Advances in Cognitive Systems 4 (2016) Submitted 5/2016; published 6/2016

A Theory of Attention for Cognitive Systems

Will Bridewell WILL.BRIDEWELL@NRL.NAVY.MIL
Paul F. Bello PAUL.BELLO@NRL.NAVY.MIL
Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory,
Washington, DC 20375 USA

Abstract

Attention provides a way to direct limited cognitive resources to a subset of available information. In the human case, this capacity enables selective mental processing, thereby shaping what we see, think, and do. Importantly, attention has the uncommon characteristic that people can direct it not only outward at objects, agents, and events in the world but also inward toward their own mental representations. In this way, attention naturally bridges perception, cognition, and action. In this paper, we describe a research program starting with six principles that comprise a core theory of attention. We then introduce an implemented, computational implementation of attention and discuss its relationship to those principles. Next, we report our progress on a computational model of visual attention, and demonstrate a task-oriented model that combines this perceptual model with executive control. We conclude with a review of related work, which emphasizes cognitive architectures, and with an overview of our current research agenda.

1. Introduction

Throughout the day, our attention constantly shifts, moving from thought to thought, attracted by a sudden noise, engaged in involving storylines, or directed at challenging tasks. Everyday language reflects the universality of attention's role in our lives. Many people struggle with disorders of attention, and we routinely blame failures on a lack of attention. We also speak solemnly of times when attention operates according to a singular directive: during ecstatic prayer, contemplative meditation, and *flow* (Csikszentmihalyi, 1990) or *being in the zone*. As these examples suggest, we experience attention as broadly influential in our mental lives, but computational models of attention that extend past the visual system are few and far between. Consequently, discussions of attention in problem solving and other cognitive activities remains surprisingly limited throughout artificial intelligence (AI).

Even so, the value of a attention to cognitive systems should be uncontroversial. As early as the mid 1970s, the designers of Hearsay-II (Erman et al., 1980; Hayes-Roth & Lesser, 1977) introduced a focus of attention as a means of search control. Their approach was aimed at exploring a large problem-space without resorting to exhaustive methods. If we limit the role of attention to controlling search, then its value is questionable. AI systems like MYCIN (Davis & Buchanan, 1984) and Soar (Laird et al., 1985) have shown that search reduction need not involve an explicit implementation of attention or strategies for directing it. However, these systems predate a considerable amount of the scientific literature on attention and fail to capture the much broader role it plays in human cognition.

In this paper, we define a research program centered on the development of a rich *computational model* of attention. Part of this project involves implementing theories of vision, audition, reasoning, and other mental capacities. To this end, we introduce a theory of attention based on general agreement in the scientific literature and then describe a computational system that implements that theory. Next, we describe our initial research on a computational model of visual attention implemented within the broader model of attention. We use the model of visual attention in an expanded computational model that integrates perception with executive control. We then look more closely at how attention appears in other cognitive systems, highlighting differences with the reported effort. Finally, we lay out an agenda for this research program, emphasizing the connection between perception and cognition.

2. Theory of Attention

Attention directs resources to a subset of available information. In this regard, a natural distinction exists between *overt attention*, which orients sensors to acquire information from a specific source, and *covert attention*, which directs mental resources to some aspect of the sensory signal. In vision, the separation between gaze and attention has been known for decades (Posner, 1980), but it exists in other modalities as well. For example, donning headphones to listen to a symphony is an overt act of attention. In contrast, following the rhythm, the melody, the violins, or the french horns are all covert acts of attention. This example illustrates that people can willfully attend to or ignore information coming from their senses, but they can also attend to information resulting from cognition through mental imagery and inner speech. Additionally, various phenomena can override volitional control and capture attention, such as shiny objects, feelings of hunger, and thoughts of home.

Assuming that attention determines how mental resources are allocated, what is the functional role of attending? Some researchers claim that attention is necessary for perception (Mack & Rock, 1998), meaning that without it people have no conscious awareness of sensory information. Others consider attention to be the "gatekeeper" to forms of short-term memory (Awh et al., 2006), playing a central part in the selection and retention of remembered information. These effects are relatively low level and subject to debate. However, few would argue against a position that attention makes information available for deliberation whether through perception, memory, or other mental faculties.

Our goal in this paper is to define a research program founded upon the relatively stable results from the scientific literature on attention. Table 1 identifies six well substantiated principles that we take to form a theory of attention in cognitive systems, human and otherwise. This list is not necessarily exhaustive, but it does serve as a hard core that appears to be resistant to refutation. We discuss each of these principles in turn, pointing to a small portion of the representative literature that supports their inclusion. Because so much of what is known about attention has resulted from research on vision and because so many of the existing computational models of attention are limited to visual attention, we illustrate the principles through their relationship to visual attention.

Table 1. Core principles in a theory of attention.

1. Attention is *limiting* and *selective*
2. Attention asymmetrically *biases* mental processing
3. Attention can be *captured* or *directed*
4. Attention can be directed *inward* or *outward*
5. Attention facilitates *integrated* mental processing
6. Attention facilitates *conscious access*

The first two principles state the immediate effects of attention. Principle 1 asserts that attention (a) is a bottleneck for information flow in a cognitive system and (b) selectively prefers some information over the rest. Research that reveals spatial and temporal limitations in visual perception illustrates these principles. *Inattentional blindness* (Mack & Rock, 1998) relates to the spatial deployment of attention and occurs when people in the middle of a task fail to see objects that would otherwise be obvious. One classic example is called *the invisible gorilla* (Chabris & Simons, 2010). In the associated experiments, people are asked to count how often teammates pass a ball to each other. While preoccupied, roughly half the people fail to see a person in a gorilla suit walk through the scene. Correspondingly, *attentional blink* (Shapiro et al., 1997b) reveals limitations in the temporal deployment of attention. This term identifies the short period of time after someone sees a task-relevant stimulus during which they fail to see a second one. In both cases, attention limits what is seen and selects one stimulus over another. Principle 2 states that there is not only a qualitative effect (a stimulus is perceived or not) but also a quantitative one: attended information receives more mental processing than surrounding, unattended information. For instance, Unseen stimuli may still induce semantic priming even though they do not receive internal attention. Multiple studies have shown that subliminal presentation enhances word recognition (Marcel, 1998; Shapiro et al., 1997a) and interfere with object categorization (Schnuerch et al., 2016) even though the stimuli are not processed to the point that subjects can verbally report on them.

The third and fourth principles address how and where attention is allocated. Principle 3 states that attention can be both captured by perceptual processing and directed by cognitive effort. In vision, attentional capture (Koch & Ullman, 1985) occurs when a particular feature, such as motion or luminance, causes a region in the visual field to pop out (e.g., Treisman, 1982). Of course, people can also direct their visual attention by manipulating their gaze based on their goals. Yarbus (1967) reported on this capability by recording eye movements while people examined images with and without specific tasks. Unsurprisingly, where people looked while freely viewing an image differed considerably from where they looked when asked to report on factual content of the image. Principle 4 clarifies that attention is not restricted to the immediate senses (visual attention, auditory attention, etc.); it can highlight internal mental representations, such as information held in working memory. Attending externally to the world determines what is processed in visual cortex, and attending inwardly to mental imagery activates some of the same neurological regions (Pearson & Kosslyn, 2015).

The fifth and sixth principles emphasize the effects of attention within a cognitive system. Principle 5 suggests that attention's selectivity enables the integration of information produced by multiple mental processes. At the lowest levels of visual processing, this principle is reflected in *feature-integration theory* (Treisman & Gelade, 1980) which posits that attention binds together visual appearances including shape and color into mental representations of objects. Further evidence suggests that attention plays a role in auditory and visual integration (e.g., Talsma et al., 2007). Principle 6 establishes attention as a gatekeeper to conscious access. This echos our earlier point that attention makes information available for deliberation and cognitive manipulation. To be more direct, verbal report is the main proxy for whether a representation is consciously accessible, and the aforementioned studies on attentional blink and inattentional blindness show that without attending to particular objects, subjects are unable to report on their appearance. Prinz (2012) offers a detailed perspective on the relationship between attention and conscious access where he argues that attention is both necessary and sufficient for consciousness.

We take these principles, although general and undoubtedly incomplete, to be definitional. On our view, a cognitive system that lacks any one of the six lacks attention. Likewise any system that exhibits these principles has attentive faculties, regardless of whether the characteristics are explicitly encoded or implicitly met. In the next section, we describe a cognitive system that was developed with these principles in mind. We refer to that system as a computational model of attention to emphasize the distinction between the scientific theory and an artifact that implements an interpretation of the theoretical principles.[1] Where possible, we point out how the theory of attention in table 1 informed the defining features of that system.

3. ARCADIA: A Computational Model of Attention

Although subsets of the principles from table 1 appear in various cognitive systems, no single system adheres to them all. As a result, we set out to build a computational model of attention that satisfies all six principles. *In this effort, we take as a working hypothesis that a modality independent mechanism of attention plays a central, integrating role for perception, cognition, and action.* This perspective encourages us to synthesize the literature on visual and auditory attention with a broader, cognitively oriented agenda. Implementing a computational model requires interpreting the central principles and specifying details beyond what they prescribe. In the rest of this section, we describe ARCADIA, a cognitive system and computational model whose organization is pictured in figure 1, and relate it to the theory of attention.

We begin describing ARCADIA by discussing its informational structures. Unlike many cognitive systems, this computational model places minimal requirements on data formats and processing. One of these requirements is that all information enters ARCADIA through *sensors*. This design defines a boundary between any particular system built with ARCADIA and the outside world. Each sensor implements an interface that connects to some low-level data provider, such as a camera or a

1. The phrase *computational model* may also refer to a cognitive system configured to carry out one or more tasks, but the difference in meaning should be clear from usage.

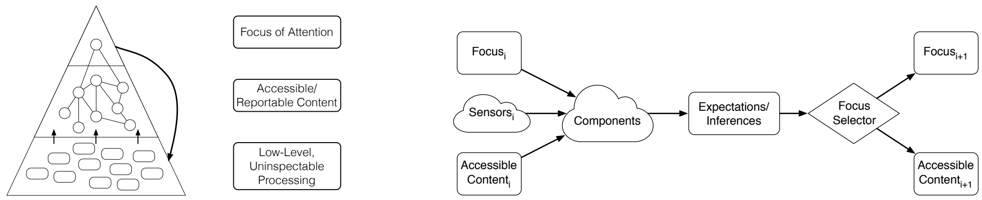

Figure 1. The informational structure and main cycle of the ARCADIA computational model.

microphone, and polls it on a routine basis. As an example, one of the implemented sensors reports images from frames produced by a video player.

Sensors make information available, but *components* request and process that information. ARCADIA imposes only a few restrictions on components. Specifically, components must be able (a) to operate as independent, black-box processors, (b) to read and write information structured into an *interlingua*, and (c) to respond to the system's cyclic requests for results. The first point implies two restrictions. First, the independence restriction means that components can run in parallel without interfering with one another, but it also prevents them from synchronizing their computation outside of ARCADIA's main cycle. Second, the black-box restriction means that components can employ any algorithms and data structures that are helpful for furnishing their output. To enable inter-component communication, results must be reported in a shared language. We refer to this language as ARCADIA's interlingua.

To motivate this design, we note that artificial intelligence has emphasized representation since its inception and researchers working on cognitive systems have followed suit. Attempts to identify a unitary representational system for encoding knowledge, carrying out inference, and supporting learning are common. This emphasis has led to the creation and maintenance of sharp divides among investigators who favor first-order logic, productions, neural networks, probabilistic graphical models, and various other formalisms. ARCADIA takes a different approach. Instead of defining a system-wide language for data representation, the computational model relies on its interlingua, a unifying schema for organizing data in a variety of formats (see Bridewell & Bello, 2015).

Not every component will have the ability to read every kind of data format in an interlingua element, but a particular component may be able to process part of an element and produce results that other components can use. For example, consider an element that stores a visual representation of an object seen in the environment. A color detection component could operate over that element's image data and report on the presence of blue or red. Another component may interpret the element's location data that would enable it to report movement. A key benefit of the interlingua is that components can exchange data in formats that best suit their processing routines.

On each cycle in ARCADIA, components report the results of their processing as interlingua elements. These results are read from and reported to a shared, volatile location called *accessible content*. This approach to inter-component communication differs from methods common to cognitive systems, such as interprocess communication and shared memory. Notably, accessible content is treated as unbuffered and practically unlimited in capacity. Also, the elements are ephemeral, lasting for a single cycle. Since this is a nonstandard communication technique for cognitive systems (and cognitive architectures), it might help to identify our motivating intuitions.

For humans, information processing is continual. We cannot stop moments from passing, and each new moment brings more information than our limited resources can fully process. Moreover, there is no special storage for all the content associated with any one moment. There is no special blackboard where task-relevant information is effortlessly organized and memorized, where that information is updated as perception, cognition, and reasoning continue. As a result, we adopt a communication strategy where the results of processing appear simultaneously on a single cycle of a cognitive system, are visible to every component, and are entirely replaced by the results for the next cycle.

Within accessible content, one element is marked as the *focus of attention*. As illustrated by the downward arrow on the left of figure 1, ARCADIA broadcasts the focus of attention to every component at the beginning of each cycle. This message tells the components that (a) accessible content has been refreshed and (b) the information in the focal element should receive priority. Not all components are focus responsive and those that are may lack routines that can process the information in a particular focal element. In those cases, components are not inactive; instead, they will proceed with their default routines: polling sensors, reading from accessible content, and processing their input. The results from that cycle will form the accessible content of the next one, but beforehand, the *focus selector* applies the current *attentional strategy* to identify a new focus of attention.

As Gopher (1993) writes, "Setting priorities is a common human experience; the question is how competent we are in establishing attention strategies and allocating processing efforts among concurrently changing task elements." In ARCADIA, attentional strategies prioritize interlingua elements by name, type, and potentially other fields. Modeling any particular task (see section 5) involves supplying a specific strategy with its own prioritization routine. Without any task, ARCADIA relies on a default strategy that currently examines objects in an undirected manner.

Returning to the theory in table 1, we claim that this computational model of attention satisfies the six principles. First, ARCADIA's focus of attention is limited in the sense that it is a single interlingua element. Although it is true that the focal element may encode complexly structured information, unpacking that structure, processing the details, and bringing the results back together requires prolonged attention. Moreover, attentional strategies explicitly implement attention's selectivity. Second, deployment of attention within the computational model influences the behavior of individual components, which biases mental processing. Third, attentional strategies may allow the results from pre-attentive components to capture attention or may direct attention to task-related information. Fourth, some ARCADIA components process sensory information whereas others operate over internal representations. Fifth, focus-responsive components will produce elements that may be bound together into integrated representation of objects or other content.

As for the sixth principle, one way of interpreting *conscious access* is through a demarcation between broadly and locally available information (Baars, 2002). Along these lines, ARCADIA distinguishes between information encapsulated within components and information made available to all components. When attentional strategies focus on one interlingua element at the expense of others, they influence which new elements will be created and therefore made broadly available.

Therefore, in a cautiously circumscribed, information-processing sense, this computational model of attention can be said to facilitate conscious access.

As a computational model, ARCADIA also relies on a stronger set of assumptions that are not part of the core principles. First among these is our working hypothesis: there is a modality-independent, attentional mechanism that plays a central, integrating role for perception, cognition, and action. Other assumptions that distinguish ARCADIA from alternative computational approaches to attention include (a) the cyclic activity of $\hookrightarrow$sense+process$\rightarrow$select$\rightarrow$broadcast, (b) the lack of restrictions on the format of component-encapsulated information, and (c) the use of an interlingua for communication among components. Perhaps the strongest assumption is the continual flushing of accessible content. As we will see in the discussion of visual attention, this evanescence has implications for how memory is incorporated within an ARCADIA-based system.

4. Visual Attention Using ARCADIA

Colloquially, we talk about attention to sights, sounds, pains, pleasures, memories, fantasies, desires, goals, and other phenomena, and rarely do we treat it as a stand-alone mechanism. With that in mind, we view ARCADIA as providing a unifying foundation for computational models across these modalities. Due to the considerable, interdisciplinary research on visual attention, our early work has emphasized that area. Although there are several computational models of human visual attention, these are by and large concerned with bottom-up, perceptually driven attentional capture. Borji & Itti (2013) and Borji and colleagues (2013) provide a thorough examination of the state-of-the-art. Notably the relationship between these models and the deliberate manipulation of attention via cognitive control remains underexplored (Baluch & Itti, 2011).

With these characteristics in mind, we are developing a computational model of *visual* attention using ARCADIA. The current implementation was designed to account for the ordinary capacity to detect and track objects in dynamic environments in a way that proceeds incrementally from sensory signals to encoded representations in working memory. Fortunately, the literature on visual attention provides considerable detail regarding the operation of the human visual system. Our main task is to take these findings and integrate them into a functioning computational model. Figure 2 shows a schematic of the computational visual system. The diagram emphasizes information flow among components and distinguishes those that are focus-driven from those that are not.

At the base level, we take it as uncontroversial that a vision system requires a way to transduce reflected light into computable formats. People have eyes, machines use cameras. In this model, a single sensor reads frames from a video file, which enables precise control over stimulus presentation. Two important characteristics of eyes are not currently integrated. First, the current implementation accounts for covert attention only. Along these lines, we have built models that include simulated camera movement, but we have not included these aspects in the computational model. Second, the computational model treats visual input as if it were uniform in pixel density. Even so, we have investigated the use of different resolutions for peripheral and foveal vision in specific task models to study how this effects which regions in the visual field may capture attention.

We also take it to be uncontroversial that there are perceptually driven mechanisms for attentional capture that operate outside cognitive control. Computational models of these mechanisms

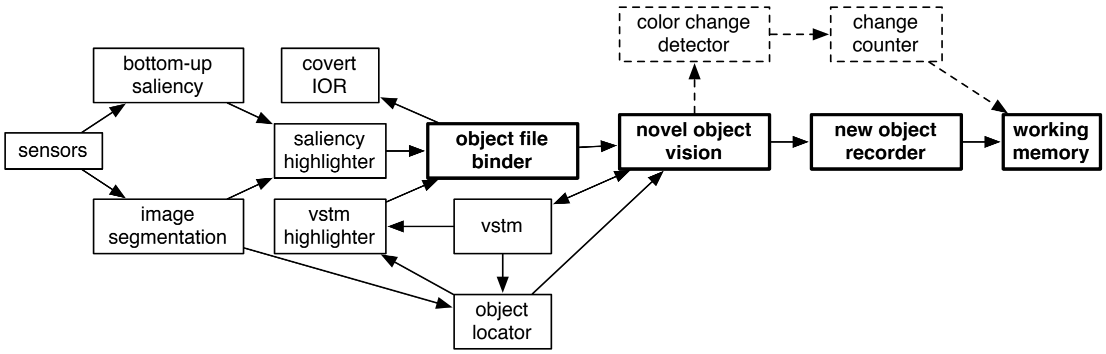

Figure 2. Component interactions in a computational model of visual attention implemented using ARCADIA, with additional support for a task-specific model of change detection. Each box represents a single component. The arrows illustrate information flow (e.g., saliency highlighter uses information produced by bottom-up saliency and image segmentation and produces information used by object file binder). Components in bold respond to the focus of attention. Dashed lines indicate the task-specific parts of the change-detection model.

are generally classed as *bottom-up saliency* approaches to visual attention, where saliency is typically defined loosely as areas in the visual field "that appear to an observer to stand out relative to their neighboring [areas]" (Borji et al., 2013, p. 185). A subset of these methods are inspired by research in neuroscience, and ARCADIA includes a reimplementation of the system reported by Itti, Dhavale, and Pighin (2004). Their system uses information about color, orientation, intensity, flicker, and motion to construct a normalized *saliency map* from image data. The maximum-valued point on the map indicates the location expected to capture attention in the absence of cognitive control.

Likewise, it is uncontroversial that the human visual system segments the visual field into object-like regions. As a result, we include an image segmentation component in the computational model. Unfortunately, there is no universally effective algorithm for computationally segmenting images. For now, we adapt the component's routines to the stimuli being processed, but as general approaches become available, we can integrate them. The good news is that as long as we work with two-dimensional objects moving across a relatively barren background, an unsophisticated method that searches for closed-contour regions is surprisingly effective. The model treats the extracted segments as proto-objects (Clark, 2004), co-located collections of visual features, that may become object representations only if attended.

The computational model uses the collection of proto-objects as a rough analog to *iconic memory* (e.g., Gegenfurtner & Sperling, 1993), but for slightly longer term storage, it relies on a separate visual short-term memory (vSTM; Luck, 2008). Although there is general agreement on the presence of vSTM in the human visual system, its structure, contents, and functionality are all matters of debate. Because the computational model is motivated by object detection and tracking, we adopt an object-file perspective on contents and structure (Vogel et al., 2001). By this, we mean that the vSTM component in the computational model stores roughly four, structured object-representations, which contain graphical depictions of an object's shape, color, and size. Because accessible content is recurrently flushed, this memory component operates by reporting its elements on each cycle.

Notably, the computational model explicitly separates object locations from object representations. This move may be controversial, but it is supported by evidence that object identity and location come apart during multiple object tracking (Pylyshyn, 2004). Our approach combines two ideas from vision research. The first idea is a theoretical construct called "fingers of instantiation" (FINSTs; Scholl & Pylyshyn, 1999), which are roughly four pointers to object locations that update almost independently of visual attention as the tracked objects move.[2] The second idea draws from Dawson's (1991) theoretical constraints for solutions to the motion-correspondence problem—the perceptual "matching" of objects across rapidly presented, static images. We ground FINSTs in the object representations stored in vSTM, so that attention is required for their construction and their existence is brittle to the redirection of visual attention. The model's object-locator component updates location information by matching proto-object locations from the visual field to the nearest-in-space previously attended objects. Object properties in vSTM are not updated until a corresponding proto-object receives attention and an object file is created.

Two components request that ARCADIA draw attention to proto-objects. The saliency highlighter combines image segments with a saliency map and selects one to four of the most salient, object-like regions as candidates for visual attention. This method is in the same vein as work by Russell and colleagues (2014) and Walther and Koch (2006), but stands out from their attempts to integrate all of visual attention through saliency maps. The vSTM highlighter suggests revisiting objects in vSTM that may have moved since they were last attended. Taking a cue from research on inhibition of return (IOR) in covert attention (McDonald et al., 2009), the computational model's covert IOR component supports inhibiting spatial regions if they were visited in the recent past.

When a proto-object receives attention, its various properties are collected together by the object-file binder into an object file in the spirit of Treisman and Gelade (1980). In the event that this object file also receives attention, the computational model's *novel object vision* component determines whether the encoded object appears in vSTM or is new. Determinations of object equality at this stage are taken to be task dependent (i.e., defined according to the needs of the model), but would operate on only those attributes available in the object file. Objects appearing in vSTM will automatically be updated with their new attributes, and others will both displace an object in vSTM and be considered for storage in working memory through the *new object recorder*.

Although access to vSTM is considered automatic once an object file receives attention, access to working memory requires two more cycles. The first cycle produces a request to memorize the object, and if that element receives focus, the second cycle includes a corresponding update to working memory. As with vSTM, the elements are reported to accessible content on each cycle. In its current form, the computational model treats working memory as a practically infinite storehouse and does not attempt to resolve equality among objects at that level of representation. However, this characteristic can be adapted to the needs of tasks as they are modeled. Developing a richer theory of working memory is left for future research.

2. Because FINSTs are supposedly limited in number, attending to a previously untracked object would presumably reassign a FINST to it.

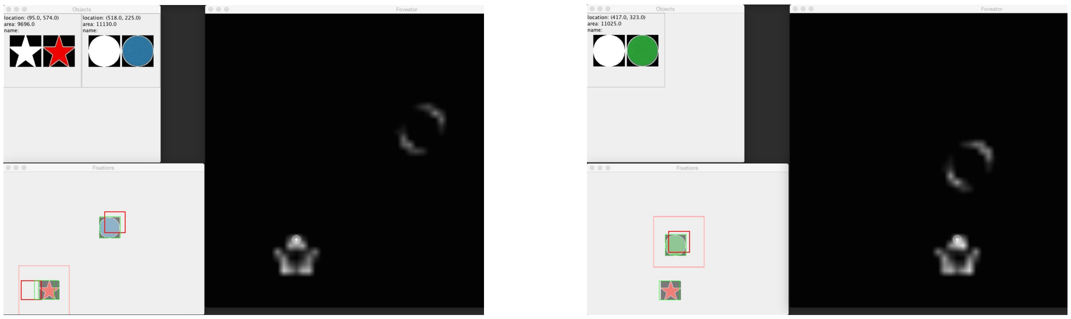

Figure 3. ARCADIA visualizations while (left) freely viewing the scene and (right) under cognitive load. The top left window in each image shows objects stored in vSTM. The bottom left window shows an iconic representation of ARCADIA's view where proto-objects are picked out. Red boxes indicate where working memory representations were last updated. The window to the right is the bottom-up saliency map. In both instances, the star is more salient than the circle.

5. A Demonstration of ARCADIA in Action

To evaluate the computational model of visual attention in ARCADIA, we have investigated a variety of tasks. We selected these tasks based on their relevance to visual attention and specifically to what they say about pre-attentive versus attentive processing. To date, we have modeled change detection (Bridewell & Bello, 2015), multiple object tracking (Bello et al., 2016), and inattentional blindness (Bridewell & Bello, 2016). We refer readers to that work for detailed treatments of each topic. Here we describe an example of a mixed perceptual–cognitive activity that combines change detection and object tracking.

The stimulus is a video that shows a circle moving against a uniform background. Over the course of the video, the circle changes color. After several seconds, a bright red star moves across the bottom of the screen. We designed this star to elicit a request for attention from ARCADIA's bottom-up saliency component. As a bit of foreshadowing, the goal of this model is to show that when the system is not configured for any task, it follows the circle, notices color changes, and is distracted by the star when it appears. In contrast, when we configure ARCADIA to count the number of times that the circle changes color, it successfully completes the task but fails to see the salient distractor, demonstrating inattentional blindness.

In ARCADIA, modeling involves (a) implementing components that provide the required functionality and (b) creating an attentional strategy that encodes specific priorities. Figure 2 shows the two components that were added to the computational model of visual attention for this demonstration. The color-change detector recognizes when a focused object's color differs from its most previous representation in vSTM and produces a change event. When a change event is the focus of attention, the change counter looks for a working memory representation of the number of changes to date. This component issues two kinds of action requests: if a counter exists, then update it; otherwise, create a new counter. Working memory updates its contents when one of these requests is the focus of attention.

When using the computational model of visual attention, we have a default attentional strategy that prioritizes elements in accessible content according to their names and types. Starting with the most important, the priorities are (1) requests for action, (2) newly created object files, (3) requests to shift covert visual attention to a proto-object, and (4) a random element. For this demonstration, we insert at the head of the list a preference to focus on the change events produced by the change detector. To represent cognitive load, we implement finer grained priorities for (3), which selects among elements produced by bottom-up saliency and vSTM as described in section 4. To simulate high cognitive load (e.g., tracking the object taxes mental resources), we have the system prefer requests from vSTM, otherwise, it selects requests as usual.

This demonstration is meant to show the configurability of ARCADIA's attentional mechanism and the resulting changes in behavior. As illustrated in figure 3, the system acts as expected. The image on the left, which was produced using the default attentional strategy, shows ARCADIA's visual attention squarely on the distractor (as indicated by the pink box) and both objects in vSTM. In contrast, the image on the right, which used the modified attentional strategy, shows that ARCADIA visually attends only to the circle throughout the scenario. Importantly, the associated saliency maps (where brighter pixels represent greater saliency) indicate that the red star coincides with the most salient region in both cases. Notice also that the distractor is always processed by low-level visual components, but never attains an object representation when there is a lack of attention.

6. Related Work

With ARCADIA, we have developed an explicit computational model of attention. Interestingly the character of attention in cognitive systems has been discussed, but rarely given the broad treatment that we present here. As a source of examples, consider the literature on cognitive architectures. Within Soar, attention is taken to be an emergent property that falls out of the operator selection process (Young & Lewis, 1999). Within ACT-R, there is no general attention mechanism, but the visual module and a potential update to it (Nyamsuren & Taatgen, 2013) encapsulate visual attention. Following this course would mean that there is necessarily a different attentional mechanism for every perceptual modality, so that a cognitive system could divide its attention without cost across each of its senses. Within EPIC, attention is also treated as an emergent phenomenon with "production rules that decide where more detail is needed, [sic] and command a corresponding eye movement" (Kieras, 2007). None of these systems satisfies all the core principles in table 1, and none are or include computational models of attention.

In terms of giving attention a central role, the closest system to ARCADIA may be LIDA (Franklin et al., 2016). Care is required in making such a comparison, though. The name "LIDA" refers to both what Franklin calls a conceptual model and an implemented computational framework. The literature on LIDA includes many broad claims about the conceptual model that are extended in the subjunctive mood toward the framework. In way of a clarification, Franklin (2016) provides an accounting of features within the canon of literature on LIDA that are either not implemented, partially implemented, or fully implemented. Because it is unclear what functionality is included in a partial implementation or even what should be accounted for in a full implementation of the items on Franklin's list, we have little upon which to base direct computational comparisons.

Even so, we can compare ARCADIA and LIDA along a few dimensions. If we look at the theoretical commitments for each model, ARCADIA has relatively few. Of course, the components that form the model of visual attention introduce several apparent commitments. However, these commitments are cordoned off in the loose organization of components and are contingent. That is, radical alterations to the visual components in ARCADIA have no ramifications on the model's core commitments. In contrast, the LIDA model makes claims about a wide variety of memory systems, action selection and execution, and the nature of consciousness. Which of these claims are central to the model and which are contingent features of the implementation is unclear. Additionally, the two research programs differ in their scientific approaches. Franklin presents LIDA as a conceptual model under development, which he expands by importing findings from the scientific literature. For that work, computational implementation is important, but not necessary. This view differs from our own in that ARCADIA is entirely computational, adhering to a few theoretical commitments. Our primary goal is not to expand the theory of ARCADIA to cover all of cognition, but to investigate and test its central claims by developing and integrating a variety of cognitive models.

Looking only at the implementation of attention in LIDA, we can be more specific in comparison. Franklin's model uses *attention codelets* which are always active processes that sift through the content of LIDA's *workspace* and form *coalitions* of elements that then compete within the *global workspace* to be broadcast. Attention codelets are relatively sophisticated processes that can, in principle, work together when forming coalitions, have a refractory period, and modify the activation of coalitions. In ARCADIA, this selective activity is encoded in an attentional strategy, which is an ordered list of preferences, and only one element receives focus at a time. Whereas LIDA makes only the winning coalition broadly available, ARCADIA ensures that all of accessible content is available for its components to inspect. We do not assume that attention can somehow determine which information hangs together and, instead, give it the narrower role of highlighting a single element at a time.

7. Concluding Remarks

The research program outlined in this paper emphasizes the computational investigation of attention, but in pursuing this program, we are working toward a cognitive-system that perceives, deliberates about, and acts within the world. To this end we are investigating auditory stream separation, possibly using an auditory saliency map (Kayser et al., 2005), and taking the first steps toward multi-sensory integration. We also intend to develop a theory of long-term memory and *attentive* recollection where memories are multi-representational in nature, containing perceptual properties (Morsella et al., 2009). Further, recollection will be modulated by internally directed attention (De Brigard, 2012). Finally, setting our sites on deliberative reasoning, we will be integrating an implementation of mental model theory (Khemlani et al., 2014).

Our goal is to develop a view of attention as a general feature of a cognitive system. This aim contrasts with much of the existing computational work, which has emphasized visual attention (e.g., Tsotsos, 2011; Itti et al., 2004) and, to a lesser degree, auditory attention (Shamma et al., 2011). To this end, we have picked out six principles that form a core theory of attention and developed ARCADIA, a computational model that adheres to these. Using ARCADIA, we have implemented a model of visual attention that can integrate information produced by perceptual and cognitive processes. However, there are many other cognitive phenomena that seem to be influenced by where we attend, either outwardly or inwardly. The research reported here takes us closer to a computational understanding of the role that attention plays in everyday life.

Acknowledgements

We thank Pat Langley, Alistair Isaac, and Lindley Darden for discussion and comments that led to improvements in the text. The authors would like to acknowledge generous support from the Office of Naval Research under grants N0001414WX20179 and N0001415WX01339. The views expressed in this paper are solely the authors and should not be taken to reflect any official policy or position of the United States Government or the Department of Defense.

References

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201–208.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, *6*, 47–52.

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in neurosciences*, *34*, 210–24.

Bello, P. F., Bridewell, W., & Wasylyshyn, C. (2016). Attentive and pre-attentive processes in multiple object tracking: A computational investigation. *Proceedings of the thirty-eighth annual meeting of the cognitive science society*. Philadelphia, PA.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, *35*, 185–207.

Borji, A., Sihite, D., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, *22*, 55–69.

Bridewell, W., & Bello, P. F. (2015). Incremental object perception in an attention-driven cognitive architecture. *Proceedings of the Thirty-Seventh Annual Meeting of the Cognitive Science Society* (pp. 279–284). Pasadena, CA.

Bridewell, W., & Bello, P. F. (2016). Inattentional blindness in a coupled perceptual–cognitive system. *Proceedings of the thirty-eighth annual meeting of the cognitive science society*. Philadelphia, PA.

Chabris, C. F., & Simons, D. J. (2010). *The invisible gorilla: and other ways our intuitions deceive us*. New York City, NY: Crown Publishers.

Clark, A. (2004). Feature-placing and proto-objects. *Philosophical Psychology*, *17*, 443–469.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal performance*. New York City, NY: Cambridge University Press.

Davis, R., & Buchanan, B. G. (1984). Meta-level knowledge. In B. G. Buchanan & E. H. Shortliffe (Eds.), *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley Publishing Company.

Dawson, M. R. (1991). The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. *Psychological review*, *98*, 569–603.

De Brigard, F. (2012). The role of attention in conscious recollection. *Frontiers in pychology*, *3*, 1–10.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty. *ACM Computing Surveys*, *12*.

Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., Snaider, J., Agrawal, P., & Chen, S. (2016). A LIDA cognitive model tutorial. *Biologically Inspired Cognitive Architectures*, *16*, 105–130.

Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 845–866.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV:Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*, (pp. 299–322). Cambridge, MA: MIT Press.

Hayes-Roth, F., & Lesser, V. (1977). Focus of attention in the Hearsay-II speech understanding system. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence* (pp. 27–35). Cambridge, MA.

Itti, L., Dhavale, N., & Pighin, F. (2004). Realistic avatar eye and head animation using a neurobiological model of visual attention. *Proceedings of the SPIE 48th Annual International Symposium on Optical Science and Technology* (pp. 64–78). Bellingham, WA.

Kayser, C., Petkov, C. I., Lippert, M., & Logothetis, K. (2005). Mechanisms for allocating auditory attention: an auditory saliency map. *Current biology*, *15*, 1943–1947.

Khemlani, S., Barbey, A., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in human neuroscience*, *8*, 1–15.

Kieras, D. (2007). Control of cognition. In W. D. Gray (Ed.), *Integrated Models of Cognitive Systems*, (pp. 327–355). Oxford, UK: Oxford University Press.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, *4*, 219–227.

Laird, J. E., Rosenbloom, P. S., & Newell, A. (1985). Chunking in Soar: The anatomy of a general learning mechanism. *Machine Learning*, *1*, 11–46.

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

Mack, A., & Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press.

Marcel, A. J. (1998). Blindsight and shape perception: Deficit of visual consciousness or of visual function? *Brain*, *121*, 1565–1588.

McDonald, J. J., Hickey, C., Green, J. J., & Whitman, J. C. (2009). Inhibition of return in the covert deployment of attention: evidence from human electrophysiology. *Journal of cognitive neuroscience*, *21*, 725–733.

Morsella, E., Lanska, M., Berger, C. C., & Gazzaley, A. (2009). Indirect cognitive control through top-down activation of perceptual symbols. *European journal of social psychology*, *39*, 1173–1177.

Nyamsuren, E., & Taatgen, N. A. (2013). Pre-attentive and attentive vision module. *Cognitive systems research*, *24*, 62–71.

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: ending the imagery debate. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 10089–10092.

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

Prinz, J. J. (2012). *The Conscious Brain*. Oxford, UK: Oxford University Press.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*, 801–822.

Russell, A. F., Mihalas, S., von der Heydt, R., Niebur, E., & Etienne-Cummings, R. (2014). A model of proto-object based saliency. *Vision Research*, *94*, 1–15.

Schnuerch, R., Kreitz, C., Gibbons, H., Memmert, D., Schnuerch, R., Kreitz, C., & Memmert, D. (2016). Not quite so blind: Semantic processing despite inattentional blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1–5.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: clues to visual objecthood. *Cognitive psychology*, *38*, 259–290.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*, 114–123.

Shapiro, K., Driver, J., Ward, R., & Sorensen, R. E. (1997a). Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psychological science*, *8*, 95–100.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997b). The attentional blink. *Trends in cognitive sciences*, *1*, 291–6.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*, 679–690.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 194–214.

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

Tsotsos, J. K. (2011). *A Computational Perspective on Visual Attention*. Cambridge, MA: MIT Press.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 92–114.

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

Yarbus, A. L. (1967). *Eye movements and vision*. New York City, NY: Plenum Press.

Young, R. M., & Lewis, R. L. (1999). The Soar cognitive architecture and human working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*, (pp. 224–256). Cambridge, England: Cambridge University Press.