TUTORIAL REVIEW

Keeping it real: Looking beyond capacity limits in visual cognition

Árni Kristjánsson[1,2] · Dejan Draschkow[3]

Abstract

Research within visual cognition has made tremendous strides in uncovering the basic operating characteristics of the visual system by reducing the complexity of natural vision to artificial but well-controlled experimental tasks and stimuli. This reductionist approach has for example been used to assess the basic limitations of visual attention, visual working memory (VWM) capacity, and the fidelity of visual long-term memory (VLTM). The assessment of these limits is usually made in a *pure* sense, irrespective of goals, actions, and priors. While it is important to map out the bottlenecks our visual system faces, we focus here on selected examples of how such limitations can be overcome. Recent findings suggest that during more natural tasks, capacity may be higher than reductionist research suggests and that separable systems subserve different actions, such as reaching and looking, which might provide important insights about how *pure* attentional or memory limitations could be circumvented. We also review evidence suggesting that the closer we get to naturalistic behavior, the more we encounter implicit learning mechanisms that operate "for free" and "on the fly." These mechanisms provide a surprisingly rich visual experience, which can support capacity-limited systems. We speculate whether natural tasks may yield different estimates of the limitations of VWM, VLTM, and attention, and propose that capacity measurements should also pass the real-world test within naturalistic frameworks. Our review highlights various approaches for this and suggests that our understanding of visual cognition will benefit from incorporating the complexities of real-world cognition in experimental approaches.

Keywords Visual attention · Visual working memory · Visual long-term memory · Virtual reality

Introduction

Lecturers in visual perception and cognitive psychology often wow undergraduates by showing them well-designed experiments that highlight the limitations of various aspects of visual cognition. How inefficient they are when searching for the yellow vertical bar amongst an array of yellow horizontals and blue verticals and how half the class misses a turbine engine disappearing and reappearing in plain sight. The students are surprised because these examples strongly contrast with their experience of being efficient visual searchers and never missing the lack of a critical plane component in their visual field. But natural behavior in natural environments is so seamless because properties of the visual system enable us to overcome all these limits. Examples of impoverished perception show how experimenters can investigate these individual processes in a pure, but because of that, isolated and artificial way. In natural interactions with our environment we effortlessly overcome situations and events that should bring us to the boundaries of our capacity limitations – in a way that seems almost trivial to us. In this tutorial review we focus on selected literature that demonstrates how these limitations can be circumvented, often by taking a step towards the real-world. Our aim is to highlight how these goals can be achieved with the various examples from the literature that are discussed here.

✉ Árni Kristjánsson
ak@hi.is

✉ Dejan Draschkow
dejan.draschkow@psych.ox.ac.uk

[1] School of Health Sciences, University of Iceland, Reykjavík, Iceland

[2] School of Psychology, National Research University Higher School of Economics, Moscow, Russia

[3] Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK

From "pure" to "messy" measures

Following the cognitive revolution of the 1950s, researchers within psychology, cognition, and neuroscience started to

investigate so-called basic mental processes. Their aim was often to study them in a *pure* sense, without any interference from other processes, such as the goals we may have at a given moment, and the particular tasks or actions we perform (see, e.g., Baddeley & Hitch, 1974; Broadbent, 2004; Neisser, 1963, 1967; Sternberg, 1969; see review in Kristjánsson & Egeth, 2020). A popular way of assessing mechanisms of visual cognition was to break them down into fundamental operations measured with simple stimuli, stepping away from real-life conditions that could otherwise contaminate the purity of the measurements. One aim, for example, was to study the capacity of visual attention and visual memory (Neisser, 1967). These measurements have undeniably been successful as they have provided novel insights into fundamental cognitive mechanisms (Kristjánsson & Egeth, 2020).

While studies using sparse artificial stimuli and tasks that focus only on snapshots of behavior have provided a foundation for understanding visual attention and visual memory representations, we highlight that in order to understand the functional nature of visual attention, visual representations, and visual memory, it is crucial to also investigate their quality and detail within the realm of active natural behavior (Draschkow, Kallmayer, & Nobre, 2020; Foulsham, Walker, & Kingstone, 2011; Malcolm, Groen, & Baker, 2016; Tatler, 2014; Tatler & Land, 2011).

It is rare for us, for example, to make a concerted effort to explicitly remember our visual surroundings, such as the location of the plates when visiting our friends for dinner at their new flat. Tasks relying on explicit memorization procedures are nevertheless commonly used in studies within visual cognition. In real life, we more commonly complete goal-directed behavior, such as setting the table, during which the location and identity representation of the surrounding objects is generated "on the fly." Recent work has shown that the representations generated through natural behavior are more reliable than those generated through explicit memorization (Draschkow, Wolfe, & Võ, 2014; Helbing, Draschkow, & Võ, 2020).

The well-known perceptual scientist James Gibson proposed the concept of "active perception," claiming that perception can only be understood in the context of actions relevant to the stimuli and conditions in each given case (Gibson, 1966, 1979; Nakayama & James, 1994). Measuring the "pure" capacity of concepts such as memory or attention was therefore seen as meaningless unless such mechanisms are tested in the context that they have evolved for. An experimental approach that ignores context runs the risk of facing similar problems to those structuralists such as Wundt and Titchener faced over 100 years ago in their attempts at assessing elementary sensations: putative elementary sensations may simply not exist without a context (Leahey, 1981). A good example of Gibson's approach is his "cookie cutter" experiment (Gibson, 1962). When a "cookie-cutter" was

statically pushed onto participants' palms, identification rates of its pattern were just under 50% (chance performance was 16.7%), while if the cutter was rotated both clockwise and counterclockwise in the observer's palm, recognition accuracy increased to about 95%. The important point here is that with increased interaction with the stimulation, the measured resolution of haptic perception increased, or to use another phrase, its capacity increased. Shying away from the more naturalistic manipulation would have left us uniformed about the recognition capacity that is actually available during real-world behavior. This accords well with most literature reviewed here: In addition to the measurements of pure capacity, any assessment of basic abilities must in the end be applied to the context they are used for.

Current goals and overview

Our aim in this tutorial review was to discuss selected recent research that demonstrates how the perceptual and cognitive limitations suggested by reductionistic approaches can be overcome with naturalistic stimuli and tasks. Our hope is that these examples can serve as templates from different domains that interested readers can use as inspiration for their own research.

The closer we get to natural behavior, the more we engage different effectors (such as eye or hand movements) and encounter perceptual mechanisms that are implicit, operating "for free" and "on the fly," i.e., quickly, efficiently and effortlessly. An informative placeholder example of how real-world tasks affect conclusions about how we orient within the world (Hayhoe, 2017) comes from Land and Hayhoe (2001), who tested the relation of eye and hand movements as participants performed natural everyday tasks such as making a cup of tea or a sandwich. They found a strong interaction between reach and gaze; gaze usually landed on the next object in the action sequence before any signs of manual action. They concluded that eye movements are planned into the motor pattern and lead each action. This suggests that studying one process without taking the other into account may not provide a complete picture of our visual and motor behavior.

Similar considerations may also apply to studies of how we construct representations of the visual world. These representations may differ according to how we interact with the visual world in each case, which can have implications for what is typically called visual long-term memory (VLTM). In many studies within this field, observers are asked to perform tasks that are very different from what they experience during everyday life. As such, estimates of their capacity may not tell the whole story. We review how testing long-term memory in a "pure" manner – without embracing the complexities of natural contexts – may change the conclusions we draw about how VLTM representations are formed and used.

An important consideration is that we rarely perform individual cognitive operations in isolation. We do not perform an "object recognition" task when drinking our cup of tea, nor explicitly recognize a door handle before opening a door. We *use* these objects to enjoy a pleasant meeting with a friend or get away from an unpleasant one. Moreover, recognizing objects does not end our engagement with them. Complex cognitive operations such as visual search are a means to an end, as we will likely interact with the object we have searched for, and inversely their function also guides our search (Castelhano & Witherspoon, 2016). Interactions with our environment can strongly influence the representation of our visual space and should therefore be considered when the goal is to understand factors guiding behavior (Schütz-Bosbach & Prinz, 2007). A common dichotomy in visual cognition and visual neuroscience is that different pathways are involved in recognition versus action (Mishkin, Ungerleider, & Macko, 1983), which means that the intended perceptual or physical acts are highly important when trying to understand the mechanisms behind the action (Gross & Graziano, 1995; Maravita & Iriki, 2004; Perry & Fallah, 2017). Engel, Maye, Kurthen, and König (2013) postulated that perception should be studied as an "enactive" process, which involves ongoing interactions with our surroundings, and that our representations subserve our interactions with the world. These processes should not be studied outside their role in action generation.

Capacity

A central concept in research of attention and visual memory is *capacity*, or, in other words, how much information those systems can process over a given time period, inspired by information theory (Shannon, 1948) and the concept of how many bits of information can be processed by a given system in a particular amount of time (Bush, 1936).

Research within cognitive psychology over the last 60 years or longer has clearly demonstrated that attentional capacity is limited (Á. Kristjánsson & Egeth, 2020). But how limited is it? This is a notoriously difficult question, and the first observation is that this obviously depends on how the question is asked. To take one example, in so-called multiple-object-tracking tasks, observers are not able to track more than approximately four items without losing track of some of the items (Pylyshyn & Storm, 1988). Studies of visual search have also been used to estimate capacity. Attentional capacity has been thought to be essentially unlimited as long as a search target stands out from the distractors on a single feature (Treisman, Sykes, & Gelade, 1977; Wolfe & Horowitz, 2004). But if targets can only be distinguished by a conjunction of features, individual items need to be assembled by attention, and the search items need to be inspected one-by-one, as proposed in the well-known feature integration theory (Treisman & Gelade, 1980). Search slopes have been used as an index of attentional engagement in a search task, but this approach has limitations (Kristjánsson, 2015, 2016; see also Wolfe, 2016).

Another good example of how a simplification of real-world complexities has been used to assess visual attention involves *cueing* studies (Posner, Nissen, & Ogden, 1978), which have uncovered interesting dynamics such as the influence of symbolic versus exogenous cues (Jonides, 1981) or different temporal components of attentional orienting (Nakayama & Mackeben, 1989). These studies provide information about how quickly visual attention is drawn to suddenly appearing stimuli. It is of note that although real-world applications of these concepts can certainly be envisioned, studies of actual implementation in dynamic real-world environments are scarce.

Attention and action

A theme in the current overview is that we should aim to also estimate the capacity of the mechanisms we use for interacting with the world with tasks that mimic our real-world interactions as closely as possible. When we interact with the world our hand and eye movements are *often* highly coupled, but this is not *always* the case: consider a musician reading music as she moves her fingers on a keyboard or a guitar neck, or a skilled typist, who types while keeping his gaze on the screen rather than on the keyboard. Are the attentional mechanisms used for eye and hand the same, or coupled? Coupling them can in some cases be useful when our goals for reach and gaze match, but this could be detrimental when they do not, as in the examples involving the musician and the typist. The study of attention over the last decades, however, has implicitly been investigated as a one-source concept – where gaze and reach are assumed to draw on the same resource.

Firstly, we should note that there is a lot of evidence that attention is tightly coupled to action (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1995; Kristjánsson, 2011; Montagnini & Castet, 2007). Eye and hand movements have been shown to link visual attention to the endpoints of their actions during movement preparation (Deubel, Schneider, & Paprotta, 1998; Rolfs, Lawrence, & Carrasco, 2013). This has usually been measured with some assessment of discrimination performance (Hanning, Deubel, & Szinte, 2019a) at the intended endpoint of the movement. While the exact nature of this relationship has been debated, this relation is probably by no means a necessary one (Hanning, Deubel, & Szinte, 2019a; Kristjánsson, 2011; Van der Stigchel & de Vries, 2015; Wollenberg, Deubel, & Szinte, 2018), as has been assumed for example in the influential *premotor theory of attention* (Rizzolatti, Riggio, Dascola, & Umiltá, 1987). Effects of action preparation have been found

when observers reach towards a target rather than moving their gaze towards it. Attentional performance is best just prior to the start of a reaching movement at the reach target (Baldauf & Deubel, 2010; Deubel et al., 1998; Rolfs et al., 2013). A number of authors have argued that the same attention mechanism is behind both the link between gaze and attention as well as reach and attention (Bekkering, Abrams, & Pratt, 1995; Huestegge & Adam, 2011; Khan, Song, & McPeek, 2011; Nissens & Fiehler, 2018; Song & McPeek, 2009).

Jonikaitis and Deubel (2011) asked observers to move their gaze or reach towards either the same or different locations. Surprisingly, they found that delaying eye movements delayed attention shifts to the gaze target without affecting attention at the locus of the reach target. In contrast with arguments for a unitary mechanism for selection, this suggests that eye and hand movements are selected by largely independent *effector-specific* attentional mechanisms. Even more importantly, their results suggest that the attentional benefit at one effector's movement target is not affected by the concurrent movement preparation of the other effector to another location. There is evidence from single-cell neurophysiology (Graziano & Gross, 1996; Perry, Sergio, Crawford, & Fallah, 2015) and functional neuroimaging (Makin, Holmes, & Zohary, 2007) that visual processing can be enhanced when a reach places the hand near the stimuli to be processed. Perry and Fallah (2017) propose that visual processing near the perceived position of the hand is amplified because of feedback from frontoparietal areas. Since attention must often be divided between visual and motor tasks, other effector systems such as the reach/grasp system may also cause attention-related vision enhancement, which would undoubtedly be beneficial in many scenarios. Hanning, Aagten-Murphy, and Deubel (2018) then tested whether targets for eye movements and targets for reach movements are selected by the same attentional mechanism by measuring visual sensitivity – an established "proxy" for motor selection – at the motor targets during the preparation of simultaneous eye and hand movements. They found that sensitivity at both the eye and the hand target locations was unaffected by the simultaneous movement preparation of the other effector. Observers were able to allocate attention simultaneously to two different targets for a movement, arguing for separate attentional mechanisms for the two effector systems. Perhaps even more important was the finding that the two selection mechanisms did not seem to compete for resources at any point during the movement preparation process, at least when the necessary resources can be freed up from irrelevant locations (Kreyenmeier, Deubel, & Hanning, 2020. Hanning et al. argued that the gaze and reach targets are represented in effector specific maps, consistent with the neural evidence from Perry et al. (2015), who showed that when monkeys placed a hand close to a visual target, orientation tuning was sharpened, providing evidence for an effector-based mechanism that improves processing of

features relevant to that effector (see also Graziano & Gross, 1996). While two attention systems can often appear as if they operate in unison, and such coupling may indeed be useful, they can also be dissociated (Graziano, 2001; Perry & Fallah, 2017). Consistent findings have been reported for patients with optic ataxia (Jackson, Newport, Mort, & Husain, 2005), where a patient's deficit was confined to reach movements with his right hand. Zelinsky and Bisley (2015) have in fact pointed out that map-based representations are ubiquitous throughout the brain and that there are separate salience maps for different effectors. Despite evidence for separate attentional systems, the systems are often studied in isolation in standard laboratory-based tasks used for capacity estimates where observers' body and eye movements are restricted. By using more complex tasks in which natural movements are not only allowed but also encouraged (Sauter, Stefani, & Mack, 2020), we may gain a fuller understanding of how our attentional systems cope with their own limitations.

Attentional capacity during visual foraging

Recent evidence from foraging tasks argues that attentional capacity may be higher than proposed in prominent attention theories such as the influential feature integration (Treisman & Gelade, 1980) and guided search theories (Wolfe, 1994, 2007). These theoretical accounts are based on findings from visual search tasks where response times (RTs) for conjunction targets (where the target is distinguished from distractors by a conjunction of two features) increased linearly with each added item to the display, since attention was required to inspect each conjunction item (see exchange on this in Kristjánsson, 2015, 2016; Wolfe, 2016). Surprisingly, Kristjánsson, Jóhannesson, and Thornton (2014) found that during a foraging task where observers were asked to select many targets of two types (defined by a conjunction of features) on a tablet touch-screen, some observers switched easily between the target types with very low switch costs. This goes against a basic tenet of visual attention research from the last four decades or so: that attention is needed to perform a time-consuming integration of features for each object. Instead, this suggests that two conjunction templates could be simultaneously active (for replications, see Clarke, Irons, James, Leber, & Hunt, 2020; Wolfe, Cain, & Aizenman, 2019), or rapid switching between templates in working memory (WM) is possible (more on that in the *Foraging and visual working memory* section).

Further work then revealed that this was not necessarily confined to a subset of observers; when time limits (5, 10, or 15 s) were imposed on how long observers had to forage for as many targets as they could, most people increased their frequency of switching between different target types during conjunction-based foraging (Kristjánsson, 2018). This shows

that under the appropriate task demands, performance can reach levels well above traditional capacity estimates from theoretical accounts inspired by the visual search literature. Capacity may be a more dynamic entity than often thought, and hand movements towards targets on tablet touch-screens reveal higher capacity than we would expect from traditional visual search studies. This is consistent with findings of higher attentional performance when the hands are near visual items that are used to assess attention (Abrams, Davoli, Du, Knapp, & Paull, 2008; Reed, Grubb, & Steele, 2006) and evidence that the visibility of the hands alters neural responses (Makin et al., 2007; Perry et al., 2015).

Note also that Kristjánsson, Thornton, Chetverikov, and Kristjánsson (2020b) showed that well-known set-size effects were only seen during selections of the last target during foraging and that selections preceding the last one were much faster. This demonstrates how our understanding of attentional mechanisms can change once we investigate it using a more interactive task – in this case finger foraging. We should note that proposals that there are separate neural mechanisms for hand and gaze selection (Makin et al., 2007; Perry et al., 2015) mean that finger foraging could differ from foraging with mouse movements (as tested, e.g., in Wolfe, 2013), but no study has directly compared finger foraging and mouse foraging within-participants.

Interestingly, the correlation in performance between gaze and finger foraging seems to be relatively low (Jóhannesson, Thornton, Smith, Chetverikov, & Kristjánsson, 2016; see also Tagu & Kristjánsson, 2020), and conjunction foraging is easier during gaze foraging, consistent with the proposal that the crucial mechanisms differ for gaze and finger selection, and that recruiting more effectors can help overcome limitations imposed by attention associated with a single effector. The constraints of the specific task at hand (Hayhoe & Rothkopf, 2011; Tatler, Hayhoe, Land, & Ballard, 2011) are critical and can even better account for observers' attentional allocation compared to external factors, such as visual salience. Relatedly, Robinson, Benjamin, and Irwin (2020) found that estimates of capacity from different tasks may not overlap much and cannot be estimated by common parameters. This is an important point showing that overall capacity estimates may not always generalize well across task and context. We believe that the results summarized above suggest that the answers about capacity could depend on how the questions are asked.

The capacity of visual working memory

Visual working memory (VWM) allows us to monitor our own mental representations and keep track of our goals as we interact with the visual world. Attention and VWM are strongly linked and share neural mechanisms to a considerable extent (Awh, Anllo-Vento, & Hillyard, 2000; Labar, Gitelman, Parrish, & Mesulam, 1999). Desimone and Duncan (1995) argued that WM elicits a neural signal that can bias selective attention, and in the Theory of Visual Attention (TVA) model (Bundesen & Habekost, 2012) our attentional goals are considered to be maintained in VWM.

In the past, VWM has often been studied with so-called change-detection tasks (e.g., Alvarez & Cavanagh, 2005; Luck & Vogel, 1997). Observers are shown an array of a number of visual items and asked to remember them. Shortly afterwards, following a blank screen or a mask, they are asked to judge whether a change occurred in the array or not. The aim with change-detection tasks has been to assess WM independently of other mechanisms. Another more recent approach to studying WM has involved *continuous reports* (Bays & Husain, 2008a; Zhang & Luck, 2008). There, instead of a change, a probe follows the mask and participants need to reproduce an item's orientation, color, or location in a continuous fashion – for example, by using a color wheel. Estimates of the capacity of VWM have focused on a fixed number of items (Luck & Vogel, 1997) or a certain amount of information (Alvarez & Cavanagh, 2005; Bays & Husain, 2008).

But while these tasks emulate a world in which visual information rapidly changes and items disappear, our surroundings tend to remain rather stable across adjacent time points. The visual system is, in fact, notoriously bad at detecting changes (Simons & Levin, 1997; Simons & Rensink, 2005), perhaps because it makes strong assumptions about continuity (Chetverikov, Campana, & Kristjánsson, 2017b; Cicchini & Kristjánsson, 2016; Fischer & Whitney, 2014; Kristjánsson & Ásgeirsson, 2019).

Motor action and visual working memory

VWM performance is influenced by our intended actions (van Ede, 2020). In Heuer, Crawford, and Schubö (2017), participants memorized a number of items and subsequently performed a pointing movement before their memory was tested at either the movement goal or an irrelevant location. Memory performance at intended movement goals was higher than at action-irrelevant locations, showing that like attention, visual memory can be bound to the actions we perform and to our goals.

Both eye movements (Bays & Husain, 2008b; Hanning, Jonikaitis, Deubel, & Szinte, 2016; Ohl & Rolfs, 2017) and hand movements (Hanning & Deubel, 2018; Heuer et al., 2017) have been found to enhance WM performance at their motor targets. This is not unexpected given that VWM and attention show strong overlap both functionally and in terms of neural mechanisms (Awh & Jonides, 2001; Jonikaitis & Moore, 2019). The findings of Hanning, Aagten-Murphy,
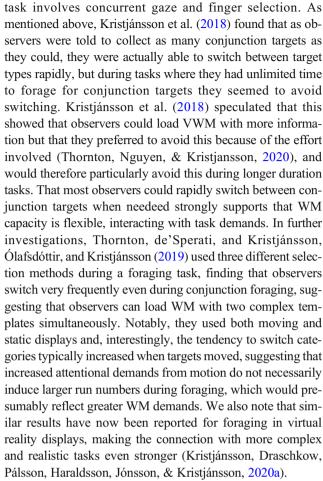
and Deubel (2018) discussed above suggest that independent mechanisms drive attention to eye and hand targets. If this is true, this raises the question of whether the two effector systems also operate separately within WM. To address this, Hanning and Deubel (2018) asked their participants to memorize several locations. Participants had to either make *single* eye or hand movements or make simultaneous eye and hand movements to two distinct memorized locations. The authors found enhanced memory at the eye and hand motor targets, with no signs of any tradeoff between the two memory processes, for gaze and reach. This shows that WM at the saccadic goal and at the reach goal can be independent of one another, and that VWM can be augmented by effector specific memory.

Further, Chetverikov et al., (2018) found that untethering hand-guided and eye-guided attention improved WM performance. In another related finding, Hanning et al. (2016) found dissociable effects of task relevance and oculomotor selection on WM. They found that task relevance on its own, without the coupled oculomotor selection, did not lead to any improvement in WM, while oculomotor selection did. Effects of task relevance and oculomotor selection on WM performance for features could be separated, in other words. These results highlight the importance of studying how VWM operates in increasingly complex behavior.

Foraging and visual working memory

Due to the close link between attention and WM, it is informative to revisit the foraging findings of Kristjánsson et al. (2014). During foraging for a given number of targets it is natural to assume that observer's attention is guided by WM templates and foraging results can therefore cast light upon the operation of VWM. A recent proposal is that while VWM can contain more than one template, only one template is accessible for attentional guidance at any given moment (Olivers, Peters, Houtkamp, & Roelfsema, 2011; Ort, Fahrenfort, & Olivers, 2017; van Moorselaar, Gunseli, Theeuwes, & Olivers, 2014). The rapid switching between target types during feature foraging (described in the section on attention above) shows how such "one template at a time" limitations could be overcome in more natural scenarios. And the rapid switching during conjunction foraging (T. Kristjánsson, Thornton, & Kristjánsson, 2018) is even more informative, since these templates would require a complex exclusion rule based on two feature dimensions (shape and color) along with very fast feature integration, yet observers seem to be able to do this. This raises the intriguing question of whether the two non-overlapping attentional systems that Hanning and Deubel (2018) found evidence for allow for higher capacity performance than the tasks used, for example, by van Moorselaar, Gunseli, Theeuwes, and Olivers (2014), since the foraging

task involves concurrent gaze and finger selection. As mentioned above, Kristjánsson et al. (2018) found that as observers were told to collect as many conjunction targets as they could, they were actually able to switch between target types rapidly, but during tasks where they had unlimited time to forage for conjunction targets they seemed to avoid switching. Kristjánsson et al. (2018) speculated that this showed that observers could load VWM with more information but that they preferred to avoid this because of the effort involved (Thornton, Nguyen, & Kristjansson, 2020), and would therefore particularly avoid this during longer duration tasks. That most observers could rapidly switch between conjunction targets when needeed strongly supports that WM capacity is flexible, interacting with task demands. In further investigations, Thornton, de'Sperati, and Kristjánsson, Ólafsdóttir, and Kristjánsson (2019) used three different selection methods during a foraging task, finding that observers switch very frequently even during conjunction foraging, suggesting that observers can load WM with two complex templates simultaneously. Notably, they used both moving and static displays and, interestingly, the tendency to switch categories typically increased when targets moved, suggesting that increased attentional demands from motion do not necessarily induce larger run numbers during foraging, which would presumably reflect greater WM demands. We also note that similar results have now been reported for foraging in virtual reality displays, making the connection with more complex and realistic tasks even stronger (Kristjánsson, Draschkow, Pálsson, Haraldsson, Jónsson, & Kristjánsson, 2020a).

Kristjánsson and Kristjánsson et al. (2018) then tested a foraging task where they varied the number of targets and distractors. They found that switch costs increased roughly linearly with the size of the memory set, consistent with load accounts of VWM. Again, this shows how active tasks can inform theoretical accounts. Foraging studies have also been used to shed new light on the development of WM, along with other executive functions (Ólafsdóttir, Gestsdóttir, & Kristjánsson, 2019, 2020).

We want to emphasize that active tasks can change how we think about VWM. Ballard, Hayhoe, Li, and Whitehead (Ballard, Hayhoe, Li, & Whitehead, 1992) asked their participants to reproduce an array of colored blocks in a certain model arrangement that was visible in an adjacent panel, picking blocks from an available pile. They expected observers to take a look at the model area, memorize the position and color of the blocks, and then place the blocks in the copy area. But instead they found that observers continually checked back and forth from the model to the copy area. In other words, they did not seem to memorize the whole area but only a small amount of information at a time. This highlights that while VWM capacity might be much higher than many estimates have posited, under more natural circumstances observers may choose to not rely on these resources (Ballard, Hayhoe,

& Pelz, 1995; Draschkow et al., 2020). This general performance pattern has also been seen in the foraging literature where observers are less likely to switch targets when given more time to forage within displays (see review in Kristjánsson et al., 2019).

The role of memory

To overcome the limits of capacity-restricted cognitive functions we can incorporate expectations and knowledge about our current behavioral context. These priors can come from different time scales (Nobre & Stokes, 2019) – from the immediate past, such as from priming, from long-term episodic and semantic representations, or from memories from an intermediate time scale, such as trial history or serial dependence effects. In this section we review some selected exemplar literature that highlights how prior knowledge can support attention and WM.

The recent past supports attention and working memory

The visual system relies on recent representations to construct current percepts as shown for example in serial dependence (J. Fischer & Whitney, 2014; Manassi, Liberman, Chaney, & Whitney, 2017; Manassi, Kristjánsson, & Whitney, 2019; Pascucci et al., 2019). Contextual information supports the integration of the recent past and present in order to enable stable percepts across time (C. Fischer et al., 2020). Critically, these recent experiences do not simply alter but can also *facilitate* perception (Cicchini, Mikellidou, & Burr, 2018). Studies of so-called feature-distribution learning (Chetverikov, Campana, & Kristjánsson, 2016, Chetverikov, Campana, & Kristjánsson, 2017c; Hansmann-Roth, Chetverikov, & Kristjánsson, Hansmann-Roth, Chetverikov, & Kristjánsson, 2019; Rafiei, Hansmann-Roth, Whitney, Kristjánsson, & Chetverikov, 2020) show that we can encode abstract details of preceding information in the environment. Priming effects in vision (Ásgeirsson, Kristjánsson, & Bundesen, 2015; Brascamp, Blake, & Kristjánsson, 2011; Maljkovic & Nakayama, 1994; for a review, see Á. Kristjánsson & Ásgeirsson, 2019) show that attention deployments are strongly determined by perceptual history. These insights highlight that we perform tasks within a temporal context and this context will in turn influence the estimates of capacity. That is, attention does not operate in a vacuum – the current event context is highly important. Curiously, the size of priming effects dwarfs many other effects such as effects of top-down attention (Á. Kristjánsson & Ásgeirsson, 2019; Kristjánsson, Wang, & Nakayama, 2002; Maljkovic & Nakayama, 1994; for a review, see Theeuwes, 2013).

Similar to attentional performance, VWM is influenced by recent experiential history (Carlisle & Kristjánsson, 2018; Cochrane, Nwabuike, Thomson, & Milliken, 2018; Cunningham & Egeth, 2016; Kristjánsson, Saevarsson, & Driver, 2013). For example, Carlisle and Kristjánsson et al. (2018) showed how priming and WM can affect one another and argued that implicit short-term memory and explicit VWM interact when they provide conflicting attentional instructions. Further, WM capacity is usually estimated by repeating the same or very similar stimuli between all trials of an experiment. This can lead to interference between consecutive trials, and in fact capacity estimates are considerably larger when such proactive interference is discouraged by using unique stimuli (Endress & Potter, 2014; Hartshorne, 2008).

Long-term representations support attention and working memory

There is rich evidence that long-term semantic (Henderson & Hayes, 2017; Torralba, Oliva, Castelhano, & Henderson, 2006; Võ & Henderson, 2010; Võ & Wolfe, 2015; Wolfe, Võ, Evans, & Greene, 2011) and episodic (Aly & Turk-Browne, 2017; Brockmole & Henderson, 2006; Chun & Jiang, 1998, 1999, 2003; Draschkow & Võ, 2016, 2017; Fan & Turk-Browne, 2016; Hutchinson & Turk-Browne, 2012; Patai, Buckley, & Nobre, 2013; Stokes, Atherton, Patai, & Nobre, 2012; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006; Võ & Wolfe, 2012) memory representations support the allocation of attention. These long-term representations are critical in enabling a seamless and continuous visual experience, because in order to overcome limitations in capacity, long-term priors extracted within the initial glimpse of an environment (Oliva, 2005) can provide clues about target appearance (Robbins & Hout, 2019) and guide attention to the most informative locations (Võ, Boettcher, & Draschkow, 2019; Wolfe et al., 2011).

With regard to WM, a striking example of how interactions between long- and short-term representation can overcome classical WM capacity limits is provided by Endress and Potter (2014). They tested capacity for briefly presented *familiar* objects (at 4 or 8 Hz) with probes following the stream of stimuli, finding very high WM capacity (up to 30 items) when all the objects were *unique* throughout the experiment (avoiding proactive interference), while if the items were recycled across trials within the experiment, capacity estimates were much lower. Since many WM experiments depend on repeating stimuli (such as colored squares, oriented bars, or locations on the screen), this finding suggests that proactive interference may explain at least some of the limits in WM capacity traditionally found in the literature.

Brady and Störmer (2020) measured VWM with real-world objects and stripped-down single-feature colored

stimuli. Their results suggest that VWM performance depends highly on whether single feature objects are used or whether real-world objects are used – not only did they find the benefit for meaningful objects but also the real-world objects benefitted from sequential presentation, which the colored patches did not – suggesting that the encoding for the two different stimuli-types may differ (Brady, Störmer, & Alvarez, 2016).

Finally, hybrid search tasks have demonstrated the remarkable efficiency of searching through visual space for any one of up to 100 targets held in memory (Wolfe, 2012). In these tasks, participants memorize upwards of 100 objects during a learning session and subsequently perform visual searches for these items amongst visual arrays of novel distractor objects. While visual search times increase linearly with an increase in items in the visual display, searching within memory for visual familiar objects (Wolfe, 2012) or words (Boettcher & Wolfe, 2015) increases logarithmically. Critically, searching for the groceries on your shopping list does not require you to search through the entire supermarket for as many times as the number of items on your list. Instead, you go through it once and perform many memory searches "on the fly" (Drew, Boettcher, & Wolfe, 2017). So, while WM memory might have limited capacity for holding the attentional template that is relevant to the current search, hybrid search results demonstrate that we can search for a number of targets with astonishing efficiency, even more so for if the objects of our search are familiar, instead of novel (Madrid, Cunningham, Robbins, & Hout, 2019).

Building and using behaviorally optimal long-term representations

While the earlier sections focused on cognitive systems which are renowned for their limitations, visual long-term memory (VLTM) is famously boundless. Early studies of VLTM showed remarkably large storage capacity as observers could determine if they had seen one of two images with over 80% accuracy even after viewing 10,000 scenes (Standing, 1973). In addition to capacity, studies have provided evidence for high VLTM detail (Brady, Konkle, Alvarez, & Oliva, 2008; Cunningham, Yassa, & Egeth, 2015; Draschkow et al., 2014; Konkle, Brady, Alvarez, & Oliva, 2010a, 2010b) and longevity (Hollingworth, 2004, 2006; Hollingworth & Henderson, 2002; Konkle et al., 2010b). Quite remarkably, even when related studies with alternative retrieval tests indicated more modest VLTM capacity, there was still no significant drop in the detail of existing representations (Cunningham et al., 2015).

In the previous section we highlighted that long-term representations are critical for the efficient guidance of attention and WM. While there is debate about the representational format of this vast storage of information, here we turn to the question of how the visual system utilizes and forms

VLTMs – as not all representations are built in the same way. Memory representations of our surroundings are closely determined by what we have seen, but also by what we have attended. Memory performance is, for example, predicted by how long (Hollingworth & Henderson, 2002) and how often we fixate an object (Tatler, Gilchrist, & Land, 2005; Tatler & Tatler, 2013). For the current topic, it is important to note that task-relevant objects are remembered better than irrelevant ones (Castelhano & Henderson, 2005; Maxcey-Richard & Hollingworth, 2013; Williams, Henderson, & Zacks, 2005) and memory representations strongly interact with behavioral goals (Droll & Hayhoe, 2007; Droll, Hayhoe, Triesch, & Sullivan, 2005; Triesch, Ballard, Hayhoe, & Sullivan, 2003), which becomes particularly evident in natural behavior (Tatler & Land, 2011).

In parallel with the previous sections, we emphasize the importance of studying active natural behavior (Draschkow et al., 2020; Foulsham et al., 2011; Malcolm et al., 2016; Tatler, 2014; Tatler et al., 2011) and how VLTMs are generated as a natural by-product of interactions with the environment (Draschkow & Võ, 2017; Helbing et al., 2020), as these representations support seamless everyday activities. In comparison to memory investigations in which *memorization* is the explicit task, during ecological behavior it is not necessary to constantly instruct ourselves to remember everything in our surroundings. In fact, an ever-growing body of literature provides strong evidence that very reliable representations are formed after incidental encoding during search (Castelhano & Henderson, 2005; Draschkow et al., 2014; Draschkow & Võ, 2016; Hout & Goldinger, 2010, 2012; Howard, Pharaon, Körner, Smith, & Gilchrist, 2011; Olejarczyk, Luke, & Henderson, 2014; Võ & Wolfe, 2012), change detection (Utochkin & Wolfe, 2018), visual discrimination (Draschkow, Reinecke, Cunningham, & Võ, 2018), or object manipulation (Draschkow & Võ, 2017; Kirtley & Tatler, 2015). Draschkow et al. (2018) investigated the capacity and detail of incidental memory, instructing participants to detect visually distorted objects among a stream of intact objects (the incidental analogue to the explicit studies of Brady et al., 2008, and Cunningham et al., 2015). In a subsequent surprise recognition memory test, they found that even after very brief exposures to thousands of isolated objects, incidental memory was above chance. Another example of incidental memory being more robust than one might intuitively assume is Pinto, Papesh, and Hout's (2020) visual search study. They employed a challenging surprise memory test that probed incidental object representations by showing participants up to as many as 16 possible alternatives (e.g., "which of these 16 butterflies did you see while searching?"). Using a quantification of object similarity via multidimensional scaling ratings, the study provides evidence that even under very adverse conditions perceptual details are being retained following incidental encoding.

Not only do we seem to be able to generate strong incidental representations, but the memories we have gathered on the fly, during natural interactions, might in fact be critical for proactively guiding our behavior. Chetverikov, Campana, and Kristjánsson (2017a) have shown how repeated searching within search arrays with particular feature distributions of orientation or color (Chetverikov et al., 2017c; Tanrikulu, Chetverikov, & Kristjánsson, 2020) enables observers to learn the probabilities of feature values and build up a probabilistic template of the set for distractor rejection (Chetverikov, Campana, & Kristjánsson, 2020a). Using a repeated-search task, Võ and Wolfe (2012) demonstrated that attentional guidance by memories from previous encounters was more effective if these memories were established when *looking for* an item (during search), compared to *looking at* targets (explicit memorization and free viewing). The task at hand is critical for the information that gets extracted from fixations in real-world environments (Tatler et al., 2013). Further, search for objects is speeded if these objects have been incidentally fixated on preceding trials both in real (Draschkow & Võ, 2016) and virtual (Draschkow & Võ, 2017) environments.

The more naturalistic a task becomes; the more incidental representations gain strength. Object handling improves the speed of subsequent object recognition over passively viewed objects (Harman, Humphrey, & Goodale, 1999; James et al., 2002). Locations were recalled better when participants made active hand movements to them compared to when the hand was passively moved (Trewartha, Case, & Flanagan, 2015). Search within naturalistic images created more robust memories for the identity of target objects than representations formed as a result of explicit memorization (Draschkow et al., 2014; Josephs, Draschkow, Wolfe, & Võ, 2016). During immersive searches in virtual reality this *search superiority* even leads to more reliable incidentally generated spatial representations when compared to memories formed under explicit instruction to memorize (Helbing et al., 2020). Critically, incidental encoding seems to strongly rely on the availability of meaningful scene semantics in the stimulus materials used (Draschkow et al., 2014; Võ et al., 2019). The search superiority effect is diminished when no semantic contextual information is provided (Draschkow et al., 2014) or participants are not given enough time to associate the context with the target (Josephs, Draschkow, Wolfe, & Võ, 2016), although it is of note that the memory representations of items searched for was no worse than those explicitly memorized even in the absence of scene semantics.

In natural behavior, new information is easily integrated with prior knowledge, as we rarely encounter items that are "new enough" to require integration effort. It is thus sensible to incorporate this information incidentally and on the fly, instead of trying to "force" new memories in explicitly. Virtual reality paves the way for studies in realistic and unconstrained task settings that can probe such dynamics, while maintaining a high degree of experimental control (David, Beitner, & Võ, 2020; Draschkow et al., 2020; Draschkow & Võ, 2017; Figueroa, Arellano, & Calinisan, 2018; Kit et al., 2014; Li, Aivar, Kit, Tong, & Hayhoe, 2016; Li, Aivar, Tong, & Hayhoe, 2018; Olk, Dinu, Zielinski, & Kopper, 2018).

Summary and general conclusions

While reductionist approaches are a cornerstone of empirical research in cognition, there are definite limits to studying real-world vision in artificially stripped-down settings. Our aim was to review selected recent findings that showcase how the basic mechanisms of visual attention and working and long-term memory operate within a framework that embraces various real-world complexities, and to highlight how such real-world paradigms can be used to inform our ideas about visual cognition.

Implications for visual attention and working memory

The evidence presented here suggests that attention may in fact be intrinsically bound to the involved effectors, as indicated by the results of dissociating reach and gaze (Chetverikov et al., 2018; Hanning et al., 2016) as well as supporting neural evidence (Gross & Graziano, 1995; Perry & Fallah, 2017). It is important to note that our aim with this claim is not to restate the well-known premotor theory of attention (Rizzolatti et al., 1987). Our claim is simply that any "pure" measurements of capacity may not encapsulate how we attend (in a general sense) in more natural tasks. Results from visual foraging tasks (Kristjánsson et al., 2014; 2020) indicate that capacity limitations can be flexibly circumvented (Kristjánsson et al., 2018; Thornton et al., 2020), and that, if needed, observers seem to behave as if they have higher capacity, but only when this is necessary due to task constraints – perhaps because of the effort involved in loading WM (Ballard et al., 1995; Draschkow et al., 2020).

We also note that attention has been thought to operate on priority maps (Itti & Koch, 2001; Koch & Ullman, 1985). Our review raises the possibility that different priority maps may exist for different action effectors, as argued by Zelinsky and Bisley (2015). This is also in line with the claims of Perry and Fallah (2017) that there are specific attentional mechanisms for each effector (such as the eye vs. the hand). Note that if separate priority maps do indeed exist, this would make the idea of pure context-free capacity suspect if each effector has its own attentional prioritization mechanism. This would, on the other hand, allow considerable flexibility. Such arrangements could, for example, enable our musician from the *Introduction* to efficiently move her fingers across the piano

keys while she independently keeps her gaze on the sheet music in front of her.

Considering the limitations in WM capacity, it is highly interesting that during combined eye and hand movements, memory can be improved at two different locations simultaneously at little or no cost (Chetverikov et al., 2018; Hanning & Deubel, 2018). In other words, WM capacity is higher when two effectors are simultaneously recruited. This calls for a different conception of WM that includes the actions we perform (Myers, Stokes, & Nobre, 2017; van Ede, 2020), and argues that putative capacity estimates need careful consideration when they are extended to natural behavior.

Finally, the limitations of attention and WM can often be overcome by incorporating representations from the recent and distant past (Nobre & Stokes, 2019). Reducing the proactive interference in the experimental approach can substantially improve estimates of capacity (Endress & Potter, 2014; Hartshorne, 2008). Testing WM performance with real objects strongly improves performance, and therefore capacity (Brady & Störmer, 2020; Endress & Potter, 2014). We also note that we can search for hundreds of targets with astonishing efficiency (Boettcher & Wolfe, 2015; Wolfe, 2012). That is, attention and WM do not operate in a vacuum, and investigations of how the past supports the future are important for our understanding of how we perform tasks efficiently despite the limits of "pure" attention and WM.

Implications for long-term memory

Natural tasks in screen-based, real or virtual settings can reveal how long-term memory representations are formed via implicit learning mechanisms. A surprising amount of information about the environment may be picked up for free (Castelhano & Henderson, 2005; Williams, 2010). Moreover, incidental memories generated through natural behavior may be more robust than those picked up during explicit memorization (Draschkow et al., 2014; Helbing et al., 2020; Josephs et al., 2016; Võ & Wolfe, 2012). In other words, when people interact with the environment, more information is accumulated than when people perform more artificial memory tasks. This highlights the importance of understanding more natural encoding conditions, as they might deviate from estimates from traditional tasks.

How might these long-term memory representations – which are so important for guiding behavior – be formed? One clue may come from recent studies of so-called feature distribution learning (for a review, see Chetverikov, Hansmann-Roth, Tanrıkulu, & Kristjánsson, 2020b). Chetverikov and colleagues (Chetverikov, Campana, & Kristjánsson, 2016, 2017a) used a novel method to investigate whether observers can encode the shape of a probability density function of distractor distributions in odd-one-out visual search tasks for orientation and color. Instead of using explicit judgments of distribution statistics, they measured observers' visual search times, which revealed observers' expectations of distractor distributions. They used slowing effects from role-reversals between the target and distractors, which occur when feature values of target and distractors used on previous search trials are swapped on the next (Á. Kristjánsson & Driver, 2008; Lamy, Antebi, Aviani, & Carmel, 2008; Wang, Kristjansson, & Nakayama, 2005). They found that observers were able to encode the statistics of feature distributions in surprising detail – much more detail than previous studies have indicated (Alvarez, 2011; Cohen, Dennett, & Kanwisher, 2016). Critically, this learning is implicit, and does not require the explicit report of the properties of the stimuli, but can nevertheless guide action (Hansmann-Roth et al., 2020). Testing such feature distribution learning in more realistic settings, including virtual reality environments, might therefore be of great value in future.

Conclusions

In this tutorial review we have provided several examples of how capacity limitations in visual cognition are overcome when attention, action, and memory cooperate, and have attempted to give examples of how such studies were implemented. Attention and memory may be intrinsically bound to the involved effectors, and our discussion highlights how long-term representations can provide the framework in which limitations might go unnoticed. Finally, natural tasks can establish representations incidentally, which subsequently become usable for proactive guidance. Taken together, we highlight the importance of investigating basic cognitive mechanisms as they unfold in increasingly complex behavior.

Note that we do not wish to claim that pure measurements are in any sense wrong – they have led to milestone discoveries concerning visual cognition. But measurements from such approaches should preferably be made to pass the real-world test; shown to apply to real-world settings as soon as possible. Yet a troubling possibility is that what we have called "pure" capacity measurements have little practical application. In some cases, they *may not exist* outside the paradigms that are used to measure them. This conclusion is probably unnecessarily pessimistic, however. A more constructive one could be that findings from reductionistic approaches generalize well, but concepts about visual mechanisms should be tested in more naturalistic conditions involving stimuli and tasks that do justice to the actual complexity of natural interactions.

References

Abrams, R. A., Davoli, C. C., Du, F., Knapp, W. H., & Paull, D. (2008). Altered vision near the hands. *Cognition*, *107*(3), 1035–1047. https://doi.org/10.1016/j.cognition.2007.09.006

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643. https://doi.org/10.1111/j.1467-9280.2005.01587.x

Aly, M., & Turk-Browne, N. B. (2017). How hippocampal memory shapes, and is shaped by, attention. In: *The hippocampus from cells to systems: structure, connectivity, and functional contributions to memory and flexible cognition* (pp. 369–403). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50406-3_12

Ásgeirsson, Á. G., Kristjánsson, Á., & Bundesen, C. (2015). Repetition priming in selective attention: A TVA analysis. *Acta Psychologica*, *160*, 35–42. https://doi.org/10.1016/j.actpsy.2015.06.008

Awh, E., Anllo-Vento, L., & Hillyard, S. A. (2000). The role of spatial selective attention in working memory for locations: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *12*(5), 840–847. https://doi.org/10.1162/089892900562444

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation - Advances in Research and Theory*, *8*(C), 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Baldauf, D., & Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Research*, *50*(11), 999–1013. https://doi.org/10.1016/j.visres.2010.02.008

Ballard, D., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *337*(1281). https://doi.org/10.1098/rstb.1992.0111

Ballard, D., 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

Bays, P. M., & Husain, M. (2008a). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Bays, P. M., & Husain, M. (2008b). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Bekkering, H., Abrams, R. A., & Pratt, J. (1995). Transfer of saccadic adaptation to the manual motor system. *Human Movement Science*, *14*(2), 155–164. https://doi.org/10.1016/0167-9457(95)00003-B

Boettcher, S. E. P., & Wolfe, J. M. (2015). Searching for the right word: Hybrid visual and memory search for words. *Attention, Perception & Psychophysics*, *77*(4), 1132–1142. https://doi.org/10.3758/s13414-015-0858-9

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, *113*(27), 7459–7464. https://doi.org/10.1073/pnas.1520027113

Brady, T., & Störmer, V. (2020). The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing. https://doi.org/10.31234/osf.io/kzvdg

Brascamp, J. W., Blake, R., & Kristjánsson, Á. (2011). Deciding Where to Attend: Priming of Pop-Out Drives Target Selection. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1700–1707. https://doi.org/10.1037/a0025636

Broadbent, D. E. (2004). *Perception and communication. Perception and communication.* Pergamon Press. https://doi.org/10.1037/10037-000

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Bundesen, C., & Habekost, T. (2012). *Principles of Visual Attention: Linking Mind and Brain. Principles of Visual Attention: Linking Mind and Brain.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570707.001.0001

Bush, V. (1936). Instrumental analysis. *Bulletin of the American Mathematical Society*, *42*(10), 649–669. https://doi.org/10.1090/S0002-9904-1936-06390-1

Carlisle, N. B., & Kristjánsson, Á. (2018). How visual working memory contents influence priming of visual attention. *Psychological Research*, *82*(5), 833–839. https://doi.org/10.1007/s00426-017-0866-6

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition: Special Issue on Real-World Scene Perception*, (12), 1017–1040.

Castelhano, M. S., & Witherspoon, R. L. (2016). How You Use It Matters: Object Function Guides Attention During Visual Search in Scenes. *Psychological Science*, *27*(5), 606–621. https://doi.org/10.1177/0956797616629130

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). 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. https://doi.org/10.1016/BS.PBR.2017.07.001

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017b). Rapid learning of visual ensembles. *Journal of Vision*, *17*(2), 21. https://doi.org/10.1167/17.2.21

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017c). Representing Color Ensembles. *Psychological Science*, *28*(10), 1510–1517. https://doi.org/10.1177/0956797617713787

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2020a). Probabilistic rejection templates in visual working memory. *Cognition*, *196*, 104075. https://doi.org/10.1016/j.cognition.2019.104075

Chetverikov, A., Hansmann-Roth, S., Tanrıkulu, Ö. D., & Kristjánsson, Á. (2020b). Feature distribution learning (FDL): A new method for studying visual ensembles perception with priming of attention shifts. In *Neuromethods* (Vol. 151, pp. 37–57). Humana Press Inc. https://doi.org/10.1007/7657_2019_20

Chetverikov, A., Kuvaldina, M., MacInnes, W. J., Jóhannesson, Ó. I. & Kristjánsson, Á. (2018). Implicit processing during change blindness revealed with mouse-contingent and gaze-contingent displays. *Attention, Perception & Psychophysics, 80*, 844–859

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (1999). Top-Down Attentional Guidance Based on Implicit Learning of Visual Covariation. *Psychological Science*, *10*(4), 360–365. https://doi.org/10.1111/1467-9280.00168

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(2), 224–234. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12696811

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1890), 20181722. https://doi.org/10.1098/rspb.2018.1722

Clarke, A. D., Irons, J. L., James, W., Leber, A. B., & Hunt, A. R. (2020). Stable individual differences in strategies within, but not between, visual search tasks. *Quarterly Journal of Experimental Psychology (2006)*, 1747021820929190. https://doi.org/10.1177/1747021820929190

Cochrane, B. A., Nwabuike, A. A., Thomson, D. R., & Milliken, B. (2018). An imagery-induced reversal of intertrial priming in visual search. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(4), 572–587. https://doi.org/10.1037/xlm0000470

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences*, *20*(5), 324–335. https://doi.org/10.1016/j.tics.2016.03.006

Cunningham, C. A., & Egeth, H. E. (2016). Taming the White Bear: Initial Costs and Eventual Benefits of Distractor Inhibition. *Psychological Science*, *27*(4), 476–485. https://doi.org/10.1177/0956797615626564

Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations on storage capacity for object details in visual long-term memory. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *22*(11), 563–566. https://doi.org/10.1101/lm.039404.115

David, E., Beitner, J., & Võ, M. L. H. (2020). Effects of transient loss of vision on head and eye movements during visual search in a virtual environment. *Brain Sciences*, *10*(11), 1–26. https://doi.org/10.3390/brainsci10110841

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

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

Deubel, H., Schneider, W. X., & Paprotta, I. (1998). Selective dorsal and ventral processing: Evidence for a common attentional mechanism in reaching and perception. *Visual Cognition*, *5*(1–2), 81–107. https://doi.org/10.1080/713756776

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2020). When Natural Behavior Engages Working Memory. *Current Biology*. https://doi.org/10.1016/j.cub.2020.11.013

Draschkow, D., Reinecke, S., Cunningham, C. A., & Võ, M. L.-H. (2018). The lower bounds of massive memory: Investigating memory for object details after incidental encoding. *Quarterly Journal of Experimental Psychology*, 174702181878372. https://doi.org/10.1177/1747021818783722

Draschkow, D., & Võ, M. L.-H. (2016). Of "what" and "where" in a natural search task: Active object handling supports object location memory beyond the object's identity. *Attention, Perception, & Psychophysics*, *78*(6), 1574–1584. https://doi.org/10.3758/s13414-016-1111-x

Draschkow, D., & Võ, M. L.-H. L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports*, *7*(1), 16471. https://doi.org/10.1038/s41598-017-16739-x

Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2014). Seek and you shall remember: scene semantics interact with visual search to build better memories. *Journal of Vision*, *14*(8), 10. https://doi.org/10.1167/14.8.10

Drew, T., Boettcher, S. E. P., & Wolfe, J. M. (2017). One visual search, many memory searches: An eye-tracking investigation of hybrid search. *Journal of Vision*, *17*(11). https://doi.org/10.1167/17.11.5

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(6), 1352–1365. https://doi.org/10.1037/0096-1523.33.6.1352

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1416–1438. https://doi.org/10.1037/0096-1523.31.6.1416

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology. General*, *143*(2), 548–565. https://doi.org/10.1037/a0033934

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17*(5), 202–209. https://doi.org/10.1016/j.tics.2013.03.006

Fan, J. E., & Turk-Browne, N. B. (2016). Incidental biasing of attention from visual long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 970–977. https://doi.org/10.1037/xlm0000209

Figueroa, J. C. M., Arellano, R. A. B., & Calinisan, J. M. E. (2018). A comparative study of virtual reality and 2D display methods in visual search in real scenes. In D. N. Cassenti (Ed.), *Advances in Human Factors in Simulation and Modeling* (pp. 366–377). Cham: Springer International Publishing.

Fischer, C., Czoschke, S., Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2020). Context information supports serial dependence of multiple visual objects across memory episodes. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-15874-w

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Gibson, J. J. (1966). The senses considered as perceptual systems. Houghton Mifflin.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Graziano, M. S. A. (2001). Is reaching eye-centered, body-centered, hand-centered, or a combination? *Reviews in the Neurosciences*. https://doi.org/10.1515/REVNEURO.2001.12.2.175

Graziano, M. S. A., & Gross, C. G. (1996). Multiple pathways for processing visual space. *Attention and Performance*.

Gross, C. G., & Graziano, M. S. A. (1995). REVIEW : Multiple Representations of Space in the Brain. *The Neuroscientist*, *1*(1), 43–50. https://doi.org/10.1177/107385849500100107

Guevara Pinto, J. D., Papesh, M. H., & Hout, M. C. (2020). The detail is in the difficulty: Challenging search facilitates rich incidental object encoding. *Memory and Cognition*, *48*(7), 1214–1233. https://doi.org/10.3758/s13421-020-01051-3

Hanning, N. M., Aagten-Murphy, D., & Deubel, H. (2018). Independent selection of eye and hand targets suggests effector-specific attentional mechanisms. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-27723-4

Hanning, N. M., & Deubel, H. (2018). Independent effects of eye and hand movements on visual working memory. *Frontiers in Systems Neuroscience*, *12*. https://doi.org/10.3389/fnsys.2018.00037

Hanning, N. M., Deubel, H., & Szinte, M. (2019a). Sensitivity measures of visuospatial attention. *Journal of Vision*, *19*(12), 1–13. https://doi.org/10.1167/19.12.17

Hanning, N. M., Jonikaitis, D., Deubel, H., & Szinte, M. (2016). Oculomotor selection underlies feature retention in visual working memory. *Journal of Neurophysiology*, *115*(2), 1071–1076. https://doi.org/10.1152/jn.00927.2015

Hanning, N. M., Szinte, M., & Deubel, H. (2019b). Visual attention is not limited to the oculomotor range. *Proceedings of the National Academy of Sciences*, 201813465. https://doi.org/10.1073/pnas.1813465116

Hansmann-Roth, S., Chetverikov, A., & Kristjánsson, Á. (2019). Representing color and orientation ensembles: Can observers learn multiple feature distributions? *Journal of Vision*, *19*(9), 2–2. https://doi.org/10.1167/19.9.2

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*(22), 1315–1318. https://doi.org/10.1016/S0960-9822(00)80053-6

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS ONE*, *3*(7). https://doi.org/10.1371/journal.pone.0002716

Hayhoe, M. M. (2017). Vision and Action. *Annual Review of Vision Science*, *3*(1), 389–413. https://doi.org/10.1146/annurev-vision-102016-061437

Hayhoe, M. M., & Rothkopf, C. A. (2011). Vision in the natural world. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(2), 158–166. https://doi.org/10.1002/wcs.113

Helbing, J., Draschkow, D., & Võ, M. L.-H. (2020). Search superiority: Goal-directed attentional allocation creates more reliable incidental identity and location memory than explicit encoding in naturalistic virtual environments. *Cognition*, *196*, 104147. https://doi.org/10.1016/j.cognition.2019.104147

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*(10), 743–747. https://doi.org/10.1038/s41562-017-0208-0

Heuer, A., Crawford, J. D., & Schubö, A. (2017). Action relevance induces an attentional weighting of representations in visual working memory. *Memory and Cognition*, *45*(3), 413–427. https://doi.org/10.3758/s13421-016-0670-3

Hollingworth, A. (2004). Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(3), 519–537. https://doi.org/10.1037/0096-1523.30.3.519

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(1), 58–69. https://doi.org/10.1037/0278-7393.32.1.58

Hollingworth, A., & Henderson, J. (2002). Accurate Visual Memory for Previously Attended Objects in Natural Scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113–136.

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, *72*(5), 1267–1282. https://doi.org/10.3758/APP.72.5.1267

Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: evidence from eye movements. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(1), 90–112. https://doi.org/10.1037/a0023894

Howard, C. J., Pharaon, R. G., Körner, C., Smith, A. D., & Gilchrist, I. D. (2011). Visual search in the real world: evidence for the formation of distractor representations. *Perception*, *40*(10), 1143–1153. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22308885

Huestegge, L., & Adam, J. J. (2011). Oculomotor interference during manual response preparation: Evidence from the response-cueing paradigm. *Attention, Perception, and Psychophysics*, *73*(3), 702–707. https://doi.org/10.3758/s13414-010-0051-0

Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: control from multiple memory systems. *Trends in Cognitive Sciences*, *16*(12), 576–579. https://doi.org/10.1016/j.tics.2012.10.003

Itti, Laurent, & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Jackson, S. R., Newport, R., Mort, D., & Husain, M. (2005). Where the eye looks, the hand follows: Limb-dependent magnetic misreaching in optic ataxia. *Current Biology*, *15*(1), 42–46. https://doi.org/10.1016/j.cub.2004.12.063

James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *34*(3), 383–390. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12395554

Jóhannesson, Ó. I., Thornton, I. M., Smith, I. J., Chetverikov, A., & Kristjánsson, Á. (2016). Visual foraging with fingers and eye gaze. *I-Perception*, *7*(2), 1–18. https://doi.org/10.1177/2041669516637279

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Attention and Performance*, 187–203. Retrieved from https://ci.nii.ac.jp/naid/20001365380

Jonikaitis, D., & Deubel, H. (2011). Independent allocation of attention to eye and hand targets in coordinated eye-hand movements. *Psychological Science*, *22*(3), 339–347. https://doi.org/10.1177/0956797610397666

Jonikaitis, D., & Moore, T. (2019). The interdependence of attention, working memory and gaze control: behavior and neural circuitry. *Current opinion in psychology*. Elsevier B.V. https://doi.org/10.1016/j.copsyc.2019.01.012

Josephs, E. L. E. L., Draschkow, D., Wolfe, J. M. J. M., & Võ, M. L.-H. M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica*, *169*, 100–108. https://doi.org/10.1016/j.actpsy.2016.05.013

Khan, A. Z., Song, J.-H., & McPeek, R. M. (2011). The eye dominates in guiding attention during simultaneous eye and hand movements. *Journal of Vision*, *11*(1), 9–9. https://doi.org/10.1167/11.1.9

Kirtley, C., & Tatler, B. W. (2015). Priorities for representation: Task settings and object interaction both influence object memory. *Memory & Cognition*. https://doi.org/10.3758/s13421-015-0550-2

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D., & Hayhoe, M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PloS One*, *9*(4), e94362. https://doi.org/10.1371/journal.pone.0094362

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227. https://doi.org/10.1007/978-94-009-3833-5_5

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology. General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*. https://doi.org/10.1016/0042-6989(94)00279-U

Kreyenmeier, P., Deubel, H., & Hanning, N. (2020). Theory of Visual Attention (TVA) in Action: Assessing Premotor Attention in Simultaneous Eye-Hand Movements. *BioRxiv*, 2020.01.08.898932. https://doi.org/10.1101/2020.01.08.898932

Kristjánsson, Á. (2011). The intriguing interactive relationship between visual atttention and saccadinc eye movements. In *Oxford handbook of eye movements*. https://doi.org/10.1002/0470018860.s00612

Kristjánsson, Á. (2015). Reconsidering Visual Search. *I-Perception*, *6*(6), 204166951561467.

Kristjánsson, Á. (2016). The slopes remain the same: Reply to Wolfe (2016). *I-Perception*, *7*(6), 2041669516673383.

Kristjánsson, Á., & Ásgeirsson, Á. G. (2019). Attentional priming: recent insights and current controversies. *Current opinion in psychology*. Elsevier B.V. https://doi.org/10.1016/j.copsyc.2018.11.013

Kristjánsson, Á., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, *48*(10), 1217–1232. https://doi.org/10.1016/j.visres.2008.02.007

Kristjánsson, Á., & Egeth, H. (2020). How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics. *Attention, perception, and psychophysics*, *82*(1), 7–23. https://doi.org/10.3758/s13414-019-01803-7

Kristjánsson, Á., Jóhannesson, Ó. I., & Thornton, I. M. (2014). Common Attentional Constraints in Visual Foraging. *PLoS ONE*, *9*(6), e100752. https://doi.org/10.1371/journal.pone.0100752

Kristjánsson, Á., Ólafsdóttir, I. M., & Kristjánsson, T. (2019). Visual foraging tasks provide new insights into the orienting of visual attention: Methodological considerations. In *Neuromethods* (Vol. 151, pp. 3–21). Humana Press Inc. https://doi.org/10.1007/7657_2019_21

Kristjánsson, Á., Saevarsson, S., & Driver, J. (2013). The boundary conditions of priming of visual search: From passive viewing through task-relevant working memory load. *Psychonomic Bulletin and Review*, *20*(3), 514–521. https://doi.org/10.3758/s13423-013-0375-6

Kristjánsson, Á., Wang, D. L., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, *85*(1), 37–52. https://doi.org/10.1016/S0010-0277(02)00074-4

Kristjánsson, T., Draschkow, D., Pálsson, Á., Haraldsson, D., Jónsson, P. Ö., & Kristjánsson, Á. (2020a). Moving foraging into 3D: Feature versus conjunction-based foraging in virtual reality. *Quarterly Journal of Experimental Psychology*, 174702182093702. https://doi.org/10.1177/1747021820937020

Kristjánsson, T., Thornton, I. M., Chetverikov, A., & Kristjánsson, Á. (2020b). Dynamics of visual attention revealed in foraging tasks. *Cognition*, *194*, 104032. https://doi.org/10.1016/j.cognition.2019.104032

Kristjánsson, T., Thornton, I. M., & Kristjánsson, Á. (2018). Time limits during visual foraging reveal flexible working memory templates. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 827–835. https://doi.org/10.1037/xhp0000517

Labar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional MRI comparison within subjects. *NeuroImage*, *10*(6), 695–704. https://doi.org/10.1006/nimg.1999.0503

Lamy, D., Antebi, C., Aviani, N., & Carmel, T. (2008). Priming of Pop-out provides reliable measures of target activation and distractor inhibition in selective attention. *Vision Research*, *48*(1), 30–41. https://doi.org/10.1016/j.visres.2007.10.009

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*(25–26), 3559–3565. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11718795

Leahey, T. H. (1981). The mistaken mirror: On Wundt'S and Titchener'S psychologies. *Journal of the History of the Behavioral Sciences*, *17*(2), 273–282.

Li, C. L., Aivar, M. P., Kit, D. M., Tong, M. H., & Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision*, *16*(8), 9. https://doi.org/10.1167/16.8.9

Li, C. L., Aivar, M. P., Tong, M. H., & Hayhoe, M. M. (2018). Memory shapes visual search strategies in large-scale environments. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-22731-w

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–284. https://doi.org/10.1038/36846

Madrid, J., Cunningham, C. A., Robbins, A., & Hout, M. C. (2019). You're looking for what? Comparing search for familiar, nameable objects to search for unfamiliar, novel objects. *Visual Cognition*, *27*(1), 8–20. https://doi.org/10.1080/13506285.2019.1577318

Makin, T. R., Holmes, N. P., & Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *Journal of Neuroscience*, *27*(4), 731–740. https://doi.org/10.1523/JNEUROSCI.3653-06.2007

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

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672. https://doi.org/10.3758/BF03209251

Manassi, M., Kristjánsson, Á., & Whitney, D. (2019). Serial dependence in a simulated clinical visual search task. *Scientific reports, 9*(1), 1–10

Manassi, M., Liberman, A., Chaney, W., & Whitney, D. (2017). The perceived stability of scenes: Serial dependence in ensemble representations. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-02201-5

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2003.12.008

Maxcey-Richard, A. M., & Hollingworth, A. (2013). The strategic retention of task-relevant objects in visual working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(3), 760–772. https://doi.org/10.1037/a0029496

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*. Elsevier Current Trends. https://doi.org/10.1016/0166-2236(83)90190-X

Montagnini, A., & Castet, E. (2007). Spatiotemporal dynamics of visual attention during saccade preparation: Independence and coupling between attention and movement planning. *Journal of Vision*, *7*(14), 8–8. https://doi.org/10.1167/7.14.8

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*, *21*(6), 449–461. https://doi.org/10.1016/J.TICS.2017.03.010

Nakayama, K. James J. (1994). Gibson-An Appreciation. *Psychological Review*, *101*(2), 329–335. https://doi.org/10.1037/0033-295x.101.2.329

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11), 1631–1647. https://doi.org/10.1016/0042-6989(89)90144-2

Neisser, U. (1963). Decision-Time without Reaction-Time: Experiments in Visual Scanning. *The American Journal of Psychology*, 76(3), 376.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Nissens, T., & Fiehler, K. (2018). Saccades and reaches curve away from the other effector's target in simultaneous eye and hand movements. *Journal of Neurophysiology*, 119(1), 118–123. https://doi.org/10.1152/jn.00618.2017

Nobre, A. C. (Kia), & Stokes, M. G. (2019). Premembering Experience: A Hierarchy of Time-Scales for Proactive Attention. *Neuron*, 104(1), 132–146. https://doi.org/10.1016/j.neuron.2019.08.030

Ohl, S., & Rolfs, M. (2017). Saccadic eye movements impose a natural bottleneck on visual short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43(5), 736–748. https://doi.org/10.1037/xlm0000338

Ólafsdóttir, I. M., Gestsdóttir, S., & Kristjánsson, Á. (2019). Visual foraging and executive functions: A developmental perspective. *Acta Psychologica*, 193, 203–213. https://doi.org/10.1016/j.actpsy.2019.01.005

Ólafsdóttir, I. M., Gestsdóttir, S., & Kristjánsson, Á. (2020). Age differences in foraging and executive functions: A cross-sectional study. *Journal of Experimental Child Psychology*, 198, 104910. https://doi.org/10.1016/j.jecp.2020.104910

Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition*, 22(7), 975–995. https://doi.org/10.1080/13506285.2014.941433

Oliva, A. (2005). Gist of the Scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256).

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*. Trends Cogn Sci. https://doi.org/10.1016/j.tics.2011.05.004

Olk, B., Dinu, A., Zielinski, D. J., & Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *Royal Society Open Science*, 5(5), 1–15. https://doi.org/10.1098/rsos.172331

Ort, E., Fahrenfort, J. J., & Olivers, C. N. L. (2017). Lack of Free Choice Reveals the Cost of Having to Search for More Than One Object. *Psychological Science*, 28(8), 1137–1147. https://doi.org/10.1177/0956797617705667

Pascucci, D., Mancuso, G., Santandrea, E., Libera, C. Della, Plomp, G., & Chelazzi, L. (2019). Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLoS Biology*, 17(3), e3000144. https://doi.org/10.1371/journal.pbio.3000144

Patai, E. Z., Buckley, A., & Nobre, A. C. (2013). Is Attention Based on Spatial Contextual Memory Preferentially Guided by Low Spatial Frequency Signals? *PLoS ONE*, 8(6), e65601. https://doi.org/10.1371/journal.pone.0065601

Perry, C. J., & Fallah, M. (2017). Effector-based attention systems. *Annals of the New York Academy of Sciences*, 1396(1), 56–69. https://doi.org/10.1111/nyas.13354

Perry, C. J., Sergio, L. E., Crawford, J. D., & Fallah, M. (2015). Hand placement near the visual stimulus improves orientation selectivity in V2 neurons. *Journal of Neurophysiology*, 113(7), 2859–2870. https://doi.org/10.1152/jn.00919.2013

Pinto, G. J. D., Papesh, M. H., & Hout, M.C. (2020). The detail is in the difficulty: Challenging search facilitates rich incidental object encoding. *Memory & Cognition, 48*, 1214–1233. https://doi.org/10.3758/s13421-020-01051-3

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: the role of set for spatial location. *Modes of Perceiving and Processing Information*. https://doi.org/10.1103/PhysRevLett.107.057601

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 179–197. https://doi.org/10.1163/156856888X00122

Rafiei, M., Hansmann-Roth, S., Whitney, D., Kristjánsson, Á., & Chetverikov, A. (2020). Optimizing perception: Attended and ignored stimuli create opposing perceptual biases. *Attention, Perception, and Psychophysics*, 1–10. https://doi.org/10.3758/s13414-020-02030-1

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: Attentional priorization of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 166–177. https://doi.org/10.1037/0096-1523.32.1.166

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25(1 PART 1), 31–40. https://doi.org/10.1016/0028-3932(87)90041-8

Robbins, A., & Hout, M. C. (2019). Scene Priming Provides Clues About Target Appearance That Improve Attentional Guidance During Categorical Search. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000707

Robinson, M. M., Benjamin, A. S., & Irwin, D. E. (2020). Is there a K in capacity? Assessing the structure of visual short-term memory. *Cognitive Psychology*, 121, 101305. https://doi.org/10.1016/j.cogpsych.2020.101305

Rolfs, M., Lawrence, B. M., & Carrasco, M. (2013). Reach preparation enhances visual performance and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628), 20130057. https://doi.org/10.1098/rstb.2013.0057

Sauter, M., Stefani, M., & Mack, W. (2020). Towards Interactive Search: Investigating Visual Search in a Novel Real-World Paradigm. *Brain Sciences, 10*(12), 927 https://doi.org/10.3390/brainsci10120927

Schütz-Bosbach, S., & Prinz, W. (2007, August). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences* https://doi.org/10.1016/j.tics.2007.06.005

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal* https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simons, D. J., & Levin, D. T. (1997, October 1). Change blindness. *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/S1364-6613(97)01080-2

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1), 16–20. https://doi.org/10.1016/j.tics.2004.11.006

Song, J. H., & McPeek, R. M. (2009). Eye-hand coordination during target selection in a pop-out visual search. *Journal of Neurophysiology*, 102(5), 2681–2692. https://doi.org/10.1152/jn.91352.2008

Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, 25(2), 207–222. https://doi.org/10.1080/14640747308400340

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6), E360-7. https://doi.org/10.1073/pnas.1108555108

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, 49(6), 905–916. https://doi.org/10.1016/j.neuron.2006.01.021

Tagu, J., & Kristjánsson, Á. (2020). Dynamics of attentional and oculomotor orienting in visual foraging tasks. *Quarterly Journal of Experimental Psychology (2006)*, 1747021820919351. https://doi.org/10.1177/1747021820919351

Tanrikulu, Ö. D., Chetverikov, A., & Kristjánsson, Á. (2020). Encoding perceptual ensembles during visual search in peripheral vision. *Journal of Vision*, 20(8), 20. https://doi.org/10.1167/jov.20.8.20

Tatler, B. W. (2014). Eye movements from laboratory to life. In M. Horsley, M. Eliot, B. A. Knight, & R. Reilly (Eds.), *Current trends*

in eye tracking research (pp. 17–35). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-02868-2

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: from fixations to object files. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 58(5), 931–960. https://doi.org/10.1080/02724980443000430

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision*, 11(5), 5. https://doi.org/10.1167/11.5.5

Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1628)20130066 https://doi.org/10.1098/rstb.2013.0066

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1564), 596–610. https://doi.org/10.1098/rstb.2010.0188

Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, 13(2), 5. https://doi.org/10.1167/13.2.5

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628). https://doi.org/10.1098/rstb.2013.0055

Thornton, Ian M., de'Sperati, C., & Kristjánsson, Á. (2019). The influence of selection modality, display dynamics and error feedback on patterns of human foraging. *Visual Cognition*, 27(5–8), 626–648. https://doi.org/10.1080/13506285.2019.1658001

Thornton, I. M., Nguyen, T. T., & Kristjánsson, Á. (2020). Foraging tempo: Human run patterns in multiple-target search are constrained by the rate of successive responses. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820961640

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*. https://doi.org/10.1016/0010-0285(80)90005-5

Trewartha, K. M., Case, S., & Flanagan, J. R. (2015). Integrating actions into object location memory: A benefit for active versus passive reaching movements. *Behavioural Brain Research*, 279, 234–239. https://doi.org/10.1016/j.bbr.2014.11.043

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, 3(1), 86–94. https://doi.org/10.1167/3.1.9

Utochkin, I. S., & Wolfe, J. M. (2018). Visual search for changes in scenes creates long-term, incidental memory traces. *Attention, Perception, & Psychophysics*, 80(4), 829–843. https://doi.org/10.3758/s13414-018-1486-y

Van der Stigchel, S., & de Vries, J. P. (2015). There is no attentional global effect: Attentional shifts are independent of the saccade endpoint. *Journal of Vision*, 15(15). https://doi.org/10.1167/15.15.17

van Ede, F. (2020). Visual working memory and action: Functional links and bi-directional influences. *Visual Cognition*. https://doi.org/10.1080/13506285.2020.1759744

van Moorselaar, D., Gunseli, E., Theeuwes, J., & Olivers, C. N. L. (2014). The time course of protecting a visual memory representation from perceptual interference. *Frontiers in Human Neuroscience*, 8. https://doi.org/10.3389/fnhum.2014.01053

Võ, M. L.-H., Boettcher, S. E. P., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*. https://doi.org/10.1016/j.copsyc.2019.03.009

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10(3), 14.1-13. https://doi.org/10.1167/10.3.14

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology. Human Perception and Performance*, 38(1), 23–41. https://doi.org/10.1037/a0024147

Võ, M. L.-H., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences*, 1339, 72–81. https://doi.org/10.1111/nyas.12667

Wang, D., Kristjansson, A., & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. *Perception and Psychophysics*. Psychonomic Society Inc. https://doi.org/10.3758/BF03206488

Williams, C. C. (2010). Incidental and intentional visual memory: What memories are and are not affected by encoding tasks? *Visual Cognition*, 18(9), 1348–1367. https://doi.org/10.1080/13506285.2010.486280

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, 67(5), 816–827. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1751468&tool=pmcentrez&rendertype=abstract

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search, 99–119. Retrieved from http://www.citeulike.org/user/mthomure/article/6719740

Wolfe, J. M. (2012). Saved by a log: how do humans perform hybrid visual and memory search? *Psychological Science*, 23(7), 698–703. https://doi.org/10.1177/0956797612443968

Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, 13(3), 10–10. https://doi.org/10.1167/13.3.10

Wolfe, J. M. (2016). Visual search revived: the slopes are not that slippery: A reply to Kristjansson (2015). *I-Perception*, 7(3), 2041669516643244.

Wolfe, J. M., Cain, M. S., & Aizenman, A. M. (2019). Guidance and selection history in hybrid foraging visual search. *Attention, Perception, and Psychophysics*, 81(3), 637–653. https://doi.org/10.3758/s13414-018-01649-5

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews. Neuroscience*, 5(6), 495–501. https://doi.org/10.1038/nrn1411

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

Wollenberg, L., Deubel, H., & Szinte, M. (2018). Visual attention is not deployed at the endpoint of averaging saccades. *PLOS Biology*, 16(6), e2006548. https://doi.org/10.1371/journal.pbio.2006548

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences*, 1339(1), 154–164. https://doi.org/10.1111/nyas.12606

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235. https://doi.org/10.1038/nature06860