuli that were already in consciousness, there would be no way for new stimuli to enter awareness. Second, there is clear evidence that object-based attention operates on cued objects even if they never reach awareness. Four studies have reported object-constrained spread of attention for objects that remained invisible. Chou and Yeh (2012) used Egly et al.'s (1994) two-rectangle method but suppressed the visibility of the rectangles with continuous flash suppression (Koch & Tsuchiya, 2007). They reported that the same-object advantage was found whether or not the participants were aware of the rectangles. Using masked two-rectangle stimuli, Norman et al. (2013) reported a significant object-based attention effect under conditions where the selected object indeed remains undetectable. Zhang and Fang (2012) also used the double-rectangle paradigm but presented them briefly at low contrast. In a control experiment, participants were at chance in reporting the orientation of the rectangles, yet the low contrast rectangles produced the classic object-specific cueing advantage that was not statistically different from that found when the rectangles were visible.

Taken together, these results provide strong evidence that the preattentive entities targeted by object-based attention can be fully developed representations of objects. They act as meaningful words or salient human figures, and their parts are unified. There is no evidence that they are processed less fully than attended objects for the purposes of establishing their spatial extent and identity. Of particular interest, the object-based advantage was found even if the object representations that guided attention never reached awareness. Clearly, if they do reach awareness, objects can

be more fully processed for the requirements of more complicated tasks, especially those requiring the scrutiny of the constituent parts of the object. For example, reporting whether the word "administer" contains an even number of vowels, or whether a book is for children or adults would only be possible once the word or the book reached awareness. For the simple purposes of object-based attention though, the evidence is compelling that the representations of preattentive objects may reach that of fully developed objects.

Is attention required to bind features into objects?

Many have argued that objects only exist once they are attended, because attention is required to put all their pieces together (Rosenholtz et al., 2012; Treisman & Gormican, 1988). According to this view, "objects" are just co-localized collections of features prior to the arrival of attention. Illusory conjunctions (Treisman & Schmidt, 1982) are evidence for this unbound state as features of one object may migrate to another under conditions of high attentional load, a case that is seen in the extreme with Balint's patients (Friedman-Hill et al., 1995). In a series of experiments, Wolfe and Bennett (1997), also reported evidence that preattentive objects are only loose collections of features. They failed to find any preattentive representation of overall shape and concluded that the representation of a visual stimulus's shape required attention to bind the features together. For example, they found that it was very hard to distinguish plus-shaped crosses made of red and green rectangles (Fig. 4, left) that differ in which color is vertical and which horizontal. Their argument is that these shapes are seen just as plus signs with some red and some green without specifying which is where, and this makes it hard to pick out the odd one unless we attend to each in turn to put the features in their correct locations. They also
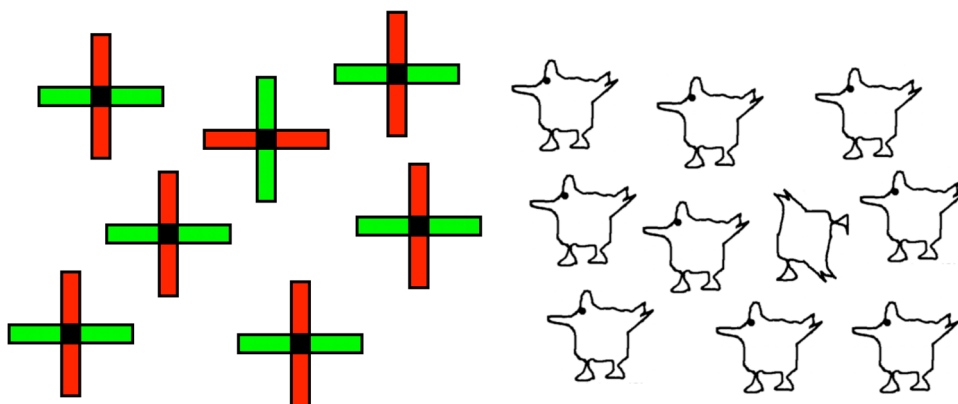

Fig. 4 Preattentive objects are feature bundles (Wolfe & Bennett, 1997). Left. The visual search for the different organization of colors on the arms of the plus signs on the left is highly inefficient (42 ms/ item). Right. Search for the non-chicken on the right is even less efficient (160 ms/item). Figures adapted from Wolfe and Bennett (1997). (Color figure online)

showed it was hard to find a non-chicken among chickens (Fig. 4, right) again indicating that putting the parts together required attention.

How can the visual search literature show that objects are just bundles of features before attention arrives, whereas the review of unconscious processing in the previous two sections suggests that objects can be fully defined preattentively? There are a number of factors that may explain this discrepancy. The first is that the search task requires the participant to report a difference between stimuli (e.g., different color assignments, or chicken vs. nonchicken; Fig. 4). Detecting the difference may be hard whether or not the representation is one of objects or bundles of features. If all the chickens are represented preattentively as chicken objects, detecting the one non-object (the jumbled chicken) is unlikely to be an easy search any more than detecting the absent diagonal feature of the *O* among *Q*s (Treisman & Gormican, 1988). Difficult discrimination is not proof that an object-level description has not yet been formed.

A second possibility is that object representations are not actually all that specific. Certainly, they cannot be a complete representation of the object's features in photorealistic detail. Perhaps some features are part of the object description without including their spatial relations to the others. So, the items in Fig. 4 on the left might represented as objects that have a plus sign shape with some red and green arms. If you want more detail, attention needs to scrutinize the object (Hochstein & Ahissar, 2002). Evidence that descriptions are only approximate is not evidence that an object-level descriptions have not been formed. An object whose shape is poorly defined

(e.g., blurry, through a mud splattered windscreen) may nevertheless be similar enough too, say, a car, to be seen as a car. It is the low-level features that are poorly registered but the configuration is sufficiently characteristic to trigger its identity as a car, a labelling with semantic properties that go far beyond those of a feature bundle.

Allocating attention to an object

Egly et al. (1994) showed that when part of an object was cued, attentional benefits were found at other locations within the object even though they had not been cued (Fig. 3). The question here is how attention was allocated to the rest of the object out to its borders. Where and how does that happen? Some authors have proposed explanations for the performance advantage within an object that invoke a spread of attention (LaBerge & Brown, 1989; Roelfsema et al., 1998). Others have mounted evidence against spreading as a mechanism. For example, attentional shifts to uncued locations may be faster within the same object, as compared with shifts to different objects (Brown & Denney, 2007; Lamy & Egeth, 2002; Luzardo et al., 2022). Alternatively, the effect of the object may be to prioritize the allocation of attention to locations within its shape (Shomstein & Yantis, 2002). In all these cases, the object's form mediates the allocation of attention. We first consider three different ways to limit the spread or shift or prioritization of attention: low-level spread; object shapes on salience maps; and control from object areas. We then examine the evidence from overlapping, moving, and imagined targets to see how these results constrain the mechanisms of allocation.
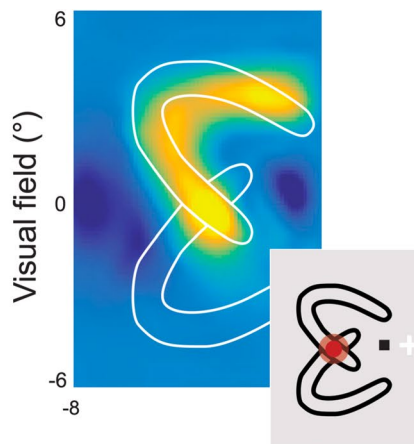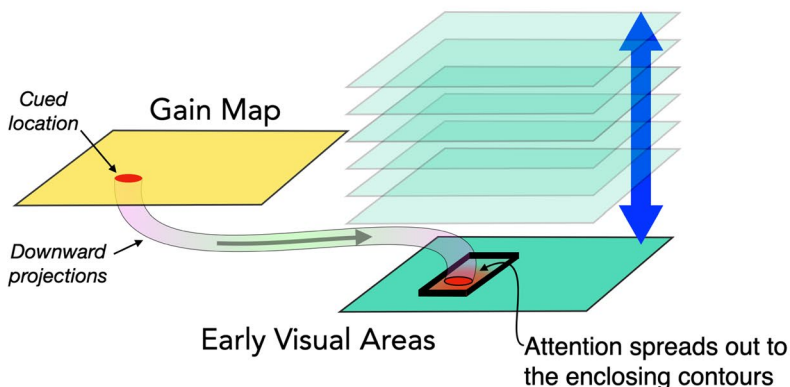

Fig. 5 Left. Gain map with boundary-limited propagation. The gain map has only local hot spots. Downward feedback to lower-level representations then spreads out to the nearest enclosing boundary. Right. Spread of attention from cued location in V1. A cue flashed at the crossing point of the two arcs where, in this case, the top arc occludes the bottom arc. fMRI activity in V1 spreads out along the top arc, not the bottom arc. From Ekman et al. (2020), with permission. (Color figure online)

Low-level spread

LaBerge and Brown (1989) among others (Roelfsema et al., 1998) proposed that when spatial attention is allocated to a part of an object, it spreads from its original focus to all parts of the object (Fig. 5, left). One explanation would be that the extra gain provided by attention can propagate over horizontal connections and that these connections are blocked once the activity, whatever it may be, reaches the object border. This is like the brush-fire or region-filling processes proposed for discovering the object's shape developed in computer and computational vision (Grossberg & Mingolla, 1985; Hojjatoleslami & Kittler, 1998; Pavlidis & Liow, 1990). This process keeps expanding the attended area across connected regions as long as the properties remain reasonably constant. Once a discontinuity is reached, say, a change in brightness, depth, or color, the process stops. As an example, Ekman et al. (2020) show how activation in V1 extends from the cued end of an arc to the opposite end, avoiding the superimposed but uncued arc (Fig. 5, right). L. Huang et al. (2022) reported a related spread of attentional benefit along a U-shaped target cued at one end. Performance improved in the uncued middle of the target before it did so at the uncued end. However, this low-level spreading proposal does not hold up as the sole mechanism for the allocation of object-based attention. For example, when one face of a cube is cued, the benefits are also found on other faces even though there is a clear edge between the two surfaces (Erlikhman et al., 2020). Moreover, attention can be allocated to one of two objects at the same location (Overlapping Targets section) and can extend across disconnected regions (Grouping and a Bridge to Feature-Based Attention section). None of these properties is possible for a region-growing process.

Attention shift or prioritization mechanisms

Indeed, several articles have presented evidence against the hypothesis that attention spreads automatically throughout the object's extent when part of the object is cued (review, Shomstein, 2012). One alternative proposed for the (Egly et al., 1994) two-rectangle task is that the object speeds the shifting of attention to another location within an object compared with a location on another object (Brown & Denney, 2007; Lamy & Tsal, 2000; Luzardo et al., 2022). A second alternative is that a higher priority is assigned to locations within the object (Lee & Shomstein, 2013; Shomstein & Johnson, 2013; Shomstein & Yantis, 2002) so that they are analyzed first. For example, once the object is cued, any search for a probe element would prioritize locations within the object rather than elsewhere. Both of these alternatives focus more on the mechanism by which attention accesses locations rather than assuming some uniform, automatic distribution of benefits throughout the object. Importantly, rather than discovering its shape by spreading activation out to the object's borders (LaBerge & Brown, 1989), these alternatives both depend on an existing representation of the object's shape to prioritize access to locations within the object.

The proposal of prioritization is an alternative to gain as a mechanism for attentional benefits—rather than boosting gain, attention would boost priority. In this case, an object's features at the early levels of visual cortex would be selected first for further processing. This is equivalent to V1 acting as a salience map where the salience was not computed locally but allocated to the object's locations by downward projections from object areas. When we say that attention boosts activity, it can be taken either as classic gain—boosting actual neural activity—or boosting priority. However, priority may be realized by increasing neural activity so the two proposals may be the same thing. Whatever the case, these alternatives depend on downward projections from object-specific areas to facilitate access to locations within the object. The following sections examine where the object's shape is represented and how it mediates the projection of attentional benefits to early visual areas.
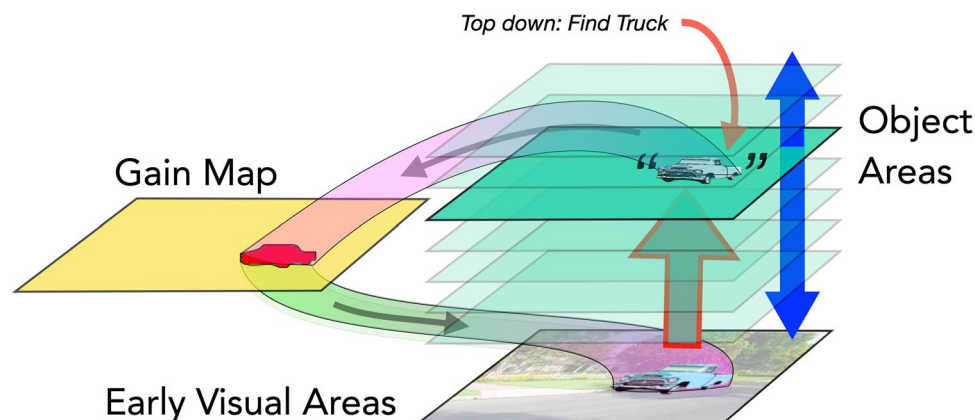

Fig. 6 Gain map with object-shaped activation. The shape of the target object is projected onto a retinotopic gain map such as the superior colliculus. Downward feedback then projects to the appropriate object location in early visual areas mirroring its shape. (Color figure online)

Object shaped activation on attentional gain maps

Is the allocation of attention guided by object-shaped activations on gain maps (Fig. 6)? If this were the case, then the downward projections that carry the attentional boost to low-level areas would fill the whole region of the object shape and directly mediate object-based attention—as suggested by Knapen et al. (2016). If this were the case, there would be a few issues to consider. First, gain maps like those in posterior parietal cortex (PPC), frontal eye fields (FEF) or subcortical areas (superior colliculus [SC]) have large receptive field sizes and so lower spatial resolution than early visual cortex. As a result, the projections from these areas to early visual cortex would cover a blurry version of the object. Attention does have low resolution (He et al., 1996; Intriligator & Cavanagh, 2001) but there have been no studies yet that explicitly look at the resolution of object-based attention measuring, for example, the profile of attentional benefits as the probe location moves from inside to outside the cued object. Second, the selection of an object not only requires enhancement of the appropriate region of space but also enhancement of the subset of features that define the object. This is critical when selecting one of potentially several overlapping (e.g., partially occluded or transparent) objects. Since the gain maps in PPC, FEF and SC are two-dimensional, retinotopic representations, they have no way to deal with overlapping objects on a purely spatial basis, an issue we address in Overlapping Targets section, below. Lastly, there would need to be some way to get the object shapes from object areas onto the salience or gain map at the right location. This is a nontrivial requirement which

we will consider in the next section and in the Anatomical Basis section.

Control from object areas

If there were object-shaped activations on the salience or gain map, where do they come from? Perhaps they are projections from ventral temporal, object recognition areas (LOC through FFA), a reasonable conjecture given that we are looking for object-based attention (Al-Aidroos et al., 2012; E. H. Cohen & Tong, 2015). Figure 7 outlines one such proposal, where input from early cortex, combined with projections (shown as the top-down red arrow in Figs. 6 and 7) from frontoparietal attention network (Liu, 2016; Serences et al., 2004; Shomstein & Behrmann, 2006) activate the appropriate object units (representing a pickup truck in this example). This high state of activation in the object area is equivalent to a gain (or salience) map, although it has only crude spatial organization. If the active object representation exceeds some threshold, it can then project directly back to early visual cortex (Fig. 7). The representation of space in the ventral areas has been characterized as rudimentary (Op De Beeck & Vogels, 2000; Sereno & Lehky, 2011). Nevertheless, there are a number of retinotopic representations within the posterior extent of the ventral temporal cortex (Arcaro et al., 2009); moreover, position information can be decoded from population responses in regions like LO, FFA, and PPA (Carlson et al., 2011; MacEvoy & Yang, 2012; Majima et al., 2017). Clearly, the recovery of the object's location to guide projections is
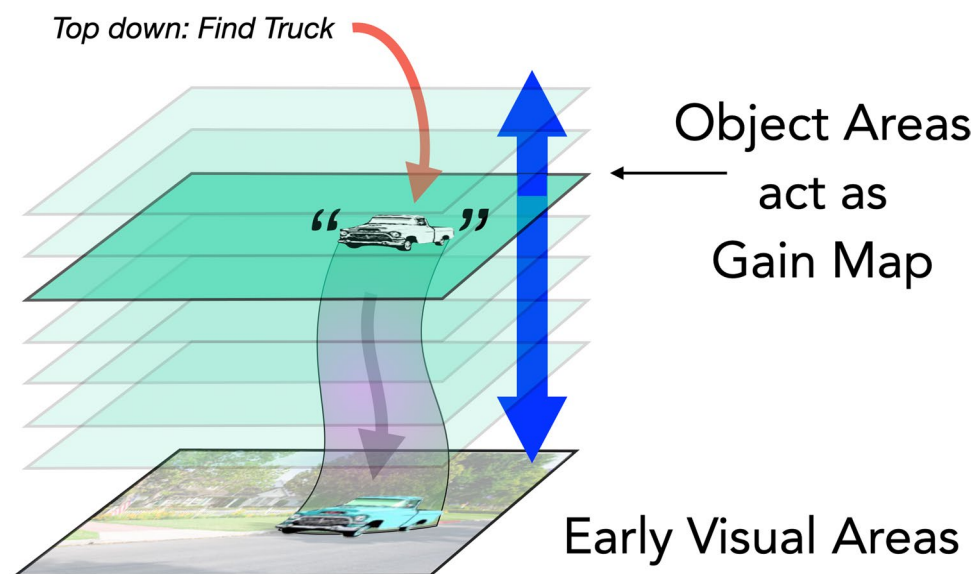

Fig. 7 Object areas project directly back to early visual areas. In object areas (and word areas) current input, aided by a top-down cue, activates the object representation. That representation preserves information about the object's approximate shape and location suf-

ficient to mediate enhanced processing within the boundaries of the shape in early areas of visual cortex. This is implemented here through re-entrant pathways directly back to early cortex. (Color figure online)

non-trivial, one that we address in the Recovering Features and Locations section, where we introduce autoencoders (Hinton, 2007; Hinton & Salakhutdinov, 2006) as a mechanism capable of this recovery. Specifically, to disentangle overlapping objects (Overlapping Targets section, below), downward projections that target the early representation of the object (Fig. 7) must be both feature and location specific (Recovering Features and Locations section). In this case, they could target one of two overlapping objects based on feature differences (E. H. Cohen & Tong, 2015; O'Craven et al., 1999).

The next sections deal with findings in object-based attention that allow us to evaluate these three allocation mechanisms.

Which allocation mechanism is it?

Here, we present evidence from overlapping, moving, imagined, and grouped targets that suggests that attentional allocation originates in the object areas of ventral visual cortex. Moreover, the results from attention allocated to groups, in the Grouping and a Bridge to Feature-Based Attention section, below, will suggest that feature-based attention and object-based attention may depend on the same mechanisms.

Overlapping targets

The findings that attention can target one of two overlapping objects (Duncan, 1984) is critical in understanding the process of selection. Even if location information could be recovered from activity at higher levels, location alone would not be sufficient to guide attention in this case. To do so, the selection process must be able to localize the object's features. In Duncan's original study on object-based attention, the two objects were only partially overlapping (Fig. 2).

To address this, Valdes-Sosa et al. (2000) and Mitchell et al. (2004) used fully overlapped motion fields (Fig. 8). They again found an object-based benefit: Selection of one plane improved the detection of a second event on the same plane, but not on the other, even though both planes were completely overlapped. These experiments might be seen as tests of feature-based attention rather than object-based attention, but even so, equivalent findings have come from fMRI studies of attention to overlapping objects, most frequently faces and houses (Cohen & Tong, 2015; O'Craven et al., 1999). So, the ability to select one of two overlapping objects is robust. How can attention be allocated to only one of two overlapping objects if the controlling map (Itti & Koch, 2001) and the receiving area were both 2D retinotopic representations? Certainly not by lateral spread on the gain map, or at the level of the activation in early visual cortex
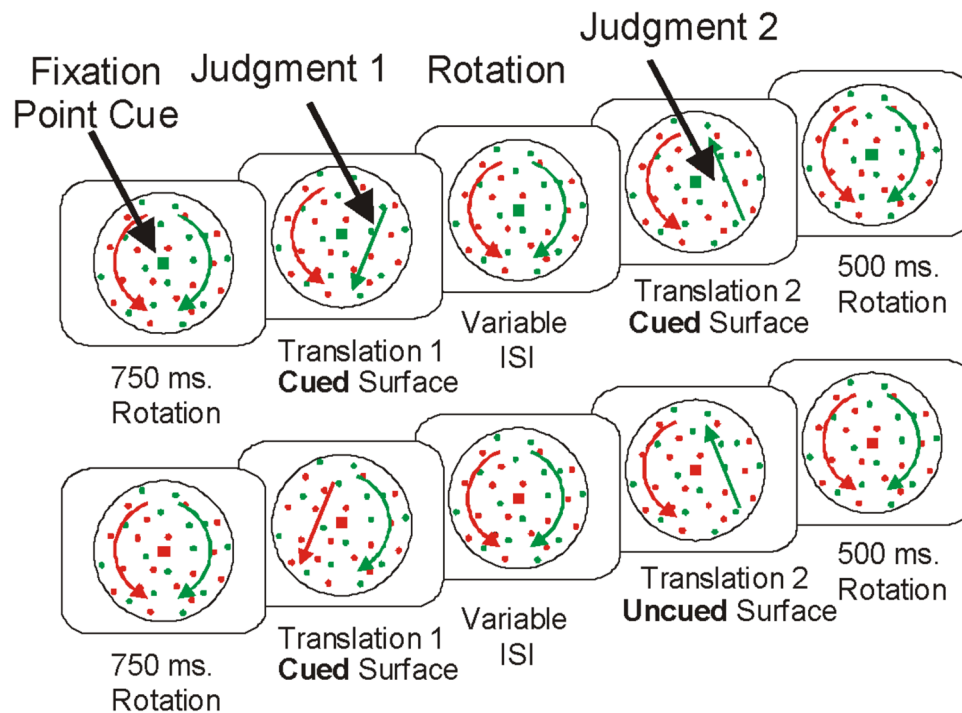

Fig. 8 Completely overlapping surfaces. Two fields of dots rotated in opposite directions. Participants reported the directions of two linear translations which were either of the same surface (top row) or of different surfaces (bottom row). Performance was better when the second translation was on the same surface as the first (Mitchell et al., 2004; Valdes-Sosa et al., 2000). From Reynolds et al. (2003), with permission. (Color figure online)

where the representations of the two objects are intermingled. One possibility is that depth or object-defining features could disambiguate the two targets. This would require that the gain map, say the superior colliculus or LIP, differentiates depth planes and/or features. There is only meagre evidence of feature specificity in areas that might be controlling attention (e.g., some preferred motion cells in LIP; Fanini & Assad, 2009) and little or none of depth specificity (Sparks, 1999). The one remaining possibility is that the downward projections to early cortex are from object areas (Cohen & Tong, 2015)—these projections can represent object features including depth. In the Cohen and Tong (2015) study, for example, the projections could be to units that code facial features versus house features. The findings with overlapping stimuli require that downward projections from object areas can target not only the object's location but also its features, a substantial challenge. This ability to project to the locations and properties of a target object is one of the key properties of autoencoders that make control from object areas a viable proposal (see Recovering Features and Locations section).

Cueing of moving objects

If an object is cued, the attentional benefits are available throughout the object (Egly et al., 1994) and its features (O'Craven et al., 1999). But what happens if the object moves before the probe is presented. Would the attentional enhancement remain at the original location of the object or follow the object to its new location? Several studies show that the benefits do move with the cued object and do not remain at the actual location of the cue. For example, Lamy and Tsal (2000) and Gonen et al. (2014) presented a pair of rectangles and cued the end of one of the rectangles. The rectangles then moved to new locations and the benefit for the uncued, within-object location moved with the cued rectangle to its new location. These results are in many ways equivalent to those of the multiple object tracking task (Cavanagh & Alvarez, 2005; Drew et al., 2009; Pylyshyn & Storm, 1988) where attention directed to an initial set of targets follows these targets around until the end of each trial. The visual system deploys attention to where the cued target is now, not to where it was, so that attention, once directed to an object, remains with the object. fMRI studies have shown that frontal and parietal areas are involved in keeping track of the attended targets in the multiple objects tracking task (Culham et al., 1998, 2001; Howe et al., 2009). This supports the claim that object selection is guided by frontal and parietal areas which then must engage the representations of the selected targets in object areas. As indicated by the overlapping target findings above, only object areas have both the location and feature information required to accurately direct attention when objects are superimposed.

Imagined targets for object-based attention

If attention is directed from representations in object areas, then we might expect that object-based attention effects would follow any activation in these areas, even in the absence of sensory input. This appears to be the case. First, we can organize ambiguous patterns into specific shapes by attending to elements in the structure that conform to the shape, which can be suggested verbally (Attneave, 1971; Huang et al., 2007). We can also impose a mental image of a shape or letter on a 3×3 or 4×4 grid and find attentional benefits for probes that fall on squares that are part of the imagined shape, as opposed to squares that are not part of the shape (Farah, 1989; Ongchoco & Scholl, 2019). Finally, results from fMRI studies show that attention to a target's location activates early cortex at the expected location even before the stimulus is present (Kastner et al., 1999) and, critically, the activation has the shape of the expected object (Silver et al., 2007). Because object-shaped effects are observed when no object is present in the image, the shape of the attentional enhancement must be driven by the downward projections themselves as opposed to an allocation out to object boundaries in early cortex (Ekman et al., 2020).

Grouping and a bridge to feature-based attention

Does object-based attention operate on groups—if attention is cued to one element in a group, is it also allocated the remaining elements? Grouping may be an initial stage in putting together an object's many adjacent bits and pieces but sometimes the pieces are not connected even though they are strongly grouped. Although connectedness is a basic property of most objects, it is not a property of groups, like a flock of birds, or a line of chorus dancers. Matsukura and Vecera (2006) suggested that object-based attention can select multiple-region objects, provided that the regions follow perceptual grouping cues. Driver and Baylis (1998) also proposed that selective attention acts similarly on objects and on unconnected elements grouped through adjacency or feature similarity. If we accept that groups can be the targets of object-based attention, then groups, like objects, must also exist preattentively. Indeed, there is strong evidence that grouping occurs in the absence of attention (Lamy et al., 2006; Moore & Egeth, 1997).

Does attention extend across the members of a group? The first evidence comes from studies where the standard adjacent rectangles were divided in two by an occluding bar (Haimson & Behrmann, 2001; Moore et al., 1998). There was no continuous, within-object space over which attention could be allocated at early levels as in Fig. 7. Nevertheless, the object advantage was still significant. Additional evidence comes from grouped items that have no connections at all. Rafal (1994) replicated the object advantage reported by
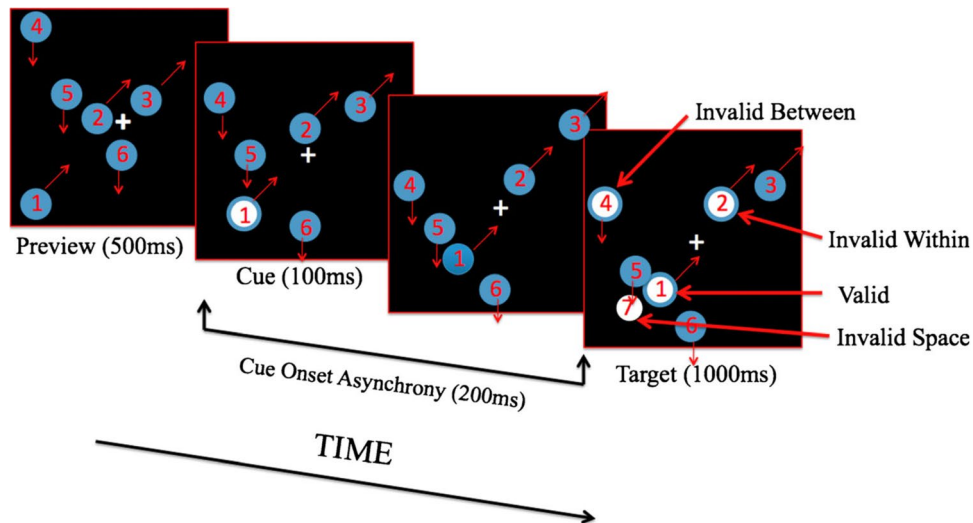
Fig. 9 Attention extends to elements that are grouped by common motion. One group of elements is moving up to the right (Disks 1, 2, and 3) while a second group moves downward (4, 5, and 6). One disk in the first group is flashed as a cue and after 200 ms either that same disc or another member of its group, Disk 2, is flashed at their new locations. Both these tests showed a reaction time advantage. For comparison, a disk from the other group may be flashed or the originally cued location, 7. None of these locations showed an advantage. From Gonen et al. (2014), with permission. (Color figure online)

Egly et al. (1994) using two groups of circles arranged into parallel rows or columns instead of using two rectangles. However, they did not test whether the attentional advantage was confined to the discs or filled the spaces between them. To answer this question, Gonen et al. (2014) created two groups of three discs where each group had a common motion direction (Fig. 9). One group was cued by briefly flashing one of its discs (second panel, Fig. 9) then 200 ms later a test was flashed and participants had to respond as quickly as possible. The cued disc (white Disc 1 in last panel of Fig. 9) had moved on but the largest benefit was seen when the test flash was at this new location. Importantly, a probe on a second disc within the cued group also showed an advantage (Disc 2, Invalid Within, in last panel, Fig. 9). The response time for probes on discs in the other group (Disc 4, Invalid Between) or on empty space (Location 7) were at baseline. These results show that an attention benefit can cross empty space to enhance processing of other group members without enhancing the processing of all locations falling within the bounds of the cued group.

Other studies have looked at the allocation of attention across the characters within a word and found a similar within-object advantage, where the word is now the object (Li & Logan, 2008; Yuan & Fu, 2014). The two characters
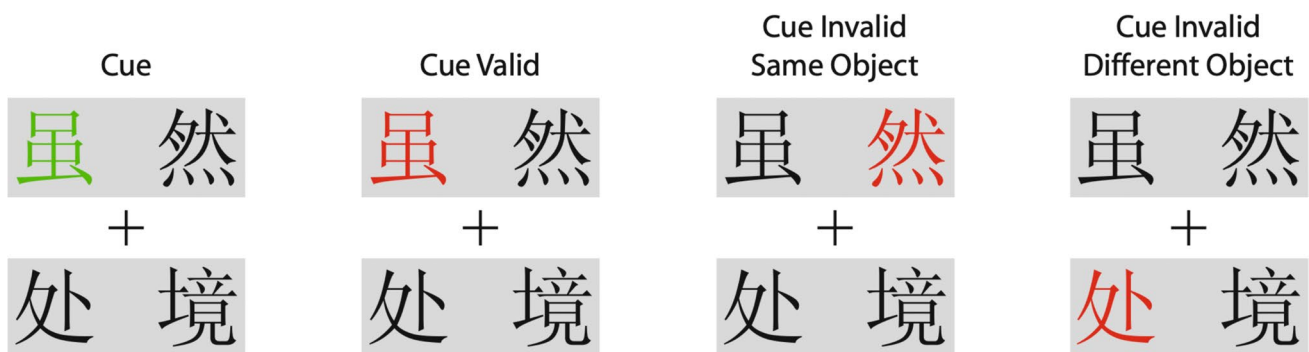

Fig. 10 Lexical object-based attention. The pairs of characters form words in Chinese. The greyed rectangles show which pairs are words. In this case, they are read left to right, but characters can also be read top down, so both organizations were used in the experiment. The two words here are "although" on the top and "situation" on the bottom). The grey rectangles were not present in the experimental display. Participants pressed a response key as soon as they saw a red character. The invalid cue, same word conditions did show a reaction-time advantage. From Li and Logan (2008), with permission. (Color figure online)

on top in Fig. 10 make the word *although*, the bottom word is *situation*. This has been called lexical object-based attention. Zhao et al. (2014) reported attention benefits also extend to learned pairs of patterns. Unlike Gonen et al. (2014) none of these studies tested whether the attentional advantage existed between the elements of the group.

These results showing that object-based attention is allocated to the members of groups suggest that object-based attention may account for the properties of feature-based attention: the allocation of attention to disconnected elements that share the same features or part of the same word independently of their locations. Indeed, some authors propose that all objects are just local spatial groups (Vecera, 1994) and others have already noted that groups and objects can have the same status for selective attention (Baylis & Driver, 1992; Driver & Baylis, 1989; Kasai et al., 2011).

If feature- and object-based attention arise from the same system, what is the underlying mechanism? Feature-based attention appears to depend on global, feature-specific activation (Boynton et al., 2006; Liu & Mance, 2011; Serences & Boynton, 2007). If the mechanism underlying feature-based attention is global, feature-specific activation (Serences & Boynton, 2007), one consequence of feature-based attention will be seen when there are several items that share a feature—a group. But the grouping is not the mechanism mediating feature-based attention, it is a byproduct. In the case of object-based attention we suggest a similar global, object-specific activation that boosts the input signals of any matching object at any location in the visual field (see Box 2). When the to-be-attended target is a single feature, this mechanism reduces to the properties of feature-based attention.

Box 2: Outstanding questions.

1. Object-based attention is proposed to generate feature-based attention when it is directed to groups. What features and objects support this grouping and carry attention to the disconnected elements of the group? Attention to a face in the left hemifield, for example, appears to extend to a task irrelevant face in the other hemifield (Störmer et al., 2019)? Would these benefits generalize to other faces, to the word "FACE"?
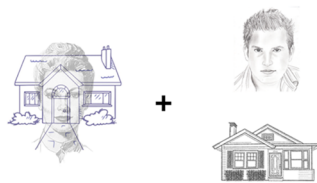

Figure B2. MRI or SSVEP study. A superimposed house/face stimulus on the left. Participant has to pay attention to either the house or the face (e.g., count its changes of hue). In the other field are task irrelevant houses or faces. Will the attention on the left modulate activity on the right?

2. Does object-based attention have suppressive surrounds for reduce sensitivity to targets that are nearby in object space, no matter where they are in the image?

3. How does autoencoder learning develop in the visual system, if it does? Would young children have less accurate feedback to object targets?

4. If space-based attention can be controlled from both object and saccade areas, could these two areas target different locations in the case of a position illusion that affects object perception but not saccades (e.g., the double-drift illusion, Lisi & Cavanagh, 2015)?

5. How do the fronto-parietal networks tag possible target objects and activate their representations in high-level object selective cortices? Is this related to the fronto-parietal representation of targets in multiple object tracking (Culham et al., 1998, 2001).

The claim that feature-based attention is just object-based attention applied to groups is plausible, but there are two factors to consider. First, the experiments we covered in this section could easily be regarded as experiments on feature-based attention, rather than object-based attention. This is more of a question about semantics because the experiments could be framed in either category without much difficulty and, to an important extent, this emphasizes the overlap between the two systems when dealing with groups as objects. The second factor is more of a practical point. If the allocation of attention over an object and the allocation of attention across disconnected group members both rely on common mechanisms, do they also originate from a common cortical area? Are groups represented at the same level as objects? Groups and objects differ in that a group is defined by its common element not by its overall shape. In contrast, shape is one of the key features defining an object. The point of the group is to reveal the spatial extent filled by the grouped elements as input to further scene segmentation. It is not yet known if different areas represent groups and objects, but it would not be surprising. Even if they are at different levels in the cortex, we may assume that similar processes generate the downward projections from both areas to early cortex that confer attentional advantages.

Recovering features and locations

The big challenge for object-based attention lies in directing the downward projections to the target objects and their features. How can the higher level representation of the object or group feed back to the locations and features that define the object in early visual areas? After all, the goal of higher level representations in object areas is to code for abstract properties and identity independently of size, orientation, pose or location (Logothetis & Sheinberg, 1996; Tanaka, 1996). However, more recent single cell studies have shown more sensitivity to position than was once thought (DiCarlo & Maunsell, 2003). Human functional magnetic resonance imaging (fMRI) studies also show that object selective areas are sensitive to position (Carlson et al., 2011; MacEvoy & Yang, 2012; Majima et al., 2017). These studies suggest that although position information is reduced for single neurons with large receptive fields in ventral areas of the visual cortex, it is retained at the population level. This is promising, but retaining some position information across the neural population in object coding areas does not reveal how this is used to target the source object in early cortex.

The development of autoencoder networks (Hinton, 2007; Hinton & Salakhutdinov, 2006) provides some insight into how a recovery of specific location and feature information might happen. Recent work in the perception of object properties, such as gloss, demonstrates the power of this unsupervised learning process in deep neural network autoencoders architectures (Fleming & Storrs, 2019; Storrs et al., 2021). The key to this architecture is the goal of matching the input and output states. The autoencoder first takes the high-dimensional input and compresses it through a series of layers into a low dimensional, coding level or latent value level (on the right in Fig. 11). This is then re-expanded through several decoding layers to reconstruct the original
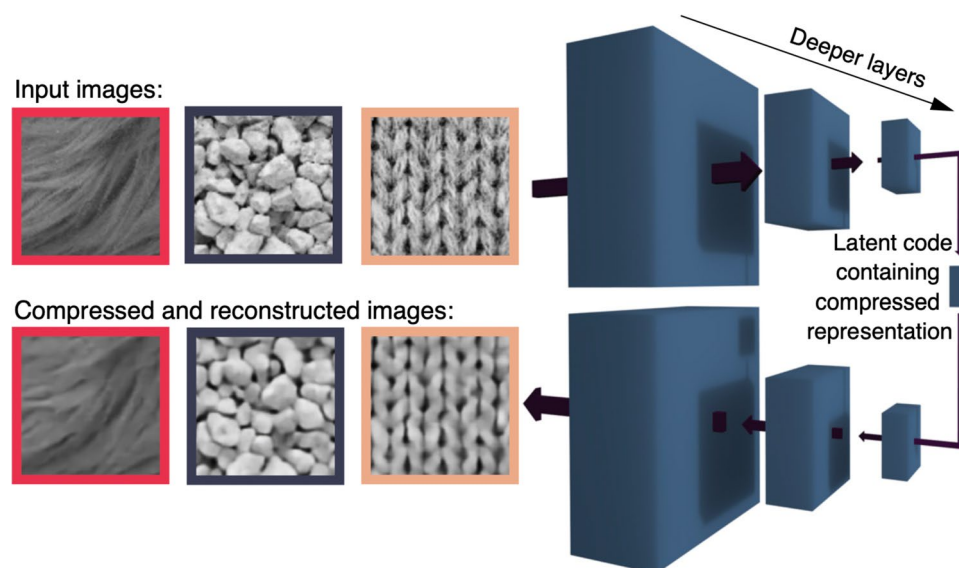

Fig. 11 Autoencoders. The network has a series of layers with fewer units at each level on the way up (encoding) and increasing numbers of units going back down (decoding). The network reconstructs the input as its goal. Once achieved, the reconstructed activity matches the input in locations and features and the topmost layer captures a low-dimensional representation of the input. From Fleming and Storrs (2019), with permission

in a high-dimensional output layer. The network connections are adjusted until the output activity matches the input, a form of unsupervised learning. In a neural implementation, we assume that the low-dimensional, latent values are the high-level, object representations. Furthermore, the only way to have comparable, high-dimensional, input and output layers is for them both to be in the same areas of cortex. This generates a closed loop from early retinotopic areas up to object areas, and then back to early retinotopic areas. As a very useful bonus, when a particular object's representation is active on the latent value level, projections from that level will target the locations and features that belong to the object, despite the impoverished representation of location at the object level. Seen in this way, the ability to modulate input activity either by a complete encoding to decoding sweep through the system, or through top-down activation of the learned latent space (object) nodes even in the absence of sensory input, can serve a second purpose: object-based attention. It is noteworthy that this property emerges not specifically to support object-based attention, but rather as a byproduct of learning to efficiently internalize representations of the external world.

Autoencoders have been used to model the processing of visual information in humans for the perception of material properties (Fleming & Storrs, 2019; Storrs et al., 2021) and for spatio-temporal and chromatic contrast sensitivity (Li et al., 2022). The results of these modeling efforts are descriptive—they can show high-level representation of image and scene properties. In a more general application to the visual system for object recognition (Han et al., 2019; Le, 2013), the higher-order, low-dimensional representation is specifically intended to be the set of familiar objects that the object area can recognize. Convolutional autoencoders have also been developed to deal with issues of encoding targets across variations of location, size, and orientation (Hinton et al., 2011; Le, 2013; Oyedotun & Dimililer, 2016).

We use autoencoders here as a placeholder or proof of existence of some network configuration that can recover the location and feature information from population level activity, filling the missing link for the operation of object-based attention. Proposing that autoencoders fill this role does not yet provide us with any novel predictions for behavior or physiology that are specific to autoencoders. They require projections from early visual areas to object-specific areas and then projections back to the early areas to target object features, as would any architecture for visual processing and object-based attentional control. There are two main differences between autoencoders and other hierarchical neural architectures. First, the autoencoder has its most compact representation (fewest units) in the middle followed by decoding to a high-dimensional representation. In typical neural nets, like those focused on recognition (e.g., AlexNet, GoogleNet), it is the final level that has the fewest units.

And second, in typical neural nets, the final compact layer is where error signals are extracted between the actual output and the desired output. In contrast, autoencoders compare the high-dimensional input and output representations and any deviation between them is used to drive the evolution of the network. In the neural implementation, the input and output representations must both be in the same early visual areas in order to make this comparison—there are no other areas that are of sufficiently high dimensionality. This comparison operation relies on the projections, feedback, from the high order representations (object areas). Typical neural nets do not have this feedback from the highest level to the input layer for error extraction and so they do not provide the possibility of targeting object features and locations that an autoencoder architecture does.

Finally, this autoencoder approach has close links to the predictive coding approach (Friston, 2010; Marino, 2022) where the brain learns to predict early activity from higher level representations by reducing prediction error. In both cases, downward projections from higher level representations of objects can target the locations and features of the object. Despite this similarity, predictive coding acts to suppress information predicted by higher level areas—whereas attention's goal is to facilitate its processing.

Computational models

We have proposed that autoencoders fill the missing link between object areas and the target object's low-level features and locations. What other advances have been made in computational modeling of attention and specifically, object-based attention? Tsotsos et al. (1995) modelled attention as a selective spatial filter that starts with the most salient upper-level unit and suppresses lower units that do not contribute to it. Later work extended the selection process to include features and objects (Tsotsos et al., 2016). Computational models generally take a biased competition approach, identifying high salience locations or objects of interest (Rao, 1997; Sun & Fisher, 2003; Wu et al., 2004; Yu et al., 2015) and then prioritizing their processing through a hierarchy of processing levels while suppressing activity that is not contributing to the salient location or object or feature. The emphasis on object-level representations matches our focus on object-specific areas as the source of object-based attention. There is a great deal of overlap between our proposal and computational modeling on the conceptual level—for the obvious reason that both approaches are attempting to explain the same performance properties. Although computational models have captured many of the properties of object-based attention, they have not yet revealed any properties that were not already known. In contrast, our architecture for object-based attention makes the following novel points: (1) the architecture unifies feature-based and object-based attention; (2) it identifies the

anatomical sites for different components of object-based attention; (3) it reveals how autoencoder networks can solve the feedback specificity problem; (4) it allows for the global projection of object-based benefits; (5) it demonstrates the importance of preattentive, identity-level object representations; and (6) it predicts suppressive surrounds in object space. Nevertheless, there will clearly be important, further advances in computational modeling of attention, some of which may follow from our proposals here.

Discussion

How does it work?

Our review has led to the following summary of likely properties for object-based attention.

1. Frontoparietal attention networks may activate a given object representation in object areas of cortex as the target for attention. This top-down activation would be based on task relevance or expectation. Alternatively, the object representation in object areas of cortex might be directly activated by bottom-up salience or cueing of part of the object.
2. Once this representation of the to-be-attended target object is active, it is a fully realized, preattentive object representation, not just a co-localized bundle of features. The object is then a target for attention with downward projections attempting to facilitate its processing (Silver et al., 2007). If it is present, the downward projections reach their target and the object is then attended.
3. These projections from the object representation to early visual cortices (and LGN) provide the boost required to improve the object's processing. Some autoencoder-like network learning enables these projections to accurately target the locations and features of the object at early levels.

What is new here? Based on the literature we reviewed, we first conclude object-based attention is controlled by activity in the object-selective areas of ventral visual cortex. This differs from some proposals of spread guided by object boundaries in early cortices (LaBerge & Brown, 1989; Roelfsema et al., 1998), but it is in line with the conclusions of fMRI studies (Cohen & Tong, 2015) where one of a pair of superimposed objects could be selected by attention. The critical point of these studies, as well as those testing superimposed moving surfaces (Mitchell et al., 2004; Valdes-Sosa et al., 2000), is that only object areas register the conjunctions of the features and their locations for a given object that would allow the targeting one of two superimposed objects. Neither the features

alone nor locations alone would be enough to separate, for example, the superimposed houses and faces of Cohen and Tong (2015) study, and the other candidate areas in frontal, parietal or subcortical regions lack sufficient feature specificity to fill this role.

Second, we suggest that preattentive objects can be fully developed rather than unbound, feature-bundles. Since object-specific units in cortex have large receptive fields, it is possible that only relatively isolated target objects can preattentively activate the appropriate object-specific units.

Third, object areas have only crude retinotopy at best so that the required information of both location and features must be recovered from across units in the object area and elsewhere. We pointed to autoencoder-like, network learning (Hinton & Salakhutdinov, 2006) in the visual cortex to accomplish this recovery. Although plausible, the extent to which these autoencoder-like properties are realized in the brain remains to be demonstrated.

Finally, the finding that object-based attention can be directed to groups shows that object-based attention can be the source of feature-based attention. In 1989, Driver and Baylis noted that attention can be directed to perceptual groups whose elements are spatially separated. They used the flanker task (Eriksen & Eriksen, 1974) and found that distant flankers could influence performance more than nearby flankers if they grouped with the target (same motion, same color). Indeed, several articles suggest that groups have the same status as objects as targets of attention (Baylis & Driver, 1992; Driver & Baylis, 1998; Duncan & Humphreys, 1989). A consequence of this attention to groups that goes unmentioned in these articles is that it offers an explanation for feature-based attention. Specifically, since the group elements can be widely spaced and attention is directed only to the group members and not the space between them (Gonen et al., 2014), this creates the signature aspect of feature-based attention, a nonspatial, feature-specific allocation. We suggest a similar broadcasting of object-specific activation to all locations in the visual field. Störmer et al. (2019) have already demonstrated this property for faces. Specifically, when a face and a house were alternated at the same location and the participants attended to the face, face-specific evoked responses (N170) were found when task irrelevant faces were presented elsewhere in the visual field. It is not yet known whether this nonspatial allocation might be found for other stimuli nor whether it would generalize to other variants of the attended stimulus (Box 2). These downward projections from an active object representation may only be triggered by the presence of a matching object (anywhere in the visual field, as in Störmer et al., 2019)—its upward projections are received and generate downward reinforcement specifically to that object via the autoencoder network. This proposal does not include

downward projections that are broadcast globally across the visual field, in effect searching for a match, as feature-based attention projections do. However, we cannot rule out that the downward projections are broadcast globally, despite how challenging the implementation would be. Indeed, such global, object-specific broadcasting would account for the reports of the decoding of a peripheral target identity from foveal ROIs when no stimulus is present there (Williams et al., 2008).

Are there parallel systems of attentional allocation and selection?

There have been many proposals to unite space-based and object-based attention. Vecera and colleagues (Mozer & Vecera, 2005; Vecera, 1994) and others (Logan, 1996; Shomstein & Behrmann, 2006) argued that objects were groups of locations so that space-based attention was simply object-based attention directed to a group that had only one element. However, these attempts to join space- and object-based do not avoid the necessity of object-level information. Scolari et al. (2014) have suggested that feature- and object-based attention are on the same continuum. Here, we extended their proposal based on the object-based attention effects seen for grouped elements (Gonen et al., 2014). Although the properties of all three systems can be seen in object-based attention, this does not mean that all attentional allocation is controlled by this one pathway. As mentioned earlier (Friedman-Hill et al., 1995), there are undoubtedly many "gain maps" that enhance performance throughout the visual system. We cannot say whether the object-based pathway dominates in the role of attention, in fact, the small effect size in the Egly et al. (1994) paradigm (Francis & Thunell, 2022) raises the question of whether it may play only a minor role.

This proposal for combined systems, while promising, runs into several strong counterarguments, especially from the patient literature. For example, patient DF has bilateral damage to the lateral occipital area that mediates object processing (de-Wit et al., 2009) and shows no evidence of object-based enhancement in two versions of the Egly et al. (1994) task. This is consistent with the control of object-based attention from object areas. DF nevertheless did show normal spatial orienting in a Posner (1980) cueing task. Finsterwalder et al. (2017) reported a patient with a thalamic lesion who was impaired in feature-based selection but not spatially based selection. Many behavioral studies also differentiate among the three systems. For example, Chou et al. (2014) used the external noise method to show that space-based attention operated through signal enhancement whereas object-based attention operated through noise exclusion. It appears that the allocation of attention through the object areas is not the only system available. As

described in the beginning of this article, the evidence for the allocation of spatial attention through activation in the saccade system (areas FEF, PPC, SC) is quite compelling (Awh et al., 2006). Whatever the merits of mediating all three types of allocation through the object areas of cortex, this must be occurring in parallel with a purely space-based system that shares processing stages with saccade control.

Evidence for cooperation between separate attention systems comes from Donovan et al. (2017) who demonstrated that spatial attention was necessary to see object-based effects in temporal order judgments, implying an interaction between the two systems. The classic Egly et al. (1994) two-rectangle stimulus also shows evidence for the activation of multiple types of attention. Hollingworth et al. (2012) and (Müller & Kleinschmidt, 2003) reported gradients of performance benefits within the cued object. Although this could be due to incomplete allocation of attention throughout the object, it could also be the result of a focused space-based attention that drops off with distance from the cue, unaffected by the object's shape, added to an object-based attentional benefit that has uniformly covered the full object shape. Kravitz and Behrmann (2011) also reported combined enhancement across all three allocation systems when the stimulus was appropriate for all three.

Anatomical basis

Throughout the review, the source of the object-specific feedback has been identified only as coming from object-specific areas of ventral visual cortex but several object-selective areas may be involved. For one, the representation of objects is associated with activity in the lateral occipital (LO) cortex (Kourtzi & Kanwisher, 2001; Malach et al., 1995). But Cohen and Tong's (2015) results with overlapping faces and houses also show activity in face-specific areas (fusiform face area, FFA) and place areas (parahippocampal place areas [PPA]) among others as correlated with the attentional selection in early cortices (V1–V4). The visual word form area (Dehaene & Cohen, 2011) (VWFA) may serve this same function when words are the targets of attention (Li & Logan, 2008). One of the earliest areas that may have object-specific units is V4 (Franken & Reynolds, 2021; Von der Heydt, 2015). An object representation may therefore be a widespread activation across several of these areas.

There is substantial evidence from fMRI studies demonstrating the role of frontoparietal networks in specifying, top-down, which object representations are active. For example, Müller and Kleinschmidt (2003) used a whole-brain analysis of the BOLD response during a two rectangle cueing task and revealed strong activation in parietal and frontal regions as well as LO. Baldauf and Desimone (2014) reported that prefrontal cortex appeared to be a common source of top-down

biasing signals for space-, object- and feature-based attention. FEF supplied signals for spatial attention while the inferior frontal junction (IFJ) supplied signals for object or feature attention. Serences et al. (2004) reported that posterior parietal and frontal regions were transiently active when attention shifted between spatially superimposed objects. Using event-related MRI in a two-rectangle cueing task, Shomstein and Behrmann (2006) showed that object-selective signals from the posterior parietal cortex control shifts of attention between object locations. Liu (2016) used the superimposed grating stimuli developed by Blaser et al. (2000) where participants are required to follow one of the two patterns with attention while both continuously changed color, orientation and spatial frequency. Liu (2016) was able to decode the attended pattern in several frontoparietal areas including the anterior intraparietal sulcus, frontal eye fields, and inferior frontal junction. Güldener et al. (2022) found that the right frontopolar cortex and other attention-related areas represented a novel stimulus prior to its arrival in awareness. These frontoparietal areas would project separately to the different object areas (e. g., LO, FFA, PPA). This hypothesis is further supported by recent findings of source-localized EEG data showing dynamic interactions between object-representations within the ventral-temporal and posterior-parietal cortex (Gurariy et al., 2022). One example of a potential projection pathway for these interactions comes from Takemura et al. (2016) who combined fMRI, diffusion MRI, and fiber tractography to show that the vertical occipital fasciculus (VOF) links dorsal and ventral visual cortex, connecting visual field maps between the two. These authors suggested that the VOF connects regions encoding object properties such form, identity, and color and to regions that register spatial information.

There is finally the aspect of surround suppression that may be common to all domains in which attention operates. In the spatial domain, attention to a target is accompanied by suppression of neighboring targets, shown behaviorally (Cutzu & Tsotsos, 2003; Mounts, 2000) and physiologically (Moran & Desimone, 1985; Schall et al., 1995). Similarly, surround suppression is seen in feature space where attention to one color, for example, will suppress responses to similar colors (Fang et al., 2019; Störmer & Alvarez, 2014; Yoo et al., 2018). Suppression of objects that are similar to an attended object may occur for object-based attention as well, although this has not yet been tested (Box 2).

Conclusions

Here we have focused on the processes underlying object-based attention: what are the targets of object-based attention; how is the selection controlled; and where does it happen? The literature in object-based attention is vast and we have undoubtedly missed articles, but the 200 papers we did survey made a compelling argument for three components. First, the targets of object-based attention are the features and locations in early visual cortices that receive downward projections from object representations in high level, object selective areas of cortex. Second, the object representations can be fully developed and activated either top-down from the frontoparietal attention network or bottom up from a salient cue located within the object's boundaries. Finally, the downward projections from these non-retinotopic object areas can nevertheless target both locations and features due to autoencoder-like learning in the visual system (Hinton & Salakhutdinov, 2006). Interestingly, the object-based attention literature also included several examples of attention to groups with the conclusion that groups have the same status as objects as targets of attention (Driver & Baylis, 1998; Kasai et al., 2011). The evidence that attention can be allocated to the elements of a group (Gonen et al., 2014) led us to suggest that mechanisms of object-based attention and feature-based attention are closely linked, offering a unification of the two. Of course, a group that is a single element or location will allocate attention to a single location, matching the properties of space-based attention (Logan, 1996; Shomstein & Behrmann, 2006; Vecera, 1994). Despite the appealing possibility that object-based attention actually unifies all three modes of allocation, there is overwhelming evidence for the independent stream of spatial allocation controlled from saccade areas (Awh et al., 2006) suggesting that space-based attention may be controlled from both the object areas and the saccade areas. Finally, our exploration of object-based attention revealed several novel predictions: the link between feature-based and object-based attention; the anatomical sites for different components of object-based attention; the role of autoencoder networks; the global projection of object-based benefits; the preattentive, identity-level object representations; and the suppressive surrounds in object space. These new points should help guide further studies revealing in more detail how object-based attention works.

References

Adamian, N., Andersen, S. K., & Hillyard, S. A. (2020). Parallel attentional facilitation of features and objects in early visual cortex. *Psychophysiology, 57*(3). https://doi.org/10.1111/psyp.13498

Al-Aidroos, N., Said, C. P., & Turk-Browne, N. B. (2012). Top-down attention switches coupling between low-level and high-level areas of human visual cortex. *Proceedings of the National Academy of Sciences, 109*(36), 14675–14680.

Ansorge, U., Kunde, W., & Kiefer, M. (2014). Unconscious vision and executive control: How unconscious processing and conscious action control interact. *Consciousness and Cognition, 27*, 268–287.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience, 29*(34), 10638–10652.

Attneave, F. (1971). Multistability in perception. *Scientific American, 225*(6), 62–71.

Awh, E., Armstrong, K. M., & Moore, T. (2006). Visual and oculomotor selection: Links, causes and implications for spatial attention. *Trends in Cognitive Sciences, 10*(3), 124–130.

Awh, E., Matsukura, M., & Serences, J. T. (2003). Top-down control over biased competition during covert spatial orienting. *Journal of Experimental Psychology: Human Perception and Performance, 29*(1), 52–63.

Bahrami, B., Carmel, D., Walsh, V., Rees, G., & Lavie, N. (2008). Spatial attention can modulate unconscious orientation processing. *Perception, 37*(10), 1520–1528.

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science, 344*(6182), 424–427.

Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Hénaff, M.-A., Michèl, F., & Barton, J. J. S. (2001). Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron, 32*(6), 985–995.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics, 51*(2), 145–162.

Bisley, J. W. (2011). The neural basis of visual attention. *The Journal of Physiology, 589*(1), 49–57.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408*(6809), 196–199.

Boynton, G. M., Ciaramitaro, V. M., & Arman, A. C. (2006). Effects of feature-based attention on the motion aftereffect at remote locations. *Vision Research, 46*(18), 2968–2976.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nature Neuroscience, 2*(4), 370–374.

Brown, J. M., & Denney, H. I. (2007). Shifting attention into and out of objects: Evaluating the processes underlying the object advantage. *Perception & Psychophysics, 69*(4), 606–618.

Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. J. (2011). Spatial coding and invariance in object-selective cortex. *Cortex, 47*(1), 14–22.

Cavanagh, P., & Alvarez, G. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349–354.

Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics, 74*(5), 784–802.

Chou, W.-L., & Yeh, S.-L. (2012). Object-based attention occurs regardless of object awareness. *Psychonomic Bulletin & Review, 19*(2), 225–231.

Chou, W.-L., Yeh, S.-L., & Chen, C.-C. (2014). Distinct mechanisms subserve location- and object-based visual attention. *Frontiers in Psychology, 5.* https://doi.org/10.3389/fpsyg.2014.00456

Cohen, E. H., & Tong, F. (2015). Neural mechanisms of object-based attention. *Cerebral Cortex, 25*(4), 1080–1092.

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences, 16*(8), 411–417.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201–215.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology, 80*(5), 2657–2670.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions. *Neuron, 32*(4), 737–745.

Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vision Research, 43*(2), 205–219.

de-Wit, L. H., Kentridge, R. W., & Milner, A. D. (2009). Object-based attention and visual area LO. *Neuropsychologia, 47*(6), 1483–1490.

Deeprose, C., & Andrade, J. (2006). Is priming during anesthesia unconscious? *Consciousness and Cognition, 15*(1), 1–23.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London: Series B: Biological Sciences, 353*(1373), 1245–1255.

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior Inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology, 89*(6), 3264–3278.

Donovan, I., Pratt, J., & Shomstein, S. (2017). Spatial attention is necessary for object-based attention: Evidence from temporal-order judgments. *Attention, Perception, & Psychophysics, 79*(3), 753–764.

Downing, C., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), Attention and performance XI (pp. 171–187). Erlbaum. http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781317246428

Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychonomic Bulletin & Review, 16*(2), 411–417.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 448–456.

Driver, J., & Baylis, G. C. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The attentive brain* (pp. 299–325). MIT Press.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*(4), 501–517.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*(2), 161–177.

Ekman, M., Roelfsema, P. R., & de Lange, F. P. (2020). Object selection by automatic spreading of top-down attentional signals in V1. *The Journal of Neuroscience, 40*(48), 9250–9259.

Elgendi, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., & Cichocki, A. (2018). Subliminal priming—State of the art and future perspectives. *Behavioral Sciences, 8*(6), 54.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149.

Erlikhman, G., Lytchenko, T., Heller, N. H., Maechler, M. R., & Caplovitz, G. P. (2020). Object-based attention generalizes to

multisurface objects. *Attention, Perception, & Psychophysics, 82*(4), 1599–1612.

Fang, M. W., Becker, M. W., & Liu, T. (2019). Attention to colors induces surround suppression at category boundaries. *Scientific Reports, 9*(1), 1–13.

Fanini, A., & Assad, J. A. (2009). Direction selectivity of neurons in the macaque lateral intraparietal area. *Journal of Neurophysiology, 101*(1), 289–305.

Farah, M. J. (1989). Mechanisms of imagery-perception interaction. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 203–211.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences, 7*(6), 252–256.

Finsterwalder, S., Demeyere, N., & Gillebert, C. R. (2017). Deficit in feature-based attention following a left thalamic lesion. *Neuropsychologia, 102*, 1–10.

Fleming, R. W., & Storrs, K. R. (2019). Learning to see stuff. *Current Opinion in Behavioral Sciences, 30*, 100–108.

Francis, G., & Thunell, E. (2022). Excess success in articles on object-based attention. *Attention, Perception, & Psychophysics, 84*(3), 700–714.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17*(3), 134–141.

Franken, T. P., & Reynolds, J. H. (2021). Columnar processing of border ownership in primate visual cortex. *Elife, 10*, e72573.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science, 269*(5225), 853–855.

Friedrich, F. J., Egly, R., Rafal, R. D., & Beck, D. (1998). Spatial attention deficits in humans: A comparison of superior parietal and temporal-parietal junction lesions. *Neuropsychology, 12*(2), 193–207.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Gazzaniga, M. S., & Ladavas, E. (1987). Disturbances in spatial attention following lesion or disconnection of the right parietal lobe. In *Advances in psychology* (Vol. 45, pp. 203–213). Elsevier.

Gegenfurtner, K. R., Kiper, D. C., & Levitt, J. B. (1997). Functional properties of neurons in macaque area V3. *Journal of Neurophysiology, 77*(4), 1906–1923.

Gonen, F. F., Hallal, H., & Ogmen, H. (2014). Facilitation by exogenous attention for static and dynamic gestalt groups. *Attention, Perception, & Psychophysics, 76*(6), 1709–1720.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35*(1), 96–111.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92*(2), 173–211.

Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2022). Frontopolar activity carries feature information of novel stimuli during unconscious reweighting of selective attention. *Cortex, 153*, 146–165.

Gurariy, G., Mruczek, R. E. B., Snow, J. C., & Caplovitz, G. P. (2022). Using high-density electroencephalography to explore spatiotemporal representations of object categories in visual cortex. *Journal of Cognitive Neuroscience, 34*(6), 967–987.

Haimson, C., & Behrmann, M. (2001). Cued visual attention does not distinguish between occluded and occluding objects. *Psychonomic Bulletin & Review, 8*(3), 496–503.

Han, K., Wen, H., Shi, J., Lu, K.-H., Zhang, Y., Fu, D., & Liu, Z. (2019). Variational autoencoder: An unsupervised model for encoding and decoding fMRI activity in visual cortex. *NeuroImage, 198*, 125–136.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334–337.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences, 11*(10), 428–434.

Hinton, G. E., Krizhevsky, A., & Wang, S. D. (2011). Transforming auto-encoders. In T. Honkela, W. Duch, M. Girolami, & S. Kaski (Eds.), *Artificial neural networks and machine learning–ICANN 2011: Lecture notes in computer science* (Vol. 6791, pp. 44–51). Springer.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504–507.

Hochstein, S., & Ahissar, M. (2002). View from the top. *Neuron, 36*(5), 791–804.

Hojjatoleslami, S. A., & Kittler, J. (1998). Region growing: A new approach. *IEEE Transactions on Image Processing, 7*(7), 1079–1084.

Hollingworth, A., Maxcey-Richard, A. M., & Vecera, S. P. (2012). The spatial distribution of attention within and across objects. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 135–151.

Howe, P. D., Horowitz, T. S., Akos Morocz, I., Wolfe, J., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision, 9*(4), 10–10.

Huang, L., Chen, Y., Shen, S., Ye, H., Ou, S., & Zhang, X. (2022). Awareness-independent gradual spread of object-based attention. *Current Psychology*, 1–13.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science, 317*(5839), 823–825.

Huang, T.-H., Yeh, S.-L., Yang, Y.-H., Liao, H.-I., Tsai, Y.-Y., Chang, P.-J., & Chen, H. H. (2015). Method and experiments of subliminal cueing for real-world images. *Multimedia Tools and Applications, 74*(22), 10111–10135.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171–216.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, 103*(45), 17048–17052.

Kasai, T., Moriya, H., & Hirano, S. (2011). Are objects the same as groups? ERP correlates of spatial attentional guidance by irrelevant feature similarity. *Brain Research, 1399*, 49–58.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751–761.

Knapen, T., Swisher, J. D., Tong, F., & Cavanagh, P. (2016). Oculomotor remapping of visual information to foveal retinotopic cortex. *Frontiers in Systems Neuroscience, 10*. https://doi.org/10.3389/fnsys.2016.00054

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11*(1), 16–22.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506–1509.

Kravitz, D. J., & Behrmann, M. (2011). Space-, object-, and feature-based attention interact to organize visual scenes. *Attention, Perception, & Psychophysics, 73*(8), 2434–2447.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*(1), 101–124.

Lamy, D., & Egeth, H. (2002). Object-based selection: The role of attentional shifts. *Perception & Psychophysics, 64*(1), 52–66.

Lamy, D., Segal, H., & Ruderman, L. (2006). Grouping does not require attention. *Perception & Psychophysics, 68*(1), 17–31.

Lamy, D., & Tsal, Y. (2000). Object features, object locations, and object files: Which does selective attention activate and when? *Journal of Experimental Psychology: Human Perception and Performance, 26*(4), 1387–1400.

Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8595–8598). IEEE.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience, 2*(4), 375–381.

Lee, J., & Shomstein, S. (2013). The differential effects of reward on space-and object-based attentional allocation. *Journal of Neuroscience, 33*(26), 10625–10633.

Li, Q., Gomez-Villa, A., Bertalmio, M., & Malo, J. (2022). Contrast sensitivity functions in autoencoders. *ArXiv:2103.00481 [q-bio].* http://arxiv.org/abs/2103.00481

Li, X., & Logan, G. D. (2008). Object-based attention in Chinese readers of Chinese words: Beyond gestalt principles. *Psychonomic Bulletin & Review, 15*(5), 945–949.

Liang, G., & Scolari, M. (2020). Limited interactions between space-and feature-based attention in visually sparse displays. *Journal of Vision, 20*(4), 5.

Lin, J. Y., Hubert-Wallander, B., Murray, S. O., & Boynton, G. M. (2011). Rapid and reflexive feature-based attention. *Journal of Vision, 11*(12), 12–12.

Lin, S.-Y., & Yeh, S.-L. (2015). Unconscious grouping of Chinese characters: Evidence from object-based attention. *Language and Linguistics, 16*(4), 517–533.

Lin, Z., & Murray, S. O. (2013). Visible propagation from invisible exogenous cueing. *Journal of Vision, 13*(11), 12–12.

Liu, T. (2016). Neural representation of object-specific attentional priority. *NeuroImage, 129*, 15–24.

Liu, T. (2019). Feature-based attention: Effects and control. *Current Opinion in Psychology, 29*, 187–192.

Liu, T., & Mance, I. (2011). Constant spread of feature-based attention across the visual field. *Vision Research, 51*(1), 26–33.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103*(4), 603–649.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19*(1), 577–621.

Lu, S., Cai, Y., Shen, M., Zhou, Y., & Han, S. (2012). Alerting and orienting of attention without visual awareness. *Consciousness and Cognition, 21*(2), 928–938.

Luzardo, F., Einhäuser, W., & Yeshurun, Y. (2022). *A continuous measure of object-based attention sheds new light on its underlying mechanisms* [preprint]. *PsyArXiv.* https://doi.org/10.31234/osf.io/7skfn

MacEvoy, S. P., & Yang, Z. (2012). Joint neuronal tuning for object form and position in the human lateral occipital complex. *NeuroImage, 63*(4), 1901–1908.

Mack, A., & Rock, I. (1998). Inattentional blindness: Perception without attention. In R. D. Wright (Ed.), *Visual attention* (pp. 55–76). Oxford University Press.

Majima, K., Sukhanov, P., Horikawa, T., & Kamitani, Y. (2017). Position information encoded by population activity in hierarchical visual areas. *Eneuro, 4*(2). https://doi.org/10.1523/ENEURO.0268-16.2017

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, 92*(18), 8135–8139.

Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General, 144*(2), 257–263.

Marino, J. (2022). Predictive coding, variational autoencoders, and biological connections. *Neural Computation, 34*(1), 1–44.

Matsukura, M., & Vecera, S. P. (2006). The return of object-based attention: Selection of multiple-region objects. *Perception & Psychophysics, 68*(7), 1163–1175.

Maunsell, J. H. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science, 1*, 373–391.

Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature, 429*(6990), 410–413.

Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance, 23*(2), 339–352.

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science, 9*(2), 104–110.

Moore, T., & Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annual Review of Psychology, 68*, 47–72.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782–784.

Mounts, J. R. (2000). Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception & Psychophysics, 62*(5), 969–983.

Mozer, M. C., & Vecera, S. P. (2005). Space- and object-based attention. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 130–134). Academic Press. https://doi.org/10.1016/B978-012375731-9/50027-6

Müller, M. M. (2014). Neural mechanisms of feature-based attention. In G. R. Mangun (Ed.), *Cognitive electrophysiology of attention: Signals of the mind* (pp. 123–135). Elsevier. https://doi.org/10.1016/B978-0-12-398451-7.00010-5

Müller, N. G., & Kleinschmidt, A. (2003). Dynamic interaction of object- and space-based attention in retinotopic visual areas. *The Journal of Neuroscience, 23*(30), 9812–9816.

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2013). Object-based attention without awareness. *Psychological Science, 24*(6), 836–843.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). FMRI evidence for objects as the units of attentional selection. *Nature, 401*(6753), 584–587.

Ongchoco, J. D. K., & Scholl, B. J. (2019). How to create objects with your mind: From object-based attention to attention-based objects. *Psychological Science, 30*(11), 1648–1655.

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *The Journal of Comparative Neurology, 426*(4), 505–518.

Oyedotun, O. K., & Dimililer, K. (2016). Pattern recognition: Invariance learning in convolutional auto encoder network. *International Journal of Image, Graphics and Signal Processing, 8*(3), 19.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*(4), 291–303.

Pavlidis, T., & Liow, Y.-T. (1990). Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*(3), 225–233.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25.

Posner, M. I., Cohen, Y., Rafal, R. D., Broadbent, D. E., & Weiskrantz, L. (1982). Neural systems control of spatial orienting.

Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 298(1089), 187–198.

Posner, M. I., Walker, J., Friedrich, F., & Rafal, R. (1984). Effects of parietal injury on covert orienting of attention. *The Journal of Neuroscience, 4*(7), 1863–1874.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509.

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

Rafal, R. (1994). Inhibitory neural mechanisms in spatial orienting. *Canadian Psychology/Psychologie Canadienne, 35*(1), 111–111.

Rao, R. (1997). Correlates of attention in a model of dynamic visual recognition. *Advances in Neural Information Processing Systems, 10*.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*(1/3), 17–42.

Reppa, I., Schmidt, W. C., & Leek, E. C. (2012). Successes and failures in producing attentional object-based cueing effects. *Attention, Perception, & Psychophysics, 74*(1), 43–69.

Reynolds, J. H., Alborzian, S., & Stoner, G. R. (2003). Exogenously cued attention triggers competitive selection of surfaces. *Vision Research, 43*(1), 59–66.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376–381.

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00013

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., & Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science, 337*(6095), 753–756.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631–632.

Schall, J. D., Hanes, D. P., Thompson, K. G., & King, D. J. (1995). Saccade target selection in frontal eye field of macaque: I. visual and premovement activation. *Journal of Neuroscience, 15*(10), 6905–6918.

Schoenfeld, M. A., Hopf, J.-M., Merkel, C., Heinze, H.-J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience, 17*(4), 619–624.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80*(1/2), 159–177.

Scolari, M., Ester, E. F., & Serences, J. T. (2014). Feature- and object-based attentional modulation in the human visual system. In A. C. (Kia) Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (Vol. 1, pp. 1–23). Oxford University Press.

Serences, J. T. (2004). Control of object-based attention in human cortex. *Cerebral Cortex, 14*(12), 1346–1357.

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55*(2), 301–312.

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex, 14*(12), 1346–1357.

Sereno, A. B., & Lehky, S. R. (2011). Population coding of visual space: Comparison of spatial representations in dorsal and ventral pathways. *Frontiers in Computational Neuroscience, 4*. https://doi.org/10.3389/fncom.2010.00159

Shin, K., Stolte, M., & Chong, S. C. (2009). The effect of spatial attention on invisible stimuli. *Attention, Perception, & Psychophysics, 71*(7), 1507–1513.

Shomstein, S. (2012). Object-based attention: Strategy versus automaticity. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(2), 163–169.

Shomstein, S., & Behrmann, M. (2006). Cortical systems mediating visual attention to both objects and spatial locations. *Proceedings of the National Academy of Sciences, 103*(30), 11387–11392.

Shomstein, S., & Johnson, J. (2013). Shaping attention with reward: Effects of reward on space-and object-based selection. *Psychological Science, 24*(12), 2369–2378.

Shomstein, S., & Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics, 64*(1), 41–51.

Shushruth, S., Ichida, J. M., Levitt, J. B., & Angelucci, A. (2009). Comparison of spatial summation properties of neurons in macaque V1 and V2. *Journal of Neurophysiology, 102*(4), 2069–2083.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *Journal of Neurophysiology, 97*(1), 229–237.

Sparks, D. L. (1999). Conceptual issues related to the role of the superior colliculus in the control of gaze. *Current Opinion in Neurobiology, 9*(6), 698–707.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*(1), 29–56.

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology, 24*(17), 1985–1988.

Störmer, V. S., Cohen, M. A., & Alvarez, G. A. (2019). Tuning attention to object categories: Spatially global effects of attention to faces in visual processing. *Journal of Cognitive Neuroscience, 31*(7), 937–947.

Storrs, K. R., Anderson, B. L., & Fleming, R. W. (2021). Unsupervised learning predicts human perception and misperception of gloss. *Nature Human Behaviour, 5*(10), 1402–1417.

Sun, Y., & Fisher, R. (2003). Object-based visual attention for computer vision. *Artificial Intelligence, 146*(1), 77–123.

Takemura, H., Rokem, A., Winawer, J., Yeatman, J. D., Wandell, B. A., & Pestilli, F. (2016). A major human white matter pathway between dorsal and ventral visual cortex. *Cerebral Cortex, 26*(5), 2205–2214.

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience, 11*(10), 697–709.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*(1), 109–139.

Tanaka, K. (2003). Columns for complex visual object features in the Inferotemporal cortex: Clustering of cells with similar but slightly different stimulus Selectivities. *Cerebral Cortex, 13*(1), 90–99.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66*(1), 170–189.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(4), 753–759.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15–48.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107–141.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*(6736), 575–579.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*(1/2), 507–545.

Tsotsos, J., Kotseruba, I., & Wloka, C. (2016). A focus on selection for fixation. *Journal of Eye Movement Research, 9*(5).

Valdes-Sosa, M., Cobo, A., & Pinilla, T. (2000). Attention to object files defined by transparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 488–505.

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin, 135*(3), 452–477.

Vecera, S. P. (1994). Grouped locations and object-based attention: Comment on Egly, Driver, and Rafal (1994). *Journal of Experimental Psychology: General, 123*(3), 316–320.

Vecera, S. P., & Behrmann, M. (2001). Attention and unit formation: A biased competition account of object-based attention. In *Advances in psychology* (Vol. 130, pp. 145–180). Elsevier.

Von der Heydt, R. (2015). Figure–ground organization and the emergence of proto-objects in the visual cortex. *Frontiers in Psychology, 6*, 1695.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407.

West, G. L., Anderson, A. A. K., & Pratt, J. (2009). Motivationally significant stimuli show visual prior entry: Evidence for attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 35*(4), 1032–1042.

White, A. L., Rolfs, M., & Carrasco, M. (2013). Adaptive deployment of spatial and feature-based attention before saccades. *Vision Research, 85*, 26–35.

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience, 11*(12), 1439–1445.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37*(1), 25–43.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences, 15*(2), 77–84.

Wu, T., Gao, J., & Zhao, Q. (2004). A computational model of object-based selective visual attention mechanism in visual information acquisition. *International conference on information acquisition, 2004. Proceedings* (pp. 405–409). https://doi.org/10.1109/ICIA.2004.1373400

Yeh, S.-L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological Science, 23*(6), 608–616.

Yoo, S.-A., Tsotsos, J. K., & Fallah, M. (2018). The attentional suppressive surround: Eccentricity, location-based and feature-based effects and interactions. *Frontiers in Neuroscience, 12*, 710.

Yu, J.-G., Xia, G.-S., Gao, C., & Samal, A. (2015). A computational model for object-based visual saliency: Spreading attention along gestalt cues. *IEEE Transactions on Multimedia, 18*(2), 273–286.

Yuan, J., & Fu, S. (2014). Attention can operate on semantic objects defined by individual Chinese characters. *Visual Cognition, 22*(6), 770–788.

Zerweck, I. A., Kao, C.-S., Meyen, S., Amado, C., von Eltz, M., Klimm, M., & Franz, V. H. (2021). Number processing outside awareness? Systematically testing sensitivities of direct and indirect measures of consciousness. *Attention, Perception, & Psychophysics, 83*(6), 2510–2529.

Zhang, X., & Fang, F. (2012). Object-based attention guided by an invisible object. *Experimental Brain Research, 223*(3), 397–404.

Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural activities in V1 create a bottom-up saliency map. *Neuron, 73*(1), 183–192.

Zhao, L., Cosman, J. D., Vatterott, D. B., Gupta, P., & Vecera, S. P. (2014). Visual statistical learning can drive object-based attentional selection. *Attention, Perception, & Psychophysics, 76*(8), 2240–2248.