Proceedings of the Annual Meeting of the Cognitive Science Society

Incremental Object Perception in an Attention-Driven Cognitive Architecture

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Bridewell, Will
Bello, Paul F

2015

Incremental Object Perception in an Attention-Driven Cognitive Architecture

Will Bridewell (will.bridewell@nrl.navy.mil)
Paul F. Bello (paul.bello@nrl.navy.mil)
U.S. Naval Research Laboratory
Washington, DC 20375 USA

Abstract

With few exceptions, architectural approaches to modeling cognition have historically emphasized what happens in the mind following the transduction of environmental signals into percepts. To our knowledge, none of these architectures implements a sophisticated, general theory of human attention. In this paper we summarize progress to date on a new cognitive architecture called ARCADIA that gives a central role to attention in both perception and cognition. First, we give an overview of the architecture, comparing it to other approaches when appropriate. Second, we present a model of incremental object construction and property binding in ARCADIA using the well known *change blindness* phenomena to illustrate the time course of object perception and its dependence on attention. Finally, we discuss near-term challenges and future plans.

Keywords: attention; change blindness; feature integration theory; salience; global workspace

Introduction

Attention plays a critical role in human cognitive-economy and bridges perception, high-level cognition, and action. In light of this importance, we note that the most complete and well studied computational cognitive architectures lack unified approaches to attention (Anderson, Matessa, & Lebiere, 1997; Laird, 2012; Meyer & Kieras, 1997). To address this gap, we are implementing a cognitive architecture that models attention as a global, configurable process that responds to top-down, cognitive and bottom-up, perceptual cues and constraints (Hollingworth, Matsukura, & Luck, 2013; Thompson & Schall, 2000).

The remainder of the paper describes this architecture, called ARCADIA,[1] and motivates a model of object perception that requires attention. We briefly discuss the change-blindness literature in psychology and show how ARCADIA is susceptible to this phenomenon under analogous circumstances. Finally, we end with a discussion of near-term plans and farther-term directions.

ARCADIA

As an architectural theory and an implemented system, ARCADIA treats attention as a central part of perception, cognition, and action. In terms of intellectual roots, the architecture shares much of the structure found in the Global Workspace Theory of consciousness (Baars, 1997), which is part of the considerable literature addressing the relationship between attention, perception, and consciousness. This relatively new area of research continues to bear fruit (Baars, Banks, & Newman, 2003; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Koch & Tsuchiya, 2007). Where appropriate, we will draw parallels between ideas from this literature and the design of ARCADIA.

Basic Architectural Framework

We take as a starting point that the vast majority of cognitive processes operate under two conditions: (1) they are not directly introspectable and (2) they can be guided via top-down control. The first condition is uncontroversial. In their oft-cited paper, Nisbett and Wilson (1977) claim that introspective access to processes associated with decision making and other forms of higher cognition is highly limited or entirely nonexistent. The claim that the contents of consciousness result from a myriad of neural processes that people can neither introspectively monitor nor verbally report on is indubitably true, regardless of any view on the limits of introspection. Since this paper is not about introspection per se, we abstain from the deeper discussions on what kind of content is introspectable and conditions under which introspection may produce veridical judgments.

The second condition is that low-level, uninspectable processing of this sort may be consciously willed—the setting of an intention, for example. In a study of visual search behavior, Alfred Yarbus (1967) demonstrated how patterns of eye movements changed in response to different task specifications. Under the experimental circumstances, subjects formed intentions in reaction to the experimenter's instructions. What is fascinating about these sorts of studies is that they reveal a subtle interplay of top-down and bottom-up processes in everyday activities like visual search. Setting a high-level intention such as "find all people wearing red shirts" does not entail consciously generating corresponding motor intentions and instantiating motor programs to move the eyeballs around. Rather, automatic processes that are guided by top-down input generate low-level eye movements.

Figure 1 illustrates the most basic set of distinctions made in ARCADIA, which are informed by these conditions. As shown, ARCADIA maintains a separate space called *accessible content* that stores ephemeral representations produced by low-level components over time. In here, the system makes available the contents of working memory, the results of perceptual processing, and other potentially reportable information.

---

[1] Adaptive, Reflective Cognition in an Attention-Driven Integrated Architecture

Similar to Baars' (1997) concept of the global workspace, accessible content is substantially larger than working memory, and we take as an assumption, subject to revision, that it corresponds to the informational contents of consciousness. Elements in this space result from the attentional process that drives ARCADIA's cognitive cycle and are produced by low-level processing in response to the focus of attention or sensory input. Notably, the theoretical relationship between accessible content and consciousness implies that verbal report is limited to the items that accessible content contains.

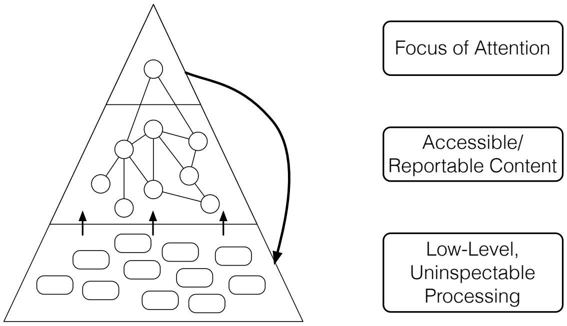

Figure 1: ARCADIA's tripartite structure

ARCADIA's *focus of attention* is a single element selected from accessible content. The general idea is that on every cognitive cycle ARCADIA broadcasts a single item in accessible content system-wide to the set of low-level processing components. By virtue of being *in focus*, an item can temporarily shape the behavior of the majority of low-level processing components. On each cycle, the system selects the next focus of attention by means of an attentional strategy that, in the abstract, operates like a prioritization list for directing attention, given the current output of the low-level components.

Components and Interlingua

ARCADIA's low-level processes are encapsulated in modules called *components*. There are no theoretical restrictions on the representational format or processing characteristics for any given component. This lack of restriction enables greater flexibility in design and the ability for modelers to rapidly prototype new capabilities.

Nevertheless, components with disparate representational formats still need to communicate with each other. To this end, each component communicates with accessible content and the current focus of attention through ARCADIA's *interlingua.* As the example in Table 1 shows, an interlingua element consists of a unique identifier, a variadic argument list, and a symbolic name for the collection of arguments. The arguments contain labeled data produced by the components and stored in formats that they can process. As a result, the interlingua can bind visual, auditory, and other sensory data to more traditional, abstract content retrieved from long-term memory structures. Moreover, each element tracks which component produced it by way of a source tag and has type (e.g., action-description or object-instance). Finally, interlingua elements are indexed to *worlds*, which describe the situation that each element refers to. For instance, elements describing aspects of ongoing perception are assigned the world "reality," whereas mental imagery might manipulate representations that describe the contents of fictional worlds.

Table 1: An example interlingua element in ARCADIA

ID	4651
name	face
arguments	{data: img[640][480]}
world	reality
source	face-detector
type	instance

The Cognitive Cycle and Attentional Strategies

On each cycle, low-level components receive the focus of attention and accessible content produced during the prior cycle. Designated pre-attentive components connect directly to sensory systems such as cameras, while others operate over only the focus and accessible content. Components automatically engage in processing and run to completion if they find input that they can respond to. The interlingua elements produced by components are deposited into what will become the next cycle's accessible content and may become the next focus of attention for the system. The nature of the elements varies. Some components produce bound object or event representations, whereas others produce abstract output (e.g., an expectation) with arguments that refer to other interlingua elements.

The key to the cognitive cycle is the *attentional strategy*, which serves as control knowledge for the system. Currently, a strategy takes the form of a priority list for selecting a focus of attention. Analogies can be drawn here to mechanisms in other architectures, especially to the role of preferences in operator selection and impasse resolution in Soar (Laird, 2012). A concrete example of an attentional strategy for object construction and tracking is given in the penultimate section of this paper.

Attention and Object Perception in ARCADIA

We now turn to giving a more detailed account of object perception and tracking in ARCADIA. We use two questions to frame the discussion: (1) how are objects constructed from raw visual input and (2) how are they maintained *qua* mental representations? We decompose both questions into sub-questions that we address in turn. In what follows, we sketch out how ARCADIA relates to well-established theoretical positions in the literature on visual attention, but we save the details of the model implementation for later in the section. Where appropriate we highlight important departures from or elaborations on extant theory.

Object Perception Pre- and Post-Fixation

How are objects constructed from raw visual input in ARCADIA? Specifically, what role does attention play in object construction? As it stands, object construction consists of two phases in ARCADIA: a pre-attentive stage and a stage involving the deployment of attention to the results of pre-attentive processing.

Pre-attentively, ARCADIA processes image features in parallel, much in the spirit of Treisman and Gelade's (1980) *feature integration theory*. The authors predicate their theory on the finding that basic features such as color, shape, movement, and orientation are computed in different areas of the brain and that these features are computed unconsciously and effortlessly. This view aligns with the implementation of ARCADIA, which associates features with *proto-objects* before a *de facto* object is constructed.

According to feature integration theory, the pre-attentive stage of visual perception can be characterized as a set of maps that roughly correspond to the dimensions of the visual field, one for each computed feature. For example, pre-attentive processing of a red object produces a color-specific map with a marking where the red object appears in the visual field. Analogous maps are computed for other features, including shape, movement, and orientation. Treisman and Gelade propose a master map that can access the locations on all feature-specific maps. When one attends to a location on the master map, all the values in the corresponding locations on the feature-specific maps are registered, resulting in the creation of an *object file*.

In keeping with Treisman and Gelade's theory, feature computation in ARCADIA is unconscious (i.e., it occurs in distributed components) and binding computed features into object files requires attention. In ARCADIA, pre-attentive processes generate candidate regions on an implicit internal map that plays the same role as the master map in feature integration. When the system fixates on a region in its internal map, that region is likely to become the focus of attention. When a fixated region is selected as the focus, other parts of the system may then report on that region. A separate component binds the resulting properties into an object instance.[2]

Object Identity and Tracking

Binding features together into object instances is an important first step, but it is only part of object perception. Maintaining object representations during continual perception is critical for cognition but is made difficult by practical concerns. Our eyes saccade between locations in the visual field on the average of three times per second. We often perceive objects that move. We frequently move while perceiving. And, if there are multiple interesting objects in a scene, we may need to look away from the first object that grabs our attention while mentally keeping track of it for later reference. As with object perception, there are two issues to be addressed. The first, which we refer to as the *continuity problem*, involves how separate instances of objects produced by ongoing perception are identified as being the same object. The second, which we call the *maintenance problem,* involves how objects are retained when unattended.

The traditional assumption is that the continuity and maintenance problems are solved by a combination of *iconic* and *visual short-term memory* (vSTM) (Luck, 2008). In general, iconic memory is understood to be an extremely volatile, high-capacity memory system that provides access to both a visual afterimage of the objects and events in the visual field tagged with limited visual information about each. Visual short-term memory has a demonstrated limit of 3–6 objects, although the nature of these capacity limitations remains contentious (Brady, Konkle, & Alvarez, 2013; Luck & Vogel, 1997; Wutz & Melcher, 2014). The relationship between these two memory systems remains a matter of debate, with some researchers suggesting a third system that shares some of the properties of both iconic memory and vSTM (Sligte, Scholte, & Lamme, 2008).

Nevertheless, there is general agreement that attention is substantially involved in determining which subset of iconic memory gets encoded into vSTM (Schmidt, Vogel, Woodman, & Luck, 2002). Mitroff and Alvarez (2007) have shown that spatiotemporal continuity imposes an unusually strong constraint on judgments of identity, with expected location information being a strong predictor of correct identity judgments. Recent findings have also emphasized constraints on cohesion, boundaries, and containment along with expectations for moving objects to traverse smooth spatiotemporal paths (Mitroff, Arita, & Fleck, 2009). This work suggests that after objects are encoded into object files, they can be tracked.

Presently, ARCADIA includes a nascent story about iconic memory, vSTM, and their interactions in the tasks of individuating, identifying, and tracking potentially moving objects. These three processes all involve attention operating over time, and stand contrary to mechanisms suggested by theories of subitization or visual indices which posit an automatic grasping of 3–4 visual objects by the perceptual system (Kaufman, Lord, Reese, & Volkmann, 1949; Pylyshyn, 2001). However, ARCADIA's approach is in line with new results suggesting a time course for individuation and identification (Wutz & Melcher, 2014).

With a discussion of some of the relevant background in place, we turn now to describing the current implementation of object perception in ARCADIA.

Components and Attentional Strategy

ARCADIA contains implementations of components corresponding to bottom-up feature computation, object individuation, feature binding, object identity, vSTM updating, and change detection. Space precludes a detailed discussion of each, but we specify the expected inputs and

---

[2] We deliberately distinguish between focus and visual fixation, since the focus of attention does not necessarily track currently fixated regions. This distinction corresponds naturally to the difference between covert and overt attention.

outputs of these components and summarize the nature of their information processing.

Component: Bottom-Up Feature Computation Bottom-up feature computation is directly fed input frames by a camera component.[3] Feature computation is carried out by a re-implementation of the Itti–Koch approach to visual saliency calculations (Itti & Koch, 2001; Itti, Dhavale, & Pighin, 2003). To this end, the component constructs feature maps to mimic the center-surround characteristics of receptive fields in the early visual system. ARCADIA relies on a standard set of feature maps: color opponency, intensity, orientation, flicker, and motion. Within-channel conspicuity, a precursor to saliency, is computed and combined into a global saliency map reminiscent of the master map in feature integration theory. This component outputs an interlingua element that contains the computed master map and a maximally salient location.

Component: Iconic Memory ARCADIA's current implementation of iconic memory takes frames directly from a camera component and applies a segmentation procedure to extract closed contours from the image. Since we are interested here in proto-objects and not segments, this component fills the interior of each computed contour providing a black-and-white image that captures shape and a bounding box that corresponds to size. This component outputs the set of these regions, associating each one with its detailed color representation[4] and retinotopic location. Consistent with results reported by Xu and Chun (2009), ARCADIA carries out this individuation prior to object construction, producing proto-objects.

Components: Fixation Generators Fixation-generation components request "eye" fixation based on information in accessible content. There are two fixation generators in ARCADIA, the first corresponds to candidate fixations produced by bottom-up, attention capture and the second is based on top-down factors. On this latter point, there is preliminary evidence for top-down, late selection in attention from work by Thompson and Schall (2000).

ARCADIA's bottom-up component scans accessible content for saliency maps and regions produced by the bottom-up feature computation and iconic memory components. When that component finds a region whose retinotopic location information matches the location of the most salient point on the saliency map, it produces a fixation request. The top-down component scans accessible content for vSTM representations and produces fixation requests for their associated locations. This characterization of top-down influence is admittedly naïve but serves to illustrate ARCADIA's ability to interleave top-down and bottom-up drivers of attention via an attentional strategy.

Component: Early Binding The binding component responds to a fixation in the focus of attention. The binder takes the proto-object target of that fixation and stores it for one extra cycle. Recall that fixations reference a region; the one-cycle wait allows other components (for example, a shape or color classifier) to post region-relevant information to accessible content. Once the intermediate cycle completes, the binder ties together the information associated with its stored region, generates a new object representation, and reports it for use during the next cycle.

Component: Identity ARCADIA's identity component tracks equality between old and new object representations. This component compares focused objects to those reported by vSTM. Presently, the comparison considers the size and location of the new object,[5] attempting to match against the last-known size and location information of objects in vSTM. If such a match is found, then the component posts an interlingua element that specifies an identity relationship between the new object and the object from vSTM. If no match is found, then the component posts an interlingua element that specifies the object as new.

Component: Visual Short Term Memory The vSTM component scans accessible content on each cycle for interlingua elements produced by the identity component. Internally, vSTM is a capacity-limited list structure. When vSTM finds an interlingua element from the identity component tagged as "new," the corresponding object representation is added and, if necessary, the least recently updated object is displaced. When vSTM finds an interlingua element generated by the identity component that signifies an update, then it carries out that update, storing the new version of the object. As output, vSTM reports its stored elements to accessible content at the end of each cycle.

Component: Change Detection For the purposes of exploring the task of change detection and the associated phenomenon of change blindness, ARCADIA includes a change detector specifically for color. This component looks through elements in accessible content for identity relationships between old and new objects (along size and location dimensions) that differ in color. Once found, the component reports the change. Upon seeing that report, a separate component displays a graphical window that contains the altered object.

Attentional Strategy The attentional strategy used for basic object perception and maintenance is admittedly

---

[3] The camera component is essentially a video or an image at present, although there is no barrier to using an actual sensor.

[4] By color, we just mean color-experience and not anything having semantic content. At this pre-attentive stage of processing it is assumed that the visual system is not actively classifying regions as being of one canonical color value or another.

[5] A simple identity-matching scheme like this is doomed to fail when size or location varies considerably across saccades. Developing a more general component that is context-sensitive is on our agenda.

unsophisticated. However, even this straight-forward strategy involves balancing the influences of bottom-up attentional capture and expectations generated in a top-down fashion. The strategy selects the focus of attention by considering in order (1) changes to objects, (2) new objects, and (3) proposed fixations, choosing arbitrarily when there are multiple elements at the same level. This strategy assumes that the system is tasked with detecting changes and collecting information about the objects in the world.

Walkthrough: Change Blindness in ARCADIA

Change blindness implies the existence of constraints that exist at the perception–cognition interface. These constraints indicate a role for attention in developing durable representations that can survive short interruptions to ongoing perception (Simons & Rensink, 2005). Change-blindness studies have played a central role in characterizing the relationship between conscious perception and attention. The consensus view is that attending to the changed object prior to an update is necessary for successfully reporting the difference. There are various well established paradigms for change-blindness experiments, including interleaving a mask between pre- and post-change images while measuring the number of exposures before subjects detect the difference. Often this can take tens of seconds for complex naturalistic images, and sometimes subjects never succeed. Verbal cues reliably improve detection, which suggests that encoding parts of the image into durable representations is a piecemeal process (Rensink, O'Regan, & Clark, 1997).

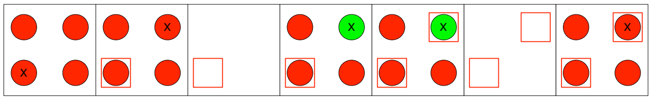

Figure 2: Progression of system responses to stimuli. Red boxes correspond to stabilized representations in vSTM and x's correspond to proposed fixations.

Because, like the human visual system, ARCADIA's perceptual system incrementally builds up scene representations over time, it is also susceptible to change-blindness. To demonstrate, we gave the progression of stimuli in Figure 2 to the system. Moving left to right, in the first box, bottom-up salience draws ARCADIA's "eyes" to the ball in the lower left quadrant, creating a fixation request. In the next box, the system has had time to attend to the fixation and encode a representation of the ball in its vSTM. During this period, eyes are drawn to the ball in the upper right quadrant. Before the ball in the upper right is attended, a visual mask suppresses visual input. As shown in box 3, the vSTM representation of the ball in the lower left quadrant survives suppression. During the masking period, the color of the ball in the upper-right quadrant changes from red to green, attracting fixation as illustrated in box 4. The change detector reports no change in color at this time since ARCADIA did not attend to the object prior to visual masking. This lack of attention left the system without an initial representation of the ball as having been previously colored red. Thus, in box 5, ARCADIA encodes the ball in the upper right quadrant as a new green colored object in vSTM. In box 6, another visual mask is presented, with both encoded vSTM representations surviving. During the masking period, the color of the ball in the upper right quadrant changes back to red. The top-down fixation-generator produces a fixation at one of the objects held in vSTM in accordance with the attentional strategy discussed at the end of the previous section. Once attended and broadcasted to the change detector, the report shown in Figure 3 is made.

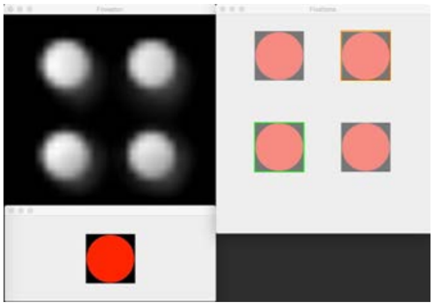

Figure 3: Output of ARCADIA for change detection. The saliency map is in the upper right, the image segments are in the upper left, and the detected change is in the lower left.

The first episode of masking led to change blindness because ARCADIA lacked the time to build a stable representation of the target object before the mask interrupted perception.

Concluding Remarks

This paper introduces the ARCADIA cognitive architecture, and motivates its commitment to attention as a central facet of cognition. Specifically, we emphasized the system's nascent implementation of object perception and tracking, with a change-blindness task serving as the backdrop for explaining system behavior at the interface between perception, attention, and conscious cognition.

In the near term, we plan to enrich the change blindness example with a model of visual search, replete with an inhibition-of-return mechanism. This step should let us capture data on change detection time as a function of both stimulus complexity and set size. We also plan to run change-blindness examples on naturalistic stimuli, although these prospects are limited by the effectiveness of segmentation algorithms and other computer vision technologies used in ARCADIA's perceptual components.

Finally, the work presented here involves only a basic attentional strategy—one that looks for new objects and changes to previously encoded ones. We have begun to look at attentional strategies for more complex tasks, such as counting the occurrence of particular event types in the world, all in the face of ongoing perception. This is the backdrop against which Simons and Chabris' (1999) well-known “invisible gorilla” study on *inattentional blindness* is set. Their study demonstrates that highly salient events, like a man in a gorilla suit walking through a scene, may go un noticed when perceivers are deeply involved in a primary task. The cognitive overload hypothesis is one of many that include greater roles for the similarity of stimuli, the shaping of perception by expectations, the ignoring of regions in the visual field, and capacity limits on representation. We do not take a position on any of these hypotheses, but it seems as if ARCADIA provides a well suited framework to compare and contrast them via implementation.

Acknowledgments

This research was funded by ONR under grant N0001414WX20179. The opinions expressed in this paper are solely the authors and should not be taken to reflect the policy or position of the United States Government or the Department of Defense.

References

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: a theory of higher level cognition and its relation to visual attention. *Human Computer Interaction*, 12, 439–462.

Baars, B. J. (1997). *In the Theater of Consciousness*. New York: Oxford University Press.

Baars, B. J., Banks, W. P., & Newman, J. B. (2003). *Essential Sources in the Scientific Study of Consciousness*. Cambridge, MA: MIT Press.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, 142, 791–808.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Science*, 10, 204–211.

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates low-level saccade target selection: evidence from rapidly generated saccades in the global effect paradigm. *Journal of Vision*, 13, 1–18.

Itti, L., Dhavale, N., & Pighin, F. (2003). Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proceedings of the SPIE 48th Annual International Symposium on Optical Science and Technology* (pp. 64–78). , Bellingham, WA: SPIE Press.

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

Kaufman, E. L., Lord, M. W., Reese, T., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology,* 62, 496–525.

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

Laird, J. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.

Luck, S. J. (2008). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual Memory* (pp. 43–85). New York: Oxford University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature,* 390, 279–281.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: part 1. basic mechanisms. *Psychological Review*, 104, 3–65.

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review*, 14, 1199–1204.

Mitroff, S. R., Arita, J. T., & Fleck, M. S. (2009). Staying in bounds: contextual constraints on object-file coherence. *Visual Cognition*, 17, 195–211.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231–259.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition,* 80, 127–158.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics*, 64, 754–763.

Simons, D., & Chabris, C. (1999). Gorillas in our midst: sustained inattentional blindess for dynamic events. *Pereception*, 28, 1059–1074.

Simons, D., & Rensink, R. (2005). Change blindness: past, present and future. *Trends in Cognitive Science*, 9, 16–20.

Sligte, I. G., Scholte, H. S., Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS One*, 3, e1699.

Thompson, K., & Schall, J. (2000). Antecedents and correlates of visual detection and awareness in macaque prefontal cortex. *Vision Research*, 40, 1523–38.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology,* 12, 97–136.

Vogel, E. K., Woodman, G. F., & Luck, S. J. The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1436–1451.

Wutz , A., & Melcher, D. (2014). The temporal window of individuation limits visual capacity. *Frontiers in Psychology.* 5.

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Science,* 13, 167–174.

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum Press.