Behavioral and Brain Sciences

cambridge.org/bbs

Visual Attention in Crisis

Ruth Rosenholtz

Department of Brain & Cognitive Sciences, CSAIL, Massachusetts Institute of Technology, USA
rruth@mit.edu

Target Article

Cite this article: Rosenholtz R. (2025) Visual Attention in Crisis. *Behavioral and Brain Sciences* **48**, e130: 1–72. doi:10.1017/S0140525X24000323

Target Article Accepted: 27 September 2022
Target Article Manuscript Online: 03 May 2024
Commentaries Accepted: 15 April 2024

Keywords:
attention; capacity limits; complexity; peripheral vision; selection

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 18) and an Author's Response (p. 66). See bbsonline.org for more information.

Corresponding author:
Ruth Rosenholtz; Email: rruth@mit.edu

Short Abstract

Recent research on peripheral vision has led to a paradigm-shifting conclusion: vision science as a field must rethink visual attention. This article reviews the evidence for a crisis in attention science and examines supposedly attentional phenomena to ask which point to additional capacity limits. Based on the resulting list of critical phenomena, and what they have in common, I propose an alternative way to think about capacity limits and the underlying mechanisms.

Long Abstract

Research on visual attention has uncovered significant anomalies, and some traditional methods may have inadvertently probed peripheral vision rather than attention. Vision science needs to rethink visual attention from the ground up. To facilitate this, for a year I banned the word "attention" in my lab. This constraint promoted a more precise discussion of attention-related phenomena, capacity limits, and mechanisms. The insights gained lead me to challenge attributing to "attention" those phenomena that can be better explained by perceptual processes, are predictable by an ideal observer model, or that otherwise may not require an additional mechanism. I enumerate a set of critical phenomena in need of explanation. Finally, I propose a unifying theory in which all perception results from performing a task, and tasks face a limit on complexity.

1. Why we need to rethink visual attention

Much of vision science attempts to explain a narrow aspect of visual perception; one might study perceptual grouping, or face perception, but not attempt to understand both simultaneously. The study of visual attention provides one significant departure from this trend. Visual attention theory proposes a critical factor in vision: human vision is faced with more information than it can process at once. If correct, understanding this *limited capacity* and the mechanisms for dealing with it could allow us to predict performance on a wide range of visual tasks.

In particular, attention theory presumes that vision must deal with limited access to higher-level visual processing. To do so, it employs a *selective attention* mechanism, which in its simplest form selects one thing at a time for access, effectively serving as a gate. Selection can be *covert,* attending away from the point of gaze, or *overt* if pointing one's eyes at the target. If attention gates access, this immediately raises several questions: At what stage does attention act? What processing requires attention, and what happens *preattentively* and automatically?

Researchers have developed various methods to answer these questions. In their seminal paper, Treisman and Gelade (1980) employed tasks in which an observer must search for a *target* among other, *distractor* items. If search displays overwhelm the observer with more items than they can process at once, then easy and difficult search tasks would distinguish between the discriminations possible with preattentive processing, as opposed to those that require attention, respectively. Their results formed the foundation of feature integration theory. According to this theory, selection occurs early in visual processing. Preattentively the visual system only has access to simple feature maps, allowing one to easily find a *salient* target defined by a unique basic feature such as orientation (θ), color, or motion (**v**). These features underlie *bottom-up processing*. On the other hand, tasks such as searching for a T among Ls, or for a target defined by a conjunction of features (white AND vertical) require selective attention (Treisman & Gelade, 1980). Binding features together, perceiving details, and most object recognition tasks require selective attention. This theory appeared consistent with demonstrations of *change blindness*, in which observers have difficulty finding the difference between two similar images; if perceiving details requires selective attention, then detecting a difference in those details will rely on moving attention from one location to another until it happens to land on the change, a relatively slow process (Rensink, O'Regan, & Clark, 1997).

However, this attention theory has obvious issues. Most notably, observers can quickly and easily get the *gist* of a scene or a set of similar items (Loftus & Ginn, 1984; Potter, 1975; Rousselet,

© The Author(s), 2024. Published by Cambridge University Press

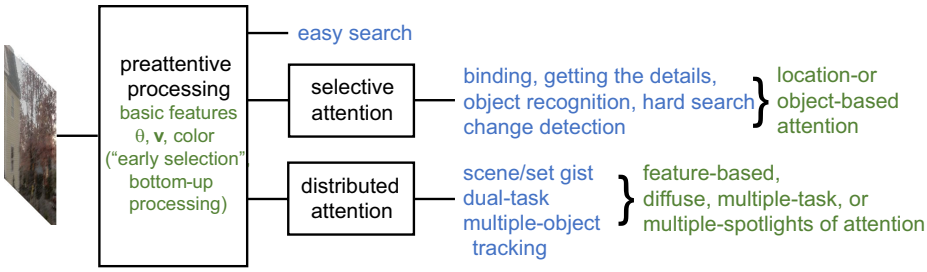


Figure 1. A subset of the mechanisms in attention theory (boxes, and green text), along with phenomena supposedly explained by those mechanisms (blue text). See, e.g. Chun et al. (2011) for a fuller account.

Joubert, & Fabre-Thorpe, 2005; Greene & Oliva, 2009; Ariely, 2001; Chong & Treisman, 2003; Chong & Treisman, 2005; Haberman & Whitney, 2009). The gist can include information such as the category or layout of a scene, the mean size of a collection of circles, or the mean emotion of a set of faces. These results appear incompatible with the idea that serial selective attention is necessary merely to combine color and shape or detect a corner. Faced with this issue, researchers proposed a mechanism that distributes or diffuses attention across an entire scene or set, allowing one to extract information about the stimulus as a whole (Treisman, 2006), or a separate, non-selective pathway for scene and set perception (Rensink, 2000; Wolfe, Vo, Evans, & Greene, 2011). Other empirical results led to additional complications in the theory (Figure 1). One might distribute attention to select a feature across space, for example selecting everything red (see Maunsell & Treue, 2006, for a review). One might divide attention across multiple tasks (VanRullen, Reddy, & Koch, 2004), if at the cost of reduced performance, or deploy a few spotlights of attention to track a small number of moving objects (Pylyshyn & Storm, 1988; Alvarez & Franconeri, 2007). Attention can be bottom-up or top-down; transient or sustained; directed voluntarily or captured involuntarily; and can modulate both the percept and neuronal firing rates (Chun, Golumb, & Turk-Browne, 2011). Researchers have criticized this proliferation of attentional mechanisms and uses of the word (Anderson, 2011; Hommel, et al., 2019; Zivony & Eimer, 2021; Chun, Golumb, & Turk-Browne, 2011). However, this might merely reflect reality; visual attention might involve a sizeable set of largely separate mechanisms. As Wu (2023) has argued, attention can refer to a "shared functional structure" in the mind even if implemented in different ways.

I argue that the science of visual attention is in crisis, and in need of a paradigm shift. In making this claim, I borrow terminology from philosopher of science Thomas Kuhn, in his seminal The Structure of Scientific Revolutions (Kuhn, 1962). Kuhn says that most of the time a given scientific field operates within a dominant paradigm. A paradigm includes agreed-upon theory, methods, and puzzles to solve. Operating within this paradigm is what Kuhn calls *normal science*. In normal science the paradigm's adherents productively solve its puzzles with the prescribed methods. However, at some point, normal science uncovers significant *anomalies*, results that one cannot easily explain using the dominant paradigm. The field enters a crisis, which persists until a new paradigm emerges, i.e., a *paradigm shift* occurs.

Kuhn gives as a canonical example the shift from the Ptolemaic earth-centric view of the solar system to the Copernican heliocentric view. The classic story demonstrates the difficulty of recognizing a crisis from within its midst. Observations of planetary motions, driven by the earth-centric paradigm, uncovered a number of anomalies. Mars appeared at times to move backwards! To account for these observations, scientists modified circular orbits around the earth by adding dozens of *epicycles*, additional circles moving on circles. For the earth-centric scientists, this was normal science; the paradigm allowed the addition of epicycles to gain better predictions. Ultimately, the paradigm shifted to heliocentrism, and this and other significant changes led to the modern view of planetary motions. (This story is oversimplified in many ways but nonetheless will serve as a useful analogy at several points in this paper. See Heliocentrism, 2023).

Kuhn describes several main signs of a crisis (Kuhn, 1962): 1. Significant anomalies that must be explained; 2. Methods no longer leading to the answers promised by the paradigm; and 3. Increasing complexity of the theory, without a corresponding increase in the ability to make accurate predictions. Here I point to a few pieces of evidence for a crisis, with additional signs described in later sections and my full list enumerated in Section 3.

First, consider the increased complexity of the theory. A proliferation of attentional mechanisms, per se, does not necessarily indicate a crisis. New phenomena led to proposals of diverse types of attention – diffuse, feature-based, multiple spotlights, etc. In that sense, these additional mechanisms provide predictive value. Epicycles improved quantitative predictions. However, the added attentional mechanisms lack specificity and integration into a coherent theory, leaving many questions unanswered. What information does diffuse attention make available? Which scene tasks does it enable, and which tasks require focal attention? Do diffuse and focal attention face the same capacity limit? If focal and diffuse attention utilize different resources, then can one deploy both at the same time? If not, what is the common capacity limit? This lack of specificity and integration suggests that the theory has added additional mechanisms without a commensurate increase in predictive value.

Second, while vision science readers may have adapted to explaining easy scene and set perception in terms of diffuse attention or a separate non-selective pathway, I argue that one should think of these phenomena as anomalies. Virtually every visual attention theory added a distinct component to account for these results. The puzzle for these theories is how to add a mechanism capable of getting the gist of a scene or set, while still predicting things like difficult search and change blindness.

Ruth Rosenholtz is a formerly a Principal Research Scientist at the Computer Science and Artificial Intelligence Laboratory (CSAIL) and the Department of Brain & Cognitive Sciences at MIT. She has published widely in the areas of visual psychophysics, computational models of human vision, and computer vision. She has done pioneering work in visual clutter, visual attention, and visual crowding. She is well known for her "Texture Tiling Model" of peripheral vision, which offers a comprehensive framework for understanding numerous visual phenomena.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press



Figure 2. (A) Original image, fixation at center. (B) Two visualizations of the information available at a glance, according to our model of peripheral vision.

Intriguingly, researchers starting with fairly different theories proposed a similar-sounding solution: *summary statistics*. Summary statistics concisely summarize a large number of observations, for instance in terms of their mean or variance, or the correlation between two different measurements. Rensink (2000) suggested that the gist of a scene might be extracted in a non-attentional pathway, based on "statistics of low-level structures such as proto-objects". Oliva and Torralba (2001) developed a summary statistic model for scene perception based on local power spectrum computed over sizeable regions of the image. Treisman (2006) suggested that the visual system might distribute attention across an entire display, making available a statistical description pooled over feature maps. Wolfe et al. (2011), like Rensink (2000), suggested an additional non-selective pathway with statistical processing abilities.

Summary statistics pooled over sizable image regions provide a form of data compression. One can think of this compression as an alternative means of dealing with limited capacity, instead of serially selecting one object at a time to feed through the bottleneck (Rosenholtz, Huang, & Ehinger, 2012). Selection and compression lead to different losses of information, with very different implications for task performance.

One of the most well-specified and tested proposals for a summary statistic encoding in vision came neither from studying attention nor scene perception, but rather from the study of peripheral vision, i.e. vision away from the point of gaze. The next subsection reviews the suggestion that peripheral vision encodes its inputs using a rich set of summary statistics. This proposal makes sense of some anomalies but also reveals additional signs of crisis. I will argue that one should view a summary statistic encoding as a paradigm shift, but that a crisis remains.

1.1. A summary statistic encoding in peripheral vision

A significant loss of information occurs in peripheral vision, particularly due to visual crowding. Crowding refers to phenomena in which peripheral vision degrades in the presence of clutter. In a common example, an observer views a peripheral word, in which each letter is closely flanked by others. Observers might perceive the letters in the wrong order, or a confusing jumble of shapes made up of parts from multiple letters (Lettvin, 1976). Lettvin (1976) suggested that the crowded letters "only [seem] to have a 'statistical' existence." Move the letters farther apart, and at some *critical spacing* letter identification improves (Bouma, 1970). Building on these observations, several researchers have suggested that crowding occurs due to the representation of peripheral information in terms of statistics summarized over sizable pooling regions (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011). These pooling regions grow linearly with eccentricity, overlap, and sparsely tile the visual field (Freeman & Simoncelli, 2011; Rosenholtz, Huang, & Ehinger, 2012; Chaney, Fischer, & Whitney, 2014). Based on these intuitions, my lab has developed and tested a computational model of peripheral crowding, known as the Texture Tiling Model (TTM). The model measures a rich set of summary statistics derived from the texture perception model of Portilla and Simoncelli (2000), with a first stage computing V1-like responses to oriented, multiscale feature detectors, and the second stage measuring a large set of primarily second-order correlations of the responses of the first stage, computed over local pooling regions (Balas et al., 2009; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; see also Freeman & Simoncelli, 2011). Figures 2 and 4 show model outputs to provide intuitions about predictions of the model.

TTM can in fact predict performance getting the gist of a scene. Given a stimulus and a fixation, TTM generates random images with the same pooled summary statistics as the original (Figure 2). These images provide visualizations of the consequences of encoding an image in this way. Aspects of the scene that one can consistently discern in these images will be readily available when fixating the specified location. Conversely, details not consistently apparent will be less available to peripheral vision. A summary statistic encoding in peripheral vision preserves a great deal of useful information for getting the gist of a scene. In Figure 2, the encoding suffices to identify a street scene, most likely a bus stop, people waiting, cars on the road, trees, and a building in the background. Ehinger and Rosenholtz (2016) demonstrated that this encoding quantitatively predicts performance on a range of scene perception tasks. Though not yet well studied, the model also seems promising for predicting the ease with which one can get the gist of a set, relative to reporting individual items of that set (Rosenholtz, Yu, & Keshvari, 2019; though see Balas, 2016).

A summary statistic encoding in peripheral vision also provides insight into a second anomaly: that different methods for distinguishing between automatic and attention-demanding tasks have not agreed on which tasks require attention. In particular, search and dual-task results have disagreed on this classification, as have attentional capture (Folk, Remington, & Johnston, 1992) and inattentional blindness experiments (Mack & Clarke, 2012; Mack & Rock, Inattentional blindness, 1998). VanRullen et al. (2004) provide a particularly useful table of search vs. dual-task results (Figure 3A).

Recall that in the traditional interpretation, easy or efficient search implies that observers can discriminate the target from

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

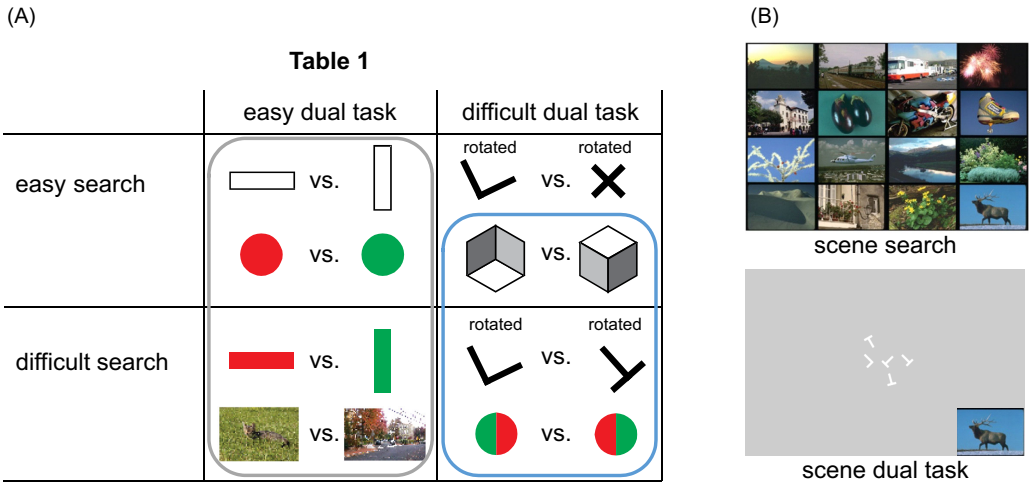


Figure 3. Example stimulus patches from three classic search conditions (top row, target specified at top). Akin to the demonstration in Figure 2, here one can use the texture analysis/synthesis model of Portilla and Simoncelli to visualize the information encoded by the model summary statistics (bottom two rows). The summary statistics capture the presence (column 1, row 2) or absence (row 3) of a uniquely tilted line. However, the information is insufficient to correctly represent the conjunction of color and orientation, producing an illusory white vertical from a target absent patch (column 2, row 3). For T among L search, no clear T appears in the target-present synthesis (column 3, row 2), whereas several appear in the target absent (row 3); this ambiguity predicts difficult search.

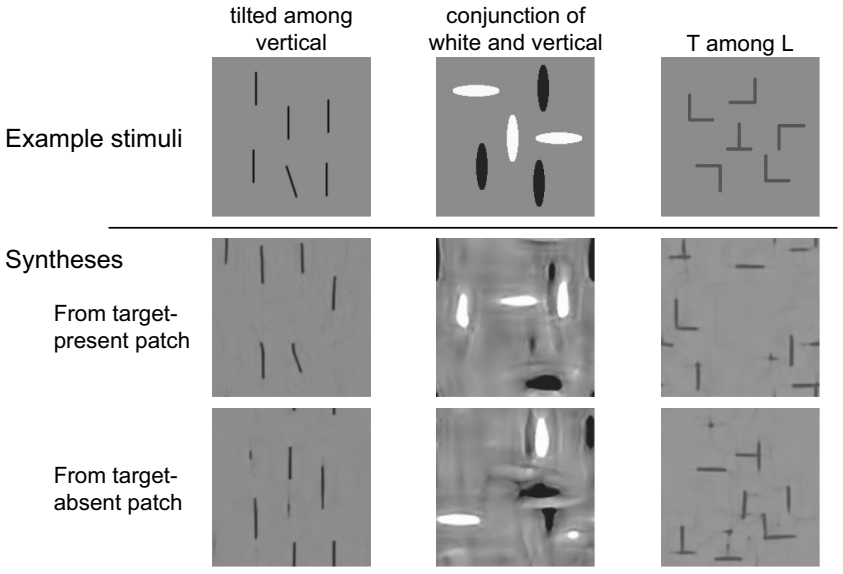


Figure 4. (A) Search and dual-task paradigms give different answers about what tasks do and do not require attention. Understanding peripheral vision resolves this conundrum by more parsimoniously explaining search, and (B) noting that search displays are often more crowded than dual-task displays. Tasks circled in gray (blue) are behaviorally or theoretically easy (hard) for peripheral vision in dual-task displays. Table adapted from VanRullen et al. (2004).

distractors preattentively, automatically, and without selective attention (Treisman & Gelade, 1980). Inefficient or difficult search, conversely, indicates that the target-distractor discrimination requires selective attention.

In dual-task experiments, an observer performs either a task in central vision, a task in the periphery, or both. Observers often perform worse when given two tasks instead of one. According to the experimental logic (VanRullen, Reddy, & Koch, 2004), easy dual tasks do not require selective attention; rather this processing happens automatically. Difficult dual tasks require attention.

Consequently, discriminations that lead to easy (hard) search tasks should also lead to easy (hard) dual tasks (upper left and lower right quadrants, Figure 3A). However, in the two remaining quadrants, the two experiments disagree on whether tasks require attention.

A summary statistic encoding in peripheral vision can make sense of these anomalous results. First, our model of visual crowding provides a more parsimonious account of critical visual search phenomena (Figure 4). Difficulty discriminating a crowded peripheral target from a distractor explains easy feature search, difficult conjunction search, and difficult search for a T among Ls (Rosenholtz, Huang, Raj, Balas, & Ilie, 2012); phenomena that originally led to feature integration theory. The model also predicts difficult search for a scene among scenes, and for a green-red bisected disk among red-green ones (Rosenholtz, Huang, & Ehinger, 2012). In addition, it explains phenomena that defy easy explanation by feature integration theory, such as easy search for a cube among differently lit cubes compared to similar 2D conditions (Zhang, Huang, Yigit-Elliot, & Rosenholtz, 2015), and effects of subtle changes to the stimuli (Chang & Rosenholtz, 2016). A summary statistic model of peripheral vision, in other words, collapses the two rows of Table 1; search may probe peripheral discriminability, not what tasks require attention.

Perhaps more surprising, peripheral vision allows us to collapse the columns of the table. A number of the easy (hard) dual tasks are easy (hard) for peripheral vision, based on either behavioral experiments or modeling (Rosenholtz, Huang, & Ehinger, 2012). Dual task difficulty, then, may also depend more on the strengths and limitations of peripheral vision than on tasks that do or do not require attention.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

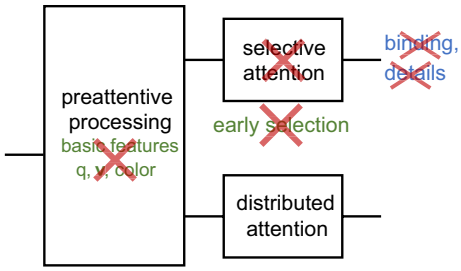


Figure 5. Understanding peripheral vision necessitates rethinking a significant amount of attention theory, including preattentive processing and selective attention mechanisms as well as the need for attention for binding and to perceive details (red X's).

One must ask why, if search and dual-task difficulty both probe peripheral vision, these tasks disagree on the off-diagonal of the table. Search displays contain considerably more clutter. This would impair scene search relative to the scene dual task (Figure 3B; Rosenholtz, Huang, & Ehinger, 2012). A less cluttered dual-task display fundamentally changes the conjunction task; the observer only needs to identify the color and orientation of two uncrowded items (Braun & Julesz, 1998). Clutter in the cube search case had the opposite effect: the dense array of cubes aligned in a way that aided search (Rosenholtz, Huang, & Ehinger, 2012).

A summary statistic encoding in peripheral vision resolves the anomaly that search and dual-task methods do not agree on what tasks require attention. Supposedly automatic tasks that did not require attention (Treisman, 2006) might simply have been inherently easy given the information available in peripheral vision. Looked at another way, this points to another sign of a Kuhnian crisis: The search and dual-task methods no longer lead to the answers promised by the paradigm. Search and dual-task experiments were seemingly fruitful, agreed-upon methods, with difficulty indicative of important aspects of the attentional mechanisms. Now it seems that one cannot easily interpret the results in that way.

This is not just a problem of methods. Rather, it necessitates rethinking a substantial amount of attention theory (Figure 5). Selection may not be early, and the information available preattentively may include more than basic feature maps. Feature binding may not require attention; the difficulty of both correctly binding features (Rosenholtz, Huang, Raj, Balas, & Ilie, 2012) and perceiving the details may arise from losses in peripheral vision rather than from inattention.

More profoundly, we need to rethink selection itself. In order for the strengths and limitations of peripheral vision to predict search performance, observers must use peripheral vision to search for the target. This leaves room for considerably more parallel processing, albeit with reduced fidelity due to crowding. This does not sound much like selection, as commonly envisioned. Recent work has also called into question supposed physiological signs of sensory selection. When monkeys perform a discrimination task with one of several items in a neuron's receptive field, the neuron responds as if only the target were present, as one would expect if the neuron selected the cued stimulus at the expense of the others (Desimone & Duncan, 1995). However, if one inactivates the superior colliculus, thought to play a critical role in visual attention, the monkey behavior displays attentional deficits, while the attentional modulations in cortex persist, suggesting that the modulations may not demonstrate a causal mechanism for sensory selection (Krauzlis, Bollimunta, Arcizet, & Wang, 2014).

One should also reconsider automaticity (which one might think of as the *dual* of selection). Evidence has not supported a distinction between tasks that do and do not require attention, nor "automatic" preattentive processing. Whether one notices a salient item or gets the gist of a scene — two common candidates for automatic processing — depends upon the difficulty of other simultaneous tasks (Joseph, Chun, & Nakayama, 1997; Rousselet, Thorpe, & Fabre-Thorpe, 2004; Matsukura, Brockmole, Boot, & Henderson, 2011; Cohen, Alvarez, & Nakayama, 2011; Mack & Clarke, 2012; Larson, Freeman, Ringer, & Loschky, 2014). In other words, these tasks are not automatic.

1.2. A paradigm shift and addressing the remaining crisis

One might reasonably call summary statistic encoding a paradigm shift, both for the science of visual attention and for vision more broadly. Researchers have long suggested that vision might use summary statistics to perform *statistical tasks*, such as texture discrimination and segmentation (Rosenholtz, 2014). This summary statistic encoding, however, would underlie all visual tasks, including getting the gist of a scene, identifying a peripheral object, or searching for a target. The underlying computations make use of neural operations like those previously described – feature detectors, nonlinearities, and pooling operations. However, pooling over substantially larger areas to compute summary statistics makes available qualitatively different information (Figure 2). Vision looks fundamentally different from within this new paradigm. The encoding captures a great deal of useful information, while lacking the details necessary for certain tasks. This may at least partially explain such diverse phenomena as change blindness (Smith, Sharan, Park, Loschky, & Rosenholtz, under revision) and difficulty noting inconsistencies in an impossible figure. The richness of the information available across the field of view means that eye trackers likely tell us less than we thought about the attentional state of the observer and has led to re-examining the subjective richness of visual perception (Rosenholtz, 2020; Cohen, Dennett, & Kanwisher, 2016). Research on summary statistic encoding has led to new proposals about what visual processing occurs in area V2 (Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013). It has as-yet-unrealized implications for all later visual processing, as vision scientists must re-envision how the visual system finds perceptually meaningful groups and computes properties such as 3D shape from the information available in the summary statistics. Researchers have even looked for evidence of similar mechanisms in auditory perception (McDermott & Simoncelli, 2011).

Despite this paradigm shift, the field of attention remains in crisis. Summary statistics have been tacked onto attention theory with little change to old notions of selection, preattentive processing, automaticity, or what tasks require attention. By analogy, the Copernican paradigm shift may have put the sun at the center, but it maintained the ideas of circular orbits and unchanging celestial spheres, and as a result, the new theory needed even more epicycles than the old (Gingerich, 1975). Sixty years passed before Kepler introduced elliptical orbits governed by physical laws. The remainder of this paper attempts to gain insights into, essentially, the epicycles, celestial spheres, and elliptical orbits of visual attention.

Kuhn suggested that new and old paradigms are *incommensurable*, meaning that scientists struggle to hold in their minds both the new and old ideas. This suggests that it helps, when looking for a new theory, to eliminate the old from one's thinking as much as

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

possible. For visual attention, we need not only rethink the theory. Given that established methods may have probed, for instance, peripheral vision rather than attention, we also need to reexamine the critical set of phenomena. One might better draw an analogy not to Copernicus – who faced general agreement on the relevant phenomena – but to Francis Bacon's reasoning about the nature of heat (Bacon, 2015).

Bacon gathered "known instances which agree in the same nature . . . without premature speculation" into three lists: "Instances Agreeing in the Nature of Heat", "Instances in Proximity Where the Nature of Heat is Absent" – examples like those on the first list, but lacking heat – and a "Table of Degrees of Comparison in Heat" – examples in which heat is present to varying degrees. By analogy, this paper re-examines phenomena attributed to attention, and which therefore seem likely to provide insights into its nature, and creates two lists of critical phenomena: The first contains phenomena that demonstrate clear evidence of additional capacity limits. If we abandoned attention theory, these phenomena would require explanation. The second contains phenomena in which human vision works well, seemingly without the same sorts of limits apparent in the first list. This list enumerates capabilities of the visual system that any viable theory must explain.

It can be difficult to reason about critical phenomena and a new theory when attention is already an overloaded term (Chun et al., 2011; Anderson, 2011; Hommel, et al., 2019), often used gratuitously (Anderson, 2011). Language can stand in the way. To help make a fresh start, for a year I banned the word "attention". Lab members avoided or attempted to clarify terminology like "selection", and I encouraged the group to be dissatisfied with explanations that relied on "available resources" without suggesting the nature of those resources. Given the difficulty measuring attention, I advised students not to blithely assume they knew a person's attentional state.

Banning "attention" provides a couple of additional benefits. Examining whether a given phenomenon demonstrates attention while simultaneously rethinking attention poses a chicken-and-egg problem. Restricting the use of the word suggests instead asking, "If we were frugal in the use of 'attention', would we use it here?" Banning "attention" also encourages us to talk about phenomena with minimal assumptions about the nature of capacity limits and the mechanisms for dealing with those limits. Examining phenomena as agnostically as possible helped in developing seeds of alternative theories.

The following section presents vignettes about phenomena that my lab pondered in search of the critical phenomena. Of course, it does not provide an exhaustive literature review, and my lists of critical phenomena are no doubt incomplete. Rather, this paper showcases specific examples that I think provide important insights into the nature of attention and/or demonstrate the process of rethinking the old paradigm. It focuses on behavioral effects rather than physiology (for discussion of additional phenomena see Rosenholtz, 2017; Rosenholtz, 2020). The section "Thoughts and a proposal" asks what the critical phenomena might have in common, and what might be the associated capacity limit(s) and mechanism(s). I will discuss a couple of alternative theories.

First, a brief example: overt attention, in which one directs attention by directing one's eyes. Terminology already exists for fixation, saccade planning, and so on; if one had to pay money every time one used "attention", one might not use it here. In deciding what goes on the list of critical phenomena, we should also ask to what degree covert and overt attention point to similar limits and mechanisms.

Humans typically only point our eyes at one location at a time. In contrast, attention theories propose that covert attention can be divided or diffused. Overt and covert attention also seem to have different consequences. Fixating provides excellent information at the point of fixation, but puts much of the visual field in the periphery, subject to crowding, reduced acuity, etc. On the other hand, after many years of studying attention, I still do not feel like I have a clear understanding of the perceptual consequences of focusing covert attention, for either the attended or unattended information. If the limits of covert and overt attention differ, as do their consequences, then what might justify calling them both attention? Perhaps the similarity lies only in how one decides what to focus on next. Common factors likely come into play: bottom-up stimulus factors, top-down task demands, prior probabilities and other contextual knowledge, the information gathered so far, reward, and so on. Several papers rethinking attention point to the importance of studying priority maps (Hommel, et al., 2019) and Bayesian decision processes (Anderson, 2011). I certainly agree with the value of such topics, although I would question whether associating them with the overused "attention" adds clarity.

2. Rethinking attention: Enumerating phenomena in need of explanation

Perhaps some previous work confused attention with peripheral vision because researchers used behavioral paradigms such as visual search that do not explicitly manipulate attention. An explicit manipulation might compare conditions with and without attention, or with attention to one item as opposed to another. This section almost exclusively considers paradigms that manipulate attention through cueing or a change in task. Loosely speaking, attention as task, object-based attention, cued search, and mental marking fall under cueing, with inattentional blindness and multitasking otherwise manipulating the task, though the dividing line between the two is fuzzy.

2.1. Attention as task

Some experiments ask the observer to "pay attention". For example, one might cue the observer to attend to and make a judgment about a target while ignoring other items, e.g. (Lavie, Hirst, de Fockert, & Viding, 2004). Or one might tell the observer that only non-targets will have a singleton color, so they should ignore those items during visual search (Theeuwes, 1992). Nonetheless, the ignored items often distract the observer, although this can come at a cost of as little as 20-40 ms (Theeuwes, 1992). Researchers often interpret the results in terms of what *captures attention*.

Selectively attending is the observer's task. Experimenters describe a task in natural language, and the brain must convert that description to its internal instruction set, making use of existing mechanisms. The visual system could do its best to perform the nominal task even if it had no mechanism for selecting an individual item. One can imagine similar results even without capacity limits. The observer perceives the entire display. Maybe it takes a few milliseconds to remember that responding based on the distractor gives the wrong answer. Or why not spend 40 ms enjoying an unusual item?

The most interesting thing about attention-as-task experiments is the observer's failure to process only the target. One might expect

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

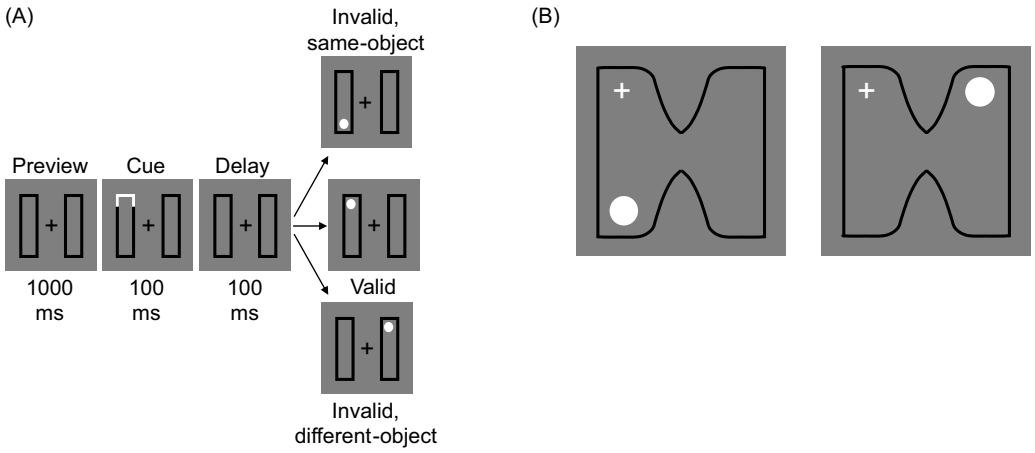


Figure 6. Object-based attention. (A) Experimental methodology. The experimenter cues one end of an object. After a delay (here, 100 ms), a target randomly appears either at the same location, the opposite end of the same object, or the equidistant location on the other object. Results often show a small (~10 ms) same-object benefit in the invalid conditions. (B) Object-based attention with a single object. Observers are faster to make judgments about the target (circle) when the object between it and the cue (+) is simple (left) than when it is complex (right). Based on stimuli from Chen et al., 2020.

distraction by salient stimuli, supposedly preattentively processed. However, results also suggest processing of not-terribly-salient numbers or letters to the extent that their category can cause response conflict (Lavie et al., 2004). Furthermore, researchers have suggested that distraction depends on whether saliency is task-relevant (Folk, Remington, & Johnston, 1992), on the perceptual task difficulty, and in a complex way on the difficulty of simultaneous cognitive tasks (Lavie et al., 2004), suggesting that distraction does not merely arise from automatic processing. These results seem puzzling from the point of view of the attention paradigm; the main mechanism for getting around a bottleneck in vision is leaky, allowing other information to pass through? Of course, if selection were perfect, vision would never work, with disastrous consequences; how would you notice the approaching tiger when you were paying attention to picking berries?

To help us rethink attention, we should try to describe phenomena in a more agnostic way that does not rely on earlier concepts of attention or selection. We might say that the experimenter asks the observer to make a judgment about the target, making sure not to respond based on any other item. Nonetheless, the observer perceives, to some degree, the task-irrelevant items, and this can cause a modest degradation in performance at judging the target. Both the modest cost of distraction and the apparent failure to select only the target seem interesting critical phenomena (see also Hommel, et al., 2019).

2.2. Object-based attention

Another set of cueing experiments tests *object-based attention* (OBA). Figure 6A shows the basic methodology. Observers respond to a target about 10-20 ms faster when it appears on a cued object than when it appears on a different object, despite controlling for the distance between the invalid cue and the target (e.g. Egly, Driver, & Rafal, 1994; Francis & Thunell, 2022). This has been taken as evidence that the observer's attention automatically spreads from the cue location to the entire cued object. At the time of the earliest OBA experiments, these phenomena required a significant shift in thinking about attention, since many models assumed that attention was directed to a location rather than an object (Kanwisher & Driver, 1992). Complicating this story, the same-object advantage applies more for horizontal objects than vertical (Al-Janabi & Greenberg, 2016; Chen & Cave, 2019; Francis & Thunell, 2022). Furthermore, when comparing two targets with no precue (Lamy & Egeth, 2002), one less often finds a same-object advantage (Chen, Cave, Basu, Suresh, & Wiltshire, 2020).

The OBA literature has complex and often seemingly conflicting results, with small effects; later work (Francis & Thunell, 2022) has called into question much of the literature, due to underpowered studies. (In this paper, any results citing Francis & Thunell, 2022, indicate OBA effects that replicated in their higher-powered study.) I nonetheless discuss OBA in the hope that it provides insights.

Peripheral vision has already explained several supposedly attentional phenomena; what about OBA? The horizontal-vertical asymmetry in OBA mirrors peripheral vision's horizontal-vertical asymmetries (Strasburger, Rentschler, & Jüttner, 2011): both acuity and crowding are worse at a given eccentricity in the vertical direction than in the horizontal direction. In addition, OBA experiments typically have more clutter between the cue and the different-object target than between the cue and the same-object target, which might make judging the latter easier due to crowding. In the two-rectangle stimuli in Figure 6A, for instance, the cue and same-object target have nothing but blank space between them, whereas the region between the cue and different-object target contains the edges of both objects. Crowding might not prohibit identification of the target but rather might necessitate additional time to integrate information from the stimulus, slowing identification.

However, drawing parallels between crowding and object-based attention presumes that the observer fixates on or near the cue. If they fixate the central "+", as directed, all targets would appear diagonal relative to the fixation, one would not expect a horizontal-vertical asymmetry due to peripheral vision, and the same- and different-object targets would usually be equally crowded. However, the vast majority of object-based attention experiments have not used an eye tracker (including Francis & Thunell, 2022), so many of the classic results could have had a peripheral confound. Observers need only fixate nearer the cue on a small number of trials for crowding to explain the effects. The experimental design typically allows more than enough time for a saccade, and researchers typically find an OBA effect only under that condition (e.g. Egly et al., 1994; Francis & Thunell, 2022). Furthermore, the observer would benefit from breaking fixation; with cue validity as high as 75-80% (e.g. Egly et al., 1994; Francis & Thunell, 2022), fixating the cued location often means fixating the target, making the task easy. Comparison tasks may less often lead to OBA effects because with two task-relevant locations the observer less obviously benefits from breaking fixation.

More recent work questions object-based attention by putting target and cue on the same object and varying the complexity of the

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

intervening object (Figure 6B; Chen et al., 2020). Observers perform better when the object between cue and target is simple, and worse when it is complex, in line with a crowding explanation.

Taken together, the so-called object-based attention phenomena may instead derive from peripheral vision, with the small effects occurring because observers break fixation on a small fraction of trials. This is certainly not to say that all processing is location-based rather than object-based. One would expect some object-based effects from the proposal in Section 3.4. Rather, it is not clear at this point that one needs to include amongst the critical phenomena the automatic spread of attention from a cued location to the rest of the object.

2.3.Cued search, object recognition, and ideal observers

The cued search methodology provides another interesting cueing manipulation. The experimenter flags a subset of display items as potential targets. Observers search faster through this subset than through the complete set; performance often appears equivalent to searching through a display containing only the cued items (Grindley & Townsend, 1968; Davis, Kramer, & Graham, 1983; Palmer, Ames, & Lindsey, 1993). In cases in which the presence of uncued items negatively impacts performance, researchers have suggested that lateral masking — a term previously used somewhat interchangeably with "crowding" — degrades search in the more cluttered display (Eriksen & Lappin, 1967; Eriksen & Rohrbaugh, 1970). Nonetheless, because cued search keeps the stimulus constant while varying the cue, the effects cannot be purely sensory.

James (1890) describes the "taking possession by the mind" of a subset of objects as happening at the *expense* of perception of others. Because the target always appears within the cued subset, these particular experiments do not provide evidence of any cost of attending to the subset. Nonetheless, Palmer et al. (1993) suggest the results provide a clear example of attention: one attends to a subset of the items, as both evidenced by and leading to faster search times. (Zivony and Eimer, 2021 have criticized this tendency to use an experimental result to infer both that attention occurred, and that attention was the cause of that result. Hommel et al., 2019, make a similar point.)

Attempting again to describe these results without reference to attention, one might say that cueing changes the priors for likely target locations (i.e. where one expects to find the target before getting any evidence from the stimulus itself), and that that effectively changes the task from "search all items for the target" to "search these items for the target." Those changes matter: people perform better when they know more about where to search. This rephrasing raises the question of whether cueing affects decisional rather than perceptual processes.

Palmer et al. (1993) asked whether better performance searching through a subset of items necessarily implies a limited capacity perceptual mechanism that samples information only from the cued subset. To do this, they utilized ideal observer analysis. An ideal observer is a model that performs a task optimally, given the information available and certain assumptions. Palmer et al. (1993) showed that an unlimited capacity ideal observer explains their results, without need for a perceptual attention mechanism like early selection. Knowing which subset contains the target reduces errors by allowing decision processes to ignore false alarms from non-cued distractors, improving performance. Whereas my lab began by examining what phenomena peripheral vision explains, Palmer et al. (1993) — and others, e.g., Anderson (2011) — suggest the important step of examining what phenomena Bayesian decision theory can explain, i.e. to what degree any intelligent system would exhibit the behavior.

Other researchers have used similar experiments and modeling to argue that search results do demonstrate limited capacity (e.g. Palmer, Fencsik, Flusberg, Horowitz & Wolfe, 2011). The question becomes whether one can make sense of the conditions under which this occurs (and occurs without a sensory confound like peripheral crowding). One hint perhaps comes from Palmer et al.'s (1993) observers, who complained about the difficulty of searching within a subset of four items. Other research has also pointed to limits to an observer's ability to respond to an arbitrarily complex cue (e.g. Eriksen & Webb, 1989; Gobell, Tseng, & Sperling, 2004; Franconeri, Alvarez, & Cavanagh, 2013). We should consider these limits as some of the critical phenomena when developing a new understanding of limited capacity and the associated mechanisms.

A related issue arises when pondering the role of attention in ordinary object recognition. Does one *select* dog-like features to recognize a dog? Modern statistical-learning-based classifiers would distinguish a dog from a cat by making an intelligent decision based on *both* dog-like and cat-like features. Dog-recognition mechanisms fundamentally involve not-dog features. The selection/attention terminology reduces clarity, as it implies particular mechanisms that allow higher-level processing of some features but not others. On the other hand, to the extent that the visual system performs sub-optimally at object recognition, in a way not explainable by purely sensory limits, such phenomena could inform our understanding of capacity limits and attention.

Both of these examples use "attention" to mean something like "use top-down knowledge to perform a task." While that concept might deserve its own word, our goal here is to collect phenomena that demonstrate capacity limits and in doing so understand the special mechanisms for dealing with those limits. As such, we clearly should not include phenomena explainable by an unlimited-capacity ideal observer. The human visual system must make use of top-down information, and any viable theory of vision would include such a mechanism. There is no need to elevate such a mechanism by calling it "attention".

2.4. Mental marking

Another set of cueing tasks falls in a category one might call *mental marking* (Figure 7). Loosely speaking, this refers to tasks in which the observer gives some sort of special status to cued locations, so as to make judgments about those locations. Huang and Pashler (2007) enumerate a number of such tasks. Multiple object tracking (MOT) provides a classic example (Figure 7E). The observer views a set of items, with a subset of them cued. The cue disappears, and the observer must track the cued subset as the objects move, ultimately identifying the tracked objects. With no visible cue during the object motion, one might think of the observer as mentally marking the items to track. Many of the classic visual cognition tasks of Ullman (1996) fall into this category, as one can interpret them as asking the observer to mentally trace along a path (Figure 7A-C).

Should we consider performance on mental marking tasks in our list of critical attentional phenomena? While most tasks examined in this paper aim to better perceive the details of the attended object(s), mental marking tasks instead typically have the goal of keeping track of a subset of location(s) and making judgments *about those locations* (Huang and Pashler, 2007, draw a

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

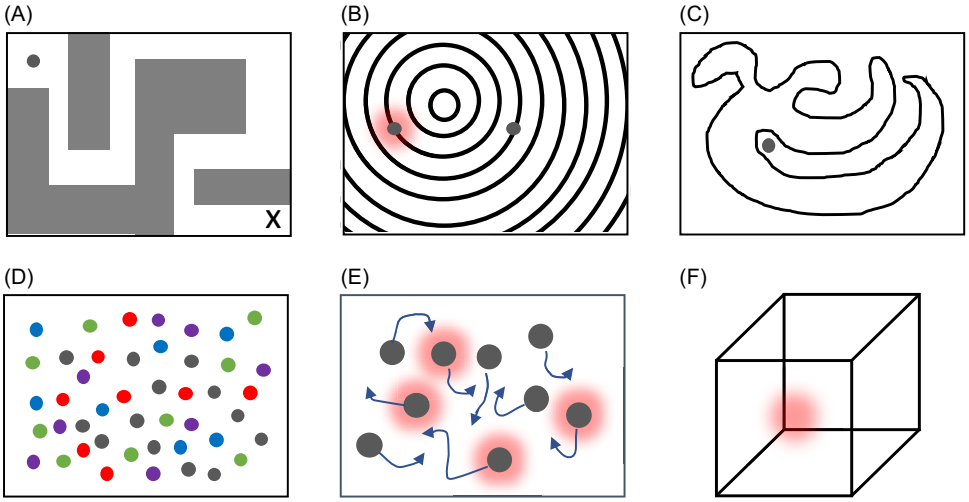


Figure 7. Mental marking tasks. (A) Is there a path between the dot and "x"? (B) Do the two dots appear on the same line? (C) Does the dot lie within the closed curve? (D) What shape do the red dots form? (E) Multiple object tracking. (F) Bistable Necker cube. Reddish glow shows loci of attention described in the text.

similar distinction). If these tasks have different goals, they might also probe different limits and mechanisms. To examine this possibility, consider two kinds of evidence for limits in mental marking tasks: changes in the percept and performance limits.

2.4.1. Changes in the percept

A mentally marked object can appear slightly higher in contrast (Carrasco, 2011), closer or farther (Huang & Pashler, 2009), or a different size (Anton-Erxleben, Henrich, & Treue, 2007), or shape (Fortenbaugh, Prinzmetal, & Robertson, 2011). Physiology research finds similar contrast effects (Reynolds, Pasternak, & Desimone, 2000; Martinez-Trujillo & Treue, 2002). Such perceptual changes might be detrimental when discriminating the features of the cued object but could benefit keeping track of or making judgments about the cued location(s). In addition to these relatively subtle perceptual changes, the locus of spatial attention to a bistable figure (Figure 1F) can strongly bias the percept in favor of one interpretation over the other (e.g. Tsal & Kolbet, 1985; Peterson & Gibson, 1991). One must, of course, take care to distinguish attentional effects from fixational; in the Necker cube (Ellis & Stark, 1978; Kawabata, Yamagami, & Noaki, 1978) and wife/mother-in-law stimuli, among others (Ruggieri & Fernandez, 1994), fixation location correlates with the dominant interpretation, suggesting a role for peripheral vision. However, changing the focus of attention (i.e., the mentally marked location) can affect the preferred interpretation of the Necker cube even when the observer cannot change fixation — for example, when the observer changes the marked location within an afterimage (see Piggins, 1979) – and experiments demonstrating physiological effects of attention monitored eye position.

It seems possible that changes in the percept might implicate different mechanisms and limits. In addition to perceiving the stimulus one might imagine the locations of the marked items, leading to subtle perceptual changes to those items. Tasks in which the observer must monitor a subset of items for an extended time might invoke mental imagery mechanisms, to help keep track of the monitored items. Implicating mental imagery might explain why effects on perceived contrast or size tend to be quite small compared to, say, dual-task or inattentional blindness effects. Larger perceptual effects could occur with bistable figures because a slight change in the internal representation might suffice to induce a large shift in the interpretation of an ambiguous figure.

2.4.2. Performance limits

In addition to perceptual effects, mental marking tasks also show performance limits. Observers have difficulty tracking more than a few objects in an MOT task (Pylyshyn & Storm, 1988). Huang and Pashler (2007) review evidence that observers can perceive only one group at a time. Observers might perceive the shape formed by just the red items (Figure 7D), just the blue items, or even by both the red and blue items, but have difficulty, Huang and Pashler argue, simultaneously perceiving both the shape formed by the red items and that formed by the blue items.

In curve-following tasks (Figure 7A-B), the observer must indicate whether a second point lies on the same curve or path as a given starting point. Completion time increases as a function of path length (Jolicoeur, Ullman, & Mackay, 1986), even when observers are forced to fixate (Houtkamp, Spekreijse, & Roelfsema, 2003; Jolicoeur et al., 1986). Furthermore, observers can better report a color lying farther along the path when it appears later in the trial, implying a shift in processing with time (Houtkamp, Spekreijse, & Roelfsema, 2003). Researchers have concluded that the tasks cannot be solved with parallel processing, and suggested a mechanism in which one places a mental mark at the first point (Figure 7B), then moves it along the path until either encountering the second point or reaching the end of the curve. This sounds a bit like moving a spotlight of attention, though Pooresmaeli and Roelfsema (2014) have instead argued for activation that spreads from the start until it mentally marks the entire perceived contour.

Peripheral vision likely plays a role in these phenomena, as they tend to utilize crowded stimuli. When cueing a subset of items (Figure 7D-E), other items often lie within the critical spacing of crowding. Intriligator and Cavanagh (2001) demonstrate that the critical spacing of crowding predicts spacing limits on their mental marking task.

Ullman (1984) suggested that difficulty on his visual cognition tasks (Figure 7A-C) might be governed at least in part by crowding. Crowding-related manipulations such as making the paths denser or more complex (e.g., more curved) lead to poorer performance (Jolicoeur, Ullman, & Mackay, 1991), and results are largely scale-invariant, i.e., independent of viewing distance (Jolicoeur & Ingleton, 1991), as expected for tasks governed by peripheral vision (Van Essen & Anderson, 1995; Rosenholtz, 2016).

Even the serial behavior in curve following does not rule out considerable parallel processing during each fixation. Both limited

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

and unlimited capacity models assume that given more time the visual system can integrate more samples of noisy observations to reduce uncertainty (Shaw, 1978). Presumably the crowded periphery induces more uncertain observations and therefore requires more integration time. Tasks that require peripheral information might lead to longer reaction times. Longer paths, being more likely to contain uncertain sections due to crowding, would take longer to process, even with parallel processing (Palmer et al., 1993 makes a similar argument about serial slopes in search). Accounting for the serial awareness of path color (Houtkamp et al., 2003) is less obvious, but perhaps awareness results from the completed processing of each path section. If so, the observer might take longer to become aware of later portions of the curve without serial processing, per se. Even if observers do serially follow curves, they might do so as one possible strategy. When forced to fixate, serially moving a mental mark may feel like a natural alternative to moving one's eyes (or pen) along the path under normal viewing conditions.

On the other hand, curve following encounters a more obvious limit: one cannot follow all curves at once. Ullman (1984) makes two arguments for why: First, it would be incredibly complex to solve curve-following by classifying all possible curves as passing or not passing through a pair of points, and one would probably not have built-in mechanisms to do this in parallel. Second, a flexible visual system should instead piece together generic components into *visual routines* that perform fairly arbitrary tasks. Some simple computations would happen in parallel across the field of view, for example estimating local orientation. However, *spatially structured processes* like curve following would be prime candidates for visual routines. Spatially structured processes require location information to specify the task; to ask about the curve passing through a given location one needs to specify that location. Ullman (1984) explicitly contrasts these limits with selective attention theory: one must select locations not to get information through the bottleneck, but rather because the locations determine the task.

Similar arguments could perhaps be made for other mental marking tasks. MOT, which one might think of as a sort of curve following, may also be an inherently complex task (Rosenholtz, 2020). Judging the shape specified by the red dots in Figure 7D in some sense requires tracing an imagined curve between neighboring dots. Huang and Pashler (2007) also seem to point to a task limit of some sort: they argue that observers cannot perform a task with the red group and simultaneously perform a separate task with the blue group. Perception of bistable figures may demonstrate an interesting version of the same limit: the visual system could theoretically provide us with both possible percepts simultaneously, but tends not to, at least if those percepts conflict in their interpretation of the stimulus (Neisser, 1976). These performance limits on mental marking seem worth considering amongst our critical phenomena.

2.5. Inattentional blindness

Inattentional blindness (IAB) experiments provide another explicit manipulation of attention. The observer performs a nominal task, and then must indicate whether they noticed an unexpected stimulus and/or make judgments about that stimulus. Often the experimenter also tests the observer in a dual-task condition, requiring them to both perform the original task and judge the now-expected stimulus (e.g., Mack and Rock, 1998). Across a variety of stimuli and tasks, observers more easily notice an expected stimulus than an unexpected stimulus.

The usual interpretation (Mack & Rock, 1998) presumes that the observer attends to the features and/or portion of the display relevant for the nominal task, and inattention to the unexpected stimulus renders them unable to perceive it. In the dual-task trials, the observer supposedly divides attention, making possible some perception of both stimuli. Inattentional blindness has been taken as evidence for little processing without attention. However, researchers have suggested other interpretations. Perhaps observers perceive the unexpected stimulus but do not become aware of it, and/or quickly forget (Wolfe, 1999). This might explain reduced inattentional blindness for meaningful stimuli like one's own name, as well as priming effects (Mack & Rock, 1998). Some IAB paradigms attempt to overcome these issues by making the unexpected stimulus sufficiently shocking (a gorilla, a unicycling clown, money on a tree) to make forgetting unlikely; nonetheless, observers often fail to become aware of the unexpected stimulus (Simons & Chabris, 1999; Hyman, Boss, Wise, McKenzie, & Caggiano, 2010; Hyman, Sarb, & Wise-Swanson, 2014). Interestingly, in some studies, even though observers fail to report the unexpected stimulus, they move to avoid a collision, suggesting that perception occurs (Tractinsky & Shinar, 2008; Hyman et al., 2014). Any viable explanation for IAB also needs to explain this perception without awareness.

Peripheral vision seems unlikely to explain the classic Mack and Rock (1998) experiments. Their primary task discriminating the lengths of two crossing lines likely encourages center fixation, and dual-task performance suggests the IAB is not purely perceptual. Peripheral vision may be one factor in the invisible gorilla phenomena (Simons & Chabris, 1999), or blindness to scene cut errors (Levin & Simons, 1997), as examples with real-world stimuli have typically not tracked the observer's gaze (Rosenholtz, 2020).

In describing these results without reference to attention, one might say that the observer has a nominal task and performs poorly at a surprise task (noticing the unexpected stimulus). When the experimenter later informs the observer about the latter task, the observer can to some degree perform both that and the nominal task. On the other hand, observers can complete tasks such as safely navigating the world – moving to avoid a collision – even when not explicitly informed of those tasks.

Framed this way, inattentional blindness does not appear particularly surprising. Surely one can often perform a task better if one knows the task. Inattentional blindness was surprising in part because observers missed supposedly automatically processed stimuli. If one automatically processes salient items, then one should notice them even when engaged in another task, but this not the case for 25% of observers (Mack & Rock, 1998). However, as previously noted, later evidence has not supported the notion of perception that occurs automatically and without requiring attention.

If no perception occurs automatically, then perhaps we should more usefully think of all perception as resulting from performing a task. If the visual system had no limits on the tasks it could simultaneously perform, then it would perform all tasks, and an observer would have no trouble noticing an unexpected item. It should come as no surprise that the visual system does have task limits. The visual system must constantly choose what task to do next among many options.

2.6. Dual task and multitasking in general

2.6.1. Illusory conjunctions

Another manipulation of task occurs in dual-task experiments. One interesting class of dual-task results concerns illusory

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

conjunctions. Experimenters show a rapidly presented display of different-colored letters and ask observers to report the identity and features (color, location) of as many as they can. Observers often report illusory conjunctions, i.e., the features of one item paired with the identity of another. Early experiments used rapid displays but no explicit attention manipulation (e.g., Snyder, 1972; Treisman & Schmidt, 1982). Some of these experiments used crowded peripheral stimuli (Snyder, 1972; Cohen & Ivry, 1989), making it possible that peripheral vision plays a role. Experiments and modeling have shown that peripheral vision can make the pairings of identity, color, and location ambiguous, leading to illusory conjunctions (Poder & Wagemans, 2007; Rosenholtz, Huang, Raj, et al., 2012; Keshvari & Rosenholtz, 2016). However, illusory conjunctions cannot purely be a peripheral phenomenon. Treisman and Schmidt (1982) briefly presented foveal, colored letters, flanked by black digits. Observers first had to accurately report the numbers, then the position, color, and identity of as many of the letters as they could. Observers incorrectly paired color and letter identity roughly a third of the time, considerably more often than they reported an absent feature, e.g., a correct letter in a non-present color. Illusory conjunctions occur even in the fovea. Treisman and Schmidt (1982) conclude that correct feature binding requires attention, consistent with feature integration theory (Treisman & Gelade, 1980).

This may, however, be a Goldilocks effect: Experimenters might initially choose a display time that is too long — observers make no errors — and conclude that this did not adequately strain attentional resources. Next, they try a display time that is too short; observers randomly guess letters. Clearly this impairs one's ability to study the effects of attention. Somewhere in between the display time is just right, and observers make in-between sorts of errors. Illusory conjunctions would likely dominate those errors. This is a variant of asking what an ideal observer would do: would any reasonable theory predict different results? In fact, Treisman and Schmidt (1982) describe just this sort of situation:

> It is worth pointing out a problem . . . in designing experiments on illusory conjunctions . . . (1) The theory claims that conjunctions arise when attention is overloaded; we therefore need to . . . use brief exposures . . . (2) However, illusory conjunctions can be formed only — from correctly identified features; . . . the briefer the exposure, the poorer the quality of the sensory information . . . Thus we were forced to trade off the need to load resources against the risk of introducing data limits . . . we controlled exposure durations separately for each subject in order to produce a feature error rate of about 10%.

This reasoning suggests caution in interpreting illusory conjunction experiments in terms of the need for attention to bind features.

2.6.2. Dual task

Considering dual-task performance more generally, one can find results more relevant to the present purpose. On the one hand, one cannot interpret easy vs. hard dual-task results in terms of whether the peripheral task does or does not require attention (Section 1.1). Rather, peripheral vision has been surprisingly predictive of dual-task difficulty. A limited-resource account might suggest that hard peripheral tasks require more resources, making performance more impaired in dual-task conditions; a satisfying explanation if only we could make the nature of the limited resources less vague.

On the other hand, peripheral vision cannot explain the more basic effect: that dual-task performance is often worse than single task. Nor would an unlimited-capacity ideal observer predict worse dual-task performance. Two different parts of the brain could do the two different tasks, with no cost to doing both. Dual tasks must encounter some additional limit.

2.6.3. Checking our understanding of multitasking in the real world: Driving

As a preliminary summary of rethinking attention, consider visual perception in the real-world task of driving a car. The literature contains a number of puzzles. Reporting in mainstream media paints a dire portrait of distracted driving, seemingly based on the classic attention paradigm: perception is poor without attention, and inattentional blindness and dual-task difficulty prove the danger of distracted driving. However, much as distracted driving is concerning, this would seem to overstate the case. Driving itself inherently involves a great deal of multitasking: one must stay in one's lane, navigate turns, avoid collisions, watch out for hazards, and notice road signs and traffic lights. Nonetheless, U.S. driving statistics from recent years indicate on average over 600,000 vehicle-miles driven per reported accident (Stewart, 2022, March), though this likely overestimates driving safety due to under-reporting. Furthermore, drivers can safely navigate familiar routes without awareness of doing so – in other words with very minimal amounts of what a lay person would call "paying attention" – a phenomenon referred to as a *zombie behavior* (Koch & Crick, 2001). Does our re-examination of attention phenomena and theory help resolve these puzzles?

Some of the direst distracted driving predictions rely on the idea of early selection. Our rethinking based on our understanding of peripheral vision suggests that vision has access to considerable information across the field of view, as opposed to the tunnel vision predicted by early selection. Of course, more recent attention theories also allow for parallel computation of the gist of the driving scene. In fact, the notion of tunnel vision remains controversial (Young, 2012; Gaspar, et al., 2016; Wolfe, Dobres, Rosenholtz, & Reimer, 2017).

The risk of distractions such as cell phone use may instead arise due to peripheral vision, as the driver fixates away from driving-relevant information. Even a non-visual distraction can lead to a change in fixation patterns (Recarte & Nunes, 2003), potentially putting critical information in the periphery. Performance of driving-relevant tasks degrades in the periphery, though drivers perform reasonably well at detecting hazards (Huestegge & Bröcker, 2016; Wolfe, Seppelt, Mehler, Reimer, & Rosenholtz, 2019) and following lane markers (Robertshaw & Wilkie, 2008). The task-relevant multitasking required by ordinary driving would likely be safer since at least drivers would keep their eyes on the road.

Both inattentional blindness and dual-task results do suggest that perception faces limits beyond sensory encoding losses, and worryingly both phenomena suggest that perception can be limited even at the fovea. This is well known in the driving literature as "looked but failed to see" (see Wolfe, Kosovicheva, & Wolfe, 2022, for a recent review). On the other hand, I have suggested reframing inattentional blindness as occurring because knowing the task matters. Surely even the most distracted driver knows that their main task is to drive safely. In addition, not all dual tasks lead to worse performance than single tasks. Observers perform well at driving-relevant dual tasks such as getting the gist of a scene or identifying a salient item. Furthermore, in-lab semantic tasks may not generalize well to action-oriented dual tasks like lane keeping or braking to avoid a collision.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Box A. Comprehensive list of indicators that visual attention is in crisis

Signs visual attention is in crisis

Theory complexity outpaces predictive power
- New components not well specified nor integrated into the theory

Anomalies not explained by the old paradigm
- Scene perception easier than expected
- Set perception easier than expected
- No categorically automatic tasks
- Apparent use of peripheral vision in parallel
- Perception outside the nominal focus of attention
- Real-world vision successful

Methods not yielding the promised results
- Because peripheral vision a significant factor:
  - Search, dual-task
  - Object-based attention
  - Illusory conjunctions
  - Inattentional blindness
  - Curve following
- Because ideal observer predicts key results:
  - Subset search
  - Object recognition
  - Illusory conjunctions
- Physiological evidence of "selection" may be an effect of attention rather than a cause
- Eye tracking may not indicate the state of attention

Box B. Critical visual attention phenomena, including both evidence of additional capacity limits and rich perception in apparent contradiction to such limits.

Evidence of additional limits and their nature:	**Proposed limit(s) and mechanisms must be consistent with:**
Dual task:	Perception outside the nominal task focus:
Often worse than single task	Zombie behaviors
Peripheral vision a factor	Awareness of rich percept
Cost to performance even for foveal tasks	Distraction in cueing tasks
No tasks automatic under all conditions	Good at scene perception, set perception
Knowing the task matters:	Awareness not needed for perception
Inattentional blindness	Success of real-world vision
Limits to cue complexity:	
Spacing, hemifield	
Arrangement	
Multiple object tracking	
Cognitive load effects	
Limits to visual marking:	
Time to trace a curve	
Complexity of marked regions	

In summary, there certainly remain reasons to worry about distracted driving, from improper fixations to other losses that can affect performance even of foveal tasks. However, the available information may be consistent with good performance at the multitasking involved in ordinary driving, and with some zombie behaviors.

3. Thoughts and a proposal

This section summarizes the results of reexamining a number of purported attentional phenomena. Reconsidering these phenomena revealed additional signs of crisis. See Box A for the complete list. Box presents the resulting lists of critical phenomena. Rephrasing these phenomena in a more mechanism-agnostic way points to a potentially coherent capacity limit on visual perception, and perhaps on cognition more broadly. I will present some thoughts about the nature of that limit, and possible underlying mechanisms.

3.1. Is there an additional limit?

I began by being open to the possibility of no additional capacity limit with an attention-like or selection-like mechanism. In examining the need for an additional limit, and the nature of that limit, I have argued against including some supposedly attentional phenomena in our list of critical phenomena, at least to begin with. In some cases, peripheral vision may have been the dominant factor, rather than attention (e.g., search, scene perception, change blindness, and object-based attention). Other phenomena could be at least partially explained by an unlimited-capacity ideal observer (e.g. some cued search), and one need not posit an additional attentional mechanism. In a related point, I have argued against

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

associating attention with mechanisms that any reasonable model of vision would include (use of top-down information and planning eye movements).

However, many phenomena seem to point to a possibly coherent additional limit. Box (left column) shows my conservative list. I have suggested that to synthesize these phenomena it helps to assume that all perception results from doing a task and that there are limits to task performance. If so, knowing the task should often matter (inattentional blindness, cued search), and dual-task performance would often be worse than single-task. Task limits could restrict the pattern of items observers can monitor or mentally mark, or the number of curves they can simultaneously follow. Visual search and change detection are likely subject to such task limits as well. Task limits are compatible with cognitive load phenomenology (e.g. Lavie et al., 2004) meaning that task limits might apply to more than just visual processing. On the other hand, many studies have suggested that attention modestly changes appearance (e.g., increases apparent contrast or size), corresponding to similar changes in physiological response. These results seem harder to think of in terms of task limits and may point to different limits or mechanisms.

The reader may be interested in comparing these two lists to those in other recent papers questioning attention (Anderson, 2011; Hommel, et al., 2019; Zivony & Eimer, 2021). Can a single additional limit explain both vision's failures and successes, and if so, what is its nature and associated mechanisms? The following subsections discuss two alternative proposals and indicate my current favorite.

3.2. Is inattention like looking away?

Contemporaneous work proposed that in the absence of attention, vision might only encode summary statistics (Rensink, 2000; Treisman, 2006; Oliva & Torralba, 2006; Wolfe et al., 2011). Arguably this misattributed summary statistic-like effects to inattention rather than peripheral vision. Nonetheless, the idea that covert and overt attentional mechanisms might share more than priority maps (Section 1.2) has a certain appeal, not least because of the success of summary statistics in making sense of diverse phenomena.

How might this work? Attention could narrow the pooling mechanisms that lead to crowding. Summarization would occur over a larger region when not attending. Some behavioral and physiological evidence seems consistent with receptive fields changing size, whether due to attention (Moran & Desimone, 1985; Desimone & Duncan, 1995), or training (Chen, et al., 2019). Researchers have even proposed that the pooling mechanisms underlying peripheral crowding might vary depending upon the perceptual organization of the stimulus (Sayim, Westheimer, & Herzog, 2010; Manassi, Sayim, & Herzog, 2012; Manassi, Lonchampt, Clarke, & Herzog, 2016). Adjusting the size of the pooling regions need not require rewiring new receptive fields on the fly. Rather, attention could reweight receptive field inputs, or change the effective size through interactions between neighbors. In the latter case, attention might enable use of multiple overlapping receptive fields to better reason about the stimulus. The presence of a cat and a boat within a single receptive field (RF) could prohibit identifying the cat from that RF alone. Another RF might contain only the boat, leading to its identification. Together, the two receptive fields could explain away the boat features, effectively – but not actually – reducing the size of the first receptive field in order to identify the cat. Vision might be limited by the arrangement of receptive fields and the complexity of the computations that combine their information, in line with suggestions from Franconeri et al. (2013).

Despite my early advocacy for this flexible pooling theory (Rosenholtz, 2011; Rosenholtz, Huang, & Ehinger, 2012), I became disillusioned for two main reasons. First, success of the theory depends greatly on the specifics regarding the size and layout of the pooling regions. Crowding experiments – with observers attending to the peripheral target – presumably probe the *minimum* pooling size. How much larger would pooling regions grow without attention, and at what cost to performance? Furthermore, Freeman and Simoncelli (2011) found that the same set of pooling regions predicted performance at both peripheral letter identification and a scene metamer task. This suggests that employing focused versus diffuse attention has minimal impact on the pooling. What happens to the overlap between neighboring pooling regions as they change size? This affects the available information (Chaney et al., 2014). The lack of details makes it hard to get intuitions. Inattention would degrade central vision, but otherwise its impact remains unclear.

Second, after a year of rethinking attention from the ground up, this theory felt like epicycles. A flexible pooling mechanism with better encoding at attended locations still essentially treats attention as spatial selection, gating the information available to some other mechanism that actually performs the task. This seems likely, yet again, to require additional mechanisms to address, for example, non-spatial selection of task-relevant features. If attention merely gates access to other processes that perform the task, do those latter processes themselves have capacity limits, as suggested by cognitive load effects?

Research into visual attention has long assumed that attention acts as a gate, but our critical phenomena do not obviously provide evidence for such a mechanism. A significant amount of processing appears to happen in parallel, and outside of the nominal task focus. This would seem surprising if the attention gates access to higher-level processing. Yet a working human visual system likely demands such parallelism and imperfect focus. Dual-task and cognitive load phenomena seem to suggest a gradation of available information rather than a gate. Nonetheless, the view of attention as a gate has persisted, even in papers challenging attention theory (Hommel, et al., 2019; Zivony & Eimer, 2021). A flexible pooling theory takes a step away from attention as gate, allowing access to some information across the field of view regardless of task relevance. The pooling mechanisms actually change the encoding of the visual information, rather than merely gating access to higher-level processes. Having pondered this step, perhaps we should consider an attention theory without any gate.

3.3. If not a gate, then what?

Without a gate, one must explain why observers do not perceive all available information, and in what way attention mechanisms address limited capacity. Perhaps perception results from doing a task, and tasks have limits. Gate theories suggest that one chooses a location, object, or feature for further processing by some separate part of the visual system. Instead, perhaps one chooses a *task*. I find it natural to equate selecting a task and setting up the classifier or mechanism to perform that task. Selection and performing a task might be *inseparable*. Executive functions, rather than choosing what information to gate based on task demands, rewards, and so on, might instead choose what task to perform. The observer asks a

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

question about the visual world – chooses and performs a task. They receive an answer: the percept.

The limit on tasks cannot simply be on the number of simultaneous tasks, nor the number of items processed, as both numbers depend on the specifics of the stimuli and nominal task (e.g. Alvarez & Franconeri, 2007; Franconeri et al., 2013; VanRullen et al., 2004). Nor can the limit be on overall task difficulty, as dual-task experiments controlled for difficulty in the component tasks by varying display time (e.g. VanRullen et al., 2004).

3.4. A limit on decision complexity

Perhaps the limit applies to a specific kind of task difficulty. Easy classification tasks look essentially the same in some high-dimensional feature space. The point clouds corresponding to chairs and rabbits will be well separated, and a simple classifier can discriminate between them with a low error rate. Hard tasks, however, can be hard for different reasons. A simple linear classifier might discriminate between a coffee table and a dining room table but overlap between the point clouds might inherently lead to errors. Such a task is data-limited but simple. At the other extreme, one might perfectly distinguish between two classes, but only if one used a complex — e.g., high-dimensional or wiggly — classification boundary. Perhaps humans are limited in the complexity of our decision boundaries.

This would immediately explain the difficulty of many dual tasks relative to their component single tasks. If each single task is two-alternative forced-choice, the dual task must distinguish between four possibilities. The decision boundary will be inherently more complex, and may often encounter limits, leading to poorer performance.

Earlier papers discuss this proposal in more detail (Rosenholtz, 2017; Rosenholtz, 2020). We can currently only speculate about what might be the nature of a limit on decision complexity. It could be the number of dimensions or neurons used to perform a task; the curvature of the decision boundary; or the number of linear hyperplanes needed to approximate the ideal boundary. If optimal task performance exceeded that limit, one would have to simplify the task, leading to errors. More complexity might also require more effort. However, in many cases, for example without time constraints, one could next perform a different task, thus further probing the available information, and eventually piece together a reasonable approximation of the more complex task.

Unlike the flexible pooling proposal, one can make reasonable qualitative guesses about decision complexity even without nailing down the details. Presumably, the visual system has developed to make many natural, ecologically important tasks simple, such as getting the gist of a scene. Hard tasks simultaneously operating on multiple items might be complex, such as following multiple curves, monitoring a complex set of cued items, or using peripheral vision to find a target in visual search. Due to the unusual nature of the peripheral encoding, hard peripheral tasks might often be complex. Gestalt grouping of stimuli might make tasks simpler, and so on.

Are these ideas merely a repackaging of the concept of "late selection"? While selecting a task based on the output of completed sensory processing does indeed sound "late," the two theories then diverge. In the classic late selection theory (as exemplified by Deutsch & Deutsch, 1963), processing proceeds until meaning is extracted, at which point the system selects relevant parts for access to short-term memory and awareness. In contrast, in the theory proposed here, meaning comes from choosing the task. Moreover, classic late selection inefficiently identifies all objects only to discard much of this information. Instead, I propose that the visual system efficiently employs sensory processing to create a general-purpose representation, applicable to multiple tasks, from which it can selectively choose.

The concept of task selection aligns with the idea that the visual system resolves competition among potential interpretations of the current visual state, rather than competing sensory data (Krauzlis, Bollimunta, Arcizet, & Wang, 2014). Additionally, we can relate this to proposals of attention for action. Neumann (1987), for instance, emphasizes physical constraints: humans have a maximum of two hands and two legs, and can only express one word at a time. Consequently, the brain faces limits in generating simultaneous action plans; it must choose "what to do and how to do it." However, the present theory primarily focuses on decision-making limits and tasks with less physical demand

3.5. Thinking about tasks

If perception always results from doing a task, then we need to rethink the nature of tasks. Tasks must not be limited to discriminations between named categories, like "bee" vs. "fly," because perception is not. Consider a useful back-pocket task, which one might employ when initially free-viewing a scene: getting the gist. The precise nature of this gist has largely eluded vision science. We should ask what task might correspond to getting the gist or getting subjective awareness of a scene. Perhaps the visual system attempts to discriminate the scene from other possible scenes. Getting the gist might equate to approximately localizing the visual input in a high-dimensional representation space. The results of that gist-localization would provide a rich percept that is difficult to describe in words. When first viewing a scene, or getting the gist without performing any other task, the localization might be fairly precise; the observer would know a great deal about the scene (Figure 8A). That knowledge would suggest new tasks to probe the details. A difficult subsequent task, such as a fine-grained judgment of object pose, performed in conjunction with precise gist-localization might exceed the task-complexity limit. The observer would need to resort to less precise localization in the high dimensional space, leading to less understanding of the scene (Figure 8B), and in extreme cases to inattentional blindness. Back-pocket tasks would not be automatic, per se, but would depend upon the complexity of the rest of the task. Shifting the primary task from moment to moment, while continuing to gather whatever gist information one could, given complexity limits, would provide stability of the percept over time. The notion of more precise localization in a higher dimensional space can be thought of as a generalization of the idea of attention leading to more precise spatial location of receptive fields (Section 3.2).

If we choose to equate choosing and performing a task with attention, then all perception requires attention. The observer may be unaware of the visual system's task, which may often differ from the nominal or conscious task. The observer may simplify a too-complex nominal task. Conversely, the observer may typically do more than the nominal task because of the benefits of perceiving one's environment and not only the prioritized object. The brain may well be built to make ignoring everything but a single object more complex, requiring more effort. The percept resulting from performing the task might also be unconscious, for example in zombie behaviors.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

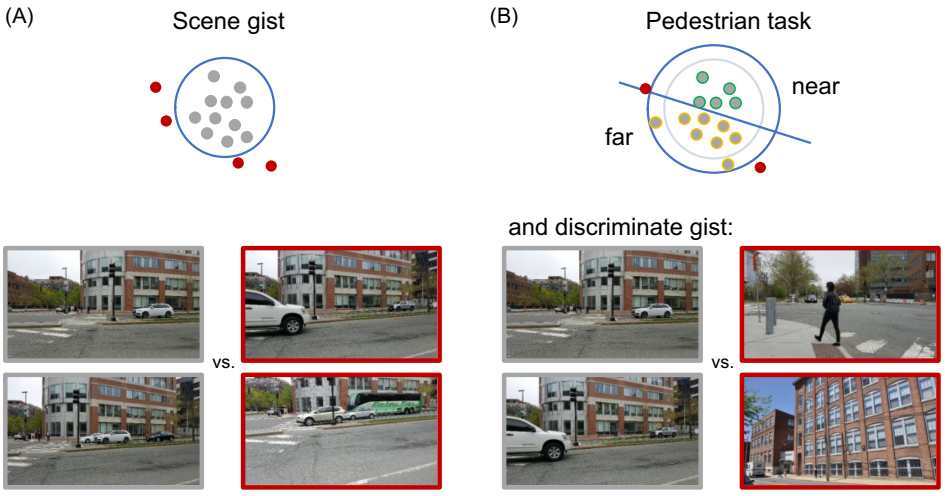


Figure 8. (A) Perhaps the gist of a scene is defined by the scenes the observer can discriminate from that scene (red) as opposed to those confused with it (gray). (B) When judging whether pedestrians are near (green outline) or far (yellow outline), the visual system might still get the gist, discriminating the current scene (gray) from others (red), albeit more coarsely than in (A).

4. Conclusions and lessons learned

Our reexamination of visual attention was originally motivated by the disruptive question: might supposedly attentional phenomena instead be explained by peripheral vision? Many of the supposed dichotomies necessitated by early selection theory may instead have arisen from easy vs. hard peripheral tasks. Tasks that utilize information readily available in peripheral vision would appear automatic and parallel, whereas tasks for which peripheral vision lacks information would appear serial and effortful. Given a continuum of difficulty for peripheral tasks, one would not expect a strong dichotomy, in agreement with other work calling these dichotomies into question (e.g. Wolfe, 1998; Anderson, 2011). At minimum, vision scientists need to worry about where the observer fixates, and what information this makes available across the field of view. Because peripheral vision appears to be a critical factor for many phenomena, we also need to worry about the specifics of the stimuli and tasks, which affect peripheral crowding in a complex way.

Rethinking visual attention from the ground up by asking another disruptive question: is there anything one might want to call "attention"? Though I eventually enumerated phenomena that do seem to show additional limits, asking disruptive questions forces one to rethink one's assumptions. I hope that by elucidating my process of identifying critical phenomena others are inspired to try it themselves if they disagree with my lists. Readers may also want to examine whether their favorite theory can account for these critical phenomena. The third disruptive question asks whether a single limit can explain these phenomena, with a single — possibly complex and flexible — mechanism. I suggest thinking of all perception as resulting from performing a task and propose new ways of thinking about back-pocket tasks like getting the gist of a scene (see also Rosenholtz, 2020). As an initial proposal, there might exist a unifying limit on task complexity. Though much work needs to be done to flesh out and test this proposal, I hope that thinking of capacity limits and "attention" in this quite different way can help move the field forward.

Acknowledgements. Thanks to several generations of lab members for their patience, efforts, and ideas. Thanks most recently to James DiCarlo and Edward Adelson for helping frame this work within the context of philosophy of science, and to Wayne Wu for many useful conversations in preparation for our phiVis conversation at VSS 2023.

Funding statement. Funded by National Science Foundation Grant BCS-1826757.

Competing interests. The author declares none.

References

Al-Janabi, S., & Greenberg, A. S. (2016). Target-object integration, attention distribution, and object orientation interactively modulate object-based selection. *Attention, Perception, & Psychophysics*, *78*(7), 1968–1984.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14.

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*, 246.

Anton-Erxleben, K. A., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision*, *7*(11), 5.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Bacon, F. (2015). *The New Organon.* Centaur Editions. Retrieved from https://www.amazon.com/New-Organon-Francis-Bacon-ebook/dp/B018KZF722/ref=sr_1_1?crid=19NE8BW2RM3CA&keywords=new+organon&qid=1703640282&sprefix=alexa+robot%2Caps%2C89&sr=8-1

Balas, B. J. (2016). Seeing number using texture: How summary statistics account for reductions in perceived numerosity in the visual periphery. *Attention, Perception, & Psychophysics*, *78*(8), 2313–2319.

Balas, B. J., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *8*(12), 13. doi: 10.1167/9.12.13

Bouma, H. (1970). Interactional effects in parafoveal letter recognition. *Nature*, *226*, 177–178.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Perception & Psychophysics*, *60*(1), 1–12.

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Research*, *51*(13), 1484–1525.

Chaney, W., Fischer, J., & Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. *Frontiers in Integrative Neuroscience*, *8*(73), 1–11. doi: 10.3389/fnint.2014.00073

Chang, H., & Rosenholtz, R. (2016). Search performance is better predicted by tileability than by the presence of a unique basic feature. *Journal of Vision*, *16*(10), 13.

Chen, N., Shin, K., Millin, R., Song, Y., Kwon, M., & Tjan, B. S. (2019). Cortical reorganization of peripheral vision induced by simulated central vision loss. *Journal of Neuroscience*, *39*(18), 3529–3536.

Chen, Z., & Cave, K. R. (2019). When is object-based attention not based on objects? *Journal of Experimental Psychology: Human Perception and Performance*, *45*(8), 1062–1082.

Chen, Z., Cave, K. R., Basu, D., Suresh, S., & Wiltshire, J. (2020). A region complexity effect masquerading as object-based attention. *Journal of Vision*, *20*(7), 24.
Chong, S.-C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.
Chong, S.-C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *66*, 1282–1294.
Chun, M. M., Golumb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.
Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 650–663.
Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psych. Science*, *22*(9), 1165–1172.
Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*(5), 324–335.
Davis, E. T., Kramer, P., & Graham, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception and Psychophysics*, *33*(1), 20–28.
Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.
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.
Ehinger, K. A., & Rosenholtz, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision*, *16*(2), 13.
Ellis, S. R., & Stark, L. (1978). Eye movements during the viewing of Necker cubes. *Perception*, *7*, 575–581.
Eriksen, C. W., & Lappin, J. S. (1967). Selective attention and very short-term recognition memory for nonsense forms. *Journal of Experimental Psychology*, *73*(3), 358–364.
Eriksen, C. W., & Rohrbaugh, J. (1970). Visual masking in multielement displays. *Journal of Experimental Psychology*, *83*(1, pt. 1), 147–154.
Eriksen, C. W., & Webb, J. M. (1989). Shifting of attentional focus within and about a visual display. *Perception & Psychophysics*, *45*(2), 175–183.
Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Perception & Performance*, *18*(4), 1030–1044.
Fortenbaugh, F. C., Prinzmetal, W., & Robertson, L. C. (2011). Rapid changes in visual-spatial attention distort object shape. *Psychonomic Bulletin & Review*, *18*, 287–294.
Fougnie, D., Cockhren, J., & Marois, R. (2018). A common source of attention for auditory and visual tracking. *Attention, Perception, & Psychophysics*, *80*, 1571–1583.
Francis, G., & Thunell, E. (2022). Excess success in articles on object-based attention. *Attention, Perception, & Psychophysics*, *84*, 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.
Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.
Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*, 974–981.
Gaspar, J. G., Ward, N., Neider, M. B., Crowell, J., Carbonari, R., Kaczmarski, H., … Loschky, L. C. (2016). Measuring the Useful Field of View during simulated driving with gaze-contingent displays. *Human Factors*, *58*(4), 630–641.
Gingerich, O. (1975). 'Crisis' versus aesthetic in the Copernican Revolution. In A. Beer (Ed.), *Vistas in Astronomy* (Vol. *17*, pp. 85–94). Oxford: Pergamon.
Gobell, J. L., Tseng, C.-H., & Sperling, G. (2004). The spatial distribution of visual attention. *Vision Research*, *44*(12), 1273–1296.
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.
Grindley, G. C., & Townsend, V. (1968). Voluntary attention in peripheral vision and its effects on acuity and differential thresholds. *Quarterly Journal of Experimental Psychology*, *20*(1), 11–19.
Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception & Performance*, *35*(3), 718–734.
Heliocentrism. (2023, December 19). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Heliocentrism
Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*, 2288–2303.
Houtkamp, R., Spekreijse, H., & Roelfsema, P. R. (2003). A gradual spread of attention during mental curve tracing. *Perception & Psychophysics*, *65*(7), 1136–1144.
Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*(3), 599–631.
Huang, L., & Pashler, H. (2009). Reversing the attention effect in figure-ground perception. *Psychological Science*, *20*(10), 1199–1201.
Huestegge, L., & Bröcker, A. (2016). Out of the corner of the driver's eye: Peripheral processing of hazards in static traffic scenes. *Journal of Vision*, *16*(2), 11.
Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2010). Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology*, *24*(5), 597–607.
Hyman, I. E., Sarb, B. A., & Wise-Swanson, B. M. (2014). Failure to see money on a tree: inattentional blindness for objects that guided behavior. *Frontiers in Psychology*, *5*, 356.
Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cogn. Psychol.*, *43*, 171–216.
James, W. (1890). *Principles of Psychology*. New York: Holt.
Jolicoeur, P., & Ingleton, M. (1991). Size invariance in curve tracing. *Memory & Cognition*, *19*, 21–36.
Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, *14*(2), 129–140.
Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 997–1022. doi: 10.1037/0096-1523.17.4.997
Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature*, *387*(6635), 805–807.
Kanwisher, N., & Driver, J. (1992). Objects, attributes, and visual attention: Which, what, and where. *Current Directions in Psychological Science*, *1*(1), 26–31.
Kawabata, N., Yamagami, K., & Noaki, M. (1978). Visual fixation points and depth perception. *Vision Research*, *18*(7), 853–854.
Keshvari, S., & Rosenholtz, R. (2016). Pooling of continuous feature provides a unifying account of crowding. *Journal of Vision*, *16*(3), 39.
Koch, C., & Crick, F. (2001). The zombie within. *Nature*, *411*, 893.
Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not cause. (457-464, Ed.) *Trends in Cognitive Sciences*, 18(9), 457-464.
Kravitz, D. J., & Behrmann, M. (2014). Space-, object-, and feature-based attention interact to organize visual scenes. *Attention, Percetpion, & Psychophysics*, *73*(8), 2434–2447.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lamy, D., & Egeth, H. (2002). Object-based selection: The role of attentional shifts. *Perception & Psychophysics*, *64*(1), 52–66.
Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *J. Exp. Psych: Human Perception & Performance*, *40*(2), 471–487.
Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psych.: General*, *133*(3), 339–354.
Lettvin, J. Y. (1976). On seeing sidelong. *The Sciences*, *16*(4), 10–20.
Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, *4*, 501–506.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 435–441.
Mack, A., & Clarke, J. (2012). Gist perception requires attention. *Visual Cognition*, *20*(3), 300–327.
Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.
Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision*, *16*(3):35, 1–13.
Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10):13, 1–14.
Martinez-Trujillo, J., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, *35*, 365–370.
Matsukura, M., Brockmole, J. R., Boot, W. R., & Henderson, J. M. (2011). Oculomotor capture during real-world scene viewing depends on cognitive load. *Vision Research*, *5*(1), 546–552.
Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neuroscience*, *29*(6), 317–322.
McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, *71*, 926–940.
Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.
Neisser, U. (1976). *Cognition and reality*. San Francisco: Freeman.
Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer, & A. F. Sanders, *Perspectives on Perception and Action*, pp. 361–394. Hillsdale, NJ: Lawrence Erlbaum.
Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision 42*, 145–175. doi: 10.1023/A:1011139631724
Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Prog. Brain Res.*, *155*, 23–36.
Palmer, E. M., Fencsik, D. E., Flusberg, S. J., Horowitz, T. S., & Wolfe, J. M. (2011). Signal detection evidence for limited capacity in visual search. *Attention, Perception, & Psychophysics*, *73*, 2413–2424.
Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *J. of Exp. Psych.: Human Perception & Performance*, *19*(1), 108–130.
Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, J. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744.
Peterson, M. A., & Gibson, B. S. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape description. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 170–182.
Piggins, D. (1979). the influence of line reduction and image stabilization on Necker cube reversal: comments upon Ellis and Stark (1978). *Perception*, *8*, 719–720.
Poder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7*(2), 23.
Pooresmaeli, A., & Roelfsema, P. R. (2014). A growth-cone model for the spread of object-based attention during contour grouping. *Current Biology*, *24*, 2869–2877. doi: 10.1016/j.cub.2014.10.007
Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49–70.
Potter, M. C. (1975). Meaning in visual search. *Science*, *187*, 965–966.
Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 1–19.
Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*(2), 119–137.
Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, *40*, 1469–1487.
Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.
Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*, 703–714.
Robertshaw, K. D., & Wilkie, R. M. (2008). Does gaze influence steering around a bend? *Journal of Vision*, *8*(4), 18.1–13.
Rosenholtz, R. (2011). What your visual system sees where you are not looking. *Proc. SPIE 7865, Hum. Vis. Electron. Imaging, XVI*. San Francisco: SPIE.
Rosenholtz, R. (2014). Texture perception. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization* (pp. 167–186). Oxford, UK: Oxford University Press.
Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Rev. of Vision Sci.*, *2*(1), 437–457.
Rosenholtz, R. (2017). Capacity limits and how the visual system copes with them. *Journal of Imaging Science and Technology (Proc. HVEI, 2017)*, *7865*, 8–23.
Rosenholtz, R. (2020). Demystifying visual awareness: Peripheral encoding plus limited decision complexity resolve the paradox of rich visual experience and curious perceptual failures. *Attention, Perception, & Psychophysics*, *82*(3), 901–925. doi: 10.3758/s13414-019-01968-1
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*:13. doi: 10.3389/fpsyg.2012.00013
Rosenholtz, R., Huang, J., Raj, A., Balas, B., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4):14, 1–17.
Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of Vision*, *19*(7), 15.
Rousselet, G. A., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes. *Visual Cognition*, *12*(6), 852–877.
Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two, or four natural scenes in humans: the limits of parallelism. *Vision Research*, *44*(9), 877–894.
Ruggieri, V., & Fernandez, M. F. (1994). Gaze orientation in perception of reversible figures. *Perceptual and Motor Skills*, *78*, 299–303.
Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, *21*(5), 641–644.
Shaw, M. L. (1978). A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 586–598.
Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074.
Smith, M. E., Sharan, L., Park, E., Loschky, L. C., & Rosenholtz, R. (under revision). Difficulty detecting changes in complex scenes depends in part upon the strengths and limitations of peripheral vision. *Journal of Vision*.
Snyder, C. R. (1972). Selection, inspection, and naming in visual search. *Journal of Experimental Psychology*, *92*(3), 428–431.
Stewart, T. (2022, March). *Overview of motor vehicle crashes in 2020*. National Highway Safety Traffic Administration.
Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13.
Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606.
Tractinsky, N., & Shinar, D. (2008). Do we bump into things more while speaking on a cell phone? *Proc. CHI '08 Extended Abstracts, Alt CHI* (pp. 2433-2442). New York: ACM.
Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*, 411–443.
Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.*, *12*, 97–136.
Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.
Tsal, Y., & Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *Quarterly Journal of Experimental Psychology*, *37*(1), 25–37.
Ullman, S. (1984). Visual routines. *Cognition*, *18*, 94–159.
Ullman, S. (1996). Visual cognition and visual routines. In S. Ullman, *High-Level Vision* (pp. 263–315). Cambridge, MA: MIT Press.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Van Essen, D. C., & Anderson, C. H. (1995). Information processing strategies and pathways in the primate visual system. In S. F. Zornetzer, J. L. Davis, C. Lau, & T. McKenna (Eds.), *An Introduction to Neural and Electronic Networks* (2nd ed., pp. 45–76). San Diego, CA: Academic.
VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *J. Cogn. Neurosci.*, *16*, 4–14.
Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the Useful Field: Considering peripheral vision in driving. *Applied Ergonomics*, *65*, 316–325.
Wolfe, B., Sawyer, B. D., Kosovicheva, A., Reimer, B., & Rosenholtz, R. (2019). Detection of brake lights while distracted: Separating peripheral vision from cognitive load. *Attention, Perception, & Psychophysics*, *81*(8), 2798–2813.
Wolfe, B., Seppelt, B. D., Mehler, B., Reimer, B., & Rosenholtz, R. (2019). Rapid holistic perception and evasion of road hazards. *Journal of Experimental Psychology: General*, *149*(3), 490–500.
Wolfe, J. M. (1998). What can 1 million trials tell us about visual search. *Psychological Science*, *9*(1), 33–39.
Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting Memories* (pp. 71–94). Cambridge, MA: MIT Press.
Wolfe, J. M., Kosovicheva, A., & Wolfe, B. (2022). Normal blindness: when we Look But Fail To See. *Trends in Cognitive Sciences*, *26*(9), 809–819.
Wolfe, J. M., Vo, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84.
Wu, W. (2023, December 15). *We know what attention is!* Retrieved from *Trends in Cognitive Sciences*: https://doi.org/10.1016/j.tics.2023.11.007
Young, R. (2012). Cognitive distraction while driving: A critical review of definitions and prevalence in crashes. *SAE International Journal of Passenger Cars – Electronic and Electrical Systems*, *5*(1), 326–342.
Zhang, X., Huang, J., Yigit-Elliot, S., & Rosenholtz, R. (2015). Cube search, revisited. *Journal of Vision*, *15*(3), 9.
Zivony, A., & Eimer, M. (2021). The diachronic account of attentional selectivity. *Psychonomic Bulletin & Review* *29*(4). doi: 10.3758/s13423-021-02023-7

Open Peer Commentary

A new algorithm of human attention

Mario Belledonne* and Ilker Yildirim*

Department of Psychology, Yale University, New Haven, CT, USA
mario.belledonne@yale.edu
ilker.yildirim@yale.edu

**Article updated 10 December 2025

*Corresponding author.

doi:10.1017/S0140525X25000196, e131

Abstract

How do our goals continually impact perceptual processing? The answer could arise from a computational specification of perception in terms of visual tasks, or perhaps several mechanisms operating over specific contexts. Here, we suggest an alternative: *adaptive computation*, a new algorithmic account of attention that rations the general resource of perceptual computations according to their impact on decision making.

What is attention?

Rosenholtz's forceful arguments for distinguishing peripheral vision from "attention" contributed to the recent resurgence of this question. Undeniably, important progress has been made in understanding some aspects of the *multi-faceted* nature of attention: for example, computational mechanisms have been developed for salience-driven attention with significant explanatory power (Bruce & Tsotsos, 2005; Tsotsos et al. 2005). However, similar progress remains elusive for another complementary form of attention that has been foundational to Cognitive Psychology (Neisser, 1967): how our goals impact what we see. Therefore, it is no coincidence that the paradigms and phenomena that populate Box B in Rosenholtz involve exactly that form of attention. To address such phenomena of attention, Rosenholtz suggests an architecture in which perception is equivalent to performing a task, and tasks are constrained by their complexity.

A new algorithmic account of attention

Here, we highlight a recent computational mechanism of attention that can address the phenomena of Rosenholtz's Box B, including multiple object tracking (MOT) and inattentional blindness (IB). This new mechanism, called "adaptive computation," explains how attention continually focuses visual processing based on our goals. Unlike many past accounts of attention, this account starts from the perspective of a cognitive architecture where attention, perception, and decision-making form distinct but dynamically interacting mental modules. Unlike Rosenholtz's proposal of defining perception as a consequence of tasks, adaptive computation does not mix the computational objective of perception with that of decision-making. Perception performs approximate inference over a world model (a generative model over objects, their interactions across time, and their appearances), $Pr(S)$, given a sequence of sensory inputs X, resulting in mental representations (i.e., a posterior distribution), $Pr(S|\ X)$. Decision-making then uses the percept to inform an action in some task (e.g., target selection in MOT, or counting specific events in an IB paradigm), denoted $Pr(\pi \mid S)$.

The attention module implements an algorithm where the task shapes dynamic patterns of perceptual processing. The challenge in formalizing such an attention module stems from the need to jointly address two disparate aspects of attention. On one hand, attention involves a general resource, with some people performing consistently better than others in a variety of tasks and phenomena (Huang, Mo, & Liu, 2012; Pashler, 1997; Treviño et al., 2021), which Rosenholtz notes almost never gets to be precisely specified. On the other hand, attention evolves at the scale of visual processing, with a fine-grained ebb-and-flow across space (He, Cavanagh, & Intriligator, 1997) and time (Nobre & Ede, 2023). In sum, the attention module must implement an algorithm with core procedures that are both general, in that they are not hard-coded to a particular paradigm or task, and also pick up on nuanced and dynamic signals as tasks and scenes evolve.

What is the general resource?

We formalize the limited resource rationed by attention as *targetable singleton computations*, denoted C. Such computations form the individual operations underlying the algorithm approximating the percept $Pr(S|\ X)$ (e.g., a unit step of the MCMC algorithm, a feedforward pass in a neural network). These perceptual computations must be (1) targetable in that they can be applied to update the approximation of particular object or latent k, $C_k(S) \rightarrow S'$ where $S' = S \,/\, \{s_k\} \cup \{s_k{}'\}$, and (2) singleton in that any such application of said computation takes nearly identical

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

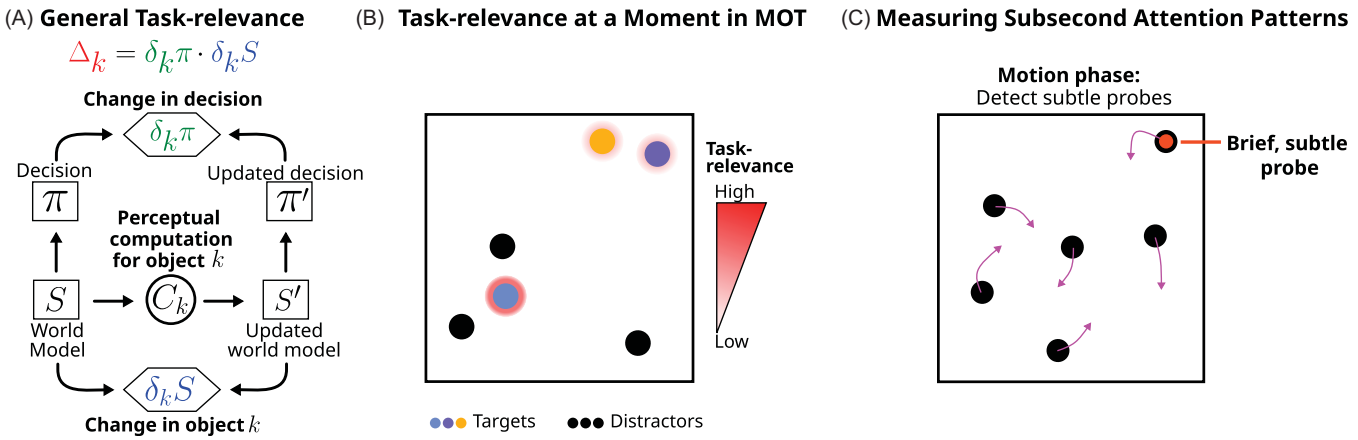


Figure 1 (Belledonne and Yildirim). (Belledonne et al. 2025). (A) Adaptive computation ratios perceptual computations to approximate object representations in a world model $S = [s_1, \ldots, s_N]$, which in turn informs decision-making outcomes π. Adaptive computation quantifies the task-relevance of a given object s_k with a general formulation: $\Delta_k = \delta_k\pi \cdot \delta_k S$, that integrates the impact of applying perceptual computation to that object (C_k) on decision outcomes ($\delta\pi$) and improvements in the world model ($\delta_k S$). Thus, an object is task-relevant if rationing additional perceptual computations can both readily improve the object's representation and doing so further inform decision-making. (B) Applied to a case study of MOT, adaptive computation generates intuitive patterns of object-level attention, with subsecond temporal resolution – all without the need for attention to explicitly access low-level details of perception or decision making (e.g., the distance between targets or distractors). (C) Probe detection adaptation of the standard MOT task to measure attention patterns throughout a single trial. In addition to the main task of target tracking, the motion phase of a trial also includes probes that briefly appeared on an object, with detection rates providing a measure of momentary attention. The patterns of task-relevance generated by adaptive computation strongly explained variability in human probe detection rates.

amounts of time regardless of the object or latent, $O(C_k) = O(C_j)$, where $O(\cdot)$ indicates computational complexity.

Critically, a targetable singleton computation has two consequences – one on perception and one on decision making. First, each C_k converges the current approximation of a percept, for example, by improving the estimate of the object's position and velocity. Importantly, some representations will require more computation to converge, with the rate of convergence denoted $\delta_k S$. Second, computations over some representations will, in turn, differentially impact decision-making, depending on how changes in the representation inform decision outcomes, denoted $\delta_k\pi$.

What is task-relevance *and how can it be tractably estimated?*

A novel formulation of task-relevance follows from this concrete specification of the general resource attention may allocate: The simple product of the two consequences of a targetable singleton computation, $\Delta_k = \delta_k\pi \cdot \delta_k S$, with $\vec{\Delta}$ denoting the task-relevance vector of all representations (Fig. 1A). Thus, a representation is only task-relevant if applying additional computations can readily improve its approximation and doing so informs decisions.

In order to be useful, attention must tractably estimate $\vec{\Delta}$ to ration imminent perceptual computations. The formulation of $\vec{\Delta}$ affords a simple estimation procedure with little overhead: aggregating the impact of recently applied computations across representations.

Accordingly, in a three-stage cycle, adaptive computation continually updates task-relevance to ration perceptual computations across objects and moments. The first stage involves a small amount of (i.e., quick) preattentive processing to initialize perceptual representations and decisions. In the second stage, adaptive computation uses the current $\vec{\Delta}$ to ration the total number of computations to use at that cycle (based on $||\vec{\Delta}||$) and to distribute these computations across representations (based on $\text{softmax}(\vec{\Delta})$), yielding a goal-conditioned approximation of S. The third stage aggregates the impacts of the rationed computations to update $\vec{\Delta}$ for the next cycle.

Every trial is an experiment unto itself

In Belledonne et al., (2025), we implemented and applied an adaptive computation architecture to the domain of MOT, explaining novel aspects of attentional dynamics.

Although adaptive computation only has access to the abstract measures of $\delta\pi$ (how much decisions change) and δS (how much object representations change), the architecture can recover rich and intuitive patterns of processing. For example, a target approaching a distractor (Fig. 1B, bottom left) will be task relevant ($\Delta_k > 0$) as additional computations will resolve the object's representation ($\delta_k S > 0$) and such updates will in turn alter confidence over selecting targets ($\delta_k\pi > 0$). However, two lone targets approaching each other (Fig. 1B, top right) would not warrant additional computations, as doing so would rarely alter target designation confidence (swapping object identities between targets does not impact expected performance). Critically, these object-level patterns of adaptive computation (determined by $\text{softmax}(\vec{\Delta})$) match the subsecond attentional dynamics in humans (measured via probe detection during the tracking phase; Fig. 1C).

Moreover, as difficulty fluctuates (e.g., via increasing object counts and speeds), the total number of computations exerted (determined by $||\vec{\Delta}||$) provides a computational account for why some trials feel more *effortful* than others: we find that those trials that humans subjectively judge more effortful are those in which adaptive computation rations more perceptual computation.

Embracing cognitive architectures to understand attention

Due to the modular nature of adaptive computation, it can extend to other domains via changing the perception and

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

decision-making modules, while preserving the core adaptive computation module and interface with targetable singleton computations C_k, and their two consequences, $\delta_k S$ and $\delta_k \pi$. Across the domains illustrated in Box B of Rosenholtz and others, adaptive computation can generate predictions of the moment-by-moment distribution of attention, perceived world models, plans to be executed, and computational signatures of subjective effort.

Acknowledgements. We thank Brian Scholl for helpful comments.

Financial support. I.Y. was supported by AFOSR grant # FA9550-22-1-0041.

Competing interests. None.

References

Belledonne, M., Butkus, E., Scholl, B.J., & Yildirim, I. (2025). Adaptive computation as a new mechanism of dynamic human attention. *Psychological Review*. Advanced online publication. https://psycnet.apa.org/doi/10.1037/rev0000572.

Bruce, N. D., & Tsotsos, J. K. (2005). Saliency based on information maximization. *Advances in Neural Information Processing Systems*, *18*, 155–162.

He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in Cognitive Sciences*, *1*(3), 115–121.

Huang, L., Mo, L., & Liu, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: an individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *38* (2), 414–428.

Neisser, U. (1967). Cognitive psychology. Appleton-Century-Crofts.

Nobre, A. C., & Ede, F. van. (2023). Attention in flux. *Neuron*, *111* (7), 971–986.

Pashler, H. (1997). *The psychology of attention*. The MIT Press. https://doi.org/10.7551/mitpress/5677.001.0001

Treviño, M., Zhu, X., Lu, Y. Y., Scheuer, L. S., Passell, E., Huang, G. C., Germine, L. T., & Horowitz, T. S. (2021). How do we measure attention? Using factor analysis to establish construct validity of neuropsychological tests. *Cognitive Research: Principles and Implications*, *6*, 1–26.

Tsotsos, J. K., Liu, Y., Martinez-Trujillo, J. C., Pomplun, M., Simine, E., & Zhou, K. (2005). Attending to visual motion. Computer Vision and Image Understanding, 100 (12), 3–40.

Starting a revolution with a refuted model?

Alban Bornet[a], Michael H. Herzog[b] and Adrien Doerig[c,d]*

[a]Data Science for Digital Health group, Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland; [b]Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; [c]Cognitive Computational Neuroscience lab, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany and [d]Bernstein Center for Computational Neuroscience, Berlin, Germany
alban.bornet@unige.ch
michael.herzog@epfl.ch
adrien.doerig@fu-berlin.de

*Corresponding author.

doi:10.1017/S0140525X25000147, e132

Abstract

Rosenholtz proposes to replace the concept of attention with peripheral vision models, such as the Texture Tiling Model (TTM). Here, we show that the TTM fails in many psychophysical studies due to its local, single-stage, feedforward, and low-level processing. Given that both attention and peripheral vision are unsettled fields, we argue that replacing one with the other is unwarranted.

Rosenholtz argues that the field of attention is in crisis and calls for a Kuhnian revolution to transform current thinking. She claims that many experimental results, usually thought to address attention, in fact come from limits in peripheral visual processing. In particular, Rosenholtz argues that the TTM (Balas, Nakano & Rosenholtz, 2009), which employs a single-stage computation of summary statistics over local image patches, can explain many phenomena traditionally attributed to attention. She suggests that "*one might reasonably call summary statistic encoding a paradigm shift, both for the science of visual attention and for vision more broadly.*"

Here, we do not debate whether there is a crisis in attention research. Rather, we review a series of experimental and computational results that directly contradict the TTM as a model of peripheral vision, and argue that the TTM is unlikely to be a good model of vision in general. While the TTM accounts for a range of psychophysical phenomena, it is far from a universal model of peripheral processing. In particular, the TTM cannot reproduce large sets of experimental results in crowding, for which it was specifically developed.

In crowding, the perception of a visual target deteriorates in the presence of flankers. In the TTM, crowding occurs when the target and flankers are combined in the same summary statistics, leading to a degraded representation of the target. A rich body of experimental and theoretical studies showed that this account is fundamentally limited because the global configuration of visual elements determines how local elements are processed (Manassi, Sayim & Herzog, 2012, 2013, Manassi et al., 2016). These effects of global configuration on visual processing are large, qualitative phenomena, and occur in diverse conditions, across multiple levels of processing (Manassi & Whitney, 2018), and even in other sensory modalities (Oberfeld & Stahn, 2012; Overvliet & Sayim, 2016). Hence, they reflect a general strategy for perceptual computations adopted by the brain. We propose that the TTM is unable to account for these crucial global effects for four main reasons:

1. *Locality.* The TTM assumes that peripheral vision is fundamentally constrained by local summary statistics computed over small patches of the visual field. Several studies showed that models based solely on local interference mechanisms, including the TTM, cannot explain global effects in crowding (Bornet, Doerig, et al., 2021a; Choung et al., 2021; Doerig, Bornet, et al., 2020; Doerig et al., 2019).
2. *Single stage.* Interference between visual elements in TTM occurs as a single stage, that is, summary statistics within local patches. However, large-scale model comparisons including the TTM found that only models incorporating, at least, a two-stage grouping and segmentation process can accurately predict global crowding effects (Bornet, Doerig, et al., 2021b; Doerig et al., 2019; Doerig, Schmittwilken, et al., 2020).

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

3. *Feedforward.* The TTM is feedforward and static in nature, with a single pass of processing. Experimental and modeling studies showed how recurrent interactions can capture dynamics of global crowding effects (Doerig, Schmittwilken, et al., 2020; Morea et al., 2024). The TTM's inability to integrate recurrent processing prevents it from capturing the dynamical processes in peripheral vision.
4. *Low-level interactions.* The TTM only models interference between low-level features via summary statistics. However, Manassi & Whitney, 2018, demonstrated that crowding occurs not only between low-level features but also between high-level features such as objects or faces. The TTM does not reproduce these high-level face crowding effects (Bornet, Choung, et al., 2021).

In reply to these concerns, Rosenholtz, Yu & Keshvari, (2019) argued on theoretical grounds that global effects in crowding might be explained by post-perceptual processes. However, Bornet, Choung, et al. (2021) provided direct and extensive empirical evidence that the post-perceptual processes proposed by Rosenholtz fail to account for global effects in crowding. In a similar vein, Rosenholtz et al. (2019) theoretically reasoned that high-level interactions in crowding may emerge from the complexity and high-dimensional nature of summary statistics computed on low-level visual features. However, Bornet, Choung, et al. (2021) showed that the crowding strength predicted by TTM correlates strongly with low-level pixel density rather than high-level interactions.

In addition to being falsified in psychophysical settings, there are also strong reasons to doubt that the TTM could be up to the task of modeling visual processing more generally. As discussed, the TTM relies on a single-stage computational process that extracts summary statistics from local low-level visual features, plus an influence of task readout. However, decades of research in neuroscience suggest that human vision depends on multiple stages of nonlinear processing, including hierarchical feature extraction, recurrent interactions, and top-down feedback. These multi-stage mechanisms are critical for extracting behaviorally relevant representations (DiCarlo, Zoccolan & Rust, 2012). By providing image-computable models of such multi-stage hierarchical processes, Artificial Neural Networks have revolutionized our ability to accurately model neural activity and visual behavior (review: Doerig et al., 2023).

Computational models should be as simple as possible, but not simpler. Simplicity avoids overfitting and allows for easy interpretation and understanding. But overly simple models miss crucial aspects and lead to poor predictive power. Therefore, rejecting more complex models in favor of the TTM to account for the richness and adaptability of peripheral vision must be supported by exceptionally strong and extensive experimental evidence, which does not exist to date. More complex, multi-stage models with recurrent grouping and segmentation mechanisms are required. The importance of multiple stages in vision is also backed up by decades of research in neuroscience and machine learning. Under this light, the TTM and other models based on simple combinations of low-level features seem unlikely to bring about a Kuhnian revolution to replace attention in our understanding of visual processing. As it stands, both attention and peripheral vision research are unsettled fields. For this reason, replacing one with the other is premature.

Financial support. Michael Herzog acknowledges funding from the SNF project "The space in between: where local meets global in human vision."

Competing interests. The author(s) declare that there are no conflicts of interest related to this work.

References

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 13.1–18.

Bornet, A., Choung, O.-H., Doerig, A., Whitney, D., Herzog, M. H., & Manassi, M. (2021a). Global and high-level effects in crowding cannot be predicted by either high-dimensional pooling or target cueing. *Journal of Vision*, *21*(12), 10.

Bornet, A., Doerig, A., Herzog, M. H., Francis, G., & Van der Burg, E. (2021b). Shrinking Bouma's window: How to model crowding in dense displays. *PLoS Computational Biology*, *17*(7), e1009187.

Choung, O.-H., Bornet, A., Doerig, A., & Herzog, M. H. (2021). Dissecting (un) crowding. *Journal of Vision*, *21*(10), 10.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron*, *73*(3), 415–434.

Doerig, A., Bornet, A., Choung, O. H., & Herzog, M. H. (2020). Crowding reveals fundamental differences in local vs. global processing in humans and machines. *Vision Research*, *167*, 39–45.

Doerig, A., Bornet, A., Rosenholtz, R., Francis, G., Clarke, A. M., & Herzog, M. H. (2019). Beyond Bouma's window: How to explain global aspects of crowding? *PLoS Computational Biology*, *15*(5), e1006580.

Doerig, A., Schmittwilken, L., Sayim, B., Manassi, M., & Herzog, M. H. (2020). Capsule networks as recurrent models of grouping and segmentation. *PLoS Computational Biology*, *16*(7), e1008017.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., van Gerven, M. A. J., Kriegeskorte, N., & Kietzmann, T. C. (2023). The neuroconnectionist research programme. *Nature Reviews. Neuroscience*, *24*(7), 431–450.

Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision*, *16*(3), 35.

Manassi, M., Sayim, B., & Herzog, M. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision*, *13*(13), 10.

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10), 13.

Manassi, M., & Whitney, D. (2018). Multi-level Crowding and the Paradox of Object Recognition in Clutter. *Current Biology: CB*, *28*(3), R127–R133.

Morea, M., Herzog, M. H., Francis, G., & Manassi, M. (2024). *Dynamics of vision: grouping takes longer than crowding.* https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=tcUjX4UAAAAJ:d1gkVwhDpl0C

Oberfeld, D., & Stahn, P. (2012). Sequential grouping modulates the effect of non-simultaneous masking on auditory intensity resolution. *PloS One*, *7*(10), e48054.

Overvliet, K. E., & Sayim, B. (2016). Perceptual grouping determines haptic contextual modulation. *Vision Research*, *126*, 52–58.

Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of Vision*, *19*(7), 15.

Development is a pathway for understanding visual attention and peripheral function

Chiara Capparini[a]*, Michelle P.S. To[b] and Vincent M. Reid[c]

[a]Center for Research in Cognition & Neuroscience (CRCN), Université libre de Bruxelles, Bruxelles, Belgium; [b]Department of Psychology, Lancaster University,

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Lancaster, UK and [c]School of Psychology, University of Waikato, Hamilton, New Zealand
chiara.capparini@ulb.be
m.to@lancaster.ac.uk
vincent.reid@waikato.ac.nz

*Corresponding author.

doi:10.1017/S0140525X25000172, e133

Abstract

Visual Attention in Crisis provides the reader with an alternative way to think about the visual attention phenomena – often interpretable in terms of perceptual processes and peripheral vision. We urge an extension of these considerations to developmental science. Infancy research underpins the foundations of mature attentional mechanisms. It may offer a critical test for evolving perceptual limits on attention.

Rosenholtz's re-examination of visual attention stems from a clear understanding of the capabilities and limits of mature peripheral vision. This marks the starting point to uncover and challenge anomalies in attention-related phenomena. Despite this, such an understanding of peripheral function is still partial over early development. Rosenholtz's position is therefore not integrated into infant research paradigms examining visual attention. It is our view that development is critical for understanding mature attentional mechanisms. Foundations of attentional networks, such as alerting and orienting, develop in infancy, and ground mature and more complex attentional and cognitive processes (Hendry, Johnson, & Holmboe, 2019). Development can also help us determine how perceptual capacities and limitations impact attention. As highlighted by Rosenholtz, perceptual mechanisms tend to be included under the same umbrella definition of attention. A better understanding of early perceptual development can provide essential information on the complexity of mature phenomena.

Currently, relatively few studies have investigated peripheral processing during development. What we know is that visual fields gradually expand, particularly over the first postnatal year (Lewis & Maurer, 1992). At birth, the peripheral retina is relatively mature compared to the central/foveal retina (Hendrickson & Drucker, 1992). Accordingly, a peripheral advantage for movement detection is observable at birth (Johnson et al., 1991) and during the third trimester of gestation (Reid et al., 2017). Of note, stimuli with evolutionary significance (e.g. face-like stimuli) can be detected at higher eccentricities compared to other stimuli (Capparini, To, & Reid, 2022a). Stimulus features such as size and distance also play a role in early peripheral functioning (de Schonen et al., 1978). Nevertheless, there is currently no comprehensive mapping of peripheral processing during different developmental stages. Most research focused on the first postnatal weeks, with very little investigations beyond 6–7 months. However, we know that peripheral vision is still developing at 9 months (Capparini, To, & Reid, 2022b). It is therefore unsurprising that only hints are present in the relationship between peripheral information processing and attention during infancy (Johnson, 1990; Johnson & de Haan, 2015).

In stark contrast to peripheral work, infancy research abounds with visual attention studies. Examples include experiments on covert and overt attention, with tasks such as the fixation shift paradigm, the gap/overlap paradigm, or other adaptations of the spatial cueing paradigm. In preverbal participants, covert attention is often revealed by physiological variations (Finlay & Ivinskis, 1984) or changes in saccadic behaviour (Clohessy et al., 1991) following the brief presentation of a peripheral cue. Further, brain activation in the anterior cingulate cortex, as indexed by the Negative component (Nc), has been proposed to mediate early attentional orienting and to distinguish periods of focused vs. unfocused processing (Richards, 2003). All the above paradigms involve the presentation of peripheral stimuli, but the limits of the developing visual field are often not considered. In parallel with Rosenholtz's reasoning, we argue that performance on attentional tasks may be explained in terms of a developing peripheral vision rather than attentional processes, and that the contributions from the perceptual and attentional systems both need to be considered. Future work is needed to clarify whether complex behaviour during development, such as rapid orientation to some emotions over others, is led by bottom-up perceptual and/or top-down attentional mechanisms.

Visual processing in the peripheral field is qualitatively different from processing in central regions. Crowding is a phenomenon that is primarily present in the periphery and impacts the detectability of objects outside fixation (Levi, 2008). Farzin, Rivera, & Whitney, (2010) demonstrated that visual experience during infancy cannot be simply predicted by developing acuity and contrast sensitivity, rather infants seem to have a coarser and more diffuse window of visual awareness than adults due to crowding. Crowding can also impact saccadic orienting behaviour towards a peripheral target and, hence, play a role in the above-mentioned visual attention studies. The effect of perceptual mechanisms such as crowding on complex attentional behaviour needs to be further explored in infancy. Infants and their developing peripheral capacities can therefore allow us to test the limits on task complexity raised by Rosenholtz and identify the nature of such a limit.

As highlighted in the target paper, a non-attentional pathway accounting for statistics of low-level visual features may explain attentional phenomena. In infancy, low-level processing has proposed cascading roles for attentional and cognitive mechanisms. An example is the relation of low-level contextual processing indicated by a slow wave ERP effect for congruent and incongruent actions (Michel et al., 2017). This then leads to an Nc attentional effect at 7 months and then an N400 semantic effect at 9 months (Reid et al., 2009; Kaduk et al., 2016). Overall, summary statistics models have had limited applications to developing populations (see Balas, 2017, with children). Whether summary statistics can explain infant behaviours, such as the preference for high-contrast information or delayed orienting responses, is still open to investigation.

On top of conceptually rethinking early attentional processes, resolving the vision-attention imbalance would be helpful in addressing newly emerging trends in developmental research. Notably, there has been increasing interest in implementing more naturalistic setups that can capture the child's naturalistic visual environment, for instance, using head-mounted cameras (Anderson, Seemiller, & Smith, 2022; Jayaraman, Fausey, & Smith, 2015). Such investigations are essential for generalizing laboratory findings to everyday environments. They currently operate with incomplete information about the child's visual capacities during development. How much visual information is available across the child's visual field, and how is this affected by peripheral limitations or crowding? These issues cannot be addressed using a simplified camera view. Another area that can benefit from a better mapping of perceptual limits over time is computational models of cognitive development (Mattern et al., 2024). Implementing immersive visual environments that capture

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

visual capabilities and limitations across different developmental stages can therefore affect further processing and decision making.

It is our view that a comprehensive mapping of peripheral vision throughout early development could lead to a better understanding of what can be defined as *attentional* rather than *perceptual*. Beyond a theoretical rethinking of attention, perceptual capabilities and limits need to be carefully considered in the developmental sciences.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. None.

References

Anderson, E. M., Seemiller, E. S., & Smith, L. B. (2022). Scene saliencies in egocentric vision and their creation by parents and infants. *Cognition*, *229*, 105256.

Balas, B. (2017). Children's use of visual summary statistics for material categorization. *Journal of Vision*, *17*(12), 22–22.

Capparini, C., To, M. P., & Reid, V. M. (2022b). Identifying the limits of peripheral visual processing in 9-month-old infants. *Developmental Psychobiology*, *64*(4), e22274.

Capparini, C., To, M. P. S., & Reid, V. M. (2022a). The detection of face-like stimuli at the edge of the infant visual field. *Brain Sciences*, *12*(4), 493.

Clohessy, A. B., Posner, M. I., Rothbart, M. K., & Vecera, S. P. (1991). The development of inhibition of return in early infancy. *Journal of Cognitive Neuroscience*, *3*(4), 345–350.

de Schonen, S., McKenzie, B., Maury, L., & Bresson, F. (1978). Central and peripheral object distances as determinants of the effective visual field in early infancy. *Perception*, *7*(5), 499–506.

Farzin, F., Rivera, S. M., & Whitney, D. (2010). Spatial resolution of conscious visual perception in infants. *Psychological Science*, *21*(10), 1502–1509.

Finlay, D., & Ivinskis, A. (1984). Cardiac and visual responses to moving stimuli presented either successively or simultaneously to the central and peripheral visual fields in 4-month-old infants. *Developmental Psychology*, *20*(1), 29.

Hendrickson, A., & Drucker, D. (1992). The development of parafoveal and mid-peripheral human retina. *Behavioural Brain Research*, *49*(1), 21–31

Hendry, A., Johnson, M. H., & Holmboe, K. (2019). Early development of visual attention: Change, stability, and longitudinal associations. *Annual Review of Developmental Psychology*, *1*(1), 251–275.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PloS One*, *10*(5), e0123780.

Johnson, M., & de Haan, M. (2015). Vision, orienting, and attention. In *Developmental cognitive neuroscience: An introduction*, (4th ed., pp. 83–109). Chichester. UK: Wiley Blackwell.

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience*, *2*(2), 81–95.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19.

Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9 months is linked to language proficiency at 9 and 18 months. *Journal of Experimental Child Psychology*, *151*, 96–108.

Levi, D. M. (2008). Crowding - An essential bottleneck for object recognition: A mini-review. *Vision research*, *48*(5), 635–654.

Lewis, T. L., & Maurer, D. (1992). The development of the temporal and nasal visual fields during infancy. *Vision Research*, *32*(5), 903–911.

Mattern, D., Schumacher, P., López, F. M., Raabe, M. C., Ernst, M. R., Aubret, A., & Triesch, J. (2024). MIMo: A multimodal infant model for studying cognitive development. *IEEE Transactions on Cognitive and Developmental Systems*, *16*(4), 1291–1301.

Michel, C., Kaduk, K., Ní Choisdealbha, Á., & Reid, V. M. (2017). Event-related potentials discriminate familiar and unusual goal outcomes in 5-month-olds and adults. *Developmental Psychology*, *53*(10), 1833.

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology*, *27*(12), 1825–1828.

Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: Evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: An ERP study. *Developmental Science*, *6*(3), 312–328.

Peripheral vision and attention: A longstanding dissociation

Marisa Carrasco*

Psychology and Neural Science, New York University, New York, USA
marisa.carrasco@nyu.edu
https://wp.nyu.edu/carrascolab/

*Corresponding author.

doi:10.1017/S0140525X2500007X, e134

Abstract

I agree that, in some literature, the term "attention" is ill-defined and the constraints of peripheral vision have been overlooked. However, I disagree with omitting the empirical studies that have emphasized the importance of defining, operationalizing, and manipulating different types of attention and have investigated their effects on perception as a function of eccentricity and polar angle in peripheral vision.

Rosenholtz discusses visual search studies, highlighting shortcomings in their interpretation, and stresses constraints of peripheral vision. I agree with this perspective; early studies often confounded stimulus factors like eccentricity with attention. However, she overlooked numerous studies dissociating visual field location and attention effects. The key was holding stimulus properties constant, including location, while manipulating attention and monitoring eye movements. This research, spanning decades, established how covert attention affects basic visual dimensions – pillars of our perception – where the best understanding of attention has been achieved.

Stimulus parameters – target eccentricity, spatial frequency, and size – were systematically manipulated to assess their role in visual search set-size effects (more distractors reduce performance). Previously attributed to attention, these effects were clarified when target eccentricity was identified as a confounding factor. Performance in both feature and conjunction search tasks worsens as targets are presented more peripherally (Carrasco et al., 1995). To investigate this eccentricity effect, we scaled stimulus size according to cortical magnification, ensuring their cortical representations were equated. This revealed stable performance across eccentricity with scaling, linking performance declines to peripheral vision's lower spatial resolution (Carrasco & Frieder, 1997; Figure 1A). We also magnified spatial frequency and orientation, showing their role in eccentricity effects and visual search asymmetries (Carrasco et al., 1998). These findings revealed that visual and physiological factors should be considered before invoking attention to explain visual

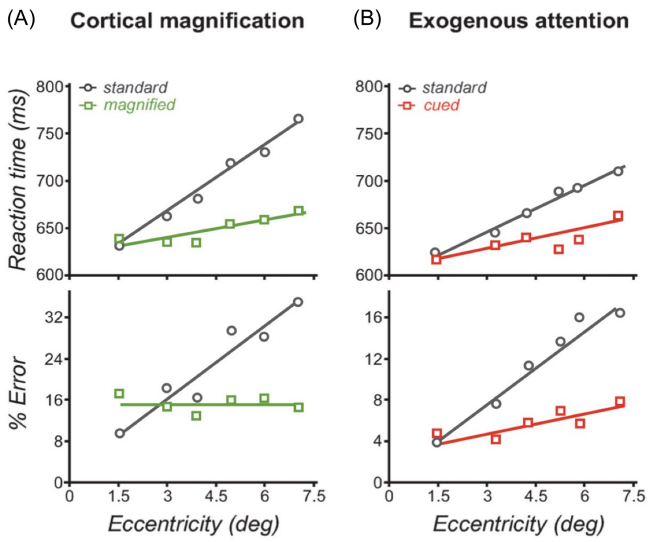


Figure 1 (Carrasco). Effects of attention in visual search tasks. (**A**) Performance in visual search decreases with eccentricity, as depicted by slower reaction times and higher error rates. Adjusting stimulus size according to the cortical magnification factor at a given eccentricity eliminates the eccentricity effect in visual search, indicating that spatial resolution constraints performance. Adapted from Carrasco & Frieder (1997). (**B**) Manipulating exogenous (involuntary) attention has a similar effect to cortical magnification, strongly reducing the eccentricity effect. These results support the idea that attention increases resolution at the attended location, restoring search performance at peripheral locations where low resolution is a limiting factor. Adapted from Carrasco & Yeshurun (1998).

search. Furthermore, signal detection theory models account for set-size effects in identification, detection, and localization tasks when targets and distractors are at iso-eccentric locations, without invoking attention (Cameron et al., 2004).

I also agree with Rosenholtz's critique of the loose use of the term "attention" in some literature. As emphasized in multiple reviews (Anton-Erxleben & Carrasco, 2013; Carrasco, 2006, 2011, 2014; Carrasco & Barbot, 2015; Carrasco & Yeshurun, 2009; Denison, 2024), attention (like any construct) must be explicitly defined, operationalized, and manipulated – with constant stimulus properties, locations, and task – rather than inferred. Indeed, in many laboratories (including mine), the term "attention" is only used if these requirements are fulfilled. However, Rosenholtz overlooks these reviews and rigorous studies on attention's selective effects on fundamental visual dimensions – spatial resolution and contrast sensitivity.

The interaction of eccentricity and attention has been thoroughly investigated. For instance, like stimulus magnification, exogenous (involuntary) covert spatial attention reduces the set-size and eccentricity effects in search (Carrasco & Yeshurun, 1998; Figure 1B). This finding that attention minimizes fovea-periphery performance differences inspired studies investigating covert attention's effects on spatial resolution in peripheral locations.

When comparing performance under neutral cues with exogenous or endogenous (voluntary) attention, both types enhance acuity (Carrasco, Williams, & Yeshurun, 2002; Golla et al., 2004; Montagna, Pestilli, & Carrasco, 2009; Yeshurun & Carrasco, 1999). However, in texture segmentation tasks, whereas endogenous attention improves performance across eccentricity (Barbot & Carrasco, 2017; Jigo & Carrasco, 2018; Yeshurun, Montagna, & Carrasco, 2008), exogenous attention impairs central performance when resolution is already too high for the texture scale (e.g., Jigo & Carrasco, 2018; Yeshurun & Carrasco, 1998, 2000). Furthermore, cueing target location reduces crowding (Grubb et al., 2013; Yeshurun & Rashal, 2010) and lowers orientation thresholds for a target among distractors to levels seen when the target appears alone (Baldassi & Burr, 2000; Morgan, Ward, & Castet, 1998). Several reviews (Anton-Erxleben & Carrasco, 2013; Carrasco, 2011, Carrasco & Barbot, 2015; Carrasco & Yeshurun, 2009) highlight the role of covert spatial attention in impacting resolution across eccentricity.

Because sensitivity varies with both eccentricity and polar angle (review, Himmelberg, Winawer, & Carrasco, 2023), psychophysical studies have compared endogenous, exogenous, and neutral attention conditions in contrast sensitivity tasks at various eccentricities (e.g., Cameron, Tai, & Carrasco, 2002; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Dosher & Lu, 2000a, 2000b; Fernández et al., 2020; Herrmann et al., 2010; Jigo & Carrasco, 2020; Ling & Carrasco, 2006a,b; Pestilli & Carrasco, 2005) and iso-eccentric locations (Cameron, Tai, & Carrasco 2002; Carrasco, Penpeci-Talgar, & Cameron, 2001; Purokayastha, Roberts, & Carrasco, 2021; Roberts et al., 2016, 2017; Talgar & Carrasco, 2002).

Diverse psychophysical methods – such as external noise (Dosher & Lu, 2000a,b; Ling, Liu, & Carrasco, 2009; Lu & Dosher, 2000, 2004), masking (Baldassi & Verghese, 2005), critical band masking (Talgar, Pelli, & Carrasco, 2004), adaptation (Barbot & Carrasco, 2017; Carrasco, Loula, & Ho, 2006; Ling & Carrasco, 2006a; Pestilli, Viera, & Carrasco, 2007), classification image analysis (Eckstein, Shimozaki, & Abbey, 2002), and reverse correlation (Fernández, Li, & Carrasco, 2019; Fernández, Okun, & Carrasco 2022) – have been used to compare attention mechanisms (e.g., signal enhancement vs. external noise reduction) and featural representations (e.g., gain vs. tuning) at matched peripheral locations. Speed-accuracy tradeoff studies further show how exogenous and endogenous attention enhance discriminability and accelerate information accrual (e.g., Carrasco & McElree, 2001; Giordano, McElree, & Carrasco, 2009) across eccentricity (Carrasco, Giordano, & McElree, 2006) and polar angle (Carrasco, Giordano, & McElree 2004).

Studies using concurrent functional magnetic resonance imaging (fMRI) and psychophysics have established differential effects of spatial (Dugué et al., 2018, 2020) and feature-based attention (Foster & Ling, 2022; Liu et al., 2007) on stimulus-evoked responses at peripheral locations across the visual hierarchy. Furthermore, studies using concurrent transcranial magnetic stimulation (TMS) and psychophysics have assessed the critical role of different brain areas in exogenous (Fernández & Carrasco, 2020; Lee, Fernández, & Carrasco, 2024) and endogenous spatial attention (Fernández, Hanning, & Carrasco, 2023) and feature-based attention (Jigo, Gong, & Liu, 2018) at matched peripheral locations. Numerous studies with nonhuman primates have also manipulated attention while keeping stimulus locations constant.

Temporal attention also improves perception, with stimulus at one foveal or peripheral location (Denison, Heeger, & Carrasco, 2017, 2021, 2024; Duyar & Carrasco, 2025; Duyar, Denison, & Carrasco, 2023; Duyar, Ren, & Carrasco, 2024; Fernández, Denison, & Carrasco, 2019; Nobre and vanEede, 2018; Zhu et al., 2024). Other studies have examined interactions between different attention types – endogenous and exogenous spatial (Grubb et al., 2014), spatial and feature (White, Rolfs, & Carrasco,

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

2015), spatial and temporal (Palmieri & Carrasco, 2024), or feature and temporal (Huang, Gao, & Chen, 2025) – at matched peripheral locations.

We have derived and tested predictions (Herrmann et al., 2010, 2012) from the normalization model of attention (Reynolds & Heeger, 2009). More recently, informed by psychophysical data in neutral and attentional conditions across the same locations, we have expanded this model to differentiate effects of spatial exogenous and endogenous attention (Jigo, Heeger, & Carrasco, 2021) and to include effects of temporal attention (Denison, Carrasco, & Heeger, 2021).

To conclude, vision models must consider how stimulus encoding changes across eccentricity and polar angle and integrate the effects of covert attention to explain visual perception. These factors must be – and for decades have been – dissociated.

Acknowledgements. I thank Antoine Barbot, Rachel Denison, Sam Ling, and Alex White for their helpful feedback on this commentary.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. None.

References

Anton-Erxleben, K., & Carrasco, M. (2013). Attentional enhancement of spatial resolution: Linking behavioral and neurophysiological evidence. *Nature Reviews Neuroscience*, *14*, 188–200. https://doi.org/10.1038/nrn3443

Baldassi, S., & Burr, D. C. (2000). Feature-based integration of orientation signals in visual search. *Vision Research*, *40*(10–12), 1293–1300. https://doi.org/10.1016/S0042-6989(00)00007-9

Baldassi, S., & Verghese, P. (2005). Attention to locations and features: Different top-down modulation of detector weights. *Journal of Vision*, *5*(6), 556–570. https://doi.org/10.1167/5.6.7

Barbot, A., & Carrasco, M. (2017). Attention modifies spatial resolution according to task demands. *Psychological Science*, *28*(3), 258–296. https://doi.org/10.1177/0956797616679632

Cameron, E. L., Tai, J., & Carrasco, M. (2002). Effects of covert attention on the psychometric function of contrast sensitivity. *Vision Research*, *42*, 949–967. https://doi.org/10.1016/S0042-6989(01)00282-0

Cameron, E. L., Tai, J. C., Eckstein, M. P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks: Identification, yes/no detection and localization. *Spatial Vision*, *17*, 295–325. https://doi.org/10.1163/1568568041920263

Carrasco, M. (2006). Covert attention increases contrast sensitivity: Psychophysical, neurophysiological, and neuroimaging studies. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Progress in Brain Research* (pp. 33–70). Elsevier.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*, 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, M. (2014). Spatial attention: Perceptual modulation. In S. Kastner, & A. C. Nobre (Eds.), *The Oxford Handbook of Attention* (pp. 183–230). Oxford University Press.

Carrasco, M., & Barbot, A. (2015). Spatial attention alters perception at multiple scales. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 149–160. https://doi.org/10.1101/sqb.2014.79.024810

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*, 1241–1261. https://doi.org/10.3758/BF03208383

Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research*, *37*, 63–82. https://doi.org/10.1016/S0042-6989(96)00107-0

Carrasco, M., Giordano, A. M., & McElree, B. (2004). Temporal performance fields: Visual and attentional factors. *Vision Research*, *44*(12), 1351–1365. https://doi.org/10.1016/j.visres.2003.11.026

Carrasco, M., Giordano, A. M., & McElree, B. (2006). Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Research*, *46*(13), 2028–2040. https://doi.org/10.1016/j.visres.2005.12.015

Carrasco, M., Loula, F., & Ho, Y. X. (2006). How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. *Perception & Psychophysics*, *68*, 1004–1012. https://doi.org/10.3758/BF03193360

Carrasco, M., & McElree, B. (2001). Covert attention speeds the accrual of visual information. *Proceedings of the National Academy of Sciences*, *98*, 5363–5367. https://doi.org/10.1073/pnas.081074098

Carrasco, M., McLean, T. L., Katz, S. M., & Frieder, K. S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation, & spatial frequency. *Vision Research*, *38*, 347–374. https://doi.org/10.1016/S0042-6989(97)00190-9

Carrasco, M., Penpeci-Talgar, C., & Cameron, E. L. (2001). Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task, and set size. *Spatial Vision*, *15*, 61–75. https://doi.org/10.1163/15685680152692015

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, *40*, 1203–1215. https://doi.org/10.1016/S0042-6989(00)00024-2

Carrasco, M., Williams, P., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, *2*, 467–479. https://doi.org/10.1167/2.6.4

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 673–692. https://doi.org/10.1037/0096-1523.24.2.673

Carrasco, M., & Yeshurun, Y. (2009). Covert attention effects on spatial resolution. In N. Srinivasan (Ed.), *Progress in Brain Research* (Vol. *176.*, pp. 65–86). Elsevier. https://doi.org/10.1016/S0079-6123(09)17604-0

Denison, R. N. (2024). Visual temporal attention from perception to computation. *Nature Reviews Psychology*, *3*, 261–274. https://doi.org/10.1038/s44159-024-00063-3

Denison, R. N., Carrasco, M., & Heeger, D. J. (2021). A dynamic normalization model of temporal attention. *Nature Human Behavior*, *5*, 1674–1685. https://doi.org/10.1038/s41562-021-01129-1

Denison, R. N., Heeger, D. J., & Carrasco, M. (2017). Attention flexibly trades off across points in time. *Psychonomic Bulletin & Review*, *24*(4), 1142–1151. https://doi.org/10.3758/s13423-016-1216-5

Denison, R. N., Tian, K. J., Heeger, D. J., & Carrasco, M. (2024). Anticipatory and evoked visual cortical dynamics of voluntary temporal attention. *Nature Communications*, *15*(1), 1–22. https://doi.org/10.1038/s41467-024-53406-y

Dosher, B. A., & Lu, Z.-L. (2000a) Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*(10–12), 1269–1292. https://doi.org/10.1016/S0042-6989(00)00015-3

Dosher, B. A., & Lu, Z.-L. (2000b) Noise exclusion in spatial attention. *Psychological Science*, *11*(2), 139–146. https://doi.org/10.1111/1467-9280.00229

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2018). Specific visual subregions of TPJ mediate reorienting of spatial attention. *Cerebral Cortex*, *28*(7), 2375–2390. https://doi.org/10.1093/cercor/bhx140

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2020). Differential impact of endogenous and exogenous attention on activity in human visual cortex. *Scientific Reports*, *10*(1), 21274. https://doi.org/10.1038/s41598-020-78172-x

Duyar, A., & Carrasco, M. (2025). Eyes on the past: Gaze stability differs between temporal expectation and temporal attention. *Journal of Vision*, *25*(4), Article 11, 1–19. https://doi.org/10.1167/jov.25.4.11pubs.rsc.org 9

Duyar, A., Denison, R., & Carrasco, M. (2023). Exogenous temporal attention varies with temporal uncertainty. *Journal of Vision*, *23*(3), 9, 1–12. https://doi.org/10.1167/jov.23.3.9

Duyar, A., Ren, S., & Carrasco, M. (2024). When temporal attention interacts with expectation. *Scientific Reports*, *14*, 4624. https://doi.org/10.1038/s41598-024-42624-8

Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). Classification images reveal the effects of attention on visual sensitivity. *Psychological Science*, *13*(3), 246–254. https://doi.org/10.1111/1467-9280.00445

Fernández, A., & Carrasco, M. (2020). Extinguishing exogenous attention via transcranial magnetic stimulation. *Current Biology*, *30*, 1–7. https://doi.org/10.1016/j.cub.2020.07.045

Fernández, A., Denison, R. N., & Carrasco, M. (2019). Temporal attention improves perception similarly at foveal and parafoveal locations. *Journal of Vision*, *19*(1), 12, 1–10. https://doi.org/10.1167/19.1.12

Fernández, A., Hanning, N. M., & Carrasco, M. (2023). Transcranial magnetic stimulation to frontal but not occipital cortex disrupts endogenous attention. *PNAS*, *120*(10), e2219635120. https://doi.org/10.1073/pnas.2219635120

Fernández, A., Li, H. H., & Carrasco, M. (2019). How exogenous visual attention affects visual representation. *Journal of Vision*, *19*(11), 4, 1–13. https://doi.org/10.1167/19.11.4

Fernández, A., Okun, S., & Carrasco, M. (2022). Differential effects of endogenous and exogenous attention on sensory tuning. *Journal of Neuroscience*, *42*(7), 1316–1327. https://doi.org/10.1523/JNEUROSCI.1316-1327.2022

Foster, J. J., & Ling, S. (2022). Feature-Based attention multiplicatively scales the fMRI-BOLD contrast-response function. *Journal of Neuroscience*, *42*(2), 221–235. https://doi.org/10.1523/JNEUROSCI.0844-21.2022

Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision*, *9*(3), 30, 1–10. https://doi.org/10.1167/9.3.30

Golla, H., Ignashchenkova, A., Haarmeier, T., & Thier, P. (2004). Improvement of visual acuity by spatial cueing: A comparative study in human and non-human primates. *Vision Research*, *44*(13), 1589–1600. https://doi.org/10.1016/j.visres.2004.01.009

Grubb, M., Behrmann, M., Egan, R., Minshew, N., Heeger, D. J., & Carrasco, M. (2013). Exogenous spatial attention: Evidence for intact functioning in adults with autism spectrum disorder. *Journal of Vision*, *13*(14), 9, 1–13. https://doi.org/10.1167/13.14.9

Grubb, M., White, A., Heeger, D. J., & Carrasco, M. (2014). Interactions between voluntary and involuntary attention modulate the quality and temporal dynamics of visual processing. *Psychonomic Bulletin & Review*, *22*(2), 437–444. https://doi.org/10.3758/s13423-014-0709-8

Herrmann, K., Heeger, D. J., & Carrasco, M. (2012). Feature-based attention enhances performance by increasing response gain. *Vision Research*, *74*, 10–20. https://doi.org/10.1016/j.visres.2012.06.006

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: Attention affects performance by contrast or response gain. *Nature Neuroscience*, *13*(12), 1554–1559. https://doi.org/10.1038/nn.2659

Himmelberg, M. M., Winawer, J., & Carrasco, M. (2023). Polar angle asymmetries in visual perception and neural architecture. *Trends in Neurosciences*, *46*(6), 445–458. https://doi.org/10.1016/j.tins.2023.03.007

Huang, D., Gao, F., & Chen, Y. (2025). Enhancing visual perception: The independent and additive effects of temporal and feature-based attention. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-025-02660-2

Jigo, M., & Carrasco, M. (2018). Attention alters spatial resolution by modulating second-order processing. *Journal of Vision*, *18*(7), 2, 1–12. https://doi.org/10.1167/18.7.2

Jigo, M., & Carrasco, M. (2020). Differential impact of exogenous and endogenous attention on the contrast sensitivity function across eccentricity. *Journal of Vision*, *20*(6), 11, 1–25. https://doi.org/10.1167/20.6.11

Jigo, M., Gong, M., & Liu, T. (2018). Neural determinants of task performance during feature-based attention in human cortex. *eNeuro*, *5*(1) 1–15. https://doi.org/10.1523/ENEURO.0375-17.2018

Jigo, M., Heeger, D. J., & Carrasco, M. (2021). An image-computable attention model on how endogenous and exogenous attention differentially alter visual perception. *PNAS*, *118*(33), e2106436118. https://doi.org/10.1073/pnas.2106436118

Lee, H. H., Fernández, A., & Carrasco, M. (2024). Adaptation and exogenous attention interact in the early visual cortex: A TMS study. *iScience*, *27*(11), Article 111155. https://doi.org/10.1016/j.isci.2024.111155

Ling, S., & Carrasco, M. (2006a). When sustained attention impairs perception. *Nature Neuroscience*, *9*, 1243–1245. https://doi.org/10.1038/nn1527

Ling, S., & Carrasco, M. (2006b). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research*, *46*, 1210–1220. https://doi.org/10.1016/j.visres.2005.05.008

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194–1204. https://doi.org/10.1016/j.visres.2008.05.025

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron*, *55*(2), 313–323. https://doi.org/10.1016/j.neuron.2007.06.030

Lu, Z. L., & Dosher, B. A. (2000). Spatial attention: Different mechanisms for central and peripheral temporal precues? *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1534–1548. https://doi.org/10.1037/0096-1523.26.5.1534

Lu, Z. L., & Dosher, B. A. (2004). Spatial attention excludes external noise without changing the spatial frequency tuning of the perceptual template. *Journal of Vision*, *4*(10), 955–966. https://doi.org/10.1167/4.10.2

Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, *49*, 735–745. https://doi.org/10.1016/j.visres.2009.02.001

Morgan, M. J., Ward, R. M., & Castet, E. (1998). Visual search for a tilted target: Tests of spatial uncertainty models. *Quarterly Journal of Experimental Psychology A*, *51*(2), 347–370. https://doi.org/10.1080/713755762

Nobre, A., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*, 34–48. https://doi.org/10.1038/nrn.2017.141

Palmieri, H., & Carrasco, M. (2024). Task demand mediates the interaction of spatial and temporal attention. *Scientific Reports*, *14*(1), Article 9228.https://doi.org/10.1038/s41598-024-58209-1

Pestilli, F., & Carrasco, M. (2005). Transient attention reduces the effect of adaptation. *Journal of Vision*, *5*(8), 165–165. https://doi.org/10.1167/5.8.165

Pestilli, F., Viera, G., & Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *Journal of Vision*, *7*(7), 1–12. https://doi.org/10.1167/7.7.9

Purokayastha, S., Roberts, M., & Carrasco, M. (2021). Voluntary attention improves performance similarly around the visual field. *Attention, Perception & Psychophysics*, *83*, 2784–2794. https://doi.org/10.3758/s13414-021-02319-x

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Roberts, M., Ashinoff, B. K., Castellanos, F. X., & Carrasco, M. (2017). When attention is intact in adults with ADHD. *Psychonomic Bulletin & Review*, *24*(6), 1. https://doi.org/10.3758/s13423-017-1236-1

Roberts, M., Cymerman, R., Smith, R. T., & Carrasco, M. (2016). Covert spatial attention is functionally intact in amblyopic human adults. *Journal of Vision*, *16*(15), 1–19. https://doi.org/10.1167/16.15.1

Rosenholtz, R. (2025). Visual attention in crisis. *Behavioral and Brain Sciences*, 1–32. https://doi.org/10.1017/S0140525X24000031

Talgar, C. P., & Carrasco, M. (2002). Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. *Psychonomic Bulletin & Review*, *9*(4), 714–722. https://doi.org/10.3758/BF03196326

Talgar, C. P., Pelli, D. G., & Carrasco, M. (2004). Covert attention enhances letter identification without affecting channel tuning. *Journal of Vision*, *4*(1), 22–31. https://doi.org/10.1167/4.1.22

White, A. L., Rolfs, M., & Carrasco, M. (2015). Stimulus competition mediates the joint effects of spatial and feature-based attention. *Journal of Vision*, *15*(14), 1–21. https://doi.org/10.1167/15.14.7

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*, 72–75. https://doi.org/10.1038/23936

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*, 293–305. https://doi.org/10.1016/S0042-6989(98)00114-X

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Yeshurun, Y., & Carrasco, M. (2000). The locus of attentional effects in texture segmentation. *Nature Neuroscience*, *3*, 622–627. https://doi.org/10.1038/75729

Yeshurun, Y., Montagna, B., & Carrasco, M. (2008). On the flexibility of sustained attention and its effects on a texture segmentation task. *Vision Research*, *48*(1), 78–93. https://doi.org/10.1016/j.visres.2007.10.014

Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision*, *10*(10), 16. https://doi.org/10.1167/10.10.16

Zhu, J., Tian, K. J., Carrasco, M., & Denison, R. (2024). Temporal attention amplifies stimulus information in fronto-cingulate cortex at an intermediate processing stage. *PNAS Nexus*, *3*(12), 535. https://doi.org/10.1093/pnasnexus/pgad235

In defense of attention: why perceptual selection cannot be replaced by decision boundaries

Angus F. Chapman[a], Douglas A. Addleman[b] and Viola S. Störmer[c]*

[a]Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA; [b]Department of Psychology, Gonzaga University, Spokane, WA, USA and [c]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA
Viola.S.Stoermer@dartmouth.edu

*Corresponding author.

doi:10.1017/S0140525X2500010X, e135

Abstract

This commentary refutes Rosenholtz's claim that visual attention lacks conceptual validity. We contend that attention remains important for elucidating capacity-limited perceptual processing and explaining phenomenological experience. Alternative frameworks centered on tasks and decision boundaries fail to account for perceptual effects that attentional mechanisms can capture. Thus, preserving attention as a theoretical construct is important, providing interpretive frameworks for empirical investigations.

Rosenholtz argues that visual attention is in crisis because phenomena typically described as attentional can be explained without appealing to attention-like mechanisms. Because of this, the field of visual cognition might be better served by abandoning, or at least significantly revising, the concept of attention. While we agree that inaccurate, incoherent, or merely inconsistent definitions of attention hold back theories of perception and cognition, as has been argued many times before (Anderson, 2011; Hommel et al., 2019; Zivony & Eimer, 2021), we believe that the concept of attention provides critical explanatory power and should be retained. Rosenholtz forwards two alternative sources of the capacity limits usually attributed to attention: peripheral vision and the drawing of decision boundaries for different tasks. We specifically challenge the effectiveness of replacing attention theory with a framework of limits on decision boundaries and instead defend at least one form of selection as "attention", in which perceptual experiences, and in some cases even the representational spaces of perception themselves, are reshaped by attentional selection.

Even if a framework focused on tasks and decision boundaries can explain perceptual effects of various capacity limits, it is unclear how it would do so in a way that avoids the very crises that Rosenholtz attributes to attention theory. The decision framework argues that tasks are solved by implementing decision boundaries that best discriminate between the features or objects that are relevant and irrelevant for a given task. Under this framework, decisions are constrained by the complexity of the decision boundaries and how well these boundaries divide the space between different items. While this appears like a promising idea that could highlight how decision-making is limited in its ability to draw complex boundaries given the high dimensionality of visual representations, we do not believe this theory can fully replace attention.

For one, focusing on task-related capacity limits alone will actually require more complex mechanisms that would fail to coherently predict results across paradigms compared to clearly defined manipulations of attention. While attention includes a multitude of principles, they all share a common, well-specified function: the selection of information that is relevant for action (Wu, 2023). Replacing this clear function of attention with a set of mechanisms that draw decision boundaries through an abstract task space, or of changing tasks, could result in an unconstrained theory capable of arbitrarily complicated predictions (but little ability to make falsifiable predictions). It also remains unclear what decisions in the high-dimensional task space achieve if not the selection of the most task-relevant information – which Rosenholtz questions the usefulness in arguing against attention as a simple gating mechanism.

What, then, *is* the rightful domain of "attention?" Rosenholtz provides a strong argument against the idea that attention resolves a bottleneck in visual processing and related views like the feature integration theory of attention. However, more recent theories conceptualize attention in other ways that are not contested by Rosenholtz and other critics. For example, many of the known capacity limits in visual processing are clearly predicted by the organization of the visual representations themselves, thus arising from perceptual stages, not decision-making stages (Alvarez & Cavanagh, 2004; Störmer, Alvarez, & Cavanagh, 2014; Störmer & Alvarez, 2014). Other patterns of behavior are similarly well predicted by attention but not decision boundaries or limits on peripheral vision. For example, auditory cues have cross-modal effects on visual phenomenology (Störmer, McDonald, & Hillyard, 2009; Störmer, 2019), and attentional manipulations affect neural signatures of and confidence about perceptual experiences, and not merely accuracy or response time (Störmer & Alvarez, 2016; Störmer et al., 2009). Furthermore, attentional manipulations can reshape entire representational spaces (Chapman, Chunharas, & Störmer, 2023), demonstrating that it is not just decision boundaries discriminating between static visual representations. Finally, selection effects also occur in the absence of any explicit task instructions: incidental learning about frequently task-relevant information influences visual search efficiency (Addleman & Lee, 2021; Golan & Lamy, 2024), and neural markers of early stages of selection like the N2pc (Duncan, Van Moorselaar, & Theeuwes, 2023; Ortego, Addleman, & Störmer, 2024). These effects can occur implicitly (Jiang, 2018), in that participants selectively process

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

frequently task-relevant information even when they can't identify what that task-relevant information is; it is ambiguous how such implicit and incidental selection biases would be reflected in a theory of decision boundaries.

To explain these and other effects, we recently proposed a theoretical framework in which attention is not wholly about the selection of discrete targets but reflects a shift in the way visual information is represented in the brain to support behavior (Chapman & Störmer, 2024). For example, one set of findings regards the effects of attention on the appearance of visual stimuli. Several studies have shown that attended items are perceived as higher in contrast than unattended items (Carrasco, Ling & Read, 2004; Gobell & Carrasco, 2005; Störmer et al., 2009), which results in attended items standing out more among competing items. Likewise, searching for a singleton color target can alter its perceived color, such that targets appear more distinct from the distractors (Chapman & Störmer, 2023; see also, Navalpakkam & Itti, 2007). This suggests that attention has effects on our visual experience that cannot be easily attributed to changes in decision boundaries. This also maps onto a large literature showing effects of attention at early cortical processing stages, where attention increases the strength of sensory signals (e.g., Hillyard & Anllo-Vento, 1998; Martinez-Trujillo & Treue, 2002; Reynolds & Chellazi, 2004).

In conclusion, as vision scientists who study attention, we advocate that attention provides significant explanatory power in understanding perceptual processing limits and phenomenology. The "attention" label may frequently be applied inappropriately to other selective processes, but any alternative theories do a poor job of explaining many phenomena in which perceptual processing is reshaped or otherwise altered by a person's "attentional" state. Instead of trying to determine which tasks or behaviors require attention, we think a more fruitful approach is to explore the role of attention in how the brain orchestrates the flow of information in support of current behavioral goals.

Financial support. Part of this work was supported by grants from the National Science Foundation (BCS-1850738) and the National Institute for Health (R01MH133689).

Competing interests. The authors declare no competing interests.

References

Addleman, D. A., & Lee, V. G. (2022). Simulated central vision loss does not impair implicit location probability learning when participants search through simple displays. *Attention, Perception, & Psychophysics*, *84*(6), 1901–1912.

Alvarez, G. A., & Cavanagh, P. (2004). Independent attention resources for the left and right visual hemifields. *Journal of Vision*, *4*(8), 29–29.

Anderson, B. (2011). There is no such thing as attention. *Frontiers in psychology*, *2*, 246.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature neuroscience*, *7*(3), 308–313

Chapman, A. F., Chunharas, C., & Störmer, V. S. (2023). Feature-based attention warps the perception of visual features. *Scientific Reports*, *13*(1), 6487.

Chapman, A. F., & Störmer, V. S. (2024). Representational structures as a unifying framework for attention. *Trends in cognitive sciences*, *28*(5), 416–427.

Duncan, D. H., Theeuwes, J., & Van Moorselaar, D. (2023). The electrophysiological markers of statistically learned attentional enhancement: evidence for a saliency-based mechanism. *Journal of Cognitive Neuroscience*, *35*(12), 2110–2125. https://doi.org/10.1162/jocn_a_02066

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Psychological Science*, *16*(8), 644–651.

Golan, A., & Lamy, D. (2024). Attentional guidance by target-location probability cueing is largely inflexible, long-lasting, and distinct from inter-trial priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(2), 244.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*, 2288–2303.

Jiang, Y. V. (2018). Habitual versus goal-driven attention. *Cortex*, *102*, 107–120.

Keefe, J. M., & Störmer, V. S. (2021). Lateralized alpha activity and slow potential shifts over visual cortex track the time course of both endogenous and exogenous orienting of attention. *NeuroImage*, *225*, 117495.

Martinez-Trujillo JC, Treue S (2002) Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, *35*, 365–370.

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*(4), 605–617.

Ortego, K., Addleman, D.A., & Störmer, V.S. (2025). Early cortical sensitivity and speeded target selection underlie incidentally learned prioritization of visual features. *The Journal of Neuroscience*, e0607252025. https://doi.org/10.1523/JNEUROSCI.0607-25.2025

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*(1), 611–647.

Störmer, V. S. (2019). Orienting spatial attention to sounds enhances visual processing. *Current Opinion in Psychology*, *29*, 193–198.

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., & Alvarez, G. A. (2016). Attention alters perceived attractiveness. *Psychological Science*, *27*(4), 563–571.

Störmer, V. S., Alvarez, G. A., & Cavanagh, P. (2014). Within-hemifield competition in early visual areas limits the ability to track multiple objects with attention. *Journal of Neuroscience*, *34*(35), 11526–11533.

Störmer, V. S., McDonald, J. J., & Hillyard, S. A. (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences*, *106*(52), 22456–22461.

Wu, W. (2024). We know what attention is! *Trends in Cognitive Sciences*, *28*(4), 304–318.

Zivony, A., & Eimer, M. (2022). The diachronic account of attentional selectivity. *Psychonomic Bulletin & Review*, *29*(4), 1118–1142.

Is attention a theory?

Rachel N. Denison*

Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA
rdenison@bu.edu

*Corresponding author.

doi:10.1017/S0140525X25000214, e136

Abstract

Rosenholtz describes "attention theory" as a scientific paradigm in crisis, in the Kuhnian sense. But is attention a theory? Here, I question this premise. Although there can be theories of attentional phenomena, attention is not a theory. Rather, like memory and emotion, attention is a psychological concept that captures a broad class of phenomena, requiring multiple mechanistic explanations.

In the target article, Rosenholtz argues that "attention theory" is a scientific paradigm in crisis (Rosenholtz, 2025). This formulation

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

comes from the philosopher of science Thomas Kuhn's influential idea that a dominant scientific theory can encounter unexplained anomalies in data, which first lead to overcomplications in the theory and eventually give way to its abandonment (Kuhn, 1962). Rosenholtz reports that she banned the use of the word "attention" in her laboratory for a year as an in-house experiment in theory shedding. But this Kuhnian move raises a question: is attention a theory?

The term "attention" is used widely and variously in psychology and neuroscience, but attention is rarely, if ever, conceptualized as a scientific theory. Rather, like memory and emotion, attention is a psychological concept that captures a broad class of phenomena. The key feature of attention is the prioritization of some items at the expense of others (Carrasco, 2011; Denison, 2024; James, 1890; Reeves, 2020). Decades of careful psychophysical work have shown that prioritization of a spatial location (Carrasco, 2011; Montagna, Pestilli, & Carrasco, 2009; Pestilli & Carrasco, 2005), point in time (Denison, Carrasco, & Heeger, 2021; Denison, Heeger, & Carrasco, 2017; Jing et al., 2022), or feature such as the color red (Liu, 2019; White, Rolfs, & Carrasco, 2015) improves perception of the attended item while impairing perception of other items, relative to a neutral baseline. These experiments have controlled for potential confounds such as eye movements and eccentricity, such that the stimulus remained constant while only instructions about what to prioritize changed from trial to trial. The findings of perceptual trade-offs—benefits for prioritized items accompanied by costs for nonprioritized ones—indicate some form of selectivity, the hallmark of attention. So, although attentional phenomena need theoretical explanations, it is simply a fact that we regularly use our minds to focus on some things and ignore others. Attention is not a theory.

What then, does Rosenholtz mean by "attention theory"? She seems to be referring to a particular class of theoretical accounts of attention based on feature integration theory (Treisman & Gelade, 1980). "Attention theory," she says, proposes a selective mechanism that gates access of visual sensory information to higher-level processing. Her Figure 1 shows a "preattentive processing" stage followed by a "selective attention" stage with a parallel "distributed attention" pathway. Rosenholtz describes the selective attention gate ("in its simplest form") as all or none, so that only the attended item would pass through to higher, cognitive stages. The distributed attention pathway, not present in the original feature integration theory, accommodates findings such as the ease with which people can report the gist of a scene (Ariely, 2001; Treisman, 2006).

This "attention theory" is a simple box-and-arrow model that makes a strong theoretical commitment to discrete stages of processing in the visual system. Too strong, it turned out. Feature integration theory was developed over decades (Treisman, 2006; Treisman & Gelade, 1980), but its notion of discrete preattentive and attentive stages is no longer prominent for three main reasons that emerged from behavioral and neuroscience research. First, attentional selection need not be all or none. Inattention can reduce the precision of visual representations without eliminating them (Anton-Erxleben & Carrasco, 2013; Kay, Weiner, & Grill-Spector, 2015). Second, attentional modulation happens concurrently at multiple levels in the visual system, starting with the earliest stages of cortical (Dugué et al., 2020; Motter, 1993) and even subcortical (Briggs, 2020; McAlonan, Cavanaugh, & Wurtz, 2008) visual processing. Indeed, attention affects basic dimensions of vision that had been considered preattentive (Carrasco, 2011). Third, research has revealed multiple mechanisms for prioritizing visual information. This research has lead the concept of attention to be subdivided based on the object of prioritization (space, time, features, etc.) and the mode of prioritization, including voluntary or involuntary (Cheal & Lyon, 1991; Müller & Rabbitt, 1989), presaccadic (Li, Hanning, & Carrasco, 2021), and experience-driven (Kim & Anderson, 2019) prioritization. So, while Rosenholtz rightly points out problems with two-stage theories of attention, most of the field has long moved beyond these early conceptual models, productive as they once were for driving attention research.

Rosenholtz worries, in the Kuhnian framework, that subdividing attention into different types is a slide into epicycles—those extra orbits tacked on to planetary motion to preserve the theory that the earth was the center of the solar system in the face of anomalous evidence. She argues that the complexity of the current attentional landscape signals a scientific paradigm in crisis. But the field of visual attention research has long embraced pluralism about attentional phenomena and recognized that different mechanisms may explain different kinds (Maunsell, 2015; Nobre & van Ede, 2023)—in the same way that memory research has identified different kinds of memory and associated mechanisms. This pluralism, in which researchers delineate multiple mechanisms of attention with increasing precision, is a sign of the healthy progress of a productive field, not epicycles of crisis. Complexity does not signal a crisis for attention, because attention is not a theory.

Although seeking a single, unified theory of diverse attentional phenomena may be counterproductive, the field has made progress in developing formal (mathematical and computational) theories that explain certain attentional phenomena. For example, the normalization model of attention (Reynolds & Heeger, 2009) predicts psychophysical and neural data from spatial and feature-based visual attention tasks (Herrmann, Heeger, & Carrasco, 2012; Herrmann et al., 2010; Ni & Maunsell, 2019) and has been extended to temporal attention (Denison et al., 2021). By combining simple computations such as gain control and divisive normalization, these models help us understand how attention modulates visual cortical activity. But Rosenholtz is right that existing theories do not capture all the attentional phenomena we would like to understand, and her catalog of phenomena to be explained is a useful guide to theory development. She also helpfully centers the question of what causes perceptual and cognitive capacity limits—which make attention useful in the first place. Hence, although attention is not a theory, the science of attention must continue to develop theoretical accounts of the variety of ways in which humans prioritize and select information.

Acknowledgments. I thank Marisa Carrasco and Michael Epstein for comments on this commentary.

Financial support. This work was supported by startup funds from Boston University to RND.

Competing interests. None.

References

Anton-Erxleben, K., & Carrasco, M. (2013). Attentional enhancement of spatial resolution: Linking behavioural and neurophysiological evidence. *Nature Reviews Neuroscience*, *14*(3), 188–200. https://doi.org/10.1038/nrn3443

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327
Briggs, F. (2020). Role of feedback connections in central visual processing. *Annual Review of Vision Science*, *6*, 313–334. https://doi.org/10.1146/annurev-vision-121219-081716
Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012
Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *The Quarterly Journal of Experimental Psychology Section A*, *43*(4), 859–880. https://doi.org/10.1080/14640749108400960
Denison, R. N. (2024). Visual temporal attention from perception to computation. *Nature Reviews Psychology*, *3*, 261–274. https://doi.org/10.1038/s44159-024-00294-0
Denison, R. N., Carrasco, M., & Heeger, D. J. (2021). A dynamic normalization model of temporal attention. *Nature Human Behaviour*, *5*, 1674–1685. https://doi.org/10.1038/s41562-021-01129-1
Denison, R. N., Heeger, D. J., & Carrasco, M. (2017). Attention flexibly trades off across points in time. *Psychonomic Bulletin & Review*, *24*(2), 1142–1151. https://doi.org/10.3758/s13423-016-1216-1
Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2020). Differential impact of endogenous and exogenous attention on activity in human visual cortex. *Scientific Reports*, *10*(1), 21274. https://doi.org/10.1038/s41598-020-78172-x
Herrmann, K., Heeger, D. J., & Carrasco, M. (2012). Feature-based attention enhances performance by increasing response gain. *Vision Research*, *74*, 10–20. https://doi.org/10.1016/j.visres.2012.04.016
Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: Attention affects performance by contrast or response gain. *Nature Neuroscience*, *13*(12), 1554–1559. https://doi.org/10.1038/nn.2669
James, W. (1890). *Principles of Psychology*. Henry Holt & Co.
Jing, C., Jin, H., Li, W., Wu, Z., Chen, Y., & Huang, D. (2022). Temporal attention affects contrast response function by response gain. *Frontiers in Human Neuroscience*, *16*, 1020260. https://doi.org/10.3389/fnhum.2022.1020260
Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology*, *25*(5), 595–600. https://doi.org/10.1016/j.cub.2014.12.050
Kim, H., & Anderson, B. A. (2019). Dissociable components of experience-driven attention. *Current Biology*, *29*(5), 841–845 e842. https://doi.org/10.1016/j.cub.2019.01.030
Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago University Press.
Li, H. H., Hanning, N. M., & Carrasco, M. (2021). To look or not to look: Dissociating presaccadic and covert spatial attention. *Trends in Neurosciences*, *44*(8), 669–686. https://doi.org/10.1016/j.tins.2021.05.002
Liu, T. (2019). Feature-based attention: Effects and control. *Current Opinion in Psychology*, *29*, 187–192. https://doi.org/10.1016/j.copsyc.2019.03.013
Maunsell, J. H. R. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science*, *1*, 373–391. https://doi.org/10.1146/annurev-vision-082114-035431
McAlonan, K., Cavanaugh, J., & Wurtz, R. H. (2008). Guarding the gateway to cortex with attention in visual thalamus. *Nature*, *456*(7220), 391–394. https://doi.org/10.1038/nature07382
Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, *49*(7), 735–745. https://doi.org/10.1016/j.visres.2009.02.001
Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*(3), 909–919. https://doi.org/10.1152/jn.1993.70.3.909
Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 315–330. https://doi.org/10.1037/0096-1523.15.2.315
Ni, A. M., & Maunsell, J. H. R. (2019). Neuronal effects of spatial and feature attention differ due to normalization. *Journal of Neuroscience*, *39*(28), 5493–5505. https://doi.org/10.1523/jneurosci.2106-18.2019
Nobre, A. C., & van Ede, F. (2023). Attention in flux. *Neuron*, *111*(7), 971–986. https://doi.org/10.1016/j.neuron.2023.02.032
Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*, *45*(14), 1867–1875. https://doi.org/10.1016/j.visres.2005.01.019
Reeves, A. (2020). Attention as a unitary concept. *Vision*, *4*, 48. https://doi.org/10.3390/vision4040048
Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002
Rosenholtz, R. (2025). Visual attention in crisis. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X24000323
Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5
White, A. L., Rolfs, M., & Carrasco, M. (2015). Stimulus competition mediates the joint effects of spatial and feature-based attention. *Journal of Vision*, *15*(14), 7, 1–21. https://doi.org/10.1167/15.14.7

No crisis when attention is the outcome of selective action

Veronica Dudarev and James T. Enns*

Department of Psychology, University of British Columbia, Vancouver, BC, Canada
vdudarev@mail.ubc.ca
jenns@psych.ubc.ca

*Corresponding author.

doi:10.1017/S0140525X25000317, e137

Abstract

Rosenholtz addresses the crisis of proliferating mechanisms for visual attention by redefining the concept in terms of (a) the known limitations of peripheral vision and (b) a proper assessment of task complexity. We argue that abandoning the see → decide → act pipeline model and the myth of a centralized gate or resource eliminates this crisis. In our view, "attention" describes an outcome—the consequence of multiple constraints on perception and action—rather than a reified cause.

We share many of Rosenholtz's concerns about attention as a gateway to perception and a limited-capacity mental resource. However, we argue that this crisis follows from adhering to a see → decide → act model of perception (Nakayama, Moher, & Song, 2023). This framework prevents her analysis from leveraging the emerging view that attention functions as an outcome—a consequence of multiple constraints on perception and action—rather than a singular, causal entity (Enns, Brennan, & Whitwell, 2017; Ristic & Enns, 2015; Kozik & Enns, 2021).

Our first point is that myriad factors shape visual selectivity, including the nonuniform resolution of the retina, neuronal selectivity in numerous cortical maps, recent experience with similar events, immediate goals, and stable individual dispositions (Kozik & Enns, 2021). Labeling only some of these factors as "attention" is as problematic as applying the term to all of them. Allport (1987) clarified this complexity by distinguishing between

selecting *objects* in the environment and selecting *actions* to interact with them. Even simple actions, like reaching for a single orange among others in a bowl, require selectivity across both dimensions. Cisek (2012) has updated this perspective, proposing that perception and action emerge from a dynamic competition between action possibilities. Rao (2024) predictive coding model of the sensory-motor cortex also distinguishes between state predictions (modeling the environment) and action predictions (modeling possible actions). Evidence from multiple species and across diverse brain regions suggests that this is the default design feature of the brain (Rao, 2024). When a cortical network becomes selective for an object or an action, it is therefore not the result of a centralized decision-maker. Rather, distributed circuits continuously integrate sensory and motor information to guide behavior toward the most viable perception and action.

These dynamics unfold even when viewers passively inspect a scene or a visual artwork. The nonuniform resolution of the retina requires that the viewer inspects the scene in a series of fixations. These fixations predictably focus on certain regions before others, guided by many factors, including the viewer's self-identified purpose (Hayhoe & Ballard, 2005), their personal history (Tanaka & Philbert, 2022), momentary physiological states (Liu et al., 2013), and of course low-level features of the image medium (DiPaola, Riebe, & Enns, 2010; DiPaola,Riebe, & Enns, 2013). "Attention" does not cause this selective experience unless we accept as many attention mechanisms as there are influencing factors. Instead, attention emerges as the outcome of all these forces combined.

Our second point is that it is a mistake to ignore the role of action, a term mentioned only twice in the target article (ms pp. 21 and 25), with the actions considered in the target article referring mostly to button presses. There is now a wealth of research showing an intimate coupling between action and perception. Merely planning an action toward an object increases the neural activity in human early visual and auditory cortices linked to those objects (Gallivan et al., 2019; Gale et al., 2021). Neurons in the premotor and motor brain regions implicated in self-action also respond when observing others perform the same actions (Rizzolatti, Fogassi, & Gallese, 2001). Neurons in the lateral interparietal cortex represent environmental locations in anticipation of an action toward the location, during perception of the location, and even in memory of it (Colby, 1998), maintaining a representation that can guide the eyes and limbs. These action-perception links seen in neurophysiology are directly translated into behavioral abilities. People are able to respond more rapidly to sensory events that are near one's hand (Reed, Grubb, Steele, 2006; Perry, Amarasooriya, & Fallah, 2016) and to spatial locations that are the target of upcoming actions (Andersen & Mountcastle, 1983; Snyder, 2000). In real-world settings, rock climbers, bird watchers, and automotive enthusiasts, all see and act on objects in their domain at a finer level of detail than novices (Tanaka & Philibert, 2022).

Actions influence perception even in the limited scenario of passively viewing a visual display. DiPaola et al. (2010, 2013) and Enns & MacDonald (2013) demonstrated that subtly blurring some parts of a display while enhancing the clarity of other regions directs fixations toward the higher resolution areas, as if the eyes adjust to the expected properties of the percept based on the non-uniform resolution of the retina. The knock-on effects of these predictable variations in gaze reliably shape personal attributions made about depicted individuals and even affect the physical proximity one is willing to maintain during simulated interactions with them (Enns, Lin-Yang, & Dudarev, 2023).

Consider also the influence that selective gaze and limb actions have on the cognitive appraisal of objects. Acting on an emotionally neutral object enhances its appeal in subsequent evaluations (Peck & Shu, 2009; Chapman, Gallivan, & Enns, 2015; Wispinski et al., 2021). Even regular involuntary interaction with objects that are initially appraised as negative contributes to their increased emotional appeal over time (Dudarev et al., 2022). And arbitrarily assigning some objects as "mine" and then moving them toward one's body leads to subsequent improved memory of those objects, compared to objects assigned as "yours" or moved away from one's body (Truong et al., 2016). All of this evidence points to iterative and bi-directional relationships between our sensory experiences and the actions we take based on those experiences.

Taking action into account would certainly give Rosenholtz's limited discussion of task complexity more concrete shape, but it would still leave important questions unanswered. For example, it remains important to distinguish between the computational complexity of a visual problem (Ullman, 1984; Neumann, 1987) from the constraints imposed by a particular physical body (Gibson, 1997; Wolpert & Flanagan, 2001). Whereas the complexity of image segmentation may be similar in a computational sense for a human, an octopus, and a large language model, their differing action capabilities will constrain their selective perception in unique ways. This embodied aspect of attention tempts us to propose replacing Rosenholtz's claim that "selectively attending is the observer's task" (ms pp. 11) with "selectively attending is what happens when someone prepares, performs, or remembers an *action* (including the act of looking), thereby privileging an event or object for perception."

Financial support. V.D. is supported by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (32562). J.T.E. is supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2024-04400).

Competing interests. V.D. None.
J.T.E. None.

References

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A. Sanders (eds.) *Perspectives on Perception and Action*, pp. 395–419. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience*, *3*(3), 532–48.

Chapman, C. S., Gallivan, J. P., & Enns, J. T. (2015). Separating value from selection frequency in rapid reaching biases to visual targets. *Visual Cognition*, *23*(1–2), 249–271.

Cisek, P. (2012). Making decisions through a distributed consensus. *Current Opinion In Neurobiology*, *22*(6), 927–936.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron*, *20*(1), 15–24.

De Vito, D., & Fenske, M. J. (2018). Affective evidence that inhibition is involved in separating accessory representations from active representations in visual working memory. *Visual Cognition*, *26*(8), 583–600.

DiPaola, S., Riebe, C., & Enns, J. T. (2010). Rembrandt's textural agency: A shared perspective in visual art and science. *Leonardo*, *43*(2), 145–151.

DiPaola, S., Riebe, C., & Enns, J. T. (2013). Following the masters: Portrait viewing and appreciation is guided by selective detail. *Perception*, *42*(6), 608–630.

Dudarev, V., Manaligod, M. G., Enns, J. T., & Todd, R. M. (2022). In the hands of the beholder: Wearing a COVID-19 mask is associated with its

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

attractiveness. *Quarterly Journal of Experimental Psychology*, *75*(4), 598–615.
Enns, J. T., Brennan, A. A., & Whitwell, R. L. (2017). Attention in action and perception: Unitary or separate mechanisms of selectivity? *Progress in Brain Research*, *236*, 25–52.
Enns, J. T., Lin-Yang, R. C., & Dudarev, V. (2023). Moving to maintain perceptual and social constancy. *Visual Cognition*, *31*(1), 43–62.
Enns, J. T., & MacDonald, S. C. (2013). The role of clarity and blur in guiding visual attention in photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 568.
Fenske, M. J., & Raymond, J. E. (2006). Affective influences of selective attention. *Current Directions in Psychological Science*, *15*(6), 312–316.
Gale, D. J., Areshenkoff, C. N., Honda, C., Johnsrude, I. S., Flanagan, J. R., & Gallivan, J. P. (2021). Motor planning modulates neural activity patterns in early human auditory cortex. *Cerebral Cortex*, *31*(6), 2952–2967.
Gallivan, J. P., Chapman, C. S., Gale, D. J., Flanagan, J. R., & Culham, J. C. (2019). Selective modulation of early visual cortical activity by movement intention. *Cerebral Cortex*, *29*(11), 4662–4678.
Gibson, J. J. (1977). *The theory of affordances*. Erlbaum Associates.
Griffiths, O., & Mitchell, C. J. (2008). Negative priming reduces affective ratings. *Cognition and Emotion*, *22*(6), 1119–1129.
Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.
Kiss, M., Goolsby, B. A., Raymond, J. E., Shapiro, K. L., Silvert, L., Nobre, A. C., ... & Eimer, M. (2007). Efficient attentional selection predicts distractor devaluation: Event-related potential evidence for a direct link between attention and emotion. *Journal of Cognitive Neuroscience*, *19*(8), 1316–1322.
Kozik, P., & Enns, J. T. (2021). Athletics and attention: Bi-directional influences in the lab and on the field. *Current Issues in Sport Science*, *6*, 002.
Liu, K., Sun, G., Li, B., Jiang, Q., Yang, X., Li, M., ... & Liu, Y. (2013). The impact of passive hyperthermia on human attention networks: An fMRI study. *Behavioural Brain Research*, *243*, 220–230.
Nakayama, K., Moher, J., & Song, J. H. (2023). Rethinking vision and action. *Annual Review of Psychology*, *74*, 59–86.
Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer, & A. F. Sanders (Eds.), *Perspectives on perception and action*. pp. 361–394. Hillsdale, NJ: Lawrence Erlbaum.
Peck, J., & Shu, S. B. (2009). The effect of mere touch on perceived ownership. *Journal of Consumer Research*, *36*(3), 434–447.
Perry, C. J., Amarasooriya, P., Fallah, M. (2016). An eye in the palm of your hand: Alterations in visual processing near the hand, a mini-review. *Frontiers in Computational Neuroscience*, *10*, 37.
Rao, R. P. (2024). A sensory–motor theory of the neocortex. *Nature neuroscience*, *27*(7), 1221–1235.
Raymond, J. E., Fenske, M. J., & Westoby, N. (2005). Emotional devaluation of distracting patterns and faces: A consequence of attentional inhibition during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1404.
Reed, C. L., Grubb, J. D., Steele, C. (2006). Hands up: Attentional prioritization of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 166–77.
Ristic, J., & Enns, J. T. (2015). The changing face of attentional development. *Current Directions in Psychological Science*, *24*(1), 24–31.
Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670.
Silver, A. M., Stahl, A. E., Loiotile, R., Smith-Flores, A. S., & Feigenson, L. (2020). When not choosing leads to not liking: Choice-induced preference in infancy. *Psychological Science*, *31*(11), 1422–1429.
Snyder, L. H. (2000). Coordinate transformations for eye and arm movements in the brain. *Current Opinion in Neurobiology*, *10*(6), 747–54.
Tanaka, J. W., & Philibert, V. (2022). *The expertise of perception: How experience changes the way we see the world*. Cambridge University Press.
Truong, G., Chapman, C. S., Chisholm, J. D., Enns, J. T., & Handy, T. C. (2016). Mine in motion: How physical actions impact the psychological sense of object ownership. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 375.
Ullman, S. (1984). Visual routines. *Cognition*, *18*, 94–159.
Wispinski, N. J., Lin, S., Enns, J. T., & Chapman, C. S. (2021). Selective attention to real-world objects drives their emotional appraisal. *Attention, Perception, & Psychophysics*, *83*(1), 122–132.
Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732.

40 ms matters

Dock H. Duncan[a,b] and Jan Theeuwes[a,b,c*]

[a]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; [b]Institute Brain and Behavior Amsterdam (iBBA), Amsterdam, Netherlands and [c]William James Center for Research, ISPA-Instituto Universitario, Lisbon, Portugal
D.H.Duncan@vu.nl
j.theeuwes@psy.vu.nl

*Corresponding author.

doi:10.1017/S0140525X25000093, e138

Abstract

Rosenholtz (2024) dismisses attentional capture, arguing that brief distractions (20–40 ms) are insignificant or intentional. However, we argue that distractions are never intentional nor negligible, and studying them is crucial both theoretically and for real-world applications.

Over the last 35 years or so, there has been a heated debate about whether we can always select the objects we want to select or whether, in some cases, the properties of the environment determine what is selected (Luck et al., 2021). It is a fundamental question: are we always in charge of what we select, or is it possible that, in some cases, we are not in control, and instead, the environment dictates what we attend to?

We and others have shown that even when objects are completely irrelevant to the task, they may, in some cases, capture attention (Theeuwes, 1992, 2025) or capture our eyes (Theeuwes et al., 1998). This research is highly relevant because it addresses the question of who is in control. It has significant consequences for theories of attention, for understanding how our brain is wired, and for explaining why people sometimes get into accidents when distracted by irrelevant events.

Rosenholtz (2024) dismisses these critical findings by arguing, *"who cares about 20 to 40 ms distraction,"* (*p.**) and suggesting that people may *"enjoy spending 40 ms attending an unusual item."* (*p.**). With these remarks, Rosenholtz dismisses a substantial literature in experimental psychology.

In most experiments investigating attentional capture, we instruct observers to selectively search for a well-defined, salient target. On some trials, an irrelevant distractor is also present. In these types of experiments, there are two possible outcomes: either the irrelevant distractor has no effect on performance, which would indicate perfect top-down selection, or the presence of the distractor causes a performance decrement, indicating bottom-

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

up capture. To argue that this performance decrement is due to observers *"enjoying spending 40 ms attending an unusual item"* invalidates the entire approach of hypothesis testing. Regardless of the outcome of such experiments, one could always argue that there is top-down control simply because people enjoy attending to something other than the target.

In the paper, Rosenholtz makes two claims: (1) the performance decrement may result from intentional "enjoyment" of a non-target stimulus, and (2) the observed interference effect is too small to be meaningful. The first claim is clearly incorrect, as it has already been demonstrated that singleton interference is driven by salience—specifically, when a target is more salient than a distractor, no interference occurs (Theeuwes, 1992). If participants were willfully choosing to dwell on and "enjoy" a salient color singleton, then why would they not do the same for a less-salient shape singleton? This explanation only holds if salience inherently captures attention, reinforcing the idea that interference effects are a result of attentional capture rather than voluntary engagement. Regarding the second point, while it is true that in classic attentional capture experiments (Theeuwes, 1992), the irrelevant distractor typically causes only a 20 to 40 ms reaction time (RT) delay, this effect is observed in experiments where a distractor appears on every trial across 2,000 or more trials. However, when a distractor appears only once in an experiment, the interference effect is 9 to 15 times larger (~370-580 ms) (Adam, Yang, & Serences, 2024), demonstrating a much stronger impact when unexpected. Additionally, eye movement studies have shown that salient events can capture gaze, directing the eyes away from the target location (Theeuwes et al., 1999), further supporting the argument that salience-driven interference is both measurable and significant.

This is why research on 40 ms distractions is important. Even a brief moment of not attending the road in front of you such as when a blinking dashboard light momentarily captures your eyes could compromise road safety, potentially leading to a rear-end collision. If a distraction causes your eyes to shift away from an oncoming bicyclist, even for a fraction of a second, the consequences could be fatal. Selecting relevant information and avoiding distractions is crucial for driving, as over 90% of driving-related information is visual, and deficits in visual attention significantly contribute to road accidents (see for a review (Theeuwes, 2021)

Financial support. This work was supported by the European Research Council (JT, advanced grant 833029 – [LEARNATTEND])

Competing interests. None

References

Adam, K. C., Yang, Z., & Serences, J. T. (2024). First encounters: Estimating the initial magnitude of attentional capture. *Visual Cognition*, 1–23. https://doi.org/10.1017/S0140525X24000323

Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W., & Theeuwes, J. (2021). Progress toward resolving the attentional capture debate. *Visual Cognition*, *29*(1), 1–21.

Rosenholtz, R. (2024). Visual attention in crisis. *Behavioral and Brain Sciences*, *76*, 251–273.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606.

Theeuwes, J. (2021). Self-explaining roads: What does visual cognition tell us about designing safer roads? *Cognitive research: principles and implications*, *6*(1), 15.

Theeuwes, J. (2025). Attentional capture and control. *Annual Review of Psychology*, *76*(Volume 76, 2025), 251–273. https://doi.org/10.1146/annurev-psych-011624-025340

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385.

Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G. J. (1999). Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1595.

(Temporal) Visual Attention NOT in Crisis

Paul Edmund Dux[a]*, Roberto Dell'Acqua[b] and Bradley Wyble[c]

[a]The University of Queensland, St Lucia, QLD, Australia; [b]University of Padova, Padova, PD, Italy and [c]The Pennsylvania State University, University Park, PA, United States
paul.e.dux@gmail.com
dar@unipd.it
bpw10@psu.edu

*Corresponding author.

doi:10.1017/S0140525X25000251, e139

Abstract

Extensive research using the *attentional blink* phenomenon illustrates, through behavioural, modelling and cognitive neuroscience approaches, that distinct selection and attention capacity limits exist. Crucially, these effects cannot reflect peripheral visual processes nor distinct task operations across conditions controlling for issues raised by Rosenholtz. Moving away from attention and selection concepts hinder rather than facilitate a mechanistic understanding of vision.

The world is highly dynamic, constantly changing across the dimensions of time and space. Consequently, in any given instant, far too much information is generated for it all to be processed up to the level of consciousness. Indeed, a number of behavioural phenomena illustrate that we only become aware of a small fraction of the information presented to us. Faced with this challenge, if humans (and many other species) are to behave adaptively they must identify and devote limited resources to information that is relevant for survival over that which is less relevant (Marois & Ivanoff, 2005)!

What cognitive and neural processes determine this prioritization of information is arguably the biggest question in brain and cognitive sciences and for decades it has been considered that attention drives selection and the capacity limits of information processing (Pashler, 1998). However, Rosenholtz questions the use of terms such as selection and attention, arguing these are not useful or necessary when attempting to describe the mechanisms underlying the capacity limits of vision. Rather these reflect properties of summary statistics, characteristics of peripheral vision and aspects of task requirements.

However, selection and attention do not just operate across space to drive perception, and not all tasks designed to tap these

operations have spatial manipulations of information. Indeed, perhaps the most inarguable contribution of selection and attention is in facilitating memory (Wyble, Bowman & Nieuwenstein, 2009). Here we focus on work employing Rapid Serial Visual Presentation (RSVP) wherein an object is selected at a moment in time. When stimuli are presented rapidly (tens of ms), subjects are only able to remember the one item that matches their goal/attentional template (Potter, 1976) with high accuracy. If this process of selecting and encoding a singular piece of information into memory according to active mental goals should not be called attention, how else can it be described?

Over the last 3 decades considerable research has been devoted to understanding attentional selection and encoding over time. Crucially, for the current arguments put forward by Rosenholtz, the relevant behavioural markers cannot be due to summary statistics, perception, the characteristics of peripheral vision and/or aspects of task requirements. Chief amongst these phenomena is the attentional blink (AB). This is a highly investigated effect: the second most highly cited article in the *Journal of Experimental Psychology: Human Perception and Performance* is the first paper describing the AB (Raymond, Shapiro & Arnell, 1992) and Dux and Marois's (2009) review on the AB is the 4th most highly cited paper in *Attention, Perception and Performance.* In the typical AB paradigm, an RSVP approach is taken with items appearing one after each other in the same central spatial location for a fraction of a second each (e.g., 100ms/item). To elicit an AB, subjects search for two targets, such as letters, among a stream of distractors, such as digits, and report them at the end of the stream, without speeded time pressure. A key finding is that if target 1 (T1) and target 2 (T2) appear within 200-600ms of one another, T2 is likely to *not* make it into working memory and consciousness. However, if T1 and T2 appear sequentially, T2 performance is not impacted or is "spared" (e.g., Lag-1 sparing). Crucially, all this happens in the same foveated spatial location and for each condition the task requirements are identical and explicit without any "cloak and dagger" experimental psychology tricks.

Numerous models have been introduced to explain Lag-1 sparing and the AB using attentional mechanisms. Across all the relevant theories there are nuances, however there are essentially 3 AB frameworks: attentional gating (e.g., Olivers & Meeter, 2008), loss of attentional control (e.g., Di Lollo et al., 2005) and attentional encoding bottlenecks (e.g., Wyble, Bowman & Nieuwenstein, 2009). Attentional gating models predict that T1 generates an attentional episode which drives the consolidation of this item into working memory, the trailing distractor causes an attentional gate to close to prevent interference for a period of ~500ms. As a result, when T2 appears in this window it is less likely to be reported. Loss of control models are similar in that it is the distractor item following T1 that triggers the AB, however in this case it causes the system to rapidly reconfigure attentional templates from target processing to distractor processing, which temporarily impairs processing. Finally, bottleneck models hypothesise that encoding information into working memory, an attentional operation, is capacity limited and takes ~500ms to complete. Consequently, at short T1–T2 lags, T2 must wait for T1 to be encoded leaving its fleeting representation open to interference/masking from distractors.

Irrespective of one's theoretical persuasion, it is difficult to characterise the AB as being perceptual in nature and no framework makes this prediction. Indeed, Lag-1 sparing is particularly difficult to imagine as a simple perceptual bottleneck since individuals are apparently selecting two items in succession when they are close in time but have more difficulty when they are farther apart in time. In addition, Lag-1 sparing can spread across multiple targets (spreading of the sparing) if these items are not separated by distractors (Di Lollo et al., 2005). It is difficult to characterise this phenomenon without reference to something like attentional selection. Similarly, the AB must represent processes between perception and memory as no T2 deficit is observed if one is presented with the same dual-target paradigm but instructed to ignore T1 (Raymond, Shapiro & Arnell, 1992). In addition, AB magnitude is increased when T1 is made more difficult to process (e.g., Dux & Harris, 2007) - the classic bottleneck effect.

Collectively, this paradigm clearly illustrates that, under conditions where summary statistics, perception, characteristics of peripheral vision and/or aspects of task requirements cannot drive effects, selection and attention are important for describing mechanisms underlying the encoding of information into memory. Moving away from these concepts will hinder rather than facilitate our understanding of vision as we would end up having to re-invent attention-like operations.

Acknowledgements. NA.

Financial support. PED was supported by funding from the Australian Research Council and AUSMURI

Competing interests. The authors have no conflicts of interest of competing interests.

References

Di Lollo, V., Kawahara, J., Ghorashi, S. M. S., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control. *Psychological Research*, *69*, 191–200.

Dux, P. E., & Harris, I. M. (2007). Viewpoint costs occur during consolidation: Evidence from the attentional blink. *Cognition*, *101*, 47–58.

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception & Psychophysics*, *71*, 1683–1700.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, *9*, 296–305.

Olivers, C. N. L., & Meeter, M. (2008). A boost and bounce theory of temporal attention. *Psychological Review*, *115*, 836–863.

Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 509–522.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 849–860.

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception & Performance*, *35*, 787–807.

Low experimental power makes a crisis in visual attention inevitable, but easy to address

Gregory Francis[a]* and Evelina Thunell[b]

[a]Purdue University, Psychological Sciences, West Lafayette, IN, USA and [b]Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden
gfrancis@purdue.edu
evelina.thunell@ki.se

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

*Corresponding author.

doi:10.1017/S0140525X25000263, e140

Abstract

We suggest that the crisis in visual attention is caused by problems prevalent in many other areas of psychology: low-power experiments and questionable research practices. We demonstrate that many attention experiments use sample sizes that are several orders of magnitude too small and argue that it is unreasonable to expect theoretical clarity based on such unreliable empirical data.

The target article noted that we (Francis & Thunell, 2022) previously reported that many multi-study publications on object-based attention seem "too good to be true" in the sense that the success rate of the experimental outcomes (e.g. finding a significant result) dramatically exceeded the estimated power of those experiments, based on the reported effects and sample sizes. Possible explanations for this inconsistency include that failed studies were not reported (publication bias) and that studies were being run incorrectly (questionable research practices), which are both causes of psychology's replication crisis over the past decade. Whatever the reason, the published studies likely exaggerate the true effect.

We further noted that *none* of the analyzed studies on object-based attention reported a power analysis; not even those performing a direct replication. Here, we perform a power analysis based on the large-sample study reported by Francis and Thunell (2022) where the object-based attention effect (the difference between response times for a same object condition and a different object condition) amounts to a statistically significant effect of 14 ms (first row of Table 1). If this finding is representative of the population object-based attention effect, a sample size of $n = 222$ participants would be needed to give 80% power for a dependent *t*-test (Francis, 2024a). The mean sample size in the studies analyzed by Francis and Thunell (2022) is 17 (last column of Table 1), which is thus nowhere near large enough to reliably show this effect. To make things worse, the studies did not aim to merely show the basic object-based effect, but rather modulations thereof, which entail lower power still. As an example, suppose some special population group has a reduced object-based attention effect that is only half as large (7 ms) as that of controls. If we compare this reduced effect to the standard 14 ms effect using an independent samples *t*-test, 1756 participants per group are required to achieve 80% power (Table 1; some scientists advocate for higher power, so we include calculations also for 90% power).

We conducted a similar analysis for four other common experimental paradigms in attention (Spatial Cueing, the Stroop Effect, the Simon Effect, and the Emotional Stroop Effect). We performed power analyses based on data from online versions of these paradigms where thousands of students participated (Francis & Neath, 2015; Francis, 2024b). Comparing the estimated sample sizes needed to achieve reasonable power (Table 1; column 5) with the sample sizes used for experiments in recent meta-analyses in the respective fields, it is clear that, also in these fields, sample sizes

Table 1 (Francis and Thunell). Power analyses for five attention paradigms. The independent test compares the basic effect in a control group against an experimental group that has a basic effect with half the magnitude of the control group. Commonly used sample sizes in experimental studies in the meta-analysis articles (far right column) are much too small to reliably show effects of interest

				N for 80% (90%) power		
Paradigm	Conditions	Mean (ms)	SD (ms)	Dependent t-test for basic effect	Independent t-test for modulation of basic effect (each group)	Mean N in published studies (range)
Object-based attention	Same Object	599	100	222 (292)	1756 (2350)	17 (6, 52)
	Different Object	613	100			
	Difference	14	74			
Spatial Cueing	Valid cue	359	105	62 (82)	480 (642)	18 (5, 49)
	Invalid cue	398	150			
	Difference	39	107			
Stroop Effect	Congruent	931	306	9 (11)	52 (70)	50 (14, 107)
	Incongruent	1116	363			
	Difference	185	167			
Simon Effect	Congruent	657	235	21 (27)	150 (200)	18 (8, 48)
	Incongruent	713	260			
	Difference	56	86			
Emotional Stroop Effect	Neutral Words	860	281	3191 (4271)	25509 (34149)	19 (5, 42)
	Emotional Words	864	284			
	Difference	4	86			

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

are often too small. The paradigm that fares best is the Stroop Effect: According to our power analysis, samples of $n = 9$ are sufficient to show the basic effect with 80% power, and a recent meta-analysis of the Stroop Effect (Šömen, *et al.*, 2023) indicates that samples of size $n = 50$ are common. However, modern studies typically do not measure just the basic Stroop Effect but instead look for modulations of the Stroop Effect, which generally reduces power: A study investigating, for example, half a difference in the effect across independent samples needs 52 participants per sample to have 80% power (Table 1).

For the three other considered paradigms, the studies included in meta-analyses (Cespón, Hommel, Korsch, & Galashan, 2020; Cisler, *et al.*, 2011; Landry & Parker, 2013) generally used samples much too small to consistently show even the basic effects (Table 1, last column). Strikingly, scientists commonly run experiments on the Emotional Stroop Effect with sample sizes similar to those used for the Stroop Effect, even though the magnitude of the Emotional Stroop Effect (4 ms) is estimated to be only about 2% of the classic Stroop Effect (185 ms). These small sample sizes should not be taken as evidence that the effects are not real, but it seems unlikely that the reported modulations have been accurately measured. We suspect that many of the conclusions derived from these studies are chasing noise, and will not replicate in new studies.

Thus, while we agree with the target article that there is a crisis in visual attention, we are not convinced that the field fits the first of Kuhn's (1962) signs of a crisis: that there be significant anomalies that must be explained. Rather, if the visual attention paradigms discussed here (Table 1) are representative, the observed "anomalies" might simply be due to noise. Indeed, the apparent consistency of some effects across multiple studies highlights a different kind of crisis: Scientists likely do not publish their (many) failed experimental results, which means that the published studies likely misrepresent the true effects. It is impossible to build good theories (based on summary statistic encoding or otherwise) until there is good experimental data with which those theories can be calibrated and compared.

Fortunately, it is relatively easy to address these problems by running informed power analyses that consider the effect sizes and the planned comparisons. Our large-sample data sets (Table 1) are good starting points to identify appropriate sample sizes for these paradigms; but, we suspect that scientists could do better with lab-based experiments, more trials per participant, and optimization of stimuli that lead to bigger effects; these methodological improvements may allow for smaller sample sizes than what is listed in Table 1. On the other hand, it is common for investigations to base their conclusions on a complex set of hypothesis test outcomes, and ensuring a high probability that all relevant outcomes are statistically significant requires larger samples. For paradigms where a single large dataset is not already available, scientists can use published effect sizes to perform power analyses. Given our concerns that the literature is biased with overestimated effects, this initial power analysis might indicate sample sizes that are still too small to give adequate power. Our hope is that new unbiased experiments will better estimate reality and allow for future experiments to identify appropriate sample sizes and thereby contribute to a better understanding of visual attention.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. None.

References

Cespón, J., Hommel, B., Korsch, M., & Galashan, D. (2020). The neurocognitive underpinnings of the Simon effect: An integrative review of current research. *Cognitive, Affective, & Behavioral Neuroscience*, *20*, 1133–1172.

Cisler, J. M., Wolitzky-Taylor, K. B., Adams, Jr., T. G., Babson, K. A., Badour, C. L. & Willems, J. L. (2011). The emotional Stroop task and posttraumatic stress disorder: A meta-analysis. *Clinical Psychological Review*, *31*(5), 817–828.

Francis, G. (2024a). *IntroStats Online*. YIP Software. https://yipstats.com/OnlineStatistics/chapters/calculators/calculators.shtml

Francis, G. (2024b). *STATLAB*. YIP Software. https://yipstats.com

Francis, G. & Neath, I. (2015). *Cognitive psychology online laboratory (CogLab)*. Cengage Publishing. https://coglab.cengage.com

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

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Landry, O., & Parker, A. (2013). A meta-analysis of visual orienting in autism. *Frontiers in Human Neuroscience*, *7*, 883. https://doi.org/10.3389/fnhum.2013.00833

Šömen, M. M., Peskar, M., Wollesen, B., Gramann, K., & Marusic, U. (2023). Does standing up enhance performance on the Stroop task in healthy young adults? A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, *20*, 2319. https://doi.org/10.3390/ijerph20032319

Pay attention to eye movement behavior

Hayward J. Godwin[a,*], Michael C. Hout[b] and Anthony S. Barnhart[c]

[a]University of Southampton, Highfield, Southampton, England; [b]New Mexico State University, Las Cruces, NM, USA and [c]Carthage College, Kenosha, WI, USA
hayward.godwin@soton.ac.uk
mhout@nmsu.edut
abarnhart@carthage.edu
https://michaelhout.com/
http://www.anthonybarnhart.com/

*Corresponding author.

doi:10.1017/S0140525X25000275, e141

Abstract

The review argues that researchers can and should disregard the examination of eye movement behavior when seeking to understand the role that peripheral vision plays in various tasks. We present specific examples to argue that eye movement behavior has and will continue to aid in confirming and disconfirming hypotheses regarding visual-cognitive information processing in a variety of tasks.

The review argues that researchers should disregard eye movement behavior when trying to understand the role that peripheral vision plays in a variety of visual cognitive tasks. In some senses, one could argue that the decision whether or not to record eye movement behavior is a mere methodological concern. However, the reality is that by studying eye movement behavior, researchers

can unlock the potential to confirm or disconfirm a wide range of hypotheses that may otherwise have taken years or even decades to resolve. Recording eye movement behavior enables researchers to gain fine-grained insights into moment-to-moment information processing as observers engage in a given task (Liversedge & Findlay, 2000; Rayner, 2009). Such insights are typically beyond the reach of behavioral measures such as response times (RTs) and accuracy alone. Eye movement metrics include fixation durations, the number of fixations, the likelihood of revisiting objects, the order of visiting objects, rate/direction of microsaccades, and many others (e.g., see Godwin et al., 2021). As a result, by recording eye movement behavior, researchers can better understand the spatiotemporal dynamics of information processing during visual cognitive tasks. Here, we provide some examples highlighting this point, focusing on how the recording of eye movement behavior enables the refinement, development, and resolution of theoretical arguments in the literature.

Decades ago, Rayner and colleagues examined the importance of peripheral vision in reading, and they did so by harnessing the insights offered to them by the recording of eye movement behavior (for a review, see Rayner, 2009). Such insights would have been impossible – quite literally – without the use of an eye tracker, because many of their studies used *gaze-contingent* displays. A gaze-contingent display changes what is presented to a participant based upon where they are fixating. A prototypical example might be a task wherein the letters in a sentence that are currently being fixated are visible, but beyond a small region (invisible to the participant), all other letters in that sentence are set to the letter 'x'. Had Rayner and colleagues used only RTs and response accuracy, it would not have been possible to measure the attentional window during reading, nor would it have been possible to determine that, in English reading, the window is skewed to the right (Rayner, 2009).

In the field of visual search, researchers have long assumed that observers focus on objects that share featural similarity to the target (Duncan & Humphreys, 1989; Wolfe, Cave & Franzel, 1989). This assumption was confirmed by eye movement research which demonstrated that searchers are more likely to fixate objects which are similar to the target object (Becker, 2011; Luria & Strauss, 1975; Rayner & Fisher, 1987; Stroud et al., 2012; Williams, 1967), and this is true both in terms of visual and semantic similarity (Godwin, Hout & Menneer, 2014). Without the recording of eye movement behavior, such confirmations of long-standing assumptions by theories and models alike would not have been possible.

So far, we have provided examples wherein utilizing eye movement analyses either confirmed or extended what was assumed by theories and models. But there are other examples which we can draw upon which demonstrate that, by paying attention to eye movement behavior, we can also adjudicate between competing theoretical arguments. For example, Wolfe and Horowitz (1998), on the basis of a paradigm wherein a search display regularly changed the location of all objects, argued that 'visual search has no memory' for already-inspected objects. This was a strong and controversial theoretical claim. In response, eye movement studies served to discount the claims made, finding that the rate of revisits to already-inspected objects was low (e.g., see Kristjánsson, 2000; Peterson et al., 2001), and that Wolfe and Horowitz's (1998) paradigm instead caused participants to 'sit and wait' – i.e., hold their eyes still – rather than to actively search for a target (Geyer, Von Mühlenen & Müller, 2007; Von Mühlenen, Müller & Müller, 2003). In this example, the study of eye movement behavior actively contributed to correcting the scientific record.

There are many other examples that demonstrate the value of paying attention to eye movement behavior, particularly in the context of inquiry into peripheral vision. For example, Tatler and Vincent (2009) reported the surprising finding that individual differences in eye-movement behavior are more predictive of information extraction from a scene than are bottom-up salience maps. Elsewhere, findings regarding change and inattentional blindness, which were deemed surprising or even shocking, were often explainable in terms of where participants were fixating (e.g., see Slattery, Angele & Rayner, 2011). In other cases, eye movements ruled out fixational (and peripheral crowding) explanations for these phenomena (Memmert, 2006; Smith, Lamont & Henderson, 2012). In all cases, unique conclusions were drawn that have a direct bearing upon theory that would not have been possible without studying eye movement behavior.

The examples we have included here highlight that the study of eye movement behavior is not simply a methodological choice and should not be disregarded, as is advocated in the review. Instead, we argue that by studying eye movement behavior, we can gain nuanced insights that can confirm or disconfirm hypotheses in a manner that would otherwise have been impossible. To take another final example, researchers explaining the function of the brain would struggle to do so without examining the brain itself; researchers studying human vision may struggle to do so to the extent that they desire without studying eye tracking. Indeed, by deciding to ignore eye movement behavior, our field opens itself up to future crises. Those crises may arise because we collectively chose not to check (at least in part) our theoretical predictions and assumptions by examining how, where, when and why observers acquired visual information from the displays presented to them. We are not advocating that all visual-cognitive researchers engage in eye-tracking. Rather, we argue that our field disregards eye movements at its own peril.

Acknowledgements. None.

Financial support. This article was not supported by any funding.

Competing interests. The authors have no competing interests to declare.

References

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLoS ONE*, *6*(3), e17740. https://doi.org/10.1371/journal.pone.0017740

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Geyer, T., Von Mühlenen, A., & Müller, H. J. (2007). What do eye movements reveal about the role of memory in visual search? *Quarterly Journal of Experimental Psychology*, *60*(7), 924–935. https://doi.org/10.1080/17470210600831119

Godwin, H. J., Hout, M. C., Alexdóttir, K. J., Walenchok, S. C., & Barnhart, A. S. (2021). Avoiding potential pitfalls in visual search and eye-movement experiments: A tutorial review. *Attention, Perception, & Psychophysics*, *83*(7), 2753–2783. https://doi.org/10.3758/s13414-021-02326-w

Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review*, *21*(3), 689–695. https://doi.org/10.3758/s13423-013-0547-4

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(6693), 575–577. https://doi.org/10.1038/29068

Kristjánsson, Á. (2000). In search of remembrance: evidence for memory in visual search. *Psychological Science*, *11*(4), 328–332. https://doi.org/10.1111/1467-9280.00265

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14. https://doi.org/10.1016/S1364-6613(99)01418-7

Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception & Psychophysics*, *17*(3), 303–308. https://doi.org/10.3758/BF03203215

Memmert, D. (2006). The effects of eye movements, age, and expertise on inattentional blindness. *Consciousness and Cognition*, *15*(3), 620–627. https://doi.org/10.1016/j.concog.2006.01.001

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, *12*(4), 287–292. https://doi.org/10.1111/1467-9280.00353

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., & Fisher, D. L. (1987). Letter processing during eye fixations in visual search. *Perception & Psychophysics*, *42*(1), 87–100. https://doi.org/10.3758/BF03211517

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1924–1938. https://doi.org/10.1037/a0024322

Smith, T. J., Lamont, P., & Henderson, J. M. (2012). The penny drops: Change blindness at fixation. *Perception*, *41*(4), 489–492. https://doi.org/10.1068/p7092

Stroud, M. J., Menneer, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 113–122. https://doi.org/10.1037/a0025887

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6–7), 1029–1054. https://doi.org/10.1080/13506280902764539

Von Mühlenen, A., Müller, H. J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, *14*(4), 309–314. https://doi.org/10.1111/1467-9280.14441

Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, *27*, 355–360. https://doi.org/10.1016/0001-6918(67)90080-7

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(3), 419–433.

Lossy processing principles in 2D and 3D vision

Aaron Hertzmann*

Adobe Research, San Francisco, CA, USA
hertzman@dgp.toronto.edu

*Corresponding author.

doi:10.1017/S0140525X2500024X, e142

Abstract

The similarities between 2D summary statistics and fragmentary 3D vision suggest common principles. Specifically, both 2D and 3D visual processing discard information whenever that information is redundant or inessential for ecologically valid vision in a consistent world. Change blindness and other illusions result from information loss without awareness, when the corresponding consistency assumptions are violated.

In the target article, Rosenholtz describes how limitations of peripheral vision can explain some visual phenomena that had previously been explained in terms of attention, including performance on visual search, as well as some change blindness illusions. She adds that her proposal "has as-yet-unrealized implications for all later visual processing, as vision scientists must re-envision how the visual system finds perceptually meaningful groups and computes properties such as 3D shape from the information available from summary statistics."

Indeed, a complementary line of work claims that human vision does not construct a single coherent 3D mental representation, but is instead *fragmentary* (Linton et al., 2023), as evidenced by many experiments in which viewers show behavior inconsistent with any single 3D mental world model. Rather, mental representations of 3D are fragmented in various ways, e.g., different mental representations used for different distance ranges (Vishwanath, 2023) or different tasks (Glennerster, Rogers, & Bradshaw, 1996; Norman et al., 2005). A possible connection to the target article is suggested by Koenderink et al. (2008) argument that visual space varies with fixation, as well as the fact that some fragmentary 3D phenomena could broadly be described in terms of change blindness, e.g., (Glennerster et al., 2006).

For example, consider redirected walking (Fan, Li, & Shi, 2022), a technique for giving virtual reality users the feeling of walking in a much larger virtual environment than they physically inhabit. In two notable redirected walking methods, the entire virtual world undergoes instantaneous, discrete rotations during eye blinks (Langbehn et al., 2018) or saccades (Sun et al., 2018). If it were the case that users maintained detailed mental world models across fixations, then this world rotation would be very jarring. But users do not notice the rotations.

I have suggested combining Rosenholtz's proposals with fragmentary 3D vision, into a hypothesis in which 3D vision is *fragmentary across fixations* (Hertzmann, 2024). At any instant, vision interprets 3D scene shape and appearance in a manner constrained by the limitations of peripheral vision. Then, rather than continually merging information into a dense 3D scene representation, the brain discards most information from each instant; only some abstracted spatial and scene awareness persists over time. Fragmentariness-across-fixations answers long-standing questions about perspective and shape perception in pictures: pictures that are inconsistent with any single fixed viewpoint can nonetheless give compelling illusions of 3D shape, so long as the illusion is effective for each individual eye fixation, and separate fixations share some limited consistency in peripheral vision and spatial memory. This explains the effectiveness of multiperspective techniques in art and photography, e.g. (Agrawala, Zorin, & Munzner, 2000; Agarwala et al., 2006; Perona, 2013), over conventional linear perspective. Even impossible figures give some illusion of reality, so long as the inconsistencies are imperceptible within individual fixations (Rosenholtz, 2011).

Together, the notions of summary statistics and fragmentary 3D vision suggest a view of visual processing as *lossy*: many stages of visual processing discard information. Information is discarded during summary statistics pooling, per-fixation scene understanding, visual memory storage and retrieval, and 3D scene

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

representation. Of course, information must be discarded; we could not hope to store and access every retinal photoreceptor activation, and summary statistics are a sort of lossy image compression. But which information can be discarded, and when? A natural principle is that information loss can occur whenever information is redundant or inessential for ecologically valid vision. Indeed, several authors have argued that viewers need not maintain consistent mental world models, since viewers can rely on the consistency of the real world (Brooks, 1990; Koenderink, 1998; O'Regan, 1992). Our evolutionary history gives us no need to detect instantaneous world rotations that occur only during eye blinks.

Viewers do maintain some 3D awareness of the world outside of the current gaze. If I look at a houseplant and then look away, I can remember the houseplant and its general shape and characteristics – I have some notion of "what it looks like" – but I will not remember every leaf, or even how many leaves there were, unless I had deliberately studied these details and committed them to memory. Likewise, if I see a houseplant in peripheral vision, I get much less information about it than if I were to look right at it.

Illusions result from unaware information loss. We are often aware of having forgotten information we "ought" to have access to ("I forgot where I put my keys"), and we can be aware of having missed something for not having "paid attention." But, most visual information loss happens without awareness, and so illusions can occur when the corresponding consistency assumptions are violated – such as in laboratory experiments that probe visual awareness. The two redirected walking methods that I mentioned above each exploit change blindness from viewers discarding visuospatial information.

In the spirit of the target article, attention represents one more way that information can be discarded, i.e., because it is not selected. Understanding attention as a distinct phenomenon requires identifying which information loss cannot be explained solely by passive, automatic information loss at some stage of processing.

Recent studies point out ways that 3D vision is fragmentary – and, often, depends on the task given to participants – but offer few explanations, and rarely take gaze or peripheral vision into account. Taking into account gaze and information loss in the stages of visual processing could help contribute to a unified understanding from the retina to 3D visual cognition.

Acknowledgments. I am very grateful to Paul Linton for extensive discussions and input on this commentary.

Financial support. None.

Competing interests. The author declares none.

References

Agarwala, A., Agrawala, M., Cohen, M., Salesin, D., & Szeliski, R. (2006). Photographing long scenes with multi-viewpoint panoramas. *ACM Transactions on Graphics*, *25*(3), 853–861.

Agrawala, M., Zorin, D., & Munzner, T. (2000). Artistic multiprojection rendering. In *Proc. Eurographics Workshop on Rendering Techniques*, 125–136.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*(1), 3–15. https://doi.org/10.1016/S0921-8890(05)80025-9.

Fan, L., Li, H., & Shi, M. (2022). Redirected walking for exploring immersive virtual spaces with HMD: A comprehensive review and recent advances. *IEEE Transactions on Visualization and Computer Graphics*, *29*(10), 4104–4123.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, *36*(21), 3441–3456.

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, *16*(4), 428–432.

Hertzmann, A. (2024). Toward a theory of perspective perception in pictures. *Journal of Vision*, *24*(4). https://jov.arvojournals.org/article.aspx?articleid=2793609.

Koenderink, J. J. (1998). Pictorial relief. *Philosophical Transactions of the Royal Society A*, *356*, 1071–1086.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., Doumen, M. J. A., & Todd, J. T. (2008). Exocentric pointing in depth. *Vision Research*, *48*(5), 716–723.

Langbehn, E., Steinicke, F., Lappe, M., Welch, G. F., & Bruder, G. 2018. In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics*, *37* (4), Article 66.

Linton, P., Morgan, M., Read, J., Vishwanath, D., Creem-Regehr, S., Domini, F. (2023), New approaches to 3d vision, *Philosophical Transactions of the Royal Society B*, *378*(1869). http://doi.org/10.1098/rstb.2021.0454.

Norman, J. F., Crabtree, C. E., Clayton, A. M., & Norman, H. F. (2005). The perception of distances and spatial relationships in natural outdoor environments. *Perception*, *34*(11), 1315–1324. https://doi.org/10.1068/p5304.

O'Regan, J. K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology. 46*(3), 461–88.

Perona, P. (2013). Far and yet close: Multiple viewpoints for the perfect portrait. *Art & Perception*, *1*(1–2), 105–120.

Rosenholtz, R. (2011). What your visual system sees where you are not looking. Proc. SPIE 7865, Hum. Vis. Electron. Imaging, XVI, San Francisco, Feb. 2.

Sun, Q., Patney, A., Wei, L.-Y., Shapira, O., Lu, J., Asente, P., & Kaufman, A. (2018). Towards virtual reality infinite walking: Dynamic saccadic redirection. *ACM Transactions on Graphics*, *37*(4), Article 67.

Vishwanath, D. (2023). From pictures to reality: Modelling the phenomenology and psychophysics of 3D perception. *Philosophical Transactions of the Royal Society B*, *378*(1869). http://doi.org/10.1098/rstb.2021.0454.

The (mis)use of the gate metaphor for attention

Johan Hulleman*

Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, The University of Manchester, Manchester, UK
johan.hulleman@manchester.ac.uk

*Corresponding author.

doi:10.1017/S0140525X25000354, e143

Abstract

The field has chosen gate as its preferred metaphor for attention. This commentary will discuss the power and the consequences of this choice. It will make the case that a better metaphor is needed, to liberate our understanding of attention from the constraints imposed by the gate metaphor.

Metaphors are important. Once chosen, they determine what makes sense and what does not. If you say that travelling by train is like travelling by plane, a world of advanced bookings, seat reservations, and dynamic pricing opens up. If, on the other hand,

you say that travelling by train is like taking the underground, things become simply a question of showing up, buying a ticket, and taking the next train that comes along. So, metaphors give rise to completely different and incompatible conceptions of the same phenomenon.

The field has cast attention as a gate. Consequently, it has become easy to imagine that attention acts as a bottleneck: some things pass through while most others do not. The gate has a certain capacity, and it is important to establish how large this capacity is and what factors influence its size. Observations that do not conform to the gate metaphor create problems and lead to complications and ad hoc solutions that have been pointed out in the target article.

The gate metaphor is an answer to the undeniable fact that there are limits to the bandwidth with which information can be transmitted in the visual system. But what if a gate is not the best metaphor for attention? What if rather than acting as a gate, attention acts as a modulator for the various parts of the visual field? Like an equalizer amplifying parts that are of interest and attenuating parts that are not. This also controls bandwidth, but in a much more gradual fashion.

In principle, the visual system covers the entire visual field (hence the latter's name). By definition, that has to mean that information from the entire visual field enters the visual system. What differs, and where attention can play a certain role is how much information about a particular location enters the visual system.

Close to fixation receptive fields are small and a lot of detail is available. But the further into the periphery, the larger the receptive fields become, and the less detail can be extracted.

Saccades move the area with a lot of detail across a visual scene. Covert attention may modulate the gain for certain parts of the visual field, but only within the limits set by the physiology of the retina. Moreover, nothing ever gets completely excluded.

Together with Chris Olivers, I actually used the author's idea about the computations of summary statistics in our model of visual search (Hulleman & Olivers, 2017). In easy search, summary statistics computed over a large area can yield enough information to determine whether a target is present. Hardly any saccades are needed, and reaction times will be fast. In difficult search, only the summary statistics computed over a small area can be used for target detection. This means a cumbersome and error prone search process where many saccades are needed to cover the search display widely enough to ensure that the target can be found. (Please note how this formulation of the visual search process, one of the main culprits in the target article, gets by without the use of the word attention.)

The target paper now dismisses this approach because it "essentially treats attention as spatial selection, gating the information available". In my view, this is a consequence of the target paper getting trapped in its own gating metaphor. The gating metaphor encourages thinking in all-or-nothing outcomes. Information either comes through the gate, or it does not. This leads to theoretical problems when information that was supposed to have been left outside the gate turns out to influence behaviour after all. Thinking about attention as a modulator where all available information enters the visual system, but at varying spatial resolutions (as a consequence of the physiology of the retina and the limited actions of attention) allows a new and probably more appropriate perspective.

Especially when considered in combination with another valuable point made in the target article: What are we really asking our participants to do? Is it something that needs the high resolution of the fovea or can it be done with the lower resolution representations?

Here again, the gate metaphor of attention has a pernicious influence. If only a very small subset of the available information enters the visual system, it is important that its quality is as high as possible. For visual search, this leads to conceptualisations of the search process where high resolution representations of individual items are compared to a target template. But a lot of the time, the participant is merely asked to indicate whether a particular search display contains a target. When the target is different enough from the distractors, this is probably something that can be achieved based on low-resolution information from across the visual field.

When thinking about attention in terms of modulation of the entire visual field, the use of information from across the search display seems like a natural strategy. However, within the gate metaphor, this strategy is not even a possibility, since the low-resolution information from the periphery will never make it through the gate in the first place.

In conclusion, I would argue that the first step towards solving the problems described in the target article is to become aware of the power of metaphors. How picking any particular metaphor determines research questions and imposes implicit assumptions (e.g. if you say that attention is a spotlight, then one implicit assumption is that it can be actively moved around because this is how spotlights work). The next (and much more difficult) step is to choose a better metaphor for the fact that there is a limit to the rate with which information can be transmitted in the visual system. Here, I proposed an equalizer, but I make no claim that it is the best.

What I would say though is that, with the equalizer metaphor in mind, the use of summary statistics is a much bigger step towards the solution of the problem of attention than the author allows.

Financial support. The author was supported by UKRI grant ES/X000443/1.

Competing interests. None.

Reference

Hulleman, J. & Olivers, C. N. L. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, *40*, 1–20. https://doi.org/10.1017/S0140525X15002794

Banishing "Attention" from the study of temporal attention

Matthew Junker* and David Huber

University of Colorado Boulder, Boulder, CO, USA.
matthew.junker@colorado.edu
david.huber@colorado.edu

*Corresponding author.

doi:10.1017/S0140525X25000287, e144

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Abstract

Attention has been used to explain performance deficits in many visual tasks, even though it lacks a clear definition or distinction from visual perception. We agree with Dr. Rosenholtz that perceptual processes may account for many phenomena previously attributed to spatial attention. We further suggest that perceptual processes may underlie the study of temporal attention, namely the "attentional blink."

Dr. Rosenholtz suggests that many attention-attributed phenomena in the visuospatial domain can be accounted for by visual crowding confounds, concluding that perception is the factor limiting performance rather than attention. We support this attempt to banish attention as an explanatory mechanism in the "spatial attention" and "object attention" literatures and think this will lead to a more accurate understanding of these phenomena (see also Di Lollo, 2018; Hommel et al., 2019). We further suggest that perceptual dynamics may account for findings in the "*temporal attention*" literature. At its core, the present "crisis" in visual attention research is an issue of circular causal inference – attention is assumed to be the cause of improved performance because performance improves. Examples of this can be found in the study of *where* attention operates (Awh, Belopolsky & Theeuwes, 2012; Posner, 1980), and *when* attention operates (Cherry, 1953; Treisman, 1964). Thus, if there is merit to banishing attention from the study of spatial attention, it should also be banished from the study of temporal attention.

Temporal attention is the selection of target stimuli for report (e.g., report "target" letters while ignoring "distractor" digits) when stimuli are presented over a brief period of time. It is often studied with the rapid serial visual presentation (RSVP) procedure in which many stimuli are presented in the same spatial location at rates faster than eight per second. Although people can easily report one target in an RSVP stream, they often fail to report a subsequent target presented shortly after the first (Broadbent & Broadbent, 1987). Because the report of the later-occurring target is preserved when participants are instructed to ignore the first target, the phenomenon was deemed an *attentional* deficit and so named the "attentional blink" (AB: Raymond, Shapiro & Arnell, 1992). This terminology reflects the common assumption that limited attentional resources are busy encoding the first target and therefore, those attentional resources cannot be deployed to a second target. In the ensuing decades, the AB literature has blossomed, revealing a highly complex set of auxiliary data patterns such as "lag 1 sparing," "the spread of sparing," and a temporal shift of intrusions with lag, to name just a few highly reliable effects. To account for this ever-growing list of effects, sophisticated mathematical models of temporal attention have been developed and refined, with many new theoretical variants including additional mechanisms to accommodate additional results. This is the way that science should progress, but we suggest that a simpler account of the AB literature might emerge from the hypothesis that high-level perceptual dynamics explain the failure to report a second target.

In keeping with Rosenholtz's suggestion to remove attention as an explanation, we refer to the basic "Attentional Blink effect" as a "second target deficit" to avoid implying a particular explanation. Rosenholtz proposes that visual crowding is the principal factor limiting performance in visual-spatial tasks, and the analog in the RSVP task is "temporal crowding" (i.e., competing items presented immediately before or after a target stimulus). Temporal crowding is measured by the inter-stimulus-interval (ISI) between the offset of one stimulus and the onset of the next. For instance, the magnitude of the second target deficit varies both as a function of the interval between the first and second target as well as the ISI between intervening distractors and the targets (e.g., Chen & Zhou, 2015). The role of temporal crowding (e.g., "visual masking") in second target deficits has been the topic of considerable debate; however, it is often discussed as interacting with temporal attention (Brehaut, Enns, & Di Lollo, 1999; Grandison, Ghirardelli, & Egeth, 1997), rather than the cause of the second target deficit.

Another form of temporal crowding is termed "repetition blindness" (Kanwisher, 1987), which is a transient inability to appreciate that a second occurrence of an item is a unique perceptual event. We refer to this as a transient repetition deficit rather than repetition blindness, considering that we cannot be certain whether this is a perceptual effect. Of note, the repetition deficit with letter stimuli exhibits a time course that is nearly identical to the second target deficit in the AB literature (Rusconi & Huber, 2018). Transient repetition deficits are observed for a wide variety of stimuli and tasks, including high-level semantic categories (Tian & Huber, 2013), as expected if they reflect neural dynamics at all levels of perceptual/conceptual processing (Huber, 2014). Thus, the second target deficit in the AB literature might reflect a failure to *perceive* that the second target belongs to the category of target items. For instance, in an RSVP task with two different target letters that occur in a stream of distracting digits, the observer might perceive the line segments and visual primitives of the second target but fail to perceive that the second target is a letter. This is the core assumption of the "perceptual wink" model of Rusconi and Huber (2018), which explains both transient repetition deficits and second target deficits. In short, it is not clear whether one needs to hypothesize a complex temporal attention system when considering the role of perceptual dynamics.

The cause of the second target deficit in the AB literature is the subject of ongoing debate, and there might not be a definitive answer to this question until such time as we have neuroimaging techniques with millisecond-level measurement of all perceptual responses. In the meantime, we should follow Dr. Rosenholtz's lead in attempting to explain as much as we can by applying well-understood perceptual mechanisms to the study of attention. Perceptual mechanisms such as visual crowding or temporal crowding cannot be "turned off" just because a study is designed to study attention rather than perception. If these perceptual mechanisms occur in studies of perception, they will also occur in studies of attention. At a minimum, we should be careful in assigning labels that entail a theoretical connotation, such as "attentional blink" or "repetition blindness." Instead, labels should simply describe the effect of interest – E.g., a second target deficit or a repetition deficit.

Acknowledgements. N/A.

Financial support. This commentary was not funded.

Competing interests. We have no competing interests to report for this commentary.

References

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Brehaut, J. C., Enns, J. T., & Di Lollo, V. (1999). Visual masking plays two roles in the attentional blink. *Perception & Psychophysics*, *61*(7), 1436–1448. https://doi.org/10.3758/BF03206192

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

Chen, X., & Zhou, X. (2015). Revisiting the spread of sparing in the attentional blink. *Atten Percept Psychophys*, *77*, 1596–1607. https://doi.org/10.3758/s13414-015-0886-5

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Di Lollo, V. (2018). Attention is a sterile concept; iterative reentry is a fertile substitute. *Consciousness and cognition*, *64*, 45–49.

Grandison, T. D., Ghirardelli, T. G., & Egeth, H. E. (1997). Beyond similarity: Masking of the target is sufficient to cause the attentional blink. *Perception & Psychophysics*, *59*(2), 266–274. https://doi.org/10.3758/BF03211894

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*, 2288–2303.

Huber, D. E. (2014). The rise and fall of the recent past: A unified account of immediate repetition paradigms. In B. Ross (Ed.), *Psychology of learning and motivation*, 60. (pp. 191–226). PLM, UK: Academic Press.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, *27*(2), 117–143.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

Rusconi, P., & Huber, D. E. (2018). The perceptual wink model of non-switching attentional blink tasks. *Psychonomic Bulletin & Review*, *25*(5), 1717–1739. https://doi.org/10.3758/s13423-017-1385-6

Tian, X. & Huber, D. E. (2013). Playing 'duck duck goose' with neurons: Change detection through connectivity reduction. *Psychological Science*, *24*(6), 819–827.

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, *3*(6), 449–459. https://doi.org/10.1016/S0022-5371(64)80015-3

Putting effort into task complexity

Damian Koevoet* and Stefan Van der Stigchel*

Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, Netherlands
d.koevoet@uu.nl
s.vanderstigchel@uu.nl

*Corresponding author.

doi:10.1017/S0140525X25000238, e145

Abstract

Understanding the limits of visual processing is at the core of understanding visual attention. Rosenholtz proposes task complexity as a potential solution to identify a putative unifying capacity limit. We argue that if task complexity is indeed used to identify a unifying limit, effort must crucially be incorporated to prevent a future crisis in the field of visual attention.

Understanding the capacity limit of how much can be processed at a given time is at the core of understanding visual attention. Rosenholtz argues that existing methodologies cannot clearly identify a unifying capacity limit and instead proposes that a unifying limit may exist in task complexity. Although Rosenholtz's proposal to consider task complexity could be promising, (1) it lacks a precise definition of task complexity and (2) it ignores the internal state of the observer and incorrectly assumes that the effort an observer is able or willing to invest is constant. We argue that pursuing task complexity as a solution to the proposed problem necessitates the consideration of the amount of effort an observer exerts.

Rosenholtz uses the example of a categorical classification task to describe task complexity (e.g. "Is this a coffee table or dining room table?"). Rosenholtz proposes that the complexity of the decision boundary may reflect a given task's complexity. This definition is problematic for two reasons. First, due to the strict focus on categorical classification tasks, it is unclear how the definition precisely extends to other visual tasks. Second, this definition does not allow one to clearly measure what a decision boundary actually looks like in a given task, making it difficult to test the proposal empirically.

How to quantify task complexity? Behavioral outcomes such as response accuracy and speed may be informative of task complexity. However, as Rosenholtz points out, task complexity cannot be equated to task difficulty when based purely on behavioral outcomes. One key factor to consider is the amount of effort observers invest in a task. Task difficulty and cognitive effort are related concepts but are certainly separable (Shenhav et al., 2017; Westbrook & Braver, 2015). For example, performance on a task with constant difficulty is modulated strongly by the amount of effort the observer is willing to exert. Put differently, if an observer is 'trying harder' on a given task, capacity may appear high and vice versa. Consequently, the complexity of any task cannot solely be quantified by task performance in isolation, and the effort observers invest in the task must be incorporated.

We highlight two ways that effort may modulate capacity estimates. First, the amount of effort an observer is willing/able to exert fluctuates across trials and conditions and also differs across individuals. Fortunately, the amount of effort an observer exerts can be reliably tracked using pupil size (Beatty, 1982; Bumke, 1911; Kahneman, 1973; Koevoet, Strauch, et al., 2024; Strauch et al., 2022; van der Wel & van Steenbergen, 2018). For example, pupil size sensitively tracks how much or how precisely material is held in working memory (Master, Li, & Curtis, 2024; Robison & Unsworth, 2019; Unsworth & Miller, 2021; Unsworth & Robison, 2017). Moreover, fluctuations in pupillary responses across trials reliably predict performance (Koevoet, Naber, et al., 2023, 2024; Koevoet, Strauch, et al., 2024; Robison & Unsworth, 2019; Unsworth & Miller, 2021). Pupil size even captures individual differences in the effort invested or necessary to solve a task (Ahern & Beatty, 1979; Robison et al., 2023; Robison & Brewer, 2020; Unsworth & Robison, 2017). These findings indicate that pupil size is a sensitive measure of exerted effort in a given task across trials, conditions, and observers and show that effort is linked but cannot be equated to task difficulty. Potentially, a combination of the

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

exerted effort with behavioral outcomes may inform task complexity.

Second, observers tend to prefer less effortful ways to solve tasks. When aiming to identify a putative unifying capacity limit, experimenters would need observers to "try as hard as possible" to circumvent underestimations of capacity. A recent accumulating body of work demonstrates that when given a free choice, observers generally prefer to solve (visual) tasks while exerting the least amount of effort as possible (Ballard, Hayhoe, & Pelz, 1995; Böing et al., 2023; Draschkow, Kallmayer, & Nobre, 2021; Hoogerbrugge et al., 2023; Sahakian et al., 2023; Somai, Schut, & Van der Stigchel, 2020). One example of such behavior comes from the copying task wherein observers copy a pattern of multiple objects. Crucially, observers are free to reinspect the example pattern as often of they would like. Despite the fact that observers could likely encode ~3 items every time they inspect the example grid (Luck & Vogel, 1997), they generally choose to encode only a single item at a time and look back each time after placing an item. Observers likely weigh the required effort necessary to load multiple items against the relatively subtle cost of reinspecting the example grid by making an eye movement (Ballard et al., 1995; Van der Stigchel, 2020). Indeed, observers prefer to choose less effortful ways to solve tasks whenever possible.

This principle even extends to the level of eye movements. We recently demonstrated that pupil size can even measure the effort of planning and executing saccades across directions (Koevoet, Strauch, et al., 2023). We then let observers freely choose between two saccade targets in different directions. Results showed that participants preferred saccading in directions associated with less effort (Koevoet, Van Zantwijk, et al., 2024). Put differently, observers prefer the "lazy" option even at the level of eye movements (see Hoppe & Rothkopf, 2016; Kadner et al., 2022; Thomas, Hoppe, & Rothkopf, 2022). Thus, a crucial consideration when incorporating effort into the quantification of a putative unifying limit on visual processing must also consider that observers tend to avoid effort exertion whenever possible.

Rosenholtz proposes that identifying a putative unifying capacity limit may be achieved by considering task complexity. The quantification of task complexity may be possible by combining behavioral outcomes and the amount of exerted effort. Pupil size could serve as a potent tool to track differences in effort exertion across trials, conditions, and observers. Moreover, tasks should be designed that make observers "try as hard as possible" to prevent underestimations of capacity limits. Together, if task complexity may be used to identify a unifying capacity limit, effort must be considered to prevent an impending crisis for the field of visual attention.

Acknowledgements. None.

Financial support. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 863732).

Competing interests. None.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*(4412), 1289–1292. https://doi.org/10.1126/science.472746

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Böing, S., Ten Brink, A. F., Hoogerbrugge, A. J., Oudman, E., Postma, A., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Eye movements as proxy for visual working memory usage: Increased reliance on the external world in Korsakoff syndrome. *Journal of Clinical Medicine*, *12*(11), Article 11. https://doi.org/10.3390/jcm12113630

Bumke, O. (1911). *Die pupillenstörungen bei geistes-und nervenkrankheiten* (2nd ed.). Fischer.

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology*, *31*(4), 869–874.e5. https://doi.org/10.1016/j.cub.2020.11.013

Hoogerbrugge, A. J., Strauch, C., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Don't hide the instruction manual: A dynamic trade-off between using internal and external templates during visual search. *Journal of Vision*, *23*(7), 14. https://doi.org/10.1167/jov.23.7.14

Hoppe, D., & Rothkopf, C. A. (2016). Learning rational temporal eye movement strategies. *Proceedings of the National Academy of Sciences*, *113*(29), 8332–8337. https://doi.org/10.1073/pnas.1601305113

Kadner, F., Thomas, T., Hoppe, D., & Rothkopf, C. A. (2022). Improving saliency models' predictions of the next fixation with humans' intrinsic cost of gaze shifts. arXiv:2207.04250.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Koevoet, D., Naber, M., Strauch, C., Somai, R. S., & Van der Stigchel, S. (2023). Differential aspects of attention predict the depth of visual working memory encoding: Evidence from pupillometry. *Journal of Vision*, *23*(6), 9. https://doi.org/10.1167/jov.23.6.9

Koevoet, D., Naber, M., Strauch, C., & Van der Stigchel, S. (2024). The intensity of internal and external attention assessed with pupillometry. *Journal of Cognition*, *7*(1), Article 1. https://doi.org/10.5334/joc.336

Koevoet, D., Strauch, C., Naber, M., & Van der Stigchel, S. (2023). The costs of paying overt and covert attention assessed with pupillometry. *Psychological Science*, *34*(8), 887–898. https://doi.org/10.1177/09567976231179378

Koevoet, D., Strauch, C., Van der Stigchel, S., Mathôt, S., & Naber, M. (2024). Revealing visual working memory operations with pupillometry: Encoding, maintenance, and prioritization. *WIREs Cognitive Science*, *15*(2), e1668. https://doi.org/10.1002/wcs.1668

Koevoet, D., Van Zantwijk, L., Naber, M., Mathôt, S., Van der Stigchel, S., & Strauch, C. (2024). Effort drives saccade selection. *eLife*, *13*, 1–25. https://doi.org/10.7554/eLife.97760.1

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), Article 6657. https://doi.org/10.1038/36846

Master, S. L., Li, S., & Curtis, C. E. (2024). Trying harder: How cognitive effort sculpts neural representations during working memory. *Journal of Neuroscience*, *44*(28), 1–12. https://doi.org/10.1523/JNEUROSCI.0060-24.2024

Robison, M. K., & Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception, & Psychophysics*, *82*(7), 3273–3290. https://doi.org/10.3758/s13414-020-02077-0

Robison, M. K., Ralph, K. J., Gondoli, D. M., Torres, A., Campbell, S., Brewer, G. A., & Gibson, B. S. (2023). Testing locus coeruleus-norepinephrine accounts of working memory, attention control, and fluid intelligence. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(4), 1014–1058. https://doi.org/10.3758/s13415-023-01096-2

Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics*, *81*(2), 407–419. https://doi.org/10.3758/s13414-018-1618-4

Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2023). Mountains of memory in a sea of uncertainty: Sampling the external world despite useful information in visual working memory. *Cognition*, *234*, 105381. https://doi.org/10.1016/j.cognition.2023.105381

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2020). Evidence for the world as an external memory: A trade-off between internal and external visual memory storage. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *122*, 108–114. https://doi.org/10.1016/j.cortex.2018.12.017

Strauch, C., Wang, C.-A., Einhäuser, W., Van der Stigchel, S., & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences*, *45*(8), 635–647. https://doi.org/10.1016/j.tins.2022.05.003

Thomas, T., Hoppe, D., & Rothkopf, C. A. (2022). The neuroeconomics of individual differences in saccadic decisions (p. 2022.06.03.494508). bioRxiv. https://doi.org/10.1101/2022.06.03.494508

Unsworth, N., & Miller, A. L. (2021). Individual differences in the intensity and consistency of attention. *Current Directions in Psychological Science*, *30*(5), 391–400. https://doi.org/10.1177/09637214211030266

Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1962–1987. https://doi.org/10.1037/xlm0000421

Van der Stigchel, S. (2020). An embodied account of visual working memory. *Visual Cognition*, *28*(5–8), 414–419. https://doi.org/10.1080/13506285.2020.1742827

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Interactions between cortical and subcortical circuits for visual attention

Richard Krauzlis*

Laboratory of Sensorimotor Research, National Eye Institute/National Institutes of Health, Bethesda, USA
richard.krauzlis@nih.gov

*Corresponding author.

doi:10.1017/S0140525X25000081, e146

Abstract

Recent electrophysiological studies of visual attention have highlighted the importance of visual circuits through evolutionarily conserved brain regions in the midbrain that target processing stages downstream from early visual cortex. These findings support the target article's emphasis on late-stage "task selection" but are also consistent with early-stage modulation of basic visual features and flexible pooling of visual signals.

The target article provides a sweeping assessment of visual attention, and lands on the general proposal that the crucial bottleneck lies in selecting which task to perform, as opposed to selecting which early visual signals are processed. For very different reasons, we arrived at a similar proposal (Krauzlis et al., 2014). Being electrophysiologists interested in the brain circuits for visual attention, our suggestion was prompted by some unexpected properties of neurons in those visual circuits, as well as broader considerations about the interactions across brain regions and their possible evolutionary origins. Here I briefly outline the logic for our interpretation and point out where our views may not be fully aligned with those expressed in the target article.

It is well established that visually responsive neurons in primate neocortex are modulated in several distinctive ways with changes in visual attention, and that these changes are linked to the state of visual attention (Maunsell, 2015). It is also established that visually responsive neurons in the midbrain superior colliculus are modulated with attention (Herman, Katz, & Krauzlis, 2018; Ignashchenkova et al., 2004), and causal manipulations demonstrate that the superior colliculus is necessary for the normal control of visual attention, especially spatial attention (Cavanaugh & Wurtz, 2004; Lovejoy & Krauzlis, 2010). The contributions of the superior colliculus to the performance of attention tasks often include controlling the eye movements required by the task, but the scope of its contributions go beyond the motor. The superior colliculus contributes to visual attention even when eye movements are not involved – causal perturbations of the superior colliculus alter the allocation of covert visual attention (Lovejoy & Krauzlis, 2010, 2017). These contributions to attention require a different explanation and presumably involve ascending projections from the midbrain through the thalamus to the forebrain.

The most parsimonious version of this idea was that the superior colliculus provides feedback to neurons in early visual cortex that are modulated with attention. However, when we directly tested this idea, we found that the profound changes in attention task performance caused by manipulations of the superior colliculus did not change the visual responsiveness in cortex but instead left the attention-related modulations completely unaltered (Zenon & Krauzlis, 2012). This dissociation between the signatures of attention in visual cortex and the ability to perform visual attention tasks was surprising and led to considerable debate and some consternation.

Here we encounter a slight disagreement. The target article interprets our findings as supporting the conclusion that attention-related modulation in visual cortex might play no meaningful role in regulating attention. We drew a more conservative interpretation that we still believe is correct – that the superior colliculus exerts its influence over visual attention through ascending projections that do not primarily target early visual cortex but instead target later stages of visual processing.

Among the potential targets downstream of the visual cortex, we pointed out that circuits through the basal ganglia were an intriguing possibility (Krauzlis et al., 2014). Beyond regulating movements, the basal ganglia are known to be important for value-based decisions about how to act based on an animal's belief about the current state of the world, invoking what is often referred to as a "decision policy". We argued that the implementation of a decision policy by the basal ganglia necessarily restricts behavioral control to be guided by some sensory signals while reducing or excluding others. This has the effect of implementing a form of sensory selection, but at the point where those sensory signals are applied to the behavioral output rather than during early sensory processing itself. This perspective on basal ganglia function outlines a form of "late selection" that I think nicely aligns with the target article's emphasis on the importance of task selection. When a subject selects a task, they adopt a belief about the current state of the world and a decision policy about how they should act – including which sensory signals should be used – given that state.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

The target article goes a bit further in its conclusions, casting doubt on the importance of flexible pooling of sensory signals, a form of "early selection". I disagree on this point and would counter that adopting different pooling strategies is a necessary component of implementing the selected task. Depending on the task being performed, the subject may be obliged to pool sensory signals very differently to reach a task-appropriate decision. How these changes in pooling strategies are implemented is a fascinating and poorly understood topic, but there is compelling evidence that task-related feedback actively changes the statistical structure of neuronal activity in visual cortex (Bondy, Haefner, & Cumming, 2018). I would suggest that the alternatives presented in the conclusion of the target article are not mutually exclusive, but instead tightly linked.

There is also an argument to be made for "selection in the middle". In addition to interactions with the basal ganglia, ascending activity from the superior colliculus has an especially large functional effect in the middle of the visual cortical hierarchy, at a stage midway between the representation of basic visual features in early visual cortex and the representation of complex objects in late visual cortex (Bogadhi et al., 2019, 2021). By modulating processing at this middle point, this circuit connects the evolutionarily conserved midbrain mechanisms for fast selection of salient objects with the more recently evolved neocortical mechanisms for extracting detailed information about specific visual objects.

Extending this point about evolution, primates are not the first animals to possess something that can be identified as visual attention. We are understandably proud of our prodigious cortical forebrain, but the emergence of visual attention predates the evolution of the neocortex and involves specializations and extensions of much older forebrain and midbrain circuits (Krauzlis et al., 2018). We should be open to the possibility that understanding visual attention will require a fuller appreciation of how these complex circuits evolved and, even more germane to resolving disorders of visual attention, a fuller appreciation of how they develop and continue to undergo plastic changes even in the adult.

Acknowledgements. This work is supported by the National Eye Institute Intramural Research Program at the National Institutes of Health ZIA EY000511.

Financial support. This work was supported by National Institutes of Health ZIA EY000511.

Competing interests. The author declares no competing interests and no conflicts of interest.

Reference

Bogadhi, A. R., Bollimunta, A., Leopold, D. A., & Krauzlis, R. J. (2019). Spatial attention deficits are causally linked to an area in macaque temporal cortex. *Curr Biol*, *29*(5), 726–736.e4. https://doi.org/10.1016/j.cub.2019.01.028

Bogadhi, A. R., Katz, L. N., Bollimunta, A., Leopold, D. A., & Krauzlis, R. J. (2021). Midbrain activity shapes high-level visual properties in the primate temporal cortex. *Neuron*, *109*(4), 690–699.e5. https://doi.org/10.1016/j.neuron.2020.11.023

Bondy, A. G., Haefner, R. M., & Cumming, B. G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. *Nat Neurosci*, *21*(4), 598–606. https://doi.org/10.1038/s41593-018-0089-1

Cavanaugh, J., & Wurtz, R. H. (2004). Subcortical modulation of attention counters change blindness. *The Journal of Neuroscience*, *24*(50), 11236–11243. https://doi.org/10.1523/JNEUROSCI.3724-04.2004

Herman, J. P., Katz, L. N., & Krauzlis, R. J. (2018). Midbrain activity can explain perceptual decisions during an attention task. *Nat Neurosci*, *21*(12), 1651–1655. https://doi.org/10.1038/s41593-018-0271-5

Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nat Neurosci*, *7*(1), 56–64. https://doi.org/10.1038/nn1169

Krauzlis, R. J., Bogadhi, A. R., Herman, J. P., & Bollimunta, A. (2018). Selective attention without a neocortex. *Cortex*, *102*, 161–175. https://doi.org/10.1016/j.cortex.2017.08.026

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, *18*(9), 457–464. https://doi.org/10.1016/j.tics.2014.05.008

Lovejoy, L. P., & Krauzlis, R. J. (2010). Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nat Neurosci*, *13*(2), 261–266. https://doi.org/10.1038/nn.2470

Lovejoy, L. P., & Krauzlis, R. J. (2017). Changes in perceptual sensitivity related to spatial cues depends on subcortical activity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(23), 6122–6126. https://doi.org/10.1073/pnas.1609711114

Maunsell, J. H. R. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science*, *1*, 373–391. https://doi.org/10.1146/annurev-vision-082114-035431

Zenon, A., & Krauzlis, R. J. (2012). Attention deficits without cortical neuronal deficits. *Nature*, *489*(7416), 434–437. https://doi.org/10.1038/nature11497

Attention is doing just fine! Just don't take it too seriously

Árni Kristjánsson[a]* and Andrey Chetverikov[b]

[a]Icelandic Vision Laboratory, University of Iceland, Reykjavík, Iceland and [b]Department of Psychosocial Science, Faculty of Psychology, University of Bergen, Bergen, Norway
ak@hi.is
Andrey.Chetverikov@uib.no
www.visionlab.is

*Corresponding author.

doi:10.1017/S0140525X25000159, e147

Abstract

We do not share Rosenholtz's central worry that visual attention is in "crisis". There are many examples of notable progress in understanding how the brain prioritizes and gathers information about the environment where "attention," as a relatively loose concept, has worked well. We also discuss how focusing on a single definition, the field can be led astray.

While we agree with Dr. Rosenholtz on many points, we do not share her central worry that visual attention is in crisis. Notable progress has recently been made in understanding how the brain prioritizes and gathers information about the environment – where relatively loose definitions of attention seem to work just fine for various paradigms. Cueing studies, visual search, rapid serial visual presentation, ensemble perception, inattentional and change blindness, and other paradigms show how wide the reach of the concept is. Thinking about attention as 'one thing' or even 'a thing' with a specific operational definition may, in fact, dampen progress.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

By focusing too narrowly on a particular definition of attention, the field can be led astray. For example, a very popular definition of visual attention was proposed in the two-stage, preattentive vs. attentive model, where feature integration (requiring attentional resources) is needed to find a target unless a particular item stands out in a 'feature map' (Neisser, 1963; Treisman & Gelade, 1980; Wolfe, 1998: see Kristjánsson & Egeth, 2020 for a historical overview).

This model is based on the highly popular present/absent visual search paradigm, and a key assumption is that search slopes of response times against distractor number can be used as a measure of to what degree attention is required for a given task. This logic has taken on a life of its own, however, used as a test of whether attention is involved in performing a particular task or not. If the slopes are around zero, it is often assumed that the target can be detected preattentively, while if search times increase with set size, an attentional spotlight moving between items in the search array is needed.

Note that this is simply a testable assumption about visual attention, and this model has been in trouble for a long time because of findings where it's central tenets are violated (e.g. McLeod, Driver, & Crisp, 1988; Nakayama & Silverman, 1986; Bravo & Nakayama, 1992; Santhi & Reeves, 2004; Schoonveld, Shimozaki, & Eckstein, 2007; Enns & Rensink, 1990; Wang, Cavanagh, & Green, 1994; Wang, Kristjánsson & Nakayama, 2005). Kristjánsson, (2015) showed that with a simple response change (from a present/absent decision to a Go/No-Go decision), search tasks with positive slopes had flat slopes. Nevertheless, the erroneous belief that visual search slopes determine whether attention is needed for a particular task persist (Kristjánsson & Kristjánsson, 2023). This shows how trying to pin attention down too narrowly might do more harm than good.

More generally, attention might simply does not need a specific definition. We do not find ourselves needing to define the term 'perception' in such a limited way, and attention can be discussed in a similarly broad and encompassing way. To paraphrase William James, everyone knows what we mean by "perception," and the same goes for attention as James originally stated. Problems arise when we seek single measures or definitions.

Despite a lack of clear definitions, considerable progress has been made in understanding the functioning of visual attention (Anderson et al., 2021; Kristjánsson, 2023). Rosenholtz discusses how summary statistics can be used to process the visual environment (Balas, Nakano, & Rosenholtz, 2009; Cohen, Dennett, & Kanwisher, 2016), referring to this as a paradigm shift. But this highlights another point where we both agree and disagree with Rosenholtz.

Research into summary statistics has taken many forms. Our work suggests that the idea of simple summary statistics, where rich data are reduced to only mean and variance (Whitney & Yamanashi-Leib, 2018), clearly underestimates how detailed representations of visual ensembles can be. Chetverikov, Campana, and Kristjánsson (2016, 2017a, 2020) used a classic attentional phenomenon – priming of pop-out (Maljkovic & Nakayama, 1994) – to test whether only simple summary statistics are encoded, but instead they found that people can quickly develop detailed probabilistic templates that guide attention and dynamically adapt to task demands. Even more interestingly, this attentional guidance accurately reflected the probabilistic nature of the input. Far more detail is therefore available in perception than simple summary statistics accounts propose (see, e.g., Chetverikov & Kristjánsson, 2024; Khvostov, Ásgeirsson & Kristjánsson, 2025).

Complex summary statistics, such as those in Rosenholtz's own work, computed with thousands of filters, can in contrast fall in line with the probabilistic template idea (Chetverikov & Kristjánsson, 2024). A summary representation created with the same statistics as those in Rosenholtz's Texture Tiling Model (Rosenholtz, Huang, & Ehinger, 2012) projected onto a single task-relevant dimension, such as orientation, can be represented as a noisy version of the true feature probability distribution (Chetverikov, Campana, & Kristjánsson 2017b).

Further progress comes with research showing how attentional templates are strategically learned or biased by perceptual and attentional history (Geng, DiQuattro, & Helm, 2017). For example, Witkowski and Geng (2019) demonstrated how the visual system prioritizes stable features over variable features. Won and Geng (2018) used a similar approach to demonstrate that so-called templates for rejection of distractors are more broadly tuned than target templates (since the latter require sharper tuning). Hansmann-Roth et al. (2022) showed how the tuning process leads to increasingly precise representations of target colour distributions. This work elucidates factors influencing the tuning of attention templates – again without a definition of attention.

Overall, this work provides important clues about how preceding probabilistic information about features and their reliability guides attention. All these insights about how the visual system picks up and learns the statistics of the environment are available without any precise definition of attention. Focusing instead on definitions of the tasks that observers perform and building computational models of their decisions (e.g., Chetverikov & Kristjánsson, 2022), following a long tradition of observer models in vision science, seem to be a more promising avenue, and here we applaud Rosenholtz's proposal.

In our opinion, visual attention research is in fine fettle and precise operational definitions are not needed. In fact, too precise operational definitions can be limiting, as examples from recent decades show.

Financial support. Funded by the Icelandic Research Fund and the research fund of the University of Iceland

Competing interests. The authors declare no competing interests

References

Anderson, B. A., Kim, H., Kim, A. J., Liao, M. R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews*, *130*, 326–350.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of vision*, *9*(12), 13–13.

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & psychophysics*, *51*(5), 465–472.

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, *153*, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017a). Representing color ensembles. *Psychological science*, *28*(10), 1510–1517.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017b). Learning features in a complex and changing environment: A distribution-based framework for visual attention and vision in general. *Progress in brain research*, *236*, 97–120.
Chetverikov, A., Campana, G., & Kristjánsson, A. (2020). Probabilistic rejection templates in visual working memory. *Cognition*, *196*, 104075.
Chetverikov, A., & Kristjánsson, A. (2022). Probabilistic representations as building blocks for higher-level vision. *Neurons, Behavior, Data Analysis, and Theory*, *6*(1), 1–32. https://doi.org/10.51628/001c.24910
Chetverikov, A., & Kristjánsson, Á. (2024). *Representing Variability: How Do We Process the Heterogeneity in the Visual Environment?*. Cambridge University Press. https://doi.org/10.1017/9781009396035
Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience?. *Trends in cognitive sciences*, *20*(5), 324–335.
Enns, J. T., & Rensink, R. A. (1990). Sensitivity to three-dimensional orientation in visual search. *Psychological Science*, *1*(5), 323–326.
Geng, J. J., DiQuattro, N. E., & Helm, J. (2017). Distractor probability changes the shape of the attentional template. *Journal of experimental psychology: human perception and performance*, *43*(12), 1993.
Hansmann-Roth, S., Þorsteinsdóttir, S., Geng, J. J., & Kristjánsson, Á. (2022). Temporal integration of feature probability distributions. *Psychological Research*, *86*(6), 2030–2044.
Khvostov, V., Ásgeirsson, Á. G., & Kristjánsson, Á. Explicit Access to Detailed Feature Distribution Representations. Open Science Framework, https://files.osf.io/v1/resources/vt98a/providers/osfstorage/66f57ed78b7d82c5b09e8728?action=download&direct&version=1
Kristjánsson, Á. (2015). Reconsidering visual search. *i-Perception*, *6*(6), 2041669515614670.
Kristjánsson, Á. (2023). Priming of probabilistic attentional templates. *Psychonomic Bulletin & Review*, *30*(1), 22–39.
Kristjánsson, Á., & Egeth, H. (2020). How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics. *Attention, Perception, & Psychophysics*, *82*, 7–23.
Kristjánsson, Á., & Kristjánsson, T. (2023). Attentional priming in Go No-Go search tasks. *Vision Research*, *213*, 108313.
Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. role of features. *Memory & cognition*, *22*(6), 657–672.
McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, *332*(6160), 154–155.
Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264–265.
Neisser, U. (1963). Decision-time without reaction-time: Experiments in visual scanning. *The American journal of psychology*, *76*(3), 376–385.
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*, 13.
Santhi, N., & Reeves, A. (2004). The roles of distractor noise and target certainty in search: A signal detection model. *Vision Research*, *44*(12), 1235–1256.
Schoonveld, W., Shimozaki, S. S., & Eckstein, M. P. (2007). Optimal observer model of single-fixation oddity search predicts a shallow set-size function. *Journal of Vision*, *7*(10), 1–1.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.
Wang, D., Kristjánsson, A., & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. *Perception & Psychophysics*, *67*, 239–253.
Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & psychophysics*, *56*, 495–500.
Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, *69*, 105–129.
Witkowski, P., & Geng, J. J. (2019). Learned feature variance is encoded in the target template and drives visual search. *Visual Cognition*, *27*(5–8), 487–501.
Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39.
Wolfe, J. M. (1998a). Visual search. In H. E. Pashler (Eds.), *Attention* (pp. 13–76). Hove, England: Psychology Press.
Won, B. Y., & Geng, J. J. (2018). Learned suppression for multiple distractors in visual search. *Journal of experimental psychology: human perception and performance*, *44*(7), 1128–1141.

Attention is still a productive framework

Tianye Ma, Inik Kim, Sahereh Varastegan, Li Yang, Jessica Golding and Weiwei Zhang*

Department of Psychology, University of California, Riverside, CA, US
mack.ma@email.ucr.edu
inik.kim@email.ucr.edu
svara010@ucr.edu
lyang147@ucr.edu
jgold065@ucr.edu
Weiwei.zhang@ucr.edu

*Corresponding author.

doi:10.1017/S0140525X25000329, e148

Abstract

The debate on attention's validity in cognitive psychology persists. However, attention remains essential beyond peripheral vision constraints, as it is a resource-limited process (Norman & Bobrow, 1975). The outright dismissal of attention proposed in the target article risks conceptual voids without superior alternatives. Instead, refining attention as a theoretical framework offers a pragmatic path for advancing cognitive research.

The concept of attention has come under scrutiny in psychology and neuroscience, with Rosenholtz and some other researchers advocating for its rejection as a meaningful construct in empirical research (e.g., Anderson, 2011, 2023, p. 202; Hommel et al., 2019). Critics argue that "attention" is too broadly applied, leading to ambiguity and circular reasoning. However, dismissing attention entirely risks discarding a framework that has driven significant progress in understanding cognition. Here, we challenge the notion that attention is an outdated or unnecessary construct and argue for its continued relevance as a pragmatic tool in cognitive science.

Is attention a redundant concept?

A major argument against attention as a theoretical construct is its purported redundancy. Critics claim that attentional effects can be explained through alternative mechanisms such as Bayesian decision processes (Anderson, 2011) or low-level constraints such as peripheral vision proposed by Rosenholtz. These perspectives suggest that empirical anomalies in attentional research result from misattributed causal mechanisms rather than failures in theoretical formulation.

However, this reasoning conflates explanatory adequacy with the overuse or misuse of the terminology. As Wu (2024) argues, attention should not be seen as an explanation in itself but rather as the effect that needs to be explained. Replacing attention with constructs such as "coarse-grained processing in peripheral vision" does not necessarily lead to greater precision in understanding cognition and behavior; rather, it shifts the label without resolving the underlying mechanisms. The assertion that attention is redundant assumes that alternative explanations are sufficiently comprehensive, yet they have not demonstrated the predictive

power necessary to fully replace attentional theories. In science history, Kepler's law generates good predictions that can be generalized from the solar system to other systems. But what about the "peripheral vision"? It is only a description of the phenomenon instead of any kind of cognitive theory that can generate testable and generalizable predictions.

The role of peripheral vision in attentional tasks: A confound, not a replacement

Rosenholtz proposes that many attentional effects can be better explained as limitations imposed by the structural and intrinsic features of peripheral vision (e.g., crowding and extraction of summary statistics) rather than an attentional bottleneck. While it is true that uncontrolled eye movements and peripheral processing influence task performance, these factors do not negate the existence of attentional mechanisms. The key question is not whether peripheral vision plays a role but rather how much variance in attentional tasks can be accounted for by peripheral factors alone.

Empirical evidence suggests that attentional processes cannot be fully reduced to sensory limitations. First, pupil size—a physiological marker of attentional effort—is associated with enhanced detection performance in peripheral vision and is modulated by cognitive demand (Eberhardt et al., 2022; Vilotijević & Mathôt, 2024). Second, studies have shown that attention can enhance neural gain even when sensory input remains unchanged (Hillyard & Anllo-Vento, 1998), reinforcing its role beyond mere sensory encoding. Third, well-documented phenomena such as temporal attention (e.g., attentional blink, Luck, Vogel, & Shapiro, 1996; Raymond, Shapiro, & Arnell, 1992) and feature-based attention (Zhang & Luck, 2009) suggest that attention functions extend beyond mere sensory constraints. Specifically, attentional blink reflects transient competition for attentional resources rather than simple perceptual masking (Chun & Potter, 1995). Four, the contrast between resource and sensory limitations in distractor processing further supports the need for an "attentional theory" (Lavie & De Fockert, 2003). Lastly, attention as a theoretical construct can and should have strong utility, for instance as an account for cognitive process such as effort (Xie & Zhang, 2023).

The approach in the target article conceptually aligns with the notion of data-limited processing, which is distinct from attentional effects (i.e., resource-limited processing), as proposed by Norman & Bobrow (1975). Specifically, cognitive performance is constrained by the quality of available perceptual information (e.g., weak or ambiguous sensory input), such that additional cognitive processes, including attention, cannot enhance performance. Consistent with this perspective, the target article argues that many phenomena traditionally attributed to attention can instead be explained by the inherently low quality of peripheral visual information, rather than by top-down attentional modulation. In this sense, visual cognition operates under data-limited constraints, with peripheral vision representing one specific form of data limitation. Norman and Bobrow (1975) clearly demonstrated that even after accounting for data-limited constraints, cognitive resources such as attention are still necessary to modulate and optimize processing beyond these inherent constraints.

The pragmatic case for attention as a cognitive construct

Attention as a multifaceted construct (Luck & Vecera, 2002) involves multiple sub-processes and mechanisms, such as sensory enhancement versus noise reduction (Hillyard & Anllo-Vento, 1998). It is clearly inconsistent with the unified account as proposed in the target article. Even if attention lacks a single, universally agreed-upon definition, it remains a useful heuristic for organizing empirical findings and guiding theoretical development. The history of science demonstrates that concepts often persist despite definitional ambiguities—consider "representation" in cognitive neuroscience (Palmer, 1978) or "capacity" in working memory research (Kane et al., 2001; Oberauer et al., 2016; Zhang & Luck, 2008). These terms continue to provide valuable insight even when their underlying mechanisms are debated.

Replacing attention with alternative paradigms does not necessarily yield theoretical elegance. Unlike Kepler's laws that shifted the field in astronomy by replacing geocentric (Earth-centered) models with a heliocentric (Sun-centered) framework, the proposed alternatives to attention remain fragmented and context-dependent. Given that attentional research has yielded significant progress in fields ranging from neuroscience (Petersen & Posner, 2012; Posner & Petersen, 1990) to translation research (e.g., ADHD, Shalev, Tsal, & Mevorach, 2007) and artificial intelligence (Soydaner, 2022), rejecting attention prematurely would introduce conceptual voids without guaranteeing superior explanatory models.

Conclusion: Attention as a productive framework

Anomalies in attentional research should be viewed not as justification for rejecting attention but as opportunities to refine the theory. The field benefits from critical discussions about its foundational concepts, but discarding attention outright is neither necessary nor advantageous. Unless a unified theory emerges that demonstrably outperforms attentional explanations in scope and predictive power, attentional "realism" remains the most pragmatic approach for advancing cognitive science.

Financial support. The work was supported by NIMH R01MH117132 to W. Zhang.

Competing interests. None.

References

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*, 246.

Anderson, B. (2023). Stop paying attention to "attention." *Wiley Interdisciplinary Reviews: Cognitive Science*, *14*(1), e1574.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109.

Eberhardt, L. V., Strauch, C., Hartmann, T. S., & Huckauf, A. (2022). Increasing pupil size is associated with improved detection performance in the periphery. *Attention, Perception, & Psychophysics*, *84*(1), 138–149.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*, 2288–2303.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Lavie, N., & De Fockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception & Psychophysics*, *65*(2), 202–212.
Luck, S. J., & Vecera, S. P. (2002). Attention. *Stevens' Handbook of Experimental Psychology*, *1*, 235–286.
Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618.
Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64.
Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758.
Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Routledge.
Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*(1), 73–89.
Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42.
Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849.
Shalev, L., Tsal, Y., & Mevorach, C. (2007). Computerized progressive attentional training (CPAT) program: Effective direct intervention for children with ADHD. *Child Neuropsychology*, *13*(4), 382–388.
Soydaner, D. (2022). Attention mechanism in neural networks: Where it comes and where it goes. *Neural Computing and Applications*, *34*(16), 13371–13385.
Vilotijević, A., & Mathôt, S. (2024). Functional benefits of cognitively driven pupil-size changes. *Wiley Interdisciplinary Reviews: Cognitive Science*, *15*(3), e1672.
Wu, W. (2024). We know what attention is! *Trends in Cognitive Sciences*, *28*(4), 304–318.
Xie, W., & Zhang, W. (2023). Effortfulness of visual working memory: Gauged by physical exertion. *Journal of Experimental Psychology: General*, *152*(7), 2074.
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.
Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*(1), 24–25.

Is allocentric neglect an attentional disorder?

Marialuisa Martelli[a], Mauro Mancuso[b] and Pierluigi Zoccolotti[a,c*]

[a]Department of Psychology, Sapienza University of Rome, Rome, Italy; [b]Physical and Rehabilitative Medicine Unit, NHS-USL Tuscany South-East, Grosseto, Italy and [c]Tuscany Rehabilitation Clinic, Montevarchi, Arezzo, Italy
marialuisa.martelli@uniroma1.it
mauro.mancuso@uslsudest.toscana.it
pierluigi.zoccolotti@uniroma1.it

*Corresponding author.

doi:10.1017/S0140525X25000202, e149

Abstract

We examine "allocentric" neglect from Rosenholtz's perspective, focusing on visual limitations in the periphery rather than attention. Studies of allocentric neglect report no or limited data on visual field defects and extinction. We propose that if visual input is degraded, patients may fail to detect the left part of the stimulus and complete it perceptually (preattentively) rather than attentionally.

Neglect is a multifactorial syndrome typically ascribed to an attentional disorder in detecting and identifying stimuli in the contralesional body or space. The concept of attention is central not only to cognitive theories but also to understand the behavioral consequences of brain damage in neuropsychological analysis and theorizing. For example, whether visuospatial neglect – commonly exhibited by patients with right-brain damage – should be considered a perceptual or attentional phenomenon has been debated since the early 20th century. Over time, an attentional interpretation has gained the upper hand, while sensory/perceptual limitations have been progressively pushed aside. Rosenholtz's article presents new insights into re-evaluating impairments in visual search, as seen in neglect patients.

The current evidence on visuospatial neglect aligns with the (prevailing) attentional interpretation of the disorder in explaining contralateral omissions in several tasks. However, some phenomena related to neglect remain challenging to address theoretically. One such example is "allocentric" or "object-centered" neglect (Behrmann & Moscovitch, 1994; Olson, 2003). In tasks such as drawing, copying, or searching for a target among multiple stimuli, patients with right-brain damage may omit the left portion of objects regardless of their location in the visual field (e.g., Gainotti, Messerli, & Tissot, 1972). This response profile contrasts with the classical egocentric form of neglect, where patients fail to search for stimuli contralateral to their lesion.

Some case reports, particularly those concerning neglect dyslexia, suggest that these patients may experience higher-level attentional disorders (Driver & Halligan, 1991; Hillis & Caramazza, 1991; Miceli & Capasso, 2001). However, other studies show error types (i.e., letter substitutions on the left part of words) aligned with what is predicted by a perceptual integration deficit (Martelli, Arduino, & Daini, 2011; Daini et al., 2013). Still, a comprehensive theoretical framework for understanding allocentric neglect remains elusive. Conversely, it has been proposed that spatial representations invariably depend on egocentric reference frames (Filimon, 2015). Therefore, drawing on Rosenholtz's proposal, an interpretation arises: that perceptual limitations characteristic of peripheral vision might provide a more straightforward explanation for the behavioral pattern.

Tests of allocentric neglect can be conducted using various tasks (Lee et al., 2022). The complexity of the target may influence performance, with more complex targets being more susceptible to the spatial arrangement of stimuli (Gainotti & Ciaraffa, 2013). Stimulus complexity impairs peripheral object recognition in brain-damaged patients (Strappini et al., 2017). Many clinical assessments use visual search tasks, where patients scan several small targets that can be identified with just one or a few eye fixations. In the Apples test (Bickerton et al., 2011) or the Broken Hearts test (Demeyere et al., 2015), allocentric neglect is indicated by the inability to detect gaps on the left side of visual stimuli. However, it is reasonable to suggest that performance could be jointly influenced by the observer's fixation pattern across multiple targets and the peripheral limitations within each fixation. Both could, in principle, indicate a perceptual or an attentional defect.

From Rosenholtz's perspective, allocentric neglect may be an interesting phenomenon for contrasting interpretations based on attention versus perceptual limitations associated with peripheral

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

vision. Notably, many neglect patients exhibit various forms of visual field defects. Some may experience extinction to double visual stimuli (with contralesional stimuli omitted when a similar stimulus is simultaneously presented in the ipsilesional visual field side). Extra-personal neglect can be double dissociated from hemianopia and extinction (Pavlovskaya, Soroker, & Bonneh, 2007). However, group studies indicate a considerable overlap among these conditions, such as that several neglect patients have field defects (Doricchi et al., 2005) or visual extinction (Vossel et al., 2011).

Assessing visual field defects and extinction in neglect patients is inherently challenging, as it can be difficult to determine whether omissions on the left side arise from a visual field defect, extinction, or exploratory difficulties (e.g., Müller-Oehring et al., 2003). Regarding visual field defects, various behavioral measures exhibit different sensitivities, with the standard confrontation procedure being the least sensitive (De Luca et al., 2024). Alternatively, evoked potentials (e.g., Keyser, Neetens, & Vissenberg, 1991) may yield conflicting results compared to behavioral measures (Watanabe et al., 2007). In large-scale studies, data on visual field defects are either absent (Marsh & Hillis, 2008) or limited to simple assessments (Bickerton et al., 2011; Demeyere & Gillebert, 2019). Similar considerations apply to extinction measurements. Simple confrontation tests can be challenging with neglect patients, and more sensitive methods have been confined to a limited number of focused studies. When methods sensitive to detecting extinction in neglect patients are used, several patients with extra-personal neglect also exhibit visual extinction (Vossel et al., 2011). Overall, neglect can be reliably dissociated from peripheral disturbances, such as visual field defects and visual extinction. However, in unselected groups of right-brain-damaged patients, several neglect patients show various forms of visual field defect or visual extinction, these patients being critical for Rosenholtz's conjecture.

Can deficits in visual processing contribute to what is described as object-based or allocentric neglect? In scanning tasks, the stimuli are relatively small, and each likely requires an eye fixation for proper processing (i.e., to avoid crossing out stimuli with a missing part on the left). Information from the left of the fixation point is crucial for carrying out such tasks. If visual input is degraded, as in the case of extinction, or absent, as in hemianopia, patients may fail to detect the missing part on the left and complete the stimulus gestaltically. Bickerton et al. (2011) considered this possibility, noting that several (although not all) patients with neglect and hemianopia exhibited an allocentric profile when assessed using the Apples test.

Overall, the perspective proposed by Rosenholtz presents a fresh and intriguing approach to the study of neglect. Traditionally, research has centered on identifying double dissociations, which have advanced our understanding of how various brain lesions can create processing asymmetries in perceiving the external world. However, recognizing the presence of dissociations does not negate that, in most patients, observable behavior is influenced by a combination of co-existing symptoms. In this vein, some studies have revealed diverse response profiles in patients with neglect, depending on the co-presence and type of field defects (Doricchi et al., 2003, 2005). While extra-personal neglect can be double dissociated from deficits such as hemianopia or extinction, the interplay between early or mid-level vision and exploratory deficits may have been overlooked in understanding allocentric forms of neglect. The theoretical framework proposed by Rosenholtz may open an interesting avenue for research in this area.

Financial support. This work was supported by PRIN (D.D. N. 0001016 del 07/07/2023 number 2022S9ZMAY, CUP B53D23014830006, PI Marialuisa Martelli).

Competing interests. None.

References

Behrmann, M., & Moscovitch, M. (1994). Object-centered neglect in patients with unilateral neglect: Effects of left-right coordinates of objects. *Journal of Cognitive Neuroscience*, *6*(1), 1–16. https://doi.org/10.1162/jocn.1994.6.1.1

Bickerton, W. L., Samson, D., Williamson, J., & Humphreys, G. W. (2011). Separating forms of neglect using the Apples Test: Validation and functional prediction in chronic and acute stroke. *Neuropsychology*, *25*(5), 567–580. https://doi.org/10.1037/a0023501

Daini, R., Albonico, A., Malaspina, M., Martelli, M., Primativo, S., & Arduino, L. S. (2013). Dissociation in optokinetic stimulation sensitivity between omission and substitution reading errors in neglect dyslexia. *Frontiers in Human Neuroscience*, *7*, 581. https://doi.org/10.3389/fnhum.2013.00581

De Luca, M., Zeri, F., Matano, A., Di Lorenzo, C., Ciurli, M. P., Mulas, M., … & Nardo, D. (2024). Sensitivity and specificity of qualitative visual field tests for screening visual hemifield deficits in right-brain-damaged stroke patients. *Brain Sciences*, *14*(3), 235. https://doi.org/10.3390/brainsci14030235

Demeyere, N., Riddoch, M. J., Slavkova, E. D., Bickerton, W. L., & Humphreys, G. W. (2015). The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive screening tool. *Psychological Assessment*, *27*(3), 883. https://doi.org/10.1037/pas0000082

Demeyere, N., & Gillebert, C. R. (2019). Ego- and allocentric visuospatial neglect: Dissociations, prevalence, and laterality in acute stroke. *Neuropsychology*, *33*(4), 490–498. https://doi.org/10.1037/neu0000527

Doricchi, F., Guariglia, P., Figliozzi, F., Magnotti, L., & Gabriele, G. (2003). Retinotopic modulation of space misrepresentation in unilateral neglect: evidence from quadrantanopia. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(1), 116–119. https://doi.org/10.1136/jnnp.74.1.116

Doricchi, F., Guariglia, P., Figliozzi, F., Silvetti, M., Bruno, G., & Gasparini, M. (2005). Causes of cross-over in unilateral neglect: Between-group comparisons, within-patient dissociations and eye movements. *Brain*, *128*(6), 1386–1406. https://doi.org/10.1093/brain/awh461

Driver, J., & Halligan, P. W. (1991). Can visual neglect operate in object-centred co-ordinates? An affirmative single-case study. *Cognitive Neuropsychology*, *8*(6), 475–496. https://doi.org/10.1080/02643299108253384

Filimon, F. (2015). Are all spatial reference frames egocentric? Reinterpreting evidence for allocentric, object-centered, or world-centered reference frames. *Frontiers in Human Neuroscience*, *9*, 648. https://doi.org/10.3389/fnhum.2015.00648

Gainotti, G., Messerli, P., & Tissot, R. (1972). Qualitative analysis of unilateral spatial neglect in relation to laterality of cerebral lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, *35*(4), 545–550. https://doi.org/10.1136/jnnp.35.4.545

Gainotti, G., & Ciaraffa, F. (2013). Is 'object-centred neglect'a homogeneous entity? *Brain and Cognition*, *81*(1), 18–23. https://doi.org/10.1016/j.bandc.2012.09.007

Hillis, A. E., & Caramazza, A. (1991). Deficit to stimulus-centered, letter shape representations in a case of "unilateral neglect". *Neuropsychologia*, *29*(12), 1223–1240. https://doi.org/10.1016/0028-3932(91)90036-8

Keyser, M. D., Neetens, A., & Vissenberg, I. (1991). Objective assessment and detection of hemianopia and hemineglect with visual evoked potentials. *Neuro-ophthalmology*, *11*(1), 49–58. https://doi.org/10.3109/01658109109009642

Lee, S. H., Lim, B. C., Jeong, C. Y., Kim, J. H., & Jang, W. H. (2022). Assessment tools for differential diagnosis of neglect: Focusing on egocentric neglect and allocentric neglect. *World Journal of Clinical Cases*, *10*(24), 8625. https://doi.org/10.12998/wjcc.v10.i24.8625

Marsh, E. B., & Hillis, A. E. (2008). Dissociation between egocentric and allocentric visuospatial and tactile neglect in acute stroke. *Cortex*, *44*(9), 1215–1220. https://doi.org/10.1016/j.cortex.2006.02.002

Martelli, M., Arduino, L. S., & Daini, R. (2011). Two different mechanisms for omission and substitution errors in neglect dyslexia. *Neurocase*, *17*(2), 122–132. https://doi.org/10.1080/13554794.2010.498382

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Miceli, G., & Capasso, R. (2001). Word-centred neglect dyslexia: evidence from a new case. *Neurocase*, *7*(3), 221–237. https://doi.org/10.1093/neucas/7.3.221
Müller-Oehring, E. M., Kasten, E., Poggel, D. A., Schulte, T., Strasburger, H., & Sabel, B. A. (2003). Neglect and hemianopia superimposed. *Journal of Clinical and Experimental Neuropsychology*, *25*(8), 1154–1168. https://doi.org/10.1076/jcen.25.8.1154.16727
Olson, C. R. (2003). Brain representation of object-centered space in monkeys and humans. *Annual Review of Neuroscience*, *26*, 331–354. https://doi.org/10.1146/annurev.neuro.26.041002.131405
Pavlovskaya, M., Soroker, N., & Bonneh, Y. (2007). Extinction is not a natural consequence of unilateral spatial neglect: evidence from contrast detection experiments. *Neuroscience Letters*, *420*(3), 240–244. https://doi.org/10.1016/j.neulet.2007.05.011
Strappini, F., Pelli, D. G., Di Pace, E., & Martelli, M. (2017). Agnosic vision is like peripheral vision, which is limited by crowding. *Cortex*, *89*, 135–155, https://doi.org/10.1016/j.cortex.2017.01.012
Vossel, S., Eschenbeck, P., Weiss, P. H., Weidner, R., Saliger, J., Karbe, H., & Fink, G. R. (2011). Visual extinction in relation to visuospatial neglect after right-hemispheric stroke: quantitative assessment and statistical lesion-symptom mapping. *Journal of Neurology, Neurosurgery & Psychiatry*, *82*(8), 862–868. https://doi.org/10.1136/jnnp.2010.224261
Watanabe, K., Shinoda, K., Kimura, I., Mashima, Y., Oguchi, Y., & Ohde, H. (2007). Discordance between subjective perimetric visual fields and objective multifocal visual evoked potential-determined visual fields in patients with hemianopsia. *American Journal of Ophthalmology*, *143*(2), 295–304. https://doi.org/10.1016/j.ajo.2006.10.042

Beyond the Blink: How Task Complexity, Temporal Crowding, and Ensemble Perception Reframe the Debate on Attention and Action

Marco Mattei*

Department of Philosophy "Piero Martinetti", University of Milan, Milan, Italy
Marco.mattei@unimi.it
https://sites.google.com/view/marcomattei/

*Corresponding author.

doi:10.1017/S0140525X25000366, e150

Abstract

Rosenholtz's framework reconciles contradictory findings in ensemble perception by attributing perceptual failures to task complexity and peripheral summary-statistic limitations rather than attentional lapses. This perspective also reframes the attentional blink (AB) as a manifestation of temporal crowding rather than a failure of selective attention. Philosophically, this challenges the idea that attention is constitutive of action, suggesting instead that task constraints shape both perception and agency.

Rosenholtz proposes that what has traditionally been attributed to "attention" may largely be explained by task demands and summary statistics in peripheral vision. I argue that this perspective not only helps reconcile longstanding inconsistencies in ensemble perception research but also offers a novel way to interpret phenomena like the attentional blink (AB). In turn, this reframing has also significant implications for the philosophy of action, particularly debates about whether attention is necessary for or constitutive of action.

Ensemble perception – the ability to extract statistical properties from groups of objects – has generated conflicting findings regarding its dependence on attention. On one side, experiments such as Li et al. (2002) and Joo et al. (2009) conclude that ensemble perception operates automatically, while on the other side, Cohen, et al. (2011) and Jackson-Nielsen et al. (2017) find results more consistent with attentional requirements. Reinterpreting these apparent contradictions through Rosenholtz's framework makes them more comprehensible.

Consider the paradigms in these conflicting studies. Li et al. asked participants to discriminate letters at fixation while briefly categorizing whether a peripheral scene contained an animal. High scene-categorization performance led to the conclusion that ensemble perception does not require attention. However, this setup strongly suggests that participants relied on peripheral summary statistics rather than detailed object-level recognition. One could in fact well predict that while they could categorize the scene, they might fail if asked to identify specific objects in it. By contrast, Cohen et al. dramatically increased task demands and complexity by employing multiple-object tracking in front of a scene, adding occlusion and elevated demands for fine-grained tracking. As predicted by Rosenholtz's approach, categorization performance dropped – likely because visual perception was disrupted by greater (and central) task constraints and occlusion. Thus, what traditionally appears to be "attentional" failure might instead reflect either how increased complexity limits visual perception of how the experimental setup itself limits the extraction of peripheral summary statistics.

These insights become even more striking when applied to the AB paradigm (Raymond, Shapiro & Arnell, 1992). People often fail to report a second target (T2) if it appears roughly 200–500 ms after a first target (T1). Surprisingly, some studies find that observers can still extract summary statistics about T2 even if they cannot explicitly report it (Dell'Acqua et al., 2006; Joo et al., 2009; McNair et al., 2017), whereas others show no such capacity (Huang, 2015; Jackson-Nielsen, Cohen & Pitts, 2017). On closer inspection, this apparent contradiction maps almost perfectly onto where T2 appears: those demonstrating preserved summary-statistic extraction generally position T2 in parafoveal or peripheral regions, whereas those reporting no extraction present it centrally. Indeed, many AB studies also use lateralized or peripheral stimuli – either to measure N2pc and SPCN in EEG (Dell'Acqua et al., 2006; Zivony & Lamy, 2022) or simply due to the type and size of T2 (Joo et al., 2009; Mcnair et al., 2017) – yet still interpret the resulting "blink" purely as an attentional failure, overlooking known limits of peripheral vision. A valuable next step is to incorporate eye-tracking, ensuring that fixation is accurately monitored and clarifying whether AB effects are better attributed to failures of attention or to the inherent difficulty of extracting precise information in the visual periphery.

Rosenholtz's framework also provides a compelling way to reinterpret the AB: rather than a strict limitation of selective attention, it may reflect "temporal crowding" (Tkacz-Domb & Yeshurun, 2021), the temporal analog of spatial crowding. Since ensemble perception is an iterative process that updates on the order of ~100ms and reaches its peak after ~500ms (Epstein et al., 2020), presenting T2 within 200–500ms of T1 can easily

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

overwhelm this process. This perspective also helps to explain "lag-1 sparing" (Hommel & Akyürek, 2005), where T2 is recognized if it immediately follows T1 and even interferes with T1 – akin to backward masking. In this view, the AB need not hinge on top-down selection mechanisms but instead reflects the limited temporal resolution of ensemble coding. By reframing the blink as a manifestation of these timing constraints, we gain a clearer understanding of why summary-statistic extraction sometimes persists – especially in peripheral viewing – while at other times T2 is simply lost.

Rosenholtz's reconceptualization also bears on the debate over whether attention is constitutive of action (Allport, 1987; Wu, 2011, 2014, 2023b) or if action can proceed without it (Dayer & Jennings, 2021; Jennings & Nanay, 2016). Traditionally, robust ensemble perception in the face of the AB has been taken as evidence that perception – and thus action – can operate independently of attention (Buehler, 2018, 2023). Conversely, cases of diminished ensemble perception have been interpreted as evidence that attention is necessary (Wu, 2023a). Rosenholtz's model suggests both conclusions rest on a flawed premise: the "selective bottleneck" may emerge not because attention is a resource that runs out but because task demands limit visual perception and how well summary statistics can be extracted. In this view, perception is structured by the requirements of the current task, reversing the usual assumption that attention is a discrete selection mechanism in service of action. If attention is reconceived in this way, philosophers of action must grapple with a new explanatory foundation. Task complexity, rather than an internal attentional "gate," determines which statistical regularities are available for guiding behavior. A failure to extract summary statistics in the AB no longer implies that attention is mandatory; a success in extracting them no longer proves attention is optional. The real driver is the interplay between task-set requirements and the perceptual system's capacity for summary-statistic extraction.

In sum, Rosenholtz's emphasis on ensemble perception bridges many gaps in visual cognition research. Rather than attributing the attentional blink and other perceptual limits to a lack of attention, we can understand them as consequences of both spatial and temporal crowding. This shift not only clarifies contradictory findings in ensemble perception studies but also reframes the longstanding "crisis" in visual attention: instead of signaling a failure in resource allocation, these phenomena may reflect deeper perceptual constraints shaped by task complexity. By embracing this perspective – and by employing methods such as eye-tracking to tease apart peripheral from genuinely attentional factors – researchers can develop a more coherent account of how, when, and why perception meets its limits.

Acknowledgments. I am grateful to Denis Buehler for the insightful discussions on the philosophy and psychology of attention. I also extend my sincere thanks to John Michael for his supervision and guidance throughout this work.

Financial support. This research was funded by the Department of Philosophy "Piero Martinetti" of the University of Milan under the Project "Departments of Excellence 2023–2027" awarded by the Ministry of University and Research (MUR).

Competing interests. None.

References

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & H. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Lawerence Erlbaum.

Buehler, D. (2018). A dilemma for 'selection-for-action.' *Thought: A Journal of Philosophy*, *7*(2), 139–149. https://doi.org/10.1002/tht3.378

Buehler, D. (2023). Seeing circles: Inattentive response-coupling. *Ergo an Open Access Journal of Philosophy*, *9*(62), 1671–1691. https://doi.org/10.3998/ergo.3587

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, *22*(9), 1165–1172. https://doi.org/10.1177/0956797611419168

Dayer, A., & Jennings, C. D. (2021). Attention in skilled behavior: An argument for pluralism. *Review of Philosophy and Psychology*, *12*(3), 615–638. https://doi.org/10.1007/s13164-021-00529-6

Dell'Acqua, R., Sessa, P., Jolicoeur, P., & Robitaille, N. (2006). Spatial attention freezes during the attention blink. *Psychophysiology*, *43*(4), 394–400. https://doi.org/10.1111/j.1469-8986.2006.00411.x

Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The outlier paradox: The role of iterative ensemble coding in discounting outliers. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(11), 1267–1279. https://doi.org/10.1037/xhp0000857

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *The Quarterly Journal of Experimental Psychology Section A*, *58*(8), 1415–1433. https://doi.org/10.1080/02724980443000647

Huang, L. (2015). Statistical properties demand as much attention as object features. *PLOS ONE*, *10*(8), e0131191. https://doi.org/10.1371/journal.pone.0131191

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, *48*, 149–160. https://doi.org/10.1016/j.concog.2016.11.007

Jennings, C. D., & Nanay, B. (2016). Action without attention. *Analysis*, *76*(1), 29–36.

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, *9*(9), 7. https://doi.org/10.1167/9.9.7

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

McNair, N. A., Goodbourn, P. T., Shone, L. T., & Harris, I. M. (2017). Summary statistics in the attentional blink. *Attention, Perception, & Psychophysics*, *79*(1), 100–116. https://doi.org/10.3758/s13414-016-1216-2

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology. Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037//0096-1523.18.3.849

Tkacz-Domb, S., & Yeshurun, Y. (2021). Temporal crowding is a unique phenomenon reflecting impaired target encoding over large temporal intervals. *Psychonomic Bulletin & Review*, *28*(6), 1885–1893. https://doi.org/10.3758/s13423-021-01943-8

Wu, W. (2011). Attention as selection for action. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays* (pp. 97–116). Oxford University Press. https://philarchive.org/rec/WUAAS

Wu, W. (2014). *Attention. Routledge.* https://doi.org/10.4324/9781315777603

Wu, W. (2023a). *Movements of the mind: A theory of attention, intention and action*. Oxford University Press.

Wu, W. (2023b). We know what attention is!. *Trends in Cognitive Sciences, 28*(4), 304–318. https://doi.org/10.1016/j.tics.2023.11.007

Zivony, A., & Lamy, D. (2022). What processes are disrupted during the attentional blink? An integrative review of event-related potential research. *Psychonomic Bulletin & Review*, *29*(2), 394–414. https://doi.org/10.3758/s13423-021-01973-2

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Attention, the homunculus, and the Greek theater effect

Ezequiel Morsella[1,2*], Sarah Brauer[1], Latoya Wright-Wilson[1] and Tala Elsabbagh[1]

[1]Department of Psychology, San Francisco State University (SFSU), San Francisco, California, United States and [2]Department of Neurology, University of California, San Francisco, California, United States
morsella@sfsu.edu
sbrauer@sfsu.edu
lwrightwilson@sfsu.edu
telsabbagh@sfsu.edu

*Corresponding author.

doi:10.1017/S0140525X25000299, e151

Abstract

The groundbreaking, anti-homuncular view presented by the author complements conclusions from diverse areas of research, including (a) research on phenomena not concerning attention (e.g., involving anosognosia, scotomata, dreams, and consciousness) and (b) the hypothesis that attention should be construed as an effect (e.g., from the activation of *priority signals* and *task sets*) rather than as a cause.

The author proposes that phenomena attributed to attention could be explained more parsimoniously by appealing to factors such as *priority signals* and *task sets* ("biasing factors," for short). Thus, no homuncular, central executive is needed to direct attention to mental representations. Instead, some representations are processed more deeply than others because of the biasing factors, which is consistent with the view of *attention as an effect* rather than as a cause (Krauzlis et al., 2014). Thus, the inferred "observing agent" of the conscious field (the "perspectival self"; Seth, 2021), which is a result of the geometry of the conscious field and cannot itself be observed in the field (Merker et al., 2021), might not be the actual agent(s) dictating which representations are processed deeply. To the "observing agent," attention seems limited only because "the pattern of weighted inputs gives the appearance of limited sensory resources, because increasing the weight of one input necessarily involves reducing the proportional weight of all of the other inputs" (Krauzlis et al., 2014, p. 458).

But is there a single observing agent? It has been proposed that the field is not apprehended (or "sampled") by a single agent, but by a collection of (unconscious) "response codes," each of which is activated differentially by the contents of the field (Morsella et al., 2016). (There is no homunculus fallacy here because each code, with its own concerns and action tendencies, does not replicate the function of the whole system.) It seems that there is a single observing agent, perhaps because of the way in which the field is formatted for *each* response code. We refer to this as the "Greek theater" effect. In Ancient Greece, theaters contained a central, circular stage. With this geometry, any spectator could believe that he or she had "the best seat in the house." But the illusion occurred equally for every spectator. Similarly, the geometry of the conscious field, which is designed that way for the purpose of adaptive action selection (Morsella et al., 2016), would seem to each response code as if the field were directed only toward it. It is adaptive for things to appear *as if* from a first-person perspective. Whether there is a single conscious field observed by the response codes or each code possesses its own field is an interesting question (for some evidence for the latter, see Gazzaniga & LeDoux, 1978; Rapp et al., 2015).

The conscious field and the Greek theater effect would still arise if, for example, the verbal, "narrative self" were absent, as might occur during anesthetization of the left hemisphere in the Wada test. Moreover, "attentional" phenomena attributed to the self (e.g., sustained selective attention) would persist in the absence of, say, the "narrative self" (Seth, 2021) or "left-hemisphere interpreter" (Gazzaniga & LeDoux, 1978). A "perspectival self" and some representations being processed more deeply than others would persist.

The field is a "closed system" in the deepest sense. One infers that the allocation of attention is occurring because of processes located spatially "inside the head." However, the head itself is a representation existing within the field (Köhler, 1947) along with, for example, percepts, urges, and "intra-psychic" conflicts (e.g., *approach-avoidance* conflicts). There is no direct conscious access to the mechanisms that generate conscious contents (e.g., percepts, urges, and memories), action conflicts, or that process some representations more than others. Conscious contents, which are usually "encapsulated" and cannot be turned off or modulated at will, wholly and exclusively determine "voluntary" action selection (Morsella et al., 2016). In voluntary action selection, there is no influence other than the contents of the field at that time. The combination of contents determines action selection. Thus, the field is unlike the "smart helmets" of fighter pilots. If one could monitor through telemetry what was projected onto a pilot's shades, one would be unable to predict consistently what the pilot decides. For example, if the pilot felt dizzy and hence returned to base, the cause of the pilot's decision would not be represented in the helmet. In contrast, voluntary action is not based on what occupies the conscious field *plus* information that exists in some other system, such as a system that could evaluate the degree to which, for example, the contents of the field are nonsensical. In the field, the absence of information does not imply information about absence, as is obvious with respect to visual scotomas, anosognosia, sensory neglect, and in the nonsensical dreams that we regularly accept as real and sensible.

The field reacts to the introduction of a stimulus, but it is the composition of the field as a whole, including memories, that determines action selection. At one moment, the field provides the "global best estimate" (Merker, 2012) for this composition, and there is never any indecision for the *action selection mechanism*. Perhaps, conflicts reflect only *thoroughness*, a freebie from encapsulation. It might be that the observer infers that there is intra-psychic conflict because the contents activated by a stimulus are incongruent with each other. One might report, "I desired the cupcake because I enjoy cupcakes and skipped breakfast, but I was conflicted about eating it because of doctor's orders" when it might be more accurate to state, "The cupcake activated in the my field thorough semantic knowledge that the stimulus was *highly desired by me then (because I was hungry and like cupcakes)* and *my doctor forbids eating it*." These encapsulated, knowledge representations determine action selection just as other kinds of knowledge representations do. But the representations are not a window into the mechanisms giving rise to inclinations and behavior, much as the fuel gauge on one's dashboard is not the actual tank of gas, and the perspectival self is not in charge of the allocation of attention.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Within this closed system, some contents are processed more deeply than others. As noted by the author, this is not due to a single homunculus directing attention toward one content versus another. Instead, this is due to biasing factors that operate outside of the conscious field.

Acknowledgments. We gratefully acknowledge the advice of Prof. Lisa Whitfield on the topic of dynamic field theories.

Financial support. This work was not funded by any agency.

Competing interests. The authors report no competing interests or conflicts of interest.

References

Gazzaniga, M. S., & LeDoux, J. E. (1978). *The integrated mind*. New York: Plenum Press.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology* (Rev. ed.). New York: Liveright.

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, *18*, 457–464. https://doi.org/10.1016/j.tics.2014.05.008.

Merker, B. (2012). From probabilities to percepts: A subcortical "global best estimate buffer" as locus of phenomenal experience. In E. Shimon, F. Tomer, & Neta, Z. (Eds.), *Being in time: Dynamical models of phenomenal experience* (pp. 37–80). Netherlands: John Benjamins.

Merker, B., Williford, K., & Rudrauf, D. (2021). The integrated information theory of consciousness: A case of mistaken identity. *Behavioral and Brain Science*, *19*(45), e41. https://doi.org/10.1017/S0140525X21000881.

Morsella, E., Godwin, C. A., Jantz, T. K., Krieger, S. C., & Gazzaley, A. (2016). Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences*, *39*, 1–17.

Rapp, B., Fisher-Baum, S., & Miozzo, M. (2015). Modality and morphology: What we write may not be what we say. *Psychological Science*, *26*, 892–902.

Seth, A. (2021). *Being you: A new science of consciousness*. New York, NY, USA: Dutton.

Seeing attention in inattentional blindness

Steven B. Most*

School of Psychology, UNSW Sydney, Sydney, Australia
s.most@unsw.edu.au
https://www.unsw.edu.au/staff/steven-most

*Corresponding author.

doi:10.1017/S0140525X25000135, e152

Abstract

Rosenholtz is right that the multiple meanings of attention hinder development of a unified attention theory, but this is not a crisis. Embracing its diversity can advance fields like clinical psychology. However, inattentional blindness challenges attempts to move beyond attention as an explanatory concept. I illustrate this by highlighting attentional set, which demonstrates selective prioritization rather than mere task constraints.

Despite William James's famous assertion that "everyone knows what attention is" (James 1890), no one can refute the diversity of phenomena embedded within the word "attention" – it includes spatial attention, temporal attention, feature-based attention, object-based attention, transient attention, sustained attention, and more. This proliferation goes even beyond that laid out in the target article because the very functions and targets of attention are myriad: attention selects, modulates, and maintains (as in working memory). You can attend outwardly to the patterns in a leaf or to your internal thoughts (e.g., Chun, Golomb, & Turk-Browne, 2011). As Rosenholtz suggests, the over-application of the notion of attention renders it nearly impossible to develop a comprehensive theory of what attention is.

It is particularly important that this message reaches beyond attention researchers, who are already well-acquainted with this state of affairs. In clinical psychology, for example, a thriving literature aims to understand not only how disorders are characterized by biases to attend to disorder-relevant stimuli (e.g., threat-related cues in anxiety or food-related cues in binge eating; MacLeod, Grafton, & Notebaert, 2019; Schmitz et al., 2014), but also how retraining of such attentional biases might serve as a form of intervention (Bar-Haim, 2010). Within this literature, discrepant results might be resolved through greater recognition that different attentional bias tasks (e.g., dot probe, visual search, emotion-induced blindness) do not tap into the same mechanism. As noted by one research team working in this area, "[a] notable psychometric weakness of the attentional bias measures is their low degree of convergent validity. Although they are all assumed to measure the same theoretical construct, different measures of attentional bias often do not correlate with one another" (Van Bockstaele et al., 2014). This assumption – that they measure the same theoretical construct – underscores the importance of Rosenholtz's point. As a corrective, work in our own lab has aimed to shine a spotlight on distinctions between mechanisms of emotion-driven attentional bias (e.g., Onie & Most, 2017, 2022), but Rosenholtz's exercise of banishing the word "attention" in order to bring, well, *attention* to actual mechanistic distinctions may be a salve that salvages this often inconsistent literature.

That said, inattentional blindness – which features heavily in the target article – is a phenomenon that is difficult to describe without reference to attention. One of Rosenholtz's alternative suggestions is that inattentional blindness reflects task limits.

However, omitted from the target article's review of inattentional blindness is the well-established role of "attentional set". This refers to the fact that people can tune their attention to prioritize some visual features over others and that this tuning accounts for tremendous variation in inattentional blindness. For example, when people tracked white shapes and ignored black ones, almost everyone noticed an unexpected novel white shape traversing the display and almost everyone failed to notice an unexpected novel black shape, with the rates of noticing a light or dark gray shape falling in between. This pattern was reversed

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

when people tracked black shapes and ignored white ones (Most et al., 2001; also see Simons & Chabris, 1999). Rates of noticing followed a non-binary linear pattern: the more similar the unexpected object was to the tracked objects, and the less similar it was to the non-tracked objects, the more it was noticed. Subsequent work showed the top-down flexibility of such prioritization, with similar effects of attentional set when people prioritized information on the basis of shape (Most et al., 2005), category membership (Koivisto & Revonsuo, 2007; Most, 2013), and number (White & Davies, 2008).

The profound degree to which people's prioritization of information determines inattentional blindness suggests factors that extend well beyond task limits. Driving this point home, in one related study, participants "drove" through a virtual cityscape and encountered a sign with blue and gold arrows at each intersection. Some participants were instructed to always turn in the direction of a gold arrow among blue arrows, whereas others always turned in the direction of a blue arrow among gold arrows. At a critical intersection, a blue or gold motorcycle veered into their path. When the motorcycle matched the color of the arrows people were following, fewer people crashed into it than when it mismatched (Most & Astur, 2007). In this study, the motorcycle was equally distinct from the task of following the arrows regardless of the color-match condition. How can we best account for such patterns without reference to the selective prioritization of some features over others – in other words, to the overarching concept of feature-based attention?

Finally, Rosenholtz notes that any viable explanation for inattentional blindness needs to account for findings that observers sometimes move to avoid collisions with unexpected stimuli they claim not to have seen. Just as we must contend with diversity in the meaning of "attention", it may be fruitful here to highlight diversity in the meaning of "perception". Famously, Goodale and Milner (1992) differentiated between visual processing pathways in the brain that subserve conscious perception and those that subserve action. A particularly striking illustration of this distinction comes from a patient with blindsight, who was able to avoid several obstacles while walking down a hallway despite not consciously seeing them (de Gelder et al., 2008). The distinction between "perception" and "action" pathways for visual processing may render the fact that people can avoid collisions with things they did not see unproblematic for accounts of inattentional blindness.

As researchers have learned more about mechanisms that might fall within the scope of the big tent we call "attention", the proliferation of such mechanisms pushes the dream of a unifying theory further out of reach. But this makes this an exciting time to be an attention researcher. The dream of a unifying theory of attention may be in crisis, but the science of attention – in all its diversity – is thriving.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interests. None.

References

Bar-Haim, Y. (2010). Attention bias modification (ABM): A novel treatment for anxiety disorders. *Journal of child psychology and psychiatry, and allied disciplines*, *51*(8), 859–870. https://doi.org/10.1111/j.1469-7610.2010.02251.x

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual review of psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

de Gelder, B., Tamietto, M., van Boxtel, G., Goebel, R., Sahraie, A., van den Stock, J., Stienen, B. M., Weiskrantz, L., & Pegna, A. (2008). Intact navigation skills after bilateral loss of striate cortex. *Current biology : CB*, *18*(24), R1128–R1129. https://doi.org/10.1016/j.cub.2008.11.002

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

James, W. (1890). The principles of psychology. New York: Henry Holt.

Koivisto, M., & Revonsuo, A. (2007). How meaning shapes seeing. *Psychological Science*, *18*(10), 845–849. https://doi.org/10.1111/j.1467-9280.2007.01989.x

MacLeod, C., Grafton, B., & Notebaert, L. (2019). Anxiety-linked attentional bias: Is it reliable?. *Annual review of clinical psychology*, *15*, 529–554. https://doi.org/10.1146/annurev-clinpsy-050718-095505

Most S. B. (2013). Setting sights higher: Category-level attentional set modulates sustained inattentional blindness. *Psychological research*, *77*(2), 139–146. https://doi.org/10.1007/s00426-011-0379-7

Most, S. B., & Astur, R. S. (2007). Feature-based attentional set as a cause of traffic accidents. *Visual Cognition*, *15*(2), 125–132. https://doi.org/10.1080/13506280600959316

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review*, *112*(1), 217–242. https://doi.org/10.1037/0033-295X.112.1.217

Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological science*, *12*(1), 9–17. https://doi.org/10.1111/1467-9280.00303

Onie, S., & Most, S. B. (2017). Two roads diverged: Distinct mechanisms of attentional bias differentially predict negative affect and persistent negative thought. *Emotion*, *17*(5), 884–894. https://doi.org/10.1037/emo0000280

Onie, S., & Most, S. B. (2022). On the relative sensitivity of spatial and nonspatial measures of attentional bias: Emotion-induced blindness, the dot probe, and gradations in ratings of negative pictures. *Emotion*, *22*(8), 1942–1951. https://doi.org/10.1037/emo0000855

Schmitz, F., Naumann, E., Trentowska, M., & Svaldi, J. (2014). Attentional bias for food cues in binge eating disorder. *Appetite*, *80*, 70–80. https://doi.org/10.1016/j.appet.2014.04.023

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological bulletin*, *140*(3), 682–721. https://doi.org/10.1037/a0034834

White, R. C., & Davies, A. A. (2008). Attention set for number: Expectation and perceptual load in inattentional blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1092–1107. https://doi.org/10.1037/0096-1523.34.5.1092

Visual attention as an integrated sensorimotor process

Sean O'Bryan and Joo-Hyun Song*

Department of Cognitive and Psychological Sciences, Brown University, RI, USA
sean_obryan@brown.edu
joo-hyun_song@brown.edu

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

*Corresponding author.

doi:10.1017/S0140525X25000226, e153

Abstract

While agreeing with Rosenholtz that the broad concept of visual attention conflates phenomena with explanatory mechanisms, leading to a crisis that demands a paradigm shift, we argue that this issue partially arises from viewing attention within a predominantly perception-focused framework. We propose reframing visual attention as part of an integrated sensorimotor system, emphasizing its role in linking vision and action.

Rosenholtz's reframing of visual attention as a process in which an observer *chooses and performs a task* raises important questions about the goals of the visual system – of the observer – and reveals how a more complete understanding of these goals can help us develop more coherent and ecologically valid theories of attention. Considering possible goals and the choices that might accompany them, a significant oversight of broader attention theory is the assumption that the perceptual tasks we choose are in service of *perceptual goals*. Instead, for embodied human observers, we argue that the primary role of perceptual selection is to support action. Beyond the laboratory, visual perception is often only useful to the extent that we can act on what we perceive: It is not enough to *see* the berries and *notice* the approaching tiger – we want to *pick* the berries and *avoid* the tiger. Thus, rather than directly adopting Rosenholtz's approach, we propose an orthogonal paradigm shift in how we conceptualize visual attention. We suggest framing it as part of an integrated sensorimotor system that links vision and action. This challenges the traditional "see → decide → act" framework and advocates for a broader view in which attention extends beyond visual perception alone. In real-world scenarios, successful interactions require not just locating target objects but also reaching for and manipulating them to achieve goals. This complex behavior depends on the seamless coordination of multiple brain systems responsible for processing sensory signals, selecting tasks consistent with our goals (i.e., allocating attention), and controlling visually guided actions. Therefore, we argue that visual attention is not passive but actively shaped by the demands of action (Nakayama, Moher, & Song, 2023; Song, 2019).

Our recent work (Moher, Anderson, & Song, 2015) highlights the complex interactions between these systems. Traditional models suggest that physically salient objects, such as flashing hazard signs, capture attention regardless of the observer's goals (Anderson, Laurent, & Yantis, 2011; Theeuwes, 1994; Theeuwes & Godljn, 2002). However, we demonstrated that these same cues can trigger a suppression mechanism during action selection, but not during perceptual tasks. Specifically, high-contrast stimuli were more distracting in perceptual tasks, in line with traditional models, but less disruptive when rapid motor actions were required. Notably, these findings also generalized to stimuli that were salient due to their association with reward. This suggests a salience-triggered suppression mechanism during goal-directed action, where highly salient distractors are quickly suppressed to minimize interference. Conversely, in perception-based visual search tasks, strong distractors caused more interference. This might seem counterintuitive as one would expect stimuli affecting conscious vision to similarly influence visually guided actions. We propose that the discrepancy arises because the cost of distraction is greater when maintaining an action course under time pressure. This leads the brain to adapt by suppressing stronger distractors more effectively during goal-directed actions. These findings underscore the need to integrate visually guided action paradigms with traditional psychophysical approaches to fully understand the coordination of perception and action. This integration would offer new insights into visual attention's role in achieving behavioral goals and resolving competing processes in complex visual environments (Song, 2017).

Research on perception and action also reveals a strong link between visual attention, hand movements, and tool use. Neurons in the monkey parietal cortex, particularly in the intraparietal sulcus, show dual receptive fields – somatic and visual. This provides a neural basis for enhanced visual processing near the active hand (Iriki, Tanaka, & Iwamura, 1996). This extends to humans, where reaction times are faster for stimuli near the hands, supporting the notion of heightened visual attention in peri-personal space (Reed, Grubb, & Steele, 2006). Crucially, this effect depends on action: visual advantages arise only when tools are actively used, not merely held (Reed et al., 2010). Further research by Thomas (Thomas, 2015, 2017) showed that different hand actions modulate visual attention through distinct pathways. Precision grips, requiring fine spatial resolution, enhanced parvocellular processing, while power grips, involving dynamic processing, engaged magnocellular pathways. Our own work (Guo & Song, 2019) demonstrated that a concurrent grasping task enhanced orientation discrimination. Furthermore, prior precision-grasping training also improved sensitivity to orientation discrimination. These findings suggest that action preparation or enhancement boosts perceptual discrimination, emphasizing the bidirectional relationship between motor and visual performance. Rather than merely relying on perception, action actively shapes it.

These findings, among many others, underscore the need to move beyond constrained models of "attention for perception" and to explicitly consider action's role in shaping the goals and selection mechanisms of the visual system. Critically, this approach does not necessitate more complex models of attention. Instead, acknowledging the pervasive interaction between perception and action could yield more *parsimonious* models by more accurately describing what selection is for, by capturing unanticipated constraints on task complexity, and by grounding the study of visual attention in more ecologically valid contexts.

This integration would also facilitate the generation of more comprehensive psychophysical datasets. Rather than reducing complex perceptual experiences to discrete keypresses, incorporating continuous action-based tasks in vision research could yield deeper insights into distractor interference, response competition, and decision confidence, revealing dynamics that traditional methods may overlook (Hayhoe, 2017; Song & Nakayama, 2009; Wöstmann et al., 2022). These insights could help reconcile previously conflicting findings and provide a clearer understanding of anomalies by framing them within a perspective that more accurately reflects the (embodied) goals of both the visual system and the observer.

In conclusion, the evolving understanding of visual attention underscores its dynamic relationship with sensorimotor integration. While Rosenholtz's critique of current attention models prompts valuable reconsideration, a broader framework that incorporates action processes provides a more comprehensive perspective. Rather than functioning solely as a perceptual filter, visual attention is dynamically shaped by the demands of action. We hope this perspective inspires refinements to attention theory and encourages a reassessment of prior findings in visual attention research.

Acknowledgments. We thank Jacob Lader for his helpful comments.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Funding statement. This work is supported by NSF BCS 2341363.

Competing interests. None.

References

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proc Natl Acad Sci U S A*, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Guo, J., & Song, J. H. (2019). Action fluency facilitates perceptual discrimination. *Psychological Science*, *30*(10), 1434–1448. https://doi.org/10.1177/0956797619859361

Hayhoe, M. M. (2017). Vision and action. *Annu Rev Vis Sci*, *3*(1), 389–413.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, *7*(14), 2325–2330.

Moher, J., Anderson, B. A., & Song, J.-H. (2015). Dissociable effects of salience on attention and goal-directed action. *Current Biology*, *25*(15), 2040–2046. https://doi.org/10.1016/j.cub.2015.06.029

Nakayama, K., Moher, J., & Song, J.-H. (2023). Rethinking vision and action. *Annu Rev Psychol*, *74*, 59–86.

Reed, C. L., Betz, R., Garza, J. P., & Roberts, R. J. (2010). Grab it! biased attention in functional hand and tool space. *Attention, Perception, & Psychophysics*, *72*, 236–245.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: attentional prioritization of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 166.

Song, J.-H. (2019). The role of attention in motor control and learning. *Curr Opin Psychol*, *29*, 261–265.

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366. https://doi.org/10.1016/j.tics.2009.04.009

Song, J. H. (2017). Abandoning and modifying one action plan for alternatives. *Phil Trans R Soc B*, *372*, 20160195. https://doi.org/10.1098/rstb.2016.0195

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J Exp Psychol Hum Percept Perform*, *20*(4), 799–806.

Theeuwes, J., & Godljn, R. (2002). Irrelevant singletons capture attention: evidence from inhibition of return. *Percept Psychophys*, *64*(5), 764–770.

Thomas, L. E. (2015). Grasp posture alters visual processing biases near the hands. *Psychological Science*, *26*(5), 625–632. https://doi.org/10.1177/0956797615571418

Thomas, L. E. (2017). Action experience drives visual-processing biases near the hands. *Psychological Science*, *28*(1), 124–131. https://doi.org/10.1177/0956797616678189

Wöstmann, M., Störmer, V. S., Obleser, J., Addleman, D. A., Andersen, S. K., Gaspelin, N., Geng, J. J., Luck, S. J., Noonan, M. P., & Slagter, H. A. (2022). Ten simple rules to study distractor suppression. *Progress in Neurobiology*, *213*, 102269.

Building attention on a firm foundation

Michael I. Posner*

Department of Psychology, University of Oregon, Eugene, OR, USA
mposner@uoregon.edu

*Corresponding author.

doi:10.1017/S0140525X25000330, e154

Abstract

Paradigm shifts as advocated by Kuhn (1962), should be rarely occurring and based upon true crises. The study of attention, however, is not in crisis requiring a paradigm shift but instead has a firm empirical foundation that can accommodate the findings in visual search cited in the target article.

What is attention?

Attention involves selection from all sensory modalities or from memory and is not purely visual. However, visual attention should be consistent with imaging and behavioral studies that have defined attention as involving three networks related to its functions irrespective of the modality of input (Posner & Petersen, 1990; Petersen & Posner 2012). An alerting network arises in the locus coeruleus, which is the source of the brain's norepinephrine, and involves obtaining and maintaining a high level of vigilance to react to incoming stimuli. The orienting network connects the frontal eye fields to the superior parietal lobe and the temporal parietal junction. The orienting network connects to the ventral visual stream (object recognition) by common oscillations in the process of selecting a visual object (Womelsdorf & Fries, 2007) and to other sensory areas which are the sites of selection by attention to auditory, touch, or other non-visual modalities. The orienting network called frontal-parietal network in imaging studies is the source of attentional effects that can influence various sensory modalities (Corbetta & Shulman, 2002). An executive network is called the cingulo-opercular network in imaging studies. The executive network is involved in more strategic functions that include the resolution of conflict between brain networks in tasks like the Stroop effect (Petersen & Posner, 2012). Much is unknown about how these networks operate separately or together.

Visual Orienting

The target paper discusses only orienting phenomena such as crowding and binding within the visual modality and correctly indicates that many studies are confounded between purely sensory and possible attention effects. Covert orienting to sensory stimuli involves portions of the orienting network that are the same irrespective of the sensory modality involved (Petersen & Posner, 2012). Thus, it is useful to distinguish between brain areas that are the source of covert shifts from those sensory sites on which orienting operates. Imaging studies show the brain areas involved include the temporal parietal junction, superior parietal lobe, and frontal eye fields which are all important in covert orienting.

Orienting of attention is most clearly involved in tasks where a cue gives priority to the location where a target may or may not occur (Posner, 1980). Evidence that foveal location is not critical to covert orienting is that a foveal cue improves reaction time (RT) by the same amount as does a peripheral cue (Posner, 1978). Studies demonstrate there is as much improvement in RT when the person moves attention from a cue 9 degrees to a target in the fovea as there is in the reverse movement. Orienting of attention to visual stimuli in primates may be overt involving head and eye movements or be done by a covert shift involving no observable motor output. In primates, an overt orienting movement usually places the fovea on the target location. However, rodents without a fovea show similar faster reaction times when a cue produces covert orienting to the target location (Beane & Marrocco, 2004).

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

This means that the fovea is not as critical for the priority of a stimulus as it is for acuity. However, even spatial resolution can be influenced covertly by cuing (Yeshurun & Carrasco, 1998).

There is controversy over whether crowding, binding, and other phenomenon are purely sensory or whether they require modulation from orienting areas of the frontal and parietal lobes. This is a purely empirical question, which is not advanced by refusing to recognize that top-down attention influences that have been shown to be critical to the priority given to a sensory stimulus (Posner & Petersen, 1990)

An important model task, cited in the target article (Treisman & Gelade, 1980), argues that the binding together of, e.g., color and form requires attention does not seem to be correct. In one study it was shown that color and form remain statistically independent during visual processing (Isenberg, Nissen, & Marchak, 1990) Moreover, cells that receive both color and form information exist even in primary visual cortex, but they do not lead to an integrated representation (Seymour et al., 2010). The integration may take place in the executive attention network or in the pathways that connect executive attention to memory (Posner et al., 2022).

To ignore what is known about attention, as suggested by the target article, would cause us to lose sight of the brain areas involved in alerting, orienting, and executive attention. Understanding attention also helps to design empirical studies that distinguish between purely sensory effects and those that involve further modulation. Moreover, we could lose our current understanding of how attention influences different sensory and memory areas and provides a foundation for understanding how top-down information influences sensation that does not occur entirely within the sensory systems.

Financial support. No funding.

Competing interests. No competing interest.

References

Beane, M. & Marrocco, R. T.. 2004. Cholinergic and noradrenergic inputs to the posterior parietal cortex modulate the components of exogenous attention. Michael I. Posner. (Eds.), In *Cognitive neuroscience of attention*. New York: Guilford, (pp. 313–25)

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Neuroscience Reviews*, *3*, 201–215. https://doi.org/10.1038/nrn755

Isenberg, L., Nissen, M.J. & Marchak, L.C. (1990). Attentional processing and the independence of color and orientation. *JEP:Human Perception and Performance*, *16*(4), 869–87

Kuhn, T.S (1962) *The structure of scientific revolutions*. Chicago:University of Chicago Press

Petersen, S.E. & Posner, M.I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*(7), 1–89

Posner, M.I. (1978). *Chronometric explorations of mind*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Posner, M.I. (1980). Orienting of attention. The 7th Sir F.C. Bartlett Lecture. *Quarterly Journal of Experimental Psychology*. *32*, 3–25.

Posner, M.I. & Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M.I., Rothbart, M.K. & Ghassemzadeh, H (2019). Restoring attention networks. *Yale Journal of Biology and Medicine*, *92*(1), 139–143.

Posner, M.I., Weible, A.P., Voelker, P., Rothbart, M.K. & Niell, C.M. (2022). Decision making as a learned skill in mice and humans. *Frontiers in Neuroscience*, *16*, 834701. https://doi.org/10.3389/fnins.2022.834701

Seymour, K. Clifford, CWG, Logothetis, NK & Bartels, A. (2010). Coding and binding of color and form in visual cortex. *Cerebral Cortex*, *20*, 1946–1954

Taylor, J. & Xu, YD (2022). Representation of color, form, and their conjunction across the human ventral visual pathway. *Neuroimage*, *251*, 118941. https://doi.org/10.1016/j.neuroimage.2022.118941

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Womelsdorf T, & Fries P. (2007)bThe role of neuronal synchronization in selective attention. *Current Opinion Neurobiology*, *17*(2),154–60. https://doi.org/10.1016/j.conb.2007.02.002.

Yeshurun, Y. & Carrasco,M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*, 72–75.

Spurious crisis versus sustainable science

Christian H. Poth[a]* and Nina L. Poth[b]

[a]Neuro-Cognitive Psychology, Department of Psychology, Bielefeld University, Bielefeld, Germany and [b]Department of Philosophy of Mind and Language, Faculty of Philosophy, Theology and Religious Studies, Radboud University, Nijmegen, The Netherlands
c.poth@uni-bielefeld.de
ninalaura.poth@ru.nl
Christian H. Poth and Nina L. Poth contributed equally.

*Corresponding author.

doi:10.1017/S0140525X25000184, e155

Abstract

Here, we argue that Rosenholtz' call for a paradigm shift in attention theory is unwarranted based on psychological evidence as well as philosophical theory and would disrupt scientific progress by preventing incremental science. To move forward, we suggest a different philosophical view on attention research that preserves the cumulative nature of scientific progress rather than waxing and waning theoretical paradigms.

Rosenholtz claims attention research was in a crisis, as defined in Kuhn's (1962) paradigm theory. To solve the crisis, Rosenholtz calls for a paradigm shift, replacing the concept of attention with other mechanisms for stimulus-driven and task-driven vision. However, from a philosophical and psychological viewpoint, attention research fails to meet the criteria of a scientific crisis. Calling for a paradigm shift seems unwarranted and would impede progress in attention science.

According to Kuhn (1962, Ch. 7), a scientific crisis occurs if the scientific community loses faith in the ability of the leading paradigm (the field's fundamental theories and methodologies) to resolve persistent conflicts ("anomalies") between theory and empirical evidence. Shifting to a new paradigm resolves the conflicts. The paradigm shift is revolutionary: Old and new paradigms are incommensurable, and there is no conceptual continuity between them. The new paradigm cannot be derived by introducing new assumptions in the old paradigm. Thus, progress becomes an illusion because with each paradigm shift, new theoretical foundations and methods are developed from scratch, cutting links to knowledge from the past.

Contrary to Rosenholtz' claims, attention research violates both the criteria of scientific crisis and those of revolution that would rule out progress in the field. Attention refers to a broad class of

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

mechanisms accommodating the brain's limited processing capacity for perception by setting processing priorities (e.g., Bundesen, 1990; Desimone & Duncan, 1995). Rosenholtz claims to have found "anomalies" revealing the failure of the "attention paradigm" to provide useful theories. Uselessness is inferred from an increasing complexity of the theories to accommodate new findings (pp. 3–4). For instance, that the gist of a visual scene is processed faster than objects within the scene has been explained by assuming additional mechanisms for scene processing that complement attention (Wolfe et al., 2011). Critically, an increase in complexity does not indicate a crisis (cf. Kuhn, 1962). To the contrary, the possibility to accommodate new findings by introducing new assumptions demonstrates a continuous progression in theory development without abandoning core posits entirely (cf. Lakatos, 1978). In scene perception, for example, newly assumed mechanisms (Wolfe et al., 2011) can be perfectly reconciled by amending previous attention theories (Wolfe, 1994). This goes against a crisis of attention research.

For a crisis, there need to be findings challenging the fundamental assumptions of a research field. For attention, as a class of priority-setting mechanisms (see above), this would be the case if findings challenged the assumption that there was any prioritization in perception and cognition. Critically, this challenge cannot be found by focusing on specific subareas of attention research, such as peripheral vision (p. 5–8), gist/scene perception (p. 3–4), or dual-tasking (p. 5–8), even if there were findings conflicting with prioritization in processing (e.g., losing information within the computation of a summary statistic as opposed to losing it because it had been filtered-out in a selection process, p. 8). This is because such a focus would necessarily neglect other areas that clearly demonstrate prioritization. For instance, that prioritization must exist in perception and cognition is evident from the fact that humans can bring their limited effectors (say, two hands) to operate on one out of several (say more than three) objects at a time and can do so, flexibly, to perform a specific (instruction-based) task (cf. Neumann, 1987). Adopting the view of Lakatos (1978) rather than Kuhn's (1962) paradigm theory, one might say that the assumption of priority was the "hard core" of attention research. Attentional mechanisms (e.g., those that select stimuli for perception [covert attention] and/or as targets for the next saccadic eye movement, bringing them onto the high-resolution fovea [overt attention], cf. Deubel & Schneider, 1996, or those that dynamically allocate neurons' receptive fields to processing a specific stimulus, Desimone & Duncan, 1995) would constitute hypotheses auxiliary to the hard core. Auxiliary hypotheses can be tested, modified, and potentially refuted or replaced without letting the core theory be challenged directly, forming a "protective belt" around the hard core (Lakatos, 1978). This seems in line with attention research, where auxiliary hypotheses of specific mechanisms performing prioritization (or selection) are tested, amended, or refuted, without questioning the principle of accommodating processing capacity limitations by prioritizing specific processes.

Now, what happens if we call for a paradigm shift without crisis? Paradigm shifts cut access to previous theories, methodologies, and findings. So, banning the concept of attention, as Rosenholtz suggests (p. 10), would disconnect new research from a plethora of theories and findings connected by the assumption of processing priorities (i.e., the broad concept of attention). This strategy risks a new replication crisis (Open Science Collaboration, 2015) because the (conceptual) replication of findings, which is the safeguard against erroneous findings (due to statistical variation), rests on these connections that allow to find, synthesize, and meta-analyze previous findings in concert with current ones. Rosenholtz suggested that "attention" was too broad and unspecific a concept to be useful for scientific progress (p. 4). However, it could be precisely the breadth of the concept that ensured stable progress. Rather than calling for a paradigm shift, the breadth of the concept could serve the important function of organizing the variety of prioritization mechanisms (auxiliary hypotheses) that compete as explanations of visual phenomena under a common superordinate principle (prioritization as the "hard core"). Adopting such an organization allows for a sustainable science of attention: Under a superordinate principle, previous findings, methodologies, and theories can be found, accessed, and understood. This ensures that they can be integrated into current revisions of attention theory, which would place the theory on a richer evidence base as compared with the situation of focusing on a specific sub-area. Thus, surprisingly, we need the broad and unspecific concept of attention (cf. Rosenholtz, p. 4) to achieve an incremental and sustainable science of attention (cf. Lakatos, 1978; Platt, 1964) rather than ruling out progress by waxing and waning Kuhn (1962) paradigms.

Acknowledgements. None.

Financial support. None.

Competing interests. Both authors declare that they have no competing interests.

References

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523. https://psycnet.apa.org/doi/10.1037/0033-295X.97.4.523

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://psycnet.apa.org/doi/10.1146/annurev.ne.18.030195.001205

Lakatos, I. (1978). Falsification and the methodology of scientific research programmes. In J. Worrall, & G. Currie, (Eds.), *The methodology of scientific research programmes: Philosophical papers*. (Vol. *1*, pp. 8–101). Cambridge University Press. https://doi.org/10.1017/CBO9780511621123.003

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer, & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 361–394). Hillsdale, NJ: Lawrence Erlbaum.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, *146*(3642), 347–353. https://doi.org/10.1126/science.146.3642.347

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238. https://doi.org/10.3758/BF03200774

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. https://doi.org/10.1016/j.tics.2010.12.001

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Crisis, contextualized: A much broader theoretical shift is needed

Seema Prasad[a] and Bernhard Hommel[b]*

[a]Cognitive Neurophysiology, Faculty of Medicine, TU Dresden, Dresden, Germany and [b]Faculty of Psychology, Shandong Normal University, Jinan, China
seema.prasad@ukdd.de
bh@bhommel.onmicrosoft.com

*Corresponding author.

doi:10.1017/S0140525X25000160, e156

Abstract

We agree that while there is a "crisis" in visual attention, the Rosenholtz' article does not offer bold enough solutions. We argue that the real crisis extends beyond attention, reflecting a broader need for theoretical integration. Addressing this requires abandoning artificial subdivisions and adopting a more ecologically valid, contextually grounded approach to cognition.

Rosenholtz' article diagnoses nothing less than a "crisis" in the science of visual attention. While we applaud the author's initiative to ban attention from her lab for a year and search for non-attentional explanations, we are both disappointed by the fact that the concept sneaks back in at the end of the article and unconvinced by the suggestion that an atheoretical list of still to-be-explained empirical phenomena and a few speculations are a bold-enough cure for a true crisis. The more so, as a closer look reveals that they are little more than re-inventions of three wheels.

First, Rosenholtz suggests that decision complexity and task factors might play a role in accounting for bottlenecks in performance. While we agree with the suggestion, this is particularly old news. Systematic research on cognitive bottlenecks by using the so-called psychological refractory period (PRP) design in combination with factorial logic has generated hundreds, if not thousands of articles over almost the last hundred years, all pointing to the same conclusion: the key bottleneck underlying human performance is response selection, whereas perceptual and otherwise "early" stages of processing reflect no or only hardware-related limitations, such as the resolution of the retina or other early-vision limitations discussed by Rosenholtz (Pashler, 1984, 1994). However, the more complex the response decision is, the more multitasking is hampered.

Second, Rosenholtz considers the possibility that the task and the purpose of perceptual/attentional activities might matter for understanding, probably even explaining processing bottlenecks. Again, we certainly agree, but this is old news as well. Various authors, including Allport (1987) and Neumann (1990), have suggested that "perception is for action" and the Theory of Event Coding (Hommel, 2004; Hommel et al., 2001), the arguably most comprehensive perception-action theory to date, considers perception as an integral part of human intentional action. This also accounts for the first observation, namely, that bottlenecks mainly reflect limitations of action, rather than of perception.

Third, Rosenholt'z article is not the first to identify a problem with the concept of attention. The second author and numerous colleagues have argued that the concept is obsolete because it lumps together numerous empirical phenomena without any theoretical reason (Anderson, 2011; Di Lollo, 2018; Hommel et al., 2019)—so that the fact that no integrative, fully comprehensive attentional theory can be found is hardly surprising. Rather than avoiding this trap and/or providing a functionally justified identification of to-be-explained phenomena, Rosenholtz applies the old-fashioned logic of defining and subdividing the research field in a purely semantic fashion. This is obvious from the choice of focusing on "visual attention," which falsely suggests that the processing and selection of visual information follows an entirely different mechanistic logic than, say, the processing of auditory, tactile, and proprioceptive information. As argued elsewhere (Hommel, 2020; Hommel et al., 2019), this is unlikely to carve nature at its true joints and likely to multiply empirical research without any hope for theoretical integration.

Hence, if there is a crisis, it is unlikely to be effectively addressed without a much more disciplined theoretical strategy. This strategy, we believe, would need to give up the widespread idea that brain mechanisms care for semantic musings and subdivisions and integrate the broader context of each performance under consideration into the theoretical analysis. It is true that scientific practice always tends to develop in the opposite direction: to-be-investigated empirical phenomena tend to undergo finer and finer subdivisions, stimulating increasingly artificial experimental designs and paradigms to study them in isolation, which in turn inflates the number of theoretical claims and working hypotheses with an ever-decreasing empirical scope (Hommel & Colzato, 2015). And yet, successful theoretical integration and the building of functioning "big theories" call for the exact opposite: a concerted effort to interconnect more and more empirical designs and tiny questions, so as to reduce, rather than to multiply, the number of subdivisions of both our phenomena and our theoretical concepts.

Another, theoretically particularly important implication of such a de-focusing strategy is that tasks should be put back into the context that they are carried out in. Experimental instructions are asking for actions, and it is actions that provide our dependent measures. Accordingly, even if our main interest is focused on one particular processing chain, we always need to consider the entire process of generating a response to a particular stimulus in our theoretical analysis. Popularizing this approach in the science of visual and other kinds of "attention" (or of the phenomena that we mistakenly or not believe to deserve this label) seems very likely to resolve most of the "crisis" that Rosenholtz diagnoses.

Our personal conclusion is therefore, yes, we do have a crisis, but it goes miles beyond the field of visual attention and rather involves cognitive (neuro)science as a whole. As Ulrich Neisser wrote in his book *Cognition and Reality* (Neisser, 1976), "... cognitive psychologists must make a greater effort to understand cognition as it occurs in the ordinary environment and in the context of natural purposeful activity. This would not mean an end to laboratory experiments, but a commitment to the study of variables that are ecologically important rather than those that are easily manageable." Along these lines, some have proposed a "cognitive ethology" approach where we do not assume that cognitive processes are invariant, stop relying only on artificial model tasks and instead collect complex real-world data (Kingstone, Smilek, & Eastwood, 2008). Others have advocated for a "synthetic approach" to cognition that involves identifying a common principle associated with a well-understood ecologically relevant basic mechanism and building an account of whatever phenomena we are interested in

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

based on this principle (Hommel et al., 2019). In any case, however, the cognitive and the neurocognitive sciences need to become much more theoretical and integrative, which, we are afraid, calls for a bolder approach than that suggested by Rosenholtz.

Financial support. SP was funded by a Maria Reiche Postdoctoral fellowship from the Graduate Academy of TU Dresden.

Competing interests. The authors declare no competing interests.

References

Allport, A. (1987). *Selection for Action: Some Behavioral and Neurophysiological Considerations of Attention and Action*. Routledge.

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*, 1–8. https://doi.org/10.3389/fpsyg.2011.00246

Di Lollo, V. (2018). Attention is a sterile concept; Iterative reentry is a fertile substitute. *Consciousness and Cognition*, *64*, 45–49. https://doi.org/10.1016/j.concog.2018.02.005

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500. https://doi.org/10.1016/j.tics.2004.08.007

Hommel, B. (2020). Pseudo-mechanistic explanations in psychology and cognitive neuroscience. *Topics in Cognitive Science*, *12*(4), 1294–1305. https://doi.org/10.1111/tops.12448

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Hommel, B., & Colzato, L. S. (2015). Learning from history: The need for a synthetic approach to human cognition. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01435

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. W.H. Freeman.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research*, *52*(2–3), 207–215. https://doi.org/10.1007/BF00877529

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 358–377. https://doi.org/10.1037/0096-1523.10.3.358

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244.

Attention in evolutionary perspective

Michael Tomasello[a,b]*

[a]Duke University, Durham, NC, USA and [b]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
michael.tomasello@duke.edu

*Corresponding author.

doi:10.1017/S0140525X25000305, e157

Abstract

Why have some organisms evolved processes of attention as distinct from perception in general? Investigation of freely behaving organisms (not in laboratory tasks) suggests that attention as distinct from perception is critical for goal-directed organisms' value-based decision making. As such, the target of attention is not punctate stimuli, but whole situations (scenes) that are relevant to that decision.

As an outsider to this field of study, I am struck by the preponderance of data from experimental studies in which subjects are in one way or another instructed on what they must attend to. This is even true to a large extent in studies with nonlinguistic animals in which there is training about which stimulus will be rewarded and then assessment of how much and how selectively they attend to exemplars (Zentall, 2005). But how does attention function in naturally behaving organisms?

In Tomasello (2022), I had occasion to think about attention from an evolutionary perspective. My primary question was what it is for, why attention as distinct from perception, evolve in the first place? Simple organisms like amoebae very likely just perceive the world directly, without attention as a distinct process. One hypothesis is that this is because they are not making agentive decisions. Generally supportive of this view, Gibson (1979) proposes that attention is best thought of as intentional perception, that is, perceptual processes that the agentive organism actively controls in pursuit of its goals. This implies that only organisms that act as goal-directed agents operate with attention as distinct from perception (very likely excluding organisms without a nervous system and perhaps others). One could even hypothesize that only in organisms that operate with an executive tier of operation – or at least with internally represented goals – is attention distinct from perception in general. In this view, attention evolved as one aspect of a control system or executive process to facilitate behavioral decision making, enabling goal-directed organisms to selectively engage with those aspects of its perceptual environment that are relevant to its decision making. The proposal that organisms naturally attend to situations relevant to their behavioral decision making is generally consistent with Rosenholtz's suggestion that what organisms are attending to is not sensory stimuli but "tasks" and with Krauzlis et al.'s (2014) focus on the role of attention in value-based decision making.

Attention may operate either top-down or bottom-up. Top-down attention evolved to facilitate the organism's perceptual engagement with situations in the environment that are relevant to its goals and values, that is, as either opportunities or obstacles. A foraging primate will naturally attend to the situation where there is edible food in this tree. It will also attend to the situations: the food is accessible, it is ripe, and there is a conspecific competitor nearby. It can attend to all of these various situations in a single visual percept of a banana in a tree – although perhaps not all at the same time. The key construct here is the *relevance* of situations to potential goal-directed (adaptive) decision making and action. Thus, when squirrels are making a decision between two paths to a goal, they often show "head switch" behavior, attending first to one and then the other alternative before choosing (Chow et al., 2015). Bottom-up attention operates in situations that are potentially more urgent, so that natural selection cannot leave it to the

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

individual organism's judgment of whether a situation is or is not relevant. For example, a loud noise in a certain direction demands attention from most organisms, as does the smell of certain foods in a given direction. It does not matter what the organism's immediate goals are; natural selection has determined that it should pay attention to some situations no matter what: natural selection has predetermined their relevance for adaptive action. Thus, in addition to attending to the fact that food is accessible, the foraging primate may have its attention drawn, bottom-up, to the loud screech of a predator as indicating that it is nearby.

It is important that what the organism attends to in natural, action-based situations is not an object or stimulus property (a prejudice that is likely a holdover from behaviorism's focus on stimulus and response). When the issue is behavioral decision making, the target of attention is a whole, fact-like situation. The primate attends to the fact-like situations that there is a banana in this tree, that it is ripe, that it is accessible, and, bottom-up, to the fact that a predator is nearby. It attends to the world in this way because its goals – as the *telos* of its actions – are formatted in this way, that is, as situations. The goal pursued by the primate is not a banana but rather the situation of having the banana in its possession. And so it must attend to relevant environmental situations as opportunities and obstacles for goal attainment. Of course, an organism may attend closely to an object to determine if it is, for example, a banana. But that act of attention is only relevant because it identifies possible food in this tree in front of me, which constitutes a situation of opportunity for foraging. Attention to objects and properties is always in the service of recognizing potentially relevant situations. The proposal that naturally behaving organisms naturally attend to whole situations or states of affairs is generally consistent with Rosenholtz's observation that humans attend to and comprehend especially quickly and easily "the gist of a scene."

Researchers may still want to focus on subjects' attention to prespecified objects or properties in controlled experiments to answer important questions of mechanism. But it might also be the case that considering why attention evolved in the first place and how it operates in its "natural habitat" of value-based decision making – its function of securing perceptual engagement with relevant situations – may open up new avenues of research for escaping from what Rosenholtz calls the "crisis" in the field.

Financial support. None.

Competing interests. None.

References

Chow, P., Leaver, L., Wang, M., & Lea, S. (2015). Serial reversal learning in grey squirrels: Learning efficiency as a function of learning and change of tactics. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*, 343–353.

Gibson, J. J. (1979). *The ecological approach to visual perception: Classic edition*. Psychology press.

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, *18*(9), 457–464.

Tomasello, M. (2022). *The evolution of agency: Behavioral organization from lizards to humans*. MIT Press.

Zentall, T. R. (2005). Selective and divided attention in animals. *Behavioural Processes*, *69*(1), 1–15.

The interplay between selective attention and summary statistics

Yonatan Vanunu[a,b,c*] and Roger Ratcliff[c]

[a]Coller School of Management, Tel Aviv University, Tel Aviv-Yafo, Israel; [b]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv-Yafo, Israel and [c]Department of Psychology, Ohio State University, Columbus, USA
yvanunu@tauex.tau.ac.ul
rogerratcliff22@gmail.com

*Corresponding author.

doi:10.1017/S0140525X25000342, e158

Abstract

Rosenholtz argues that summary statistics explain attentional phenomena via peripheral vision. While we acknowledge their role, we challenge the claim that they serve as an alternative mechanism. Instead, we argue that summary statistics and selective attention are interdependent, shaping visual perception under limited capacity, as evidenced by perceptual biases in numerosity judgments and mean estimations of shape size, color, and position.

Rosenholtz (2024) argues that summary statistics, processed through peripheral vision, can account for various attentional phenomena, such as change blindness and object-based attention. While we broadly agree, we challenge the notion that summary statistics and selective attention are alternative mechanisms necessitating a "paradigm shift." Instead, we propose that they are inherently interdependent – extracting summary statistics from a visual display often requires selective attention, depending on task demands and display complexity.

For instance, de Fockert and Marchant (De Fockert & Marchant, 2008) found that when observers estimated the average size of a set of circles, they did not always integrate information across the entire set. Instead, their estimates were often biased by the features of the more attended item or based on a smaller subset of items. Similarly, Myczek and Simons (2008) demonstrated that apparent evidence for mean size perception could be explained by various focused-attention strategies, such as sampling only one or two items from the set or averaging the largest and smallest items. Therefore, just as summary statistics can explain visual perception beyond focused attention, selective attention contributes to the formation of summary statistics under limited capacity.

Summary statistic processing, guided by attention, operates across diverse contexts, suggesting a generalized mechanism in which selective attention plays a key role in extracting summary information. For example, when participants estimated the average color of a set of colored objects, their judgments exhibited a "robust averaging" effect, in which the impact of the most extreme colors in the spectrum of the set on the average estimation was under-weighted – potentially as a strategy to enhance the signal over noise (de Gardelle & Summerfield, 2011). In contrast, the opposite pattern emerged when participants estimated the average of numerical values: extreme values received greater weight

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

(Rosenbaum, de Gardelle, & Usher, 2021; Spitzer, Waschke, & Summerfield, 2017; Tsetsos, Chater, & Usher, 2012; Vanunu et al., 2020; Vanunu, Pachur, Usher, 2019), presumably because they capture more attention (Kunar et al., 2017; Vanunu et al., 2021). Further, a similar pattern has been observed in spatial perception, where visually salient objects disproportionately influence centroid estimations, shifting perceived locations toward them (i.e., "The Weighted Average Illusion"; Hong, Witt, & Szafir, 2021).

We argue that while summary statistics reduce information load, they are often derived through selective attention, with their interplay shaped by task demands that would cause processing to exceed capacity if all information were to be used. For example, recent studies have shown that when discriminating between object numerosities – such as determining whether an intermingled array contains more blue or yellow dots – people tend to rely on a subset of information from the display's center, a phenomenon termed the "proximity-to-center bias" (Vanunu & Ratcliff, 2025; Vanunu & Donnelly, 2024). As a result, there is a bias to report objects near the center as more numerous, even when they are less prevalent across the entire display. Notably, the proximity-to-center bias was absent when participants compared the numerosity of two spatially separated arrays, suggesting that this bias is specific to tasks requiring high visual sensitivity. Because visual sensitivity peaks in the central region of the visual field, participants likely rely on a subsample from the center to distinguish between object types in briefly presented, intermingled displays.

To conclude, we challenge Rosenholtz's call for a "paradigm shift" and argue that summary statistics cannot replace selective attention (or inattention); rather, they function as complementary mechanisms operating under limited capacity. That is, while summary statistics can explain perception in the absence of focal attention, they are themselves constrained by capacity limits. Thus, rather than serving as an alternative to selective attention, summary statistics and attention work in tandem, shaping visual processing in an interdependent manner.

Financial support. This work was supported by funding from the National Institute on Aging (Grant numbers R01-AG041176 and R01-AG057841).

Competing interests. The authors declare no competing interests.

References

De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*, 789–794.

De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346.

Hong, M. H., Witt, J. K., & Szafir, D. A. (2021). The weighted average illusion: Biases in perceived mean position in scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 987–997.

Kunar, M. A., Watson, D. G., Tsetsos, K., & Chater, N. (2017). The influence of attention on value integration. *Attention, Perception, & Psychophysics*, *79*, 1615–1627.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & psychophysics*, *70*, 772–788.

Rosenbaum, D., de Gardelle, V., & Usher, M. (2021). Ensemble perception: Extracting the average of perceptual versus numerical stimuli. *Attention, Perception, & Psychophysics*, *83*, 956–969.

Rosenholtz, R. (2024). Visual attention in crisis. *Behavioral and brain sciences*, 1–32.

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature human behaviour*, *1*(8), 0145.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664.

Vanunu, Y., & Donnelly, K. (2024, April 16). *Center of Attention: Spatial Position Affects Quantity Judgments and Product Preference*. https://doi.org/10.31234/osf.io/mn24r

Vanunu, Y., Hotaling, J. M., Le Pelley, M. E., & Newell, B. R. (2021). How top-down and bottom-up attention modulate risky choice. *Proceedings of the National Academy of Sciences*, *118*(39), e2025646118. https://www.pnas.org/doi/full/10.1073/pnas.2025646118.

Vanunu, Y., Hotaling, J. M., & Newell, B. R. (2020). Elucidating the differential impact of extreme-outcomes in perceptual and preferential choice. *Cognitive Psychology*, *119*, 101274.

Vanunu, Y., Pachur, T., & Usher, M. (2019). Constructing preference from sequential samples: The impact of evaluation format on risk attitudes. *Decision*, *6*(3), 223–236.

Vanunu, Y., & Ratcliff, R. (2025). A selective sampling account of forming numerosity representations. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000575

Attention and visual search: No crisis here

Jeremy M. M. Wolfe*

Visual Attention Lab, Department of Surgery, Brigham & Women's Hospital/Harvard Medical School, Boston, MA 02215, USA
jwolfe@bwh.harvard.edu
https://search.bwh.harvard.edu/new/index.html

*Corresponding author.

doi:10.1017/S0140525X25000123, e159

Abstract

Rosenholtz's paper is a solid addition to a long tradition of throwing out babies with attentional bathwater (Di Lollo, 2012) (Wolfe, 2012) (Hulleman & Olivers, 2017) (Wolfe, 2017). She is correct that the term "attention" has been used in profligate and often ill-defined ways. That said, I argue that any plausible model of visual search must include visual selective attention.

Ruth Rosenholtz believes that visual attention is in crisis. Her argument makes several valid points. There is a tendency to reify "attention" as a single thing when, in fact, attention refers to a large family of selective processes in the nervous system. If I ask you to "pay attention" to your toes in your shoes, that is a different "attention" than the attention you would use to look for the letter "x" in the previous sentence. Turning to my favored topic of visual search, Rosenholtz is undoubtedly correct that several puzzles in the visual search literature might be understood as peripheral vision/crowding/clutter phenomena. Her work on these effects makes important contributions to the search literature. But Ruth would like to go further. Her initial hope was to ban "attention," at least as a term in discussion in her lab. However, as the bulk of her article shows, there is a range of tasks that appear to involve selective processes we could call "attentional." I want to argue that

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

visual search is one of those tasks that involves a central role for selective visual attention, even if the properties of peripheral vision do play their part in search.

The synthesized images in Figure 4 in Rosenholtz's article are a nice illustration of how peripheral vision might make chopped salad out of the preattentive features in a display. Indeed, the figures evoke Anne Treisman's claim that, in preattentive vision, features are unbound, floating free from their objects (Treisman & Gelade, 1980). However, the existence of this jumble of features fails to explain how it is that observers can find and identify targets in peripheral vision if those observers successfully deploy attention to that object. Could the target emerge from the jumble without attention? Perhaps features just recombine randomly, allowing you to successfully detect the target if you wait. Suppose, however, that we flash a ring of letters around fixation and ask if there is a "T" present. Let us assume that the letters are big enough and widely separated enough that there is no acuity/crowding effect blocking the identification of the letter in the periphery. With a short flash, say – 100 msec, performance will be poor. However, if we cue the potential location of the T, perhaps 100 msec before the onset of ring of letters, then performance will be near perfect. The task has not changed, but the elimination of spatial uncertainty changed appearance of the T. In that 100 msec, something has been deployed to the cued location, allowing the jumble of features to become a recognizable T. I wrote that sentence without the use of the terms "binding" or "attention," but that is a mere word game. The standard way to describe this situation is to say that the exogenous cue (Klein, 2009) directed your selective attention to the subsequently presented object. Attention enabled the binding of the object's features into a recognizable letter, "T." In the absence of the cue, either attention is deployed to the wrong item, by chance, or it is diffused over multiple items, and, most of the time, you fail to locate the T.

If you do not like the idea of covert attention, a different objection can be raised to the notion that peripheral vision fully accounts for the difficulty of some search tasks, even if we agree, peripheral factors are important. In most real-world search tasks, the eyes move around because, as Rosenholtz notes, object recognition is not very good in the periphery. If eye movements are allowed and observers search for that letter T, they will typically miss 5–10% of those targets. Moreover, if we track their eye movements, we will find that, in many cases, those observers fixated on or were very close to the missed target. That is, the target fell in the fovea and was missed, nevertheless (Wu & Wolfe, 2022) (Wolfe et al., 2021). Why are fixated targets missed on some trials but not on others (Li, Wolfe, & Hulleman, 2024)? It seems unlikely that basic visual processes changed. This seems more like a failure that involves something that we might call "attention" (Wolfe, Kosovicheva, & Wolfe, 2022).

If you look at the black fixation cross in the figure, you can get an intuitive feel for the processes discussed here and in the main Rosenholtz article. You immediately see a ring of two-color plusses. Call this the "gist" of the scene; the product of what we have labeled as a "non-selective pathway" (Wolfe, Vo, Evans, & Greene, 2011). You might also sense some instability as implied in the illustration of peripheral summary statistics shown in Figure 4 of Rosenholtz's target article. Now, looking at the plusses presented in this comment, without moving your eyes, search for purple-vertical, green-horizontal plusses. You seem to do this by "attending" (for want of a better word) from item to item or location to location, binding features until you find the target. While still fixating, direct your attention to the bottom item. If you found the target at the upper left, you might not have noticed that there are two targets. Now, with no change in the stimulus, you are aware that bottom item looks like a target, too. You could say that you shrank your Functional Visual Field around that item, allowing you to limit processing to a small set of features (Hulleman & Olivers, 2017), but that sounds quite a lot like attending to that item.

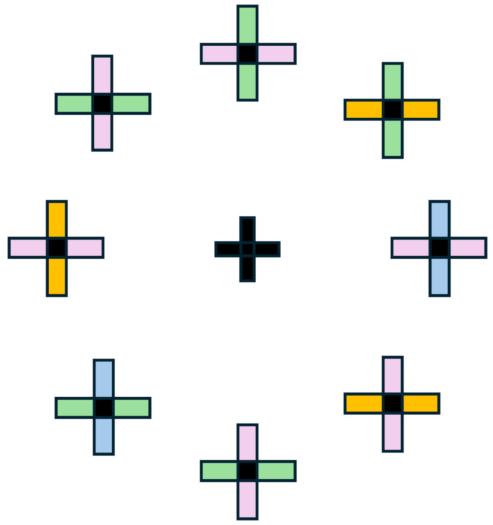


Figure 1 (Wolfe). What do you see when you look at the black cross? How can what you see change without a change in the stimulus?.

In sum, Ruth Rosenholtz does not much care for the term "attention." As Miranda says of the "monster," Caliban, in Shakespeare's Tempest, Ruth might say, "'Tis a villain, sir, I do not love to look on." (*Tempest 1:2:370–371*). Like Prospero, we might respond "But, as 'tis, we cannot miss him. He . . . serves in offices that profit us" (*Tempest 1:2:372–375*). There is much work to be done on visual search and visual attention, more generally, but visual attention is not in crisis, and no model of visual search will be plausible without a role for visual selective attention, whatever one may call it.

Financial support. JMW was supported by NSF-2146617, NIH-CA207490, and NIH-EY01700115.

Competing interests. No competing interests.

Reference

Di Lollo, V. (2012). The feature- binding problem is an ill-posed problem *Trends in Cognitive Sciences*, *16*(6), 317–321.

Hulleman, J., & Olivers, C. N. L. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, *40*, e132. https://doi.org/10.1017/S0140525X15002794

Klein, R. M. (2009). On the control of attention *Canadian Journal of Experimental Psychology*, *63*(3), 240–252.

Li, A., Wolfe, J. M., & Hulleman, J. (2024). Errors in visual search: Are they stochastic or deterministic? *Cognitive Research: Principles and Implications*, *9*, 15. https://doi.org/10.1186/s41235-024-00543-z

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Wolfe, J. M. (2012). The binding problem lives on: Comment on Di Lollo. *Trends in Cognitive Sciences*, *16*(6), 307–308. http://dx.doi.org/10.1016/j.tics.2012.04.013

Wolfe, J. M. (2017). "I am not dead yet!" – The item responds to Hulleman and Olivers. *Behavioral and Brain Sciences* 48–e161. https://doi.org/10.1017/S0140525X16000303

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Wolfe, J. M., Kosovicheva, A., & Wolfe, B. A. (2022). Normal blindness: When we look but fail to see. *Trends in Cognitive Sciences*, *26*, 809–819. https://doi.org/10.1016/j.tics.2022.06.006

Wolfe, J. M., Vo, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Wolfe, J. M., Wu, C. C., Li, J., & Suresh, S. B. (2021). What do experts look at and what do experts find when reading mammograms? *Journal of Medical Imaging (Bellingham)*, *8*(4), 045501. https://doi.org/10.1117/1.Jmi.8.4.045501

Wu, C. C., & Wolfe, J. M. (2022). The Functional Visual Field(s) in simple visual search. *Vision Research*, *190*, 107965. https://doi.org/10.1016/j.visres.2021.107965.

Mechanistic disunity as attention in crisis

Wayne Wu*

Department of History and Philosophy of Science, Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, USA
waynewu@pitt.edu

*Corresponding author.

doi:10.1017/S0140525X25000111, e160

Abstract

While I agree with Rosenholtz that attention as mechanism should often be "banned"—this conception is confused and often explanatorily useless—I suggest that the real crisis is the proliferation of different, too often underspecified, mechanisms as attention. Attention is not an explainer. It is what we are trying to explain. Confusion on this point leads to unnecessary theoretical disunity.

Ruth Rosenholtz trenchantly argues that there is a Kuhnian crisis regarding visual attention. Yet is there a paradigm, an overarching theory of attention, such that *it* faces anomalies? I'm unsure. For example, the Treismanian framework (target article, Figure 1) does not figure centrally in the neuroscience of visual search with nonhuman primates. Arguably, there is no paradigm that everyone agrees with, and absent a unifying account, we find fragmentation as exemplified in the proliferation of postulated mechanisms said to be attention (Rosenholtz, target, *p.* 3). Disunity is the crisis.

Disunity leads to skepticism: there is no such thing as attention and no one knows what attention is (B. Anderson, 2011; Hommel et al., 2019). Yet this crisis is self-imposed due to our failure to recognize a shared methodological commitment to a functional conception of attention (Wu, 2024). This commitment, which provides the functional structure of attentional tasks, reflects attention as the folk know it. In everyday actions, when a person mentally selects something to respond to, they attend to it (animals too, Hommel et al. op. cit.). So, we visually attend to a target when we visually select it to inform behavior. In visually selecting something to remember, to grasp, to avoid, to report, or to ponder, we visually attend to it to respond appropriately.

This *common conception* of ordinary attention engenders the shared methodological commitment. To study visual attention, we design appropriate tasks that require specific visual selection for correct performance. Selection is set by task instructions that manipulate how subjects attend. Subjects perform experimental tasks akin to mundane actions, so science studies attention as we all know it. Rosenholtz is right that "selection and performing a task might be *inseparable*" (24). They *are* inseparable. Accordingly, visual attention as what guides a person's doing something is what we're trying to explain.

The science of attention is anchored on experimental tasks built from the common conception. Consequently, the fragmentation of attention into distinct, often underspecified, mechanisms that rightly bothers Rosenholtz amounts to science changing the topic from the shared conception found in the folk and (methodologically) among scientists. *Neural mechanisms are a different thing*. They are not attention but explain it. Worse, this fragmentation of attention into distinct internal mechanisms, when not due to careless labeling, results from bad logic: reasoning from a process's being selective to concluding that it is attention. Because many selective processes are not attention, the inference is a *non sequitur*, the conclusion a category mistake.

This allows us to recast Rosenholtz's useful exercise of banning attention since her emphasis on eliminating the word 'attention' obscures its force. The issue is not terminology but explanation and ontology. After all, if "attention" successfully refers to a mechanism, nothing is gained by treating it as "that which should not be named". Rather, her exercise aims at *explanatory parsimony*. We do without attention by showing that the referenced mechanism is not needed to explain performance. For example, the spotlight of attention as (a stand in for) a mechanism to explain behavioral effects in spatial cueing can be eliminated in favor of appealing to expectations (statistical beliefs) and task representations (intentions) (Wu, 2025, Chapter 4).

These points bear on Rosenholtz's discussion of inattentional blindness as a sign of crisis. Inattentional blindness informs debates about phenomenal consciousness (a mistake, see (Wu, 2025, Chapter 10)), but the *basic point* consistent with the data is this: when attention is not directed at a target, subjects will fail to access it for memory or report. Attention, as what the subject does ("focus"), is necessary for access. It gates. Inattentional blindness tests this by experimentally eliminating attention to salient stimuli, say a gorilla, and recording failure to report it. This deploys the shared methodological commitment: attention is directed away from the gorilla by requiring subjects to visually select a distinct target, a ball, for a counting task.

Rosenholtz suggests that to "rethink attention, we should try to describe phenomena in a more agnostic way that does not rely on earlier concepts of attention or selection" (target article, 12). I agree if attention is misidentified with internal mechanisms (cf. her earlier mechanistic suggestion that "attention could narrow the pooling mechanisms that lead to crowding", target article, 23). Yet inattentional blindness fundamentally concerns selective attention in the ordinary sense. Indeed, we learn the basic point at a young age when teachers admonish us that in focusing on one thing (Tik-Tok), we miss other things (the lesson). One hardly needs experiments to establish this.

What needs rethinking is not whether inattentional blindness concerns attention (it does), but its significance. Such paradigms probe *distractability* during performance, and distraction, as

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

William James noted, is essentially contrasted with (task-directed) attention. Further, we can explain distractability without appeal to spotlights (attention as mechanism). For example, we can deploy priority maps that integrate various biases including task representations (goals), expectation, value, saliency maps and other factors. Attention as pulled from task is explained by the influence of various biases such as value (B. A. Anderson et al., 2021). No attentional mechanism (a "neural spotlight") is needed as priority maps and historical factors might suffice, and a priority map is not attention because it explains the subject's (shifts of) attention (Bisley & Mirpour, 2019).

Despite agreement on "banning", I disagree with Rosenholtz concerning the main crisis. A healthy science clashes about mechanistic explanations of core phenomena such as search. Her commentators will likely debate her crossing out moves depicted in (target article, Figure 5). This is "normal" science anchored on a paradigmatic experimental task. The real crisis is the continued proliferation of attention into underspecified mechanisms and processes, leading to fallacies, confusion, skepticism and pessimism. This fragmentation blocks us from seeing that a common conception of attention unifies decades of empirical work (Wu, 2025, Chapters 4–5). We should eliminate attention as an internal mechanism. It is not (B. Anderson, 2011; Desimone & Duncan, 1995; Hommel et al., 2019; Krauzlis et al., 2014; Wu, 2023). Attention is an activity we undertake as part of doing things in the world.

Acknowledgements. I'm grateful to Ruth Rosenholtz for many fruitful theoretical conversations.

Financial support. There is no funding source to report

Competing interests. There are no competing interests

References

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Theoretical and Philosophical Psychology*, *2*, 246. https://doi.org/10.3389/fpsyg.2011.00246

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews*, *130*, 326–350. https://doi.org/10.1016/j.neubiorev.2021.09.004

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current Opinion in Psychology*, *29*, 108–112. https://doi.org/10.1016/j.copsyc.2019.01.002

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception & Psychophysics*, *81*(7), 2288–2303.

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, *18*(9), 457–464. https://doi.org/10.1016/j.tics.2014.05.008

Wu, W. (2023). *Movements of the Mind*. Oxford University Press.

Wu, W. (2024). We know what attention is!. *Trends in Cognitive Sciences*, *28*(4), 304–318. https://doi.org/10.1016/j.tics.2023.11.007

Wu, W. (2025). *Attention*. (2nd ed.) Routledge.

Author's Response

Toward a paradigm shift in visual attention

Ruth Rosenholtz*

Massachusetts Institute of Technology, Cambridge, MA, USA
rruth@mit.edu

*Corresponding author

doi:10.1017/S0140525X2510143X, e161

Abstract

The commentaries debated numerous points in the target article. Many questioned the existence of a crisis and the benefits of a paradigm shift, even though none countered the listed signs of a crisis. Paradigm shifts are an important way that science progresses. There remains hope for a unifying theory, and for a reawakening of an ambitious science of visual attention.

Introduction

I am grateful to the commentators for their thoughtful and insightful responses to my article. They have provided valuable and often surprising perspectives and enriched the discussion on the current state of visual attention research. Here, I will highlight important points of agreement or extensions of the ideas in the target article, as well as lingering points of confusion or disagreement. The main points of contention concerned whether – whatever the perceived problems with the state of attention research – we should call it a crisis and work towards a paradigm shift. First, the next section provides a brief refresher on the main points of the target article and clarifies those points based on commentary.

Clarifying the main points

In the target article, I argued that visual attention is in a Kuhnian crisis for a number of reasons. First, a number of supposedly attentional phenomena can instead be explained by the rich but lossy encoding of peripheral vision in terms of summary statistics. Due to the quirks of peripheral vision, one must not only worry about the observer's eye movements (a point amplified by **Godwin, Hart, & Barnhart**), but also about the details of the stimulus and the task. Still other supposedly attentional phenomena might be predicted by an ideal observer. These results are problematic for an influential theory, Feature Integration Theory (FIT), and necessitate rethinking what we have learned about visual attention. Second, this points to problems with the methodologies of the visual attention paradigm not yielding the promised results. Third, this situation has facilitated a proliferation of types of attention and often vague proposed mechanisms. In short, the signs that visual attention is in crisis (Box A, target article) extend beyond merely non-attentional explanations

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

for some supposedly attentional phenomena (**Chapman, Addleman, & Störmer**).

Nonetheless, a number of phenomena seem to require explanation by an attention-like mechanism. I proposed a list of such phenomena (Box B, target article) and, based on that analysis, suggested two candidates for a new theory. My preferred option reframes all perception as requiring performing a task, with limits on task complexity. Making sense of this requires rethinking tasks. One may be unaware of the task, and rather than corresponding exactly with one's nominal task, it may often involve a sort of multitasking. If a task exceeds the complexity limit, the observer's brain must simplify it, leading to poorer performance. One might want to use "attention" to refer to the flexible mechanism that chooses and implements a task, subject to this limit.

In the end, the target article does not outright dismiss a mechanism one might want to call "attention" (**Ma et al.**), and, in fact, suggests a new way to think about an attention-like limit and the mechanisms for dealing with that limit. I agree that inattentional blindness necessitates an additional attention-like mechanism (**Most**); that is why it appears in Box B. Similarly, as the target article states, search tasks can encounter task limits. In this sense, I agree with **Wolfe** that "any plausible model of visual search must include visual selective attention," though I remain unclear what Wolfe or other commentators mean by "selective attention." Wolfe also raises the topic of attention-like limits on perception even at the fovea; this already appears in the list in Box B under "Dual task." I am not surprised that these errors are stochastic, given the noisy nature of vision. Perhaps Wolfe would like to add other phenomena to Box B, which seems reasonable. I do see, however, several difficulties. It is not clear to me at this point what lessons we can learn about attention from visual search. I am particularly wary of deriving mechanisms from the *subjective experience* of searching item-by-item. Considerable processing in peripheral vision seems to happen without awareness. Even if this processing proceeds in parallel, the different items might accumulate evidence at slightly different rates, leading to awareness of their identities at different times, a bit like popcorn popping.

A number of commentators point out that attention is not just about vision (**Posner**, **Dudarev & Enns**, Chapman, Addleman, & Störmer, and **Prasad & Hommel**). Prasad & Hommel, in particular, see this as a critical error, overly subdividing the field. I largely agree and merely attempted to stick to what I know and limit the scope of the target article. See Rosenholtz (2020), for instance, where I relate task complexity limits to working memory. Similarly, I wholeheartedly agree with commentators who point out that we need to worry about both sensory and attentional aspects of perception over time, not just space (**Dux, Dell'Acqua, & Wyble; Junker & Huber; Belledonne & Yildirim; Denison**). Again, this was not a theoretical division, merely limiting the scope to aspects I better understand. In the same vein, **Koevoet & Van der Stigchel** rightly protest my focus on classification tasks; this is simply easiest for me to conceptualize and easiest to explain to a broad audience. (I disagree that the target article is limited to *categorical* classification tasks; see Figure 8).

The nature of the visual representation would fundamentally determine the complexity of a given task. Stimuli with evolutionary importance might have developed more advantageous representations (**Capparini, To, & Reid**). Here, I have far more agreement with Chapman, Addleman, & Störmer than they suggest. Under limited task complexity, capacity limits in visual processing would be "predicted by the organization of the visual representations themselves," *plus the nature of the limit*. If the visual representation makes two classes well separated by a linear classifier, one would imagine that discriminating between those classes would be simple by most definitions of task complexity. If one were limited in the number of neurons one could use, then performance would depend on how well the best *N* neurons discriminated between cats and dogs, rather than how well the entire brain could discriminate between the two. In addition, I have argued for the utility of thinking of the visual system as almost never merely doing tasks like cat vs. dog discrimination. Particularly in the real world, but also in the laboratory, vision at the bare minimum attempts to monitor the situation and keep track of the general layout of the scene. It is this compound task that potentially encounters complexity limits. Performance would also depend on the usual suspects, such as the strength of the observer's prior information and the effort they exert (Koevoet & Van der Stigchel).

In a related note, Most describes the role of *attentional set* as opposed to task limits in inattentional blindness. On the surface, task limits might merely be a constraint and attentional set the implementation of task requirements subject to that constraint. However, in rethinking the lessons from FIT, we must rethink whether attentional set might include far more sophisticated features (**Kristjánsson & Chetverikov**). Most points to experimental results in which observers better react to a blue motorcycle than a gold one when following blue arrows; surely, he suggests, this is feature-based attention. As a description of results, sure. As evidence that the visual system can only attend to simple features, I would wager not. The observer likely simultaneously gives preference to processing features relevant to the driving task. The task set might appear to include simpler features because monitoring blue stuff generally requires less effort than monitoring blue *arrows*.

Agreement on a number of issues

Commentaries expressed general agreement with a number of the target article's main points. This includes the importance of peripheral vision and ideal observers as potential explanatory factors, the state of Feature Integration Theory, and the proliferation of ill-defined or underspecified attentional mechanisms. In addition, several commentators resonated with the second half of the target article: the suggested list of attentional phenomena, thoughts about tasks, and task limits.

Vanunu & Ratcliff broadly agree that peripheral vision accounts in part for change blindness and inattentional blindness. Wolfe concurs that at least some search puzzles may have been due to peripheral vision confounds. Posner agrees broadly that the field has sometimes confounded sensory and attentional effects. **Carrasco, Chapman, Addleman, & Störmer**, and Kristjánsson & Chetverikov reiterate the need to control for both perceptual factors and what can be explained by an ideal observer. Carrasco rightly points out that her laboratory has spent decades controlling for perceptual factors in order to study attention.

In spite of **Wu**'s prediction that "Commentators will likely debate her crossing out moves . . . ," essentially no essays obviously defended Feature Integration Theory. Kristjánsson & Chetverikov and Denison describe FIT as having been in trouble for a long time. From the commentaries, we might infer that others agree; Carrasco, for instance, concurs with past shortcomings in interpreting search experiments. Posner, in fact, cites evidence that a lack of binding can occur even with visual attention. To me, this looks like a Goldilocks effect, as described in the target article.

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

Despite this lack of disagreement, FIT lives on in the field in the form of the assumptions that often go unquestioned. We should be alert to ideas about automatic processing, the availability of only simple features without attention, the need for attention for binding, and FIT-like notions of the purpose and mechanisms of selection.

A number of commentaries agreed with the focus of previous essays on the detrimental effects of the proliferation of ill-defined (Carrasco), underspecified (Chapman, Addleman, & Störmer; Wu) types of attention or attentional mechanisms (Anderson, 2011; Hommel et al., 2019; Zivony & Eimer, 2021; Chun, Golumb, & Turk-Browne, 2011). But authors also argued for the correctness of their favorite distinctions, for example: Posner for alerting, orienting, and executive function; Dudarev & Enns for the correspondence between multiple reasons to prioritize and multiple kinds of selectivity; Carrasco for the exogenous (involuntary)/endogenous (voluntary) and feature/spatial/temporal distinctions; and Denison for multiple mechanisms generally. The sections on "**Paradigm Shift = Bad?"** and "**Paradigm Shift Good"** debate whether this precludes a unifying theory.

Several commentaries found the list of critical attentional phenomena valuable (Box B). **Belledonne & Yildirim** appreciate the connection between the phenomena listed and a fundamental part of attention: how goals impact our perception. Denison agrees that more modern theories of attention, for example, the normalization model and extensions thereof (Reynolds & Heeger, 2009; Jigo, Heeger, & Carrasco, 2021), do not attempt to capture all of the attentional phenomena one might want, and the list provides "a useful guide to theory development." However, in some cases, reading between the lines, a number of commentaries seem to want to add phenomena to that list. I will address this along the way, and in more detail in the subsection on "**Paradigm Shifts and Ambitious Science**."

Some of the more interesting agreement concerns how we might think about tasks, task-based selection, and even task limits. First, in terms of tasks and complex templates, Kristjánsson & Chetverikov cite evidence for summary-statistic-based, complex attentional templates, a welcome change from FIT's claim that one can only select simple features. **Tomasello** resonates with complex tasks in discussing attention to the *situation*, rather than to objects. **Morsella et al.** describe what I would consider a task-based view of attention, which they call "priority-based," in which one is likely unaware of the complex tasks performed by the system. **Duncan & Theeuwes** concur that their results show that one is not in full conscious control of the task set but rather processes information outside of the nominal task (see Box A, target article). Similarly, Chapman, Addleman, & Störmer describe attentional effects without an explicit nominal task.

In terms of formulating attention in terms of task-based selection and task limits, we can see agreement on several points. Wu finds it reasonable to think of task performance and selection as inseparable. Koevoet & Van der Stigchel stress the importance of considering effort if one is to study decision complexity, and I wholeheartedly agree. Belledonne & Yildirim describe a theory with a moment-to-moment evolution of attention over time, consisting of mini tasks or "targetable singleton computations," each of which takes unit time. One might think of this time requirement as a different form of complexity limit. **Poth & Poth** point to number of end effectors as an important additional limit, and I agree that this may have guided the development of the attentional system. Along these lines, several commentators from the domain of perception and action resonate with a number of the claims in the target article, from thinking of attention as selecting tasks (**O'Bryan & Song**) to notions of decision complexity; Prasad & Hommel describe this as "old news," which I will take as a form of agreement. **Krauzlis** independently arrived at a task-based selection proposal based on physiology research, but cautions that those results also support the flexible pooling account.

The fact that similar ideas have arisen from thinking about perception for action (Dudarev & Enns, Prasad & Hommel), consciousness (Morsella et al.), evolution (Tomasello), and physiology (Krauzlis) gives me hope that we are on the right track.

Suggested extensions

A number of commentaries went beyond agreement with ideas in the target article to suggest extensions to those ideas. This includes the role of perceptual factors, the utility of banning "attention," and thinking about task complexity. Many such suggestions lie outside of my expertise. However, I will speculate about attentional blink, the topic of three commentaries.

Francis & Thunell, for instance, mention that their concerns about statistical power apply to other attentional paradigms, not just object-based attention. They point in particular to studies that examine modulations of already small effects, e.g., looking for some sort of group difference in a spatial cueing effect.

Capparini, To, & Reid report that both peripheral vision and attention change a great deal during development. They suggest it is important to extend understanding of peripheral vision to development, where many studies look at attention but far fewer examine peripheral vision. One needs to understand possible confounds in the same way as in the target paper.

Martelli, Mancuso, & Zoccolotti note that allocentric neglect leads to errors that sound a bit like crowding. They ask whether there might be a peripheral vision component. This question has particular relevance since search tasks are a popular experimental paradigm for studying the topic.

In an interesting collection of commentaries, **Mattei** and **Junker & Huber** suggest that attentional blink (AB) might be due at least in part to temporal crowding. Dux, Dell'Acqua, & Wyble, on the other hand, describe AB as obviously non-perceptual and teaching us something about the time course of attention.

I do think that Mattei and Junker & Huber have a valid point that some AB phenomena, such as lag-1 sparing, may have perceptual explanations. Experimenters create stimuli by presenting a sequence of discrete symbols, but the visual system inputs not a list of symbols but a spatiotemporal signal. Early in vision, it represents this as energy in multiple bands, not as items for selection. We know far less about temporal crowding than about spatial, but it seems likely that vision pools over significant portions of time as well as over space, even at the fovea. This could lead to temporal crowding with familiar signatures. In fact, the temporal phenomena of AB can sound eerily like spatial crowding: The stimulus appears to contain separate items, and yet, puzzlingly, it can be difficult to individuate and report them under certain circumstances. Crowding has well-known dependence on stimulus similarity, which could conceivably make discerning two letter targets separated by a more distinct number harder than discerning two adjacent letter targets. Even without crowding, sequential, similar target letters, such as a K, H, or B, may be grouped by motion into a K-H morph, or even a K-H-B morph, like a flip book animation. Retrospectively, the observer may be able to parse these chunks into a reasonable guess as to the

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

individual letters, whereas a K-6-H sequence with more complex motions may not create such a comprehensible chunk.

That said, I agree with Dux, Dell'Acqua, & Wyble that better performance when one has to report only the second target is a clear sign of an attention-like mechanism, and probably deserves to be listed in Box B. It does seem like we learn something about the timing of some kind of attentional or task-switching mechanism, though it might be helpful to understand temporal crowding to disentangle the timings of perceptual as opposed to attentional mechanisms.

Hertzmann, meanwhile, takes note of a different point in the target article: a summary-statistic encoding has still-unrealized implications for understanding much of later visual processing. In particular, he rethinks 3D vision across eye movements through the lens of both a lossy encoding in peripheral vision and preferentially preserving ecologically valid information across fixations. This seems like a promising approach.

Most finds value in the exercise of banning the word "attention" when studying individual biases and anxiety disorders, though for a somewhat opposed reason. He suggests that researchers in that field may be finding a confusing pattern of results because they are examining different kinds of attention. One might also want to do Francis & Thunell's analysis on what power is necessary to use a dot-probe task to study what sounds like individual differences or impact of training.

Finally, Mattei deserves a shout-out for reasoning about task complexity and about perception occurring as the result of a task. He focuses on ensemble perception and, by extension, whether attention is needed for action. Vanunu & Ratcliff also discuss ensemble perception, and correctly suggest that the ability to report the mean size of a set of circles cannot serve as an alternative to attention nor as a paradigm shift. However, this confuses the rich image statistic representation in my peripheral vision model with the task of reporting ensemble properties; see also Kristjánsson & Chetverikov, who clarify the difference between simple ensemble statistics and the summary statistics in the model.

Two other sources of confusion worth discussing

Several commentators associate attention with awareness. Dux, Dell'Acqua, & Wyble, as an example, describe the observer as faced with far too much information "for it all to be processed up to the level of consciousness." The possible association between attention and awareness has been much debated (see Koch & Tsuchiya, 2007; Cohen, Dennett, & Kanwisher, 2016; Rosenholtz, 2020). I would presume that consciousness faces limits beyond those that govern attention. We know from perception for action that quite a lot of perception can occur without awareness. Along these lines, Hertzmann might be cautious about assuming that when observers fail to notice redirected walking in VR that is because their visual systems have thrown away inconsistent information; apparently researchers in that field note that even observers who fail to notice the trick appear more likely to experience motion sickness (N. Williams, personal communication, March 15, 2025). Personally, I would not attempt to fix the mismatch between awareness and action by proposing two separate systems (Most), as this adds considerable complexity to the theory.

More benignly, Duncan & Theeuwes have mistaken my point about the 40-ms effects in their distraction paradigm. They say that in the real world, 40 ms can matter a lot. I wholeheartedly agree, particularly in applications such as vision in driving (Wolfe et al., 2021). The target article did not actually say that *I* thought that 40 ms is not important, but rather that the *observer* might be unaware of or unconcerned by that additional time. In other words, the effect might depend on the effort the observer was willing to put in to ignore the distraction (Koevoet & Van der Stigchel). Along these lines, O'Bryan & Song report that, in real-world tasks, observers may not be distracted by irrelevant salient stimuli, and, in fact, sometimes show the opposite effect.

"All models are wrong, some are useful"

This famous quote from George Box (Box, 1979) is apt for several of the commentaries. These ask both about the quality of the peripheral vision model and the falsifiability of task complexity.

First, **Bornet, Herzog, & Doerig** provocatively ask whether we should "start a revolution with a refuted model." Work by these researchers has demonstrated that the Texture Tiling Model (TTM) for peripheral vision appears not to preserve enough information to predict performance on some Mooney face tasks. They have also argued that TTM cannot explain grouping effects in Vernier acuity tasks. I argue that Vernier acuity is a special hyperacuity task, and quibble with whether they have performed the definitive test, but I necessarily leave this detailed discussion for another venue. Their suggestion echoes the concern that attention influences crowding (Carrasco), though those results are consistent with cueing improving performance rather than changing the peripheral representation. Bornet et al. also state that TTM is unlikely to be a good model of vision in general. TTM is not intended as a model of vision in general; one would say the same of a model of the retina, or of V1.

Even if we take as a given that TTM is wrong in these and no doubt other ways, it is nonetheless useful. In particular, the claim that visual attention is in crisis relies predominantly on predictions for visual search, scene perception, and dual-task performance. We have backed up all of these predictions with peripheral vision experiments, plus, in some cases, intuitions from TTM to help us understand how the sparse dual-task displays relate to the crowded search ones. **Hulleman**, together with Olivers (2017), provide corroborating evidence. FIT is in trouble – and attentional mechanisms need rethinking – without predictions from TTM. What TTM does provide is the glue that allows us to see that a single model can predict all of these results; in other words, that peripheral vision more parsimoniously explains that entire set of phenomena than attention does. Furthermore, a model can clearly be "wrong" and support a paradigm shift; just look at Copernicus' version of heliocentrism.

Ma et al. ask whether the peripheral vision explanation for some phenomena gives us "greater precision in understanding cognition and behavior" than the old attentional explanation. They suggest asking "how much variance in [supposedly] attentional tasks can be accounted for by peripheral factors alone." I certainly agree that we should not accept a new handwavy theory in place of the old, but TTM is a useful model in just this way. Unlike old word-model predictions for things like visual search, with ill-defined attention mechanisms, TTM can make quantitative predictions. Zhang et al. (2015), for instance, show that peripheral mechanisms explain 75% of the variance for the 10 search conditions studied. (We should not, I think, presume that all the remaining 25% of the variance is attention.) In addition, Chang & Rosenholtz (2016) showed the model could predict in advance the results for five new pairs of search conditions.

On the other hand, Koevoet & Van der Stigchel and Chapman, Addleman, & Störmer worry that task complexity remains ill-

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

defined, and the latter ask whether that theory is testable. I agree that we would need to nail down how to operationalize task complexity, though I have found it useful to get intuitions even without that precision. Rosenholtz (2017) discusses testability in more detail. In brief, for a given measure of complexity, and a candidate visual representation, such as from a pre-trained neural network, one would want to ask whether, for example, hard dual tasks correspond to higher complexity than easy dual tasks. Here, Koevoet & Van der Stigchel have, I think, a critical point: one should do one's best to normalize the effort exerted by the observers across conditions.

Paradigm shift = Bad?

Overwhelmingly, commentators most agreed with each other either in declaring visual attention not in crisis or not in need of a paradigm shift, and/or in suggesting that such a paradigm shift would be pointless or detrimental to science. Here, I review those arguments and respond briefly. In Paradigm Shift Good, I argue for the feasibility and benefits of a paradigm shift.

Someone already fixed the problems, and the field is making progress

In this category, commentaries suggest that there is no crisis and no need for a paradigm shift because, for example, Wolfe's theory (Poth & Poth) has resolved the theoretical issues, or researchers have developed normalization theories of attention (Carrasco, Denison), or bottleneck models (Dux, Dell'Acqua, & Wyble) to explain a subset of behavioral and physiological effects. I believe the field can develop a coherent theory that explains a wider range of phenomena. These researchers also point to extensive, existing attempts to control perceptual factors by keeping the stimulus constant while varying the task. Such methodological changes do represent an important step. Better understanding of perceptual factors can help, by allowing one to study a broader set of phenomena while accounting for stimulus differences using a model. Controlling for decision effects – i.e., what would an ideal observer do – poses greater difficulties. Cueing does change the task (Wolfe). When a cue or prior knowledge reduces uncertainty about the location or identity of a target that can improve performance even in an unlimited capacity system; as a result, performance improvements may not result from selection in the traditional sense. Studying attention is hard.

Commentaries also objected to calling for a crisis or a paradigm shift because the field continues to make progress. Kristjánsson & Chetverikov point to improved understanding of how the brain prioritizes and gathers information. They suggest that loose definitions have been sufficient to guide a diverse set of experiments. Posner and Ma et al. point to progress in neuroscience, Dux, Dell'Acqua, & Wyble in attentional blink, and Carrasco regarding effects of selection on spatial frequency and contrast sensitivity. The list could go on. Herein, however, lies a misunderstanding. Labeling a field "in crisis" does not mean the field has not progressed. Ptolemaic astronomers productively made celestial measurements, and these were useful for guiding later theory. The questions are whether the old paradigm has fatal issues – a claim that no commentary really disagreed with – and whether a new paradigm consisting of theory, methods, and puzzles to solve has truly replaced it.

Contrary to these defenders of the current paradigm, Francis & Thunell argue against a paradigm shift because the anomalies might not be real, but rather the result of underpowered experiments. This certainly sounds to me like a sign of crisis, albeit of a different sort. I doubt this is the entire story, though it plausibly suggests delaying integration of some empirical results while waiting for the field to sort this out.

Concern about a paradigm shift

Several commentaries express what seems to me excessive concerns about the consequences of a paradigm shift. They worry that it would cause us to lose understanding of the different brain areas involved, of how attention influences sensory and memory processing, and of how the brain makes use of top-down information (Posner). It would risk a replication crisis and disrupt scientific progress by "preventing incremental science" (Poth & Poth). Re-interpreting existing experimental results "invalidates the entire approach of hypothesis testing" (Duncan & Theeuwes). A paradigm shift does not require a collective amnesia of past work, and healthy science does rethink the meaning of past experiments based on new understanding.

Attention is an apple and paradigm shifts are for oranges

A number of commentaries appear to argue that attention is simply not the kind of thing that can require a paradigm shift. Is attention simply self-evident? "Essential" because "it is a resource-limited process" (Ma et al.)? Commentaries state the obviousness of prioritization (Poth & Poth) and a need to focus – when someone selects something to respond to, they attend to that thing; "one hardly needs experiments" (Wu). (Some of these comments may simply reflect the misconception that the target article did away with attention entirely.) Commentators describe attention as effect or outcome, what we need to explain rather than the cause (Morsella et al., Dudarev & Enns, Wu). Kristjánsson & Chetverikov suggest we consider attention a general concept, like perception, and Denison posits that attention might not be a theory. To all of this, I ask: what non-obvious concept worthy of scientific study does require a theory? Let's discuss that instead.

No point in a unifying theory

Other commentaries suggest that seeking a unifying theory is misguided. Wu asks whether there has ever been an agreed-upon attention paradigm, noting differences between dominant paradigms in behavior as opposed to physiological studies. Many different reasons to prioritize information supposedly correspond to many kinds of selectivity (Dudarev & Enns). There simply exist different kinds of attention and many different kinds of mechanisms, as indicated by use of different brain areas (e.g., Posner, Carrasco), with different timing and performance characteristics (Carrasco). Pinning down attention too narrowly can lead the field astray (Kristjánsson & Chetverikov). The next section addresses this apparent proliferation and asks what it means to have a unifying theory.

Paradigm shift good

Having gone through a multi-year exercise of rethinking visual attention, I understand that it looks like a significant amount of work. Discarding old understandings can be difficult and even painful, especially if we have put substantial effort into the old paradigm. The reader might question whether the rewards justify the effort. Having seen the benefits in understanding a broad array

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

of phenomena and making predictions, I believe the effort is well worthwhile. Let us start by addressing the question of whether there might yet be hope for a unifying theory. I argue that we should not worry too much about the apparent diversity of mechanisms.

Is the proliferation of mechanisms important?

What does it mean for there to be a unifying theory, with one capacity limit and one flexible mechanism for dealing with that limit? At its most basic level, it means that we get predictive value from thinking that way. In vision science, linear systems analysis provides a classic example. One can make quite good predictions about detection thresholds for a diverse set of patterns – circles, checkerboards, noise – by measuring thresholds for narrow spatial frequency bands, and reasoning about how one combines those spatial frequencies to form the more complex patterns (e.g., Watson, 2000). To a first approximation, one need not worry about circle-detection, checkerboard-detection, and noise-detection mechanisms. Trained neural networks provide a more modern example. If one can train a network to extract generally useful features, and based on those features predict recognition of squirrels, apples, and toys, that provides a unifying theory of object recognition (a model that is, of course, wrong, but useful). Note that this does not imply that squirrel recognition occurs in the same brain area as apple recognition, nor that performance is the same for both tasks.

There are a number of reasons why the apparent diversity of attentional mechanisms might not be real. Purported attention mechanisms may have proliferated for a number of reasons:

(1) The dominant theory at the time may have predicted the existence of multiple mechanisms
(2) Lumping together perceptual, attentional, and decision-making phenomena may have made it appear that more attentional mechanisms were needed
(3) Confusing empirical results with mechanisms
(4) Not adequately considering inherent task demands

Consider the exogenous/involuntary vs. endogenous/voluntary distinction as an example (for simplicity, I will refer to this as exogenous vs. endogenous). I do not intend this discussion as a debunking of that distinction; Carrasco, for instance, argues for it on the basis of a differential effect on the contrast sensitivity function, and I will not address that here. Rather, this distinction provides a useful example because it demonstrates several of the potential pitfalls.

FIT predicted a difference between automatic, preattentive processing and processes that required attention. As a result, early work may not have adequately questioned the exogenous/endogenous subdivision. Now that we are rethinking FIT and automatic processes, this deserves a second look.

One can change the task from spatial attention to feature attention, but that does not mean that observers use different mechanisms to perform those tasks. Similarly, one instructs observers to follow an endogenous arrow cue, whereas internal goals might drive them to follow an exogenous cue. But this does not mean that the two do not have shared mechanisms. If they appear to utilize different brain regions, that might be because of non-attentional processes the two tasks do not share, such as arrow perception or interpreting the experimenter's instructions.

One often sees arguments for an exogenous/endogenous distinction on the basis of difference in timing (e.g., Cheal & Lyon, 1991). However, exogenous vs. endogenous cues arguably have inherently different time courses. Observing response to an exogenous cue after 50 ms, and to endogenous cues after 400 ms might not mean the two require different mechanisms. If one endogenous arrow pointed to a second endogenous arrow, which finally pointed to the target location, that complex cueing might require 600 ms. Most of us would not consider that difference in timing evidence of a new attentional mechanism.

Paradigm shifts and ambitious science

A call for a paradigm shift is an appeal for a return to a bold, ambitious science of visual attention. This would mean a return to the promise that understanding a single capacity limit might allow one to predict performance for a wide range of tasks, from simple laboratory experiments through complicated, ecologically valid real-world perception for action. Ultimately, this is the promise of science: that we not merely test hypotheses and describe empirical results but build a deeper understanding that enables intuitions and extrapolation to experiments not yet run. In addition, it can be deeply satisfying to ponder a field's worth of empirical results and attempt to make sense of them all.

Of course, not everyone needs to work towards this ambitious goal. Some may prefer to collect the empirical data that others use to build a theory. Other scientists may elect to develop a theory of a more targeted set of phenomenology. Science benefits from a diversity of approaches, skills, backgrounds, and personalities.

Those who do want to aim higher do not need to take as given my list of phenomena in need of explanation. Dudarev & Enns may want to add neural and behavioral results of planning and executing an action. Duncan & Theeuwes may want to add more distraction phenomena, Wolfe examples from search. Godwin, Hout, & Barnhart may want to add some aspects of overt attention back into the fold. Carrasco and Chapman, Addleman, & Störmer might argue for including shifts in representation and appearance as a result of attention. Importantly, note that the intent of the list is not to capture all phenomena that might involve attention. (Recall, after all, that in the task-complexity theory, all perception requires attention, i.e., requires performing a task.) Rather, the list aims to be conservative by excluding phenomena that might have important sensory or decision confounds or otherwise be primarily governed by another mechanism. In so doing, we hope to minimize the chances of going down a blind alley in search of a theory.

Similarly, I have found the task complexity way of thinking quite useful. But others may prefer a different unifying theory. Bornet, Herzog, & Doerig and Hulleman seem to prefer something closer to the flexible pooling idea, which Krauzlis also supports. Others may wish to (continue to) develop their own ambitious theory, such as Belledonne & Yildirim, Chapman, Addleman, & Störmer, or Wolfe.

I invite other researchers to tackle the crisis in visual attention head-on and work together to drive the paradigm shift forward. This ambitious endeavor needs diverse perspectives and contributions from throughout the scientific community. Together, we

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press

can push the boundaries of our understanding and make significant strides in cognitive science.

Acknowledgments. I would like to thank Jen Corbett, Niall Williams, Zoe Xu, and Robin Adelson for helpful discussions. Special thanks to Ted Adelson for his unwavering support and insightful conversations. Additionally, I am grateful to my fellow speakers at the 2024 Vision Sciences Society Symposium, "Attention: accept, reject, or major revisions?" Alon Zivony, Britt Anderson, Wayne Wu, and Sarah Shomstein, for their camaraderie and moral support. I greatly appreciate their shared enthusiasm for discussing attention.

References

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in Statistics* (pp. 201–236). Academic Press.

Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express*, *6*(1), 12–33.

Wolfe, B., Kosovicheva, A., Stent, S., et al. (2021) Effects of temporal and spatiotemporal cues on detection of dynamic road hazards. *Cognitive research*, *6*, 80. https://doi.org/10.1186/s41235-021-00348-4

https://doi.org/10.1017/S0140525X24000323 Published online by Cambridge University Press