CrossMark

A R T I C L E

Giving Eyesight to the Blind: Towards Attention-Aware AIED

Sidney K. D'Mello[1]

Abstract There is an inextricable link between attention and learning, yet AIED systems in 2015 are largely blind to learners' attentional states. We argue that next-generation AIED systems should have the ability to monitor and dynamically (re)direct attention in order to optimize allocation of sparse attentional resources. We present some initial ideas towards achieving this goal, starting with a 2×2 (direction of attention × content of thoughts) organizational framework that encapsulates a range of attentional states including overt inattention, covert inattention, zone outs, tune outs, and focused, alternating, and divided attention. We then sketch out a three component attentional computing architecture consisting of: (1) devices to monitor where attention appears to be directed; (2) mechanisms for real-time attentional state diagnosis; and (3) interventions to dynamically (re)direct attention. We describe two closed-loop attention-aware AIED systems to serve as concrete renditions of these ideas. We conclude by arguing that AIED can achieve the dual goals of advancing basic research on the science of learning while simultaneously developing highly-effective AIED systems by "attending to attention."

Introduction

It is 2030. You are learning about some new statistical technique while being driven to school in your self-driving car. The medium is an interactive text (yes, people still read in 2030) with a multimedia graphics panel projected on the car's display. The technique is confusing, so you really have to concentrate. But it's

---

✉ Sidney K. D'Mello
  sdmello@nd.edu

[1] University of Notre Dame, Notre Dame, IN, USA

hard to stay focused. You think, "Despite all of this technology, why can't they figure out a way to directly implant the knowledge into my brain—like in the Matrix movies that were popular 25 years ago. Wouldn't that be more useful than making a car that drives and a robot that cooks? I used to like driving and cooking, but not learning…" Suddenly, the system presents you with an interactive simulation of the statistical technique and asks whether you would like to see it in action on some sample data. You agree, and a simulation ensues. Now you are really getting into it and things start going well. A bit later, you then try to replicate the analysis on your own data, but something does not appear to be right. The system suggests that you review a certain part of the simulation in slow motion, and interactively highlights key aspects of the content. Essentially, it directs your attention and you realize that you were missing one key step. You close your eyes, lower your head, and deeply reflect on what you have just learned. The system is silent. It knows you are thinking of the learning task and were not tuned out, despite all outward appearances. In other words, the system was responsive to thoughts and feelings in addition to your words, actions, and knowledge.

This hypothetical scenario might seem like science fiction today, but could be routine in the next 25 years. Some components are already in place, as current intelligent systems already model knowledge (knowledge tracing and item-response theory), actions (educational data mining and learning analytics), words (natural language processing and discourse analytics), and even feelings and emotions (affective computing and semantic mining). Attention, however, has not been modeled to nearly the same extent. This is a critical omission, because learning requires attention (Olney et al. 2015). Cognitive processes, such as prior knowledge activation, maintenance and elaborative rehearsal, inference generation, causal reasoning, and comprehension, all demand attentional resources. A lack of attention counters these processes and leads to radically different behaviors and outcomes. Learners who cannot sustain attentional focus are more likely to partake in self-distracting and other unproductive behaviors (Damrad-Frye and Laird 1989; Forbes-Riley and Litman 2011). Involuntary lapses in attention (or mind wandering—Smallwood and Schooler (2015)) can occur even when learners make a concentrated effort to sustain attention. A lack of attentional focus, either in the form of overt off-task behaviors (Baker et al. 2004) or more covert attentional lapses in the form of mind wandering, lead to superficial understanding rather than deep comprehension.

Sustaining attentional focus is not sufficient in and of itself. Learners must also effectively allocate limited attentional resources in a manner that aligns with changing task demands, and with the dynamics of the learning environment. For example, learners must effectively alternate attention between the text and diagram when learning from illustrated texts (Hegarty and Just 1993; Schnotz 2005). When diagnosing problems with complex systems, learners must allocate information to critical components to deeply comprehend the mechanisms (Graesser et al. 2005). Processing animations requires learners to allocate attention in a manner that aligns with changes in the animation and in concert with any accompanying narration (van Gog and Scheiter 2010). Effective problem solving also demands the appropriate allocation of attentional resources (Knoblich et al. 2005; van Gog et al. 2009). The list goes on. The ability to sustain and appropriately allocate limited attentional resources is critical for effective learning.

We argue that next-generation AIED technologies should include mechanisms to model and respond to learners' attention in real-time. This is not a new idea; the idea of

attention-aware user interfaces was proposed almost a decade ago in a special issue by Roda and Thomas (2006). There was even an article on futuristic applications of attention-aware systems in educational application (Rapp 2006). Prior to this, Gluck et al. (2000) discussed the use of eye tracking to increase the bandwidth of information available to an intelligent tutoring system (ITS) in an aptly titled paper "Broader Bandwidth in Student Modeling: What if ITS Were "Eye" TS?" Similarly, Anderson (2002) followed up on some of these ideas by demonstrating how particular beneficial instructional strategies could only be launched via a real-time analysis of eye gaze. Most of the recent work on leveraging eye gaze to increase the bandwidth of learner models has been pioneered by Conati and colleagues (Bondareva et al. 2013; Conati et al. 2013; Conati and Merten 2007; Jaques et al. 2014; Kardan and Conati 2012; Muir and Conati 2012).

Conati et al. (2013) provide an excellent review of much of the existing work in this area. We can group the research into three categories: (1) offline-analyses of eye gaze to understand attentional processes, (2) modeling of attentional states, and (3) closed-loop systems that respond to attention in real-time. Offline-analysis of eye movements has enjoyed considerable attention in AIED, cognitive psychology, and educational psychology for several decades (e.g. Graesser et al. 2005; Hegarty and Just 1993; Mathews et al. 2012; Muir and Conati 2012; Ponce and Mayer 2014), so this area of research is relatively healthy. Online models of learner attention are just beginning to emerge (e.g., Bixler and D'Mello 2014, 2015; Blanchard et al. 2014; Bondareva et al. 2013; Conati and Merten 2007; Drummond and Litman 2010; Kardan and Conati 2012). Closed-loop attention-aware are few and far between (for a more or less exhaustive list, see D'Mello et al. 2012; Gluck et al. 2000; Sibert et al. 2000; Wang et al. 2006).

In summary, despite earlier calls for the importance of incorporating eye gaze in AIED learning environments (henceforth referred to as learning environments), uptake has been slow. We think that this is due to complexities in modeling an evasive construct like attention, the prohibitive costs of research-grade eye tracking, and the sheer number of other important problems that need to be solved. However, things are looking up in 2015. Recent research highlighting the central role of attention in learning coupled with advances in mental state estimation and low-cost eye tracking suggests that AIED can finally "attend to attention." However, it may need a roadmap to get there. Towards this end, we sketch out initial ideas towards attention-aware learning environments within the near term (5–10 years). We also suggest a few speculative long-term (10–25 years) ideas. Thus, this paper is part theoretical, part technical, and part forward-looking.

Organizing Framework

Attention can be thought of as a filter in that it has an object or location of focus. It can be driven by top-down goal-directed control or captured by bottom-up stimulus-driven processing (Egeth and Yantis 1997; Kinchla 1992). However, the locus of attention is not synonymous with where attention appears to be directed, because a person can be looking at one thing but thinking about something else entirely. In line with this, Table 1 outlines a 2×2 (direction of attention × content of thoughts) framework to organize attentional states during learning with technology. At a very basic level, we

Table 1 Organizing framework to differentiate between various attentional states

Content of thoughts	Direction of attention	
	Learning environment	Elsewhere

Content-related	*Overt attention* (sustained attention)
	    Focused attention
	    Alternating attention
	    Divided attention
		Covert attention
		    On-task conversation
		    Help seeking
		    Concentrating with eyes closed
		    Others….

Content-unrelated	*Covert inattention* (mind wandering)
	    Tune outs
	    Zone outs
		Overt inattention
		    Off-task
		    Distracted

can distinguish between attention that *appears* to be directed towards the learning environment or elsewhere. *Overt inattention* occurs when the learner directs attention elsewhere, as when the learner intentionally goes off-task or is distracted by external stimuli. Inattention can also be *covert*, when attention drifts away from the learning task to content-unrelated thoughts even though the learner may appear to be concentrating. These content-unrelated thoughts can be directed towards external factors (e.g., "the temperature in the room"), task-factors (e.g., "this tutor agent looks funny"), or something else entirely (e.g., "I wonder what's for dinner") (Stawarczyk et al. 2011). Mind wandering can occur both *intentionally* (e.g., tuning out) and *unintentionally* (zoning out).

An interesting situation arises when attention appears to be directed away from the learning environment, but the focus of thoughts is content-related; we refer to this as *covert attention*. For example, a learner could talk to a peer about a particular problem (on-task conversation), could engage in help seeking behaviors, or could close one's eyes and deeply reflect on the content. In contrast, *overt attention* occurs when attention is both directed toward the learning environment and consists of content-related thoughts. This is also referred to as *sustained attention*, which can take on different forms. *Focused attention* occurs when attention is directed towards a particular component of the learning environment. *Alternating attention* consists of rapidly switching attention between different interface components; for example, reading a sentence from the text, then looking at the image, back to the text, and so on. Finally, *divided attention* is the highest level of attention and involves simultaneously attending to multiple components of the environment (e.g., attending to the narration of a multimedia presentation while simultaneously processing an accompanying animated image).

We can now begin to think of a series of questions on attention during learning:

1. Does attention appear to be directed towards the learning environment or is it directed elsewhere?
2. If attention is directed elsewhere, are the thoughts content-unrelated (*overt inattention*) or are they related to the learning task (*covert attention*)?
3. If attention appears to be directed to the learning environment, is the learner actually attending to the learning task (*overt attention*), or is the learner focusing on off-task thoughts (*covert inattention*)?

4. In the case of covert inattention, is mind wandering intentional (*tuning out*) or unintentional (*zoning out*)?
5. In the case of overt attention, is attention *focused*, *alternating*, *divided*, and what is the *object or location of focus* at any given time?
6. Is the current attentional state beneficial with respect to task demands? If not, how can attention be redirected?

Architecture

We define an *attention-aware* learning environment as one that "detects and (re)directs the learner's attention in near real-time in order to achieve some desired outcome." This requires an *attentional computing* (computing involving attention) layer with the following three basic components: (1) a mechanism to monitor where attention appears to be directed, (2) a mechanism to diagnose the underlying attentional state, and (3) strategies to (re)direct attention as needed. We now sketch out basic mechanisms to implement these three components with respect to covert inattention, overt inattention, and overt attention. Covert attention is discussed later on in the paper.

Monitoring Where Attention Appears to be Directed

Eye tracking is perhaps the most direct method to identify where *visual attention* appears to be directed. Decades of scientific evidence has supported an *eye-mind link* that suggests a tight coupling between external information, attention, and eye gaze (Deubel and Schneider 1996; Hoffman and Subramaniam 1995; Rayner 1998). In the case of reading, this link has been demonstrated when lexical and semantic properties of words on a page predict, for example, which words will be fixated on and which will be skipped (Engbert et al. 2005; Just and Carpenter 1980; Rayner 1998; Reichle et al. 2003). In the context of scene processing, similar predictions can be made regarding which (and when) objects in a scene will be fixated on (Brockmole and Boot 2009; Brockmole and Henderson 2005, 2008; Currie et al. 2000). Although there are other indicators, such as physiology, gestures, and so on, these are all undifferentiated signals that encode information other than attention (e.g., arousal, communicative intent). As such, eye gaze is the best near-term indicator of where visual attention appears to be directed. We suggests that it might be strategic for near-term attentional-aware AIED systems to focus on eye gaze, and we do so in the remainder of this article. In the long-term, it is likely that brain-computer interfaces (BCIs) will have progressed to a point to complement or replace eye gaze.

We can broadly categorize eye movements as being gaze-stabilizing (*fixations* on an object) or gaze-shifting. Gaze-shifting movements can be further subdivided into *saccades* (quick jerky movements from one object to another) or *smooth pursuits* (movements tracking an object over time). There are other types of eye movements, such as vergence, vestibular, drifts, and microsaccades, but these are of less relevance to information processing tasks (Rayner 1998). Eye movements can be further divided into *global* vs. *local* movements. Global movements are independent of any specific visual stimulus (e.g., number of fixations, mean fixation duration), while local eye movements are stimulus-dependent (e.g., first pass fixations—the first time a word is

fixated on during reading). We can consider content-specific eye movements, such as the number of fixations on a particular content word, as being a special subset of local movements.

An eye tracker is a device that yields a series of eye gaze positions relative to some location on a display. We need fixation filtering algorithms to convert these raw eye gaze positions into a time series of fixations, saccades, and smooth pursuits (if applicable). Figure 1 displays fixations (circles) and saccades (lines connecting circles) obtained with the EyeTribe (see below) overlaid on a screenshot of Guru, a learning environment for high-school biology (Olney et al. 2012).

The high cost of eye trackers has traditionally relegated gaze tracking to the lab. This has rapidly changed with the recent advent (as late as 2013) of consumer-grade eye tracking devices that retail at a fraction of the cost of research-grade eye trackers (e.g., the EyeTribe for $99 and the Tobii EyeX for $150 compared to the thousands of dollars for a research-grade eye tracker). These new technologies afford the exciting possibility of applying decades of lab-based research on eye gaze, attention, and learning to develop attention-aware learning technologies that can be scaled for real-world use. This unprecedented technological trigger is yet another reason to focus on eye tracking in the short-term.

Diagnosing the Underlying Attentional State

Our goal is to diagnose the learner's attentional state from time series of fixations, saccades, and smooth pursuits. We consider different strategies/methods for the three main categories of attention—overt inattention, mind wandering, and sustained attention.
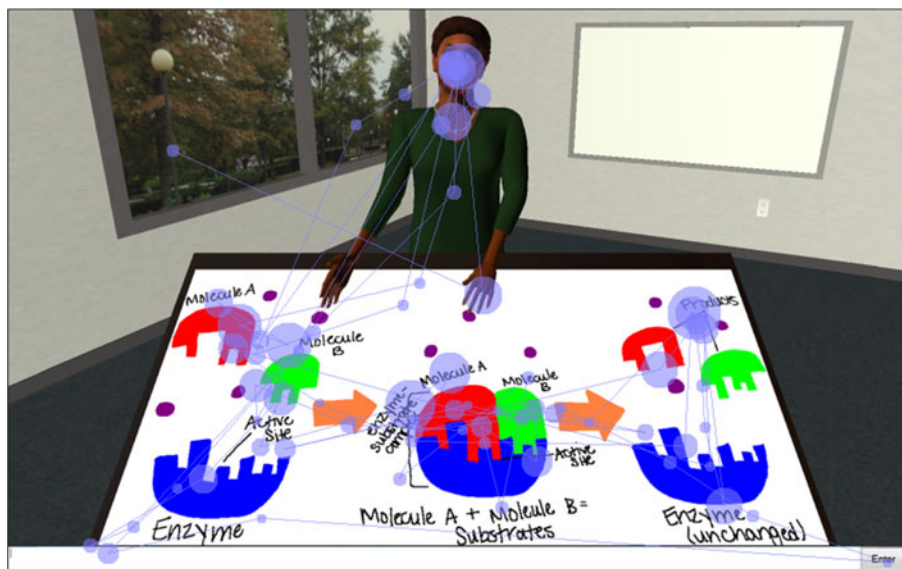

Fig. 1 Eye gaze obtained via the EyeTribe eye tracker overlaid on Guru tutor. Fixations are shown in circles and saccades are shown as *lines* connecting fixations

Overt Inattention We can diagnose overt inattention if eye gaze cannot be tracked for a certain period of time. This simple rule is based on three assumptions—the lack of valid eye gaze data is not due to gaze tracking errors (e.g., poor eye tracker calibration), the eyes are open, there is a need for visual attention. It does not take much imagination to conjure up counter-examples to these assumptions. For example, a misdiagnosis would occur if a learner is concentrating with his or her eyes closed, especially if there were no visual attention demands (e.g., the action is in the auditory channel). Nevertheless, this simple rule is likely to suffice in most (but not all) circumstances, as most learning environments do require focused visual attention.

Mind Wandering Detecting mind wandering is challenging as it is a form of 'looking without seeing' in that the eyes might be appropriately *externally* fixated but very little is being processed because attention is directed *internally*. Eye tracking is attractive for mind wandering detection because well-known relationships between eye movements and the external stimulus tend to break down during mind wandering. For example, participants are less likely to fixate, re-fixate, and regress (i.e., look backward through previously read text) when mind wandering compared to normal reading (Reichle et al. 2010). Blink-rates are also higher during mind wandering while reading (Smilek et al. 2010), ostensibly due to a reduction in the processing of external information during reading (because eyes are closed during blinks) (Bristow et al. 2005; Volkmann 1986). In line with this, researchers have had some success is using supervised learning approaches to detect mind wandering from eye gaze (Bixler and D'Mello 2014, 2015). Some of the other modalities considered include: peripheral physiology (Blanchard et al. 2014), speech patterns (Drummond and Litman 2010), as well as interaction and context cues (Franklin et al. 2011; Mills and D'Mello 2015). To date, mind wandering has been treated as a binary outcome, but it might be useful to distinguish between intentional (tune outs) versus unintentional (zone outs) mind wandering, as they need to be addressed somewhat differently.

Sustained Attention How can we detect if sustained attention is focused, alternating, or divided? We can discriminate between the first two types depending on whether gaze is *focused* on one component of the interface, or *alternates* among multiple components across a window of time. We would need to divide the visual display into multiple regions of interest (ROIs), and study the distribution of eye gaze on the ROIs across short time windows. Our selection of ROIs will depend on the level of granularity desired. For example, considering Fig. 2, we could select coarse-grained ROIs like the tutor vs. image vs. background or more fine-grained ones, such as the tutor's head, tutor's arms (for gesturing), an image chunk corresponding to enzyme, molecule A, etc. We can similarly ascertain if the learner is attending to the appropriate ROI or is correctly alternating between multiple ROIs.

In contrast to alternating attention among items, divided attention involves attending to multiple items at the same time. Humans have difficulty dividing attention within the same modality (e.g., looking at two items at the same time or listening to two sounds at once). Hence, divided attention is mostly cross-modal and typically is audio-visual in learning environments. We can track divided attention by the extent to which attention is synchronized across modalities. For example, in the case of the Guru interface shown in Fig. 2, the learner must attend to the speech of the animated tutor agent (auditory) in

tandem with the information displayed on the multimedia panel (visual). This would require shifting eye gaze when the tutor verbally refers to an image component or makes an explicit effort to redirect attention ("e.g., look at the enzyme on the far left"). The learner does not need to direct attention to every auditory cue of the tutor, but missing a large number of cues more than likely signals divided attention failures.

Attention (Re)direction Strategies

We must finally close the loop by (re)directing attention if we determine that the learner is in a suboptimal attentional state. This begs the question of what is 'suboptimal' vs. 'ideal' or 'optimal.'

Like everything in learning, the ideal attentional state varies as a function of the learner, learning task, and learning environment; hence, strategies to (re)direct attention are likely domain-dependent. However, there are also general group-level effects, so we can prescribe some high-level domain-independent strategies.

Overt Inattention There are many reasons why a learner might be overtly inattentive—they may be bored, disinterested, distracted, and so on. We can address momentary cases of inattention with simple audio-visual cues, say by appealing to the auditory channel if the visual channel appears to be distracted. For example, the use of the learner's first name (e.g., "Mary, what do you think about this problem") in spoken dialog systems should be effective in capturing attention a la the cocktail-party effect (Cherry 1953). Of course, triggering too many of these attentional reorientation cues in too short a period of time is likely to be annoying and even disruptive, so we would recommend a "less is more" strategy. If inattention is persistent, then the learning environment can suspend the current activity and suggest a new activity, a new topic, or even offer the learner a choice of what to do next. If all else fails, the system might even suggest that the learner take a break.

Mind Wandering Numerous studies have linked mind wandering with reduced performance as reported in a recent meta-analysis (Randall et al. 2014). One initial effect of mind wandering is that the learner fails to attend to a piece of information or a salient event in the learning environment. This knowledge deficiency can impair subsequent comprehension, so it should be corrected in the near-term. We can take a direct approach by reasserting the missed information (e.g., "Let me repeat that…"), or by directing attention to specific areas of the display. We could also ask a content-specific question (e.g., "What happens to the chromosomes when they duplicate" in the case of Guru), or ask the learner to complete a mini-activity. This form of interleaved questions and embedded activities can reduce mind wandering (Szpunar et al. 2013) as can asking learners to generate self-explanations (Moss et al. 2013). In general, the strategies to combat mind wandering share the common goals of: (a) capturing attention, (b) giving the learner an opportunity to reflect on the content/activity, and (c) providing an opportunity to correct any comprehension deficits due to mind wandering.

Sustained Attention Attention is a limited resource, so it should be beneficial to (re)direct the locus of attentional focus if suboptimal patterns are detected. For example,

when asked to diagnose malfunctions from descriptions of everyday devices (e.g., toasters, door locks), knowledgeable learners are more apt to focus attention on critical components compared to their less knowledgeable counterparts (Graesser et al. 2005). In this case, we could (re)direct attention to critical ROIs if learners appear to be focusing on tangential ones. Similarly, learners often ignore or only shallowly process hints provided by the learning environment (Muir and Conati 2012). We could address this by presenting the hint via a different modality (e.g., i.e., auditory) (Anderson 2002).

We can also (re)direct attentional patterns as they unfold over time. In particular, when processing an illustrated text, learners must alternate attention between the text and diagram in order to construct a coherent mental model that integrates the two representations (Hegarty and Just 1993). Their comprehension might suffer if they attend to one component (text or image) at the expense of the other for an extended period of time. Similarly, a breakdown in the temporal synchronization of auditory and visual attention when processing animations with narration should negatively influence comprehension. In either case, we can use simple cues to engender appropriate attentional patterns. For example, we can explicitly link keywords in a text with corresponding areas in an image in order to effectively cue alternate attention (Scheiter and Eitel 2010).

In general, if we can theoretically specify which eye gaze patterns correspond to 'ideal' or 'desired' attentional states, then deviations can be detected and corrected (Anderson 2002; Conati and Merten 2007). However, it is sometimes difficult to specify how attention should be deployed, especially in complex interfaces or when there are strong individual differences. In these situations, we can learn *how* attention should be deployed by performing a post-hoc analysis of the attentional patterns of successful vs. unsuccessful learners (Bondareva et al. 2013; Kardan and Conati 2012), presumably crossed with low vs. high domain knowledge. We can subsequently use the learned model to detect and address deviations from optimal paths when new learners use the system.

Case Studies

We now turn to two case studies that highlight key components of the attentional computing layer in attention-aware learning environments. The first focuses on overt inattention while the second addresses covert inattention or mind wandering.

Case Study 1: Addressing Momentary Episodes of Inattention

GazeTutor (D'Mello et al. 2012) is learning environment for biology. It has an animated conversational agent that provides spoken explanations on biology topics which are synchronized with annotated images (see Fig. 2a). The system uses a Tobii T60 eye tracker to detect inattention, which is assumed to occur when gaze was not on the tutor or image for at least five consecutive seconds. When this occurs, the system (a) interrupts its speech mid utterance, (b) directs learners to reorient their attention (e.g., "I'm over here you know"), and (c) repeats speaking from the start of the current utterance.

We conducted a small study to evaluate the effectiveness of GazeTutor. Forty eight learners (undergraduate students) completed a learning session on four biology topics with the attention-aware components enabled (experimental group) or disabled (control

group). We found that GazeTutor was successful in dynamically reorienting learners' attentional patterns towards the interface (see Fig. 2b). Importantly, learning gains for deep reasoning questions were significantly higher for the experimental group compared to the control group, but only for high aptitude learners. The results are important because they suggest that even the most basic attention-aware system can be effective in improving learning, at least for a subset of learners.

Case Study 2: Detecting and Responding to Mind Wandering

We recently developed an intelligent computerized reading interface that detects and corrects mind wandering in real-time. We used a supervised learning approach to detect mind wandering. Data used to train the detector was collected as 98 learners (Kopp et al. 2015) read a 57-page scientific text on surface tension in liquids (Boys 1895). Learners used the arrow key to navigate forward. Their gaze was tracked with a Tobii TX 300 eye tracker. Learners self-reported when they realized they were mind wandering throughout the reading session. A support vector machine was used to discriminate between mind wandering (pages with a self-report—32 %) and normal reading from eye-gaze using methods discussed in Bixler and D'Mello (2015). Importantly, we designed the model to generalize to new learners rather than optimizing to individual learners. The model had a precision of 69 % and a recall of 67 %, which we deemed to be sufficiently accurate for our purposes.

The mind wandering detector was then integrated into the computerized reading interface so as to provide real-time page-by-page estimates of the likelihood of mind wandering for new learners. The main strategy consisted on asking comprehension questions on the page where mind wandering was detected and providing opportunities to re-read if necessary. In line with this, two multiple choice questions were created for each of the 57 pages. Mind wandering detection occurred when the learner attempted to navigate to the next page. Eye gaze data from the previous page (the one just read) was submitted to the mind wandering detector, which provided an estimate of the likelihood that the learner was mind wandering. If the likelihood was determined to be sufficiently high (based on a probabilistic prediction), one of the questions (randomly selected) was presented to the learner. If the learner answered the question correctly, feedback was provided, and the learner could advance to the next page. If the learner answered incorrectly, the system encouraged the learner to re-read the page. The learner was then provided with a second
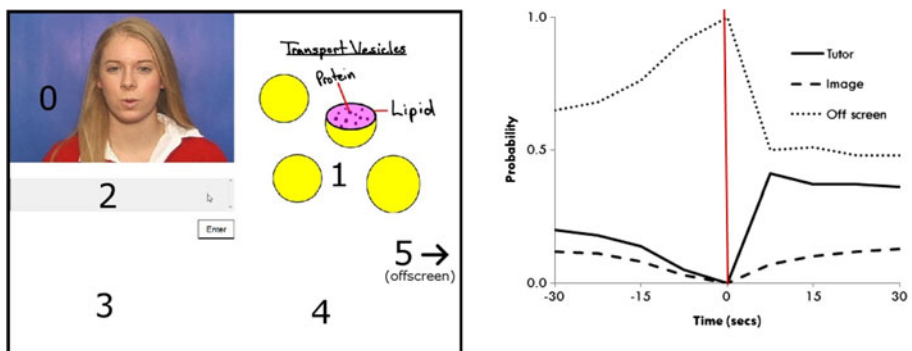

Fig. 2 **a** Gaze-tutor. Screen shot of interface on *left*. **b** Gaze before and after intervention on *right*

(randomly selected) question after re-reading. This second question could either be the same question or the alternate question for that page. Regardless of whether or not the learner answered the second question correctly, the system presented the learner with the next page of text. The efficacy of the intervention is currently being tested in an experiment that compares learners who received the intervention to a yoked-control condition. Preliminary results suggest that the system is effective in correcting comprehension deficits when the probability of mind wandering is high.

General Discussion

Attention is one of the core facets of human intelligence. The ability to monitor, share, and direct attention is a hallmark of human-human communication. Communication essentially breaks down when there is a lack of joint attention amongst communicators. The importance of attention also extends beyond human-human communication into the realm of human learning. Learners must have the ability to dynamically allocate attentional resources throughout the learning process if they are to learn effectively from 21st century learning environments, which are increasingly complex in a world mired with distractors (e.g., Facebook, twitter, email). However, learners are notoriously inadequate at sustaining and allocating scarce attentional resources in a manner that optimally meshes with the affordances of the learning environment and the learning task. AIED systems have come a long way in delivering individually-optimized instruction by modeling various aspects of the learner (e.g., knowledge, affect, disengagement, persistence—see edited volume Sottilare et al. (2013)), but they have yet to meaningfully model learner attention.

Building on the ideas of early visionaries (Anderson 2002; Gluck et al. 2000; Sibert et al. 2000), and on recent work on learner attentional modeling (Bondareva et al. 2013; Conati et al. 2013; Kardan and Conati 2012), we propose one foundational vision for the next 25 years of AIED. This vision consists of attention-aware learning environments that monitors and dynamically adapts to learner attention, thereby coordinating tacit (what the learner knows), external (what the learner does), and internal (what the learner attends to) behaviors. However, a futuristic vision without a plan to get there is not very useful. Therefore, we offered the following two basic contributions to scaffold the field of attentional-aware AIED: (1) we proposed a multicomponential organization of attention that integrates where attention appears to be directed with the content of internal thoughts, and (2) we fleshed out a three layer attentional-computing architecture consisting of eye tracking, in order to monitor where attention appears to be directed, computational techniques for real-time attentional state diagnosis, and interventions to (re)direct attentional focus.

We note two points of caution. The first is that the entire endeavor—from eye tracking, to attentional diagnosis, to identifying the ideal attentional state for a given situation, to prescribing the correct strategy—is fraught with ambiguity. Hence, attention (re)direction strategies should be used sparingly (i.e., when there is high confidence that the attentional state diagnosis is correct), should be 'fail-soft' in that they are not disruptive or harmful if delivered incorrectly, and should be implemented within probabilistic frameworks that can make decisions under uncertainty. Second, there is a temptation to hyper-optimize models to individual interfaces and tasks. This can lead to highly accurate models in the short term, but very few generalizable insights.

Therefore, generalizability and broad transferability of principles need to be key design constraints, not casual afterthoughts.

The paper has only scratched the surface on what is possible in an attentional-aware AIED systems. A small set of open questions and issues are listed below.

1. Can we further increase scalability by replacing already scalable consumer-grade eye trackers with even more scalable web-cam based eye tracking (Sewell and Komogortsev 2010)?

2. How do we integrate eye gaze, which provides information on where attention might be focused, with measures of physiological arousal or alertness (e.g., electrodermal activity)?

3. Can we incorporate information on the external context (i.e., measured via microphones and cameras) to discriminate between covert attention and overt inattention? (e.g., can we use speech recognition to discriminate between off-task vs. on task conversations?)

4. Can covert inattention be detected by directly monitoring brain signals? More broadly, can brain signals be used to complement or even replace existing modalities to track attention?

5. How do we integrate physiological sensing, context modeling, behavioral sensing, action dynamics, language and discourse to obtain unified multicomponential attentional models?

6. How can we make even finer-grained distinctions between the different forms of mind wandering (e.g., zone outs vs. tune outs)?

7. What is the best way to integrate models of attention with models of knowledge, affect, motivation, and metacognition?

8. How do we incorporate individual differences in "ideal" attentional patterns to increase adaptivity by tailoring attention (re)orientation strategies to the individual or to groups of individuals? And can these individual differences be automatically detected on the fly?

9. How do we design domain-independent attention diagnosis and (re)direction strategies for integration into generalized AIED frameworks?

10. Instead of merely reacting to attentional states, how can we leverage lessons from fields that are very successful at capturing, directing, and maintaining attention (e.g., film, art, games, and literature)?

11. As the models become increasingly complex, they run the risk of regulating too many aspects of learner behavior. What is the most effective way to balance the trade-off between external regulation and self-regulation of learning?

12. What can be learned from attentional management strategies of expert teachers, and how do we incorporate these insights into our models?

13. What is the best way to leverage recent advances in classroom learning analytics, such as modeling attention of entire classes of students (Raca et al. 2015), or automated analyses of teacher instruction (D'Mello et al. 2015), into attention-aware AIED?

We believe that addressing these questions while implementing attention-aware learning environments should yield fundamental insights on attentional processes during learning. Thus, in addition to the practical goals of improving learning with

innovative technologies, research on attention-aware AIED will also make foundational theoretical contributions to the science of learning. Indeed many interesting discoveries await discovery once we give eyesight (attentional computing) to the blind (current learning environments).

References

Anderson, J. R. (2002). Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science, 26*(1), 85–112.

Baker, R., Corbett, A., Koedinger, K., & Wagner, A. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system". *Proceedings of ACM CHI 2004: Computer-Human Interaction Conference* (pp. 383–390). New York: ACM.

Bixler, R., & D'Mello, S. (2014). Toward fully automated person-independent detection of mind wandering. In V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, & G.-J. Houben (Eds.), *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization* (pp. 37–48). Switzerland: Springer International Publishing.

Bixler, R., & D'Mello, S. K. (2015). Automatic gaze-based detection of mind wandering with metacognitive awareness. In F. Ricci, K. Bontcheva, O. Conlan, & S. Lawless (Eds.), *Proceedings of the 23rd International Conference on User Modeling, Adaptation, and Personalization (UMAP 2015)* (pp. 31–43). Switzerland: Springer International Publishing.

Blanchard, N., Bixler, R., & D'Mello, S. K. (2014). Automated physiological-based detection of mind wandering during learning. In S. Trausan-Matu, K. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014)* (pp. 55–60). Switzerland: Springer International Publishing.

Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., & Bouchet, F. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning. In K. Yacef, C. Lane, J. Mostow, & P. Pavlik (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED 2013)* (pp. 229–238). Berlin: Springer.

Boys, C. V. (1895). *Soap bubbles, their colours and the forces which mold them: Society for promoting christian knowledge.*

Bristow, D., Frith, C., & Rees, G. (2005). Two distinct neural effects of blinking on human visual processing. *NeuroImage, 27*(1), 136–145.

Brockmole, J. R., & Boot, W. R. (2009). Should I stay or should I go? Attentional disengagement from visually unique and unexpected items at fixation. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 808.

Brockmole, J. R., & Henderson, J. M. (2005). Object appearance, disappearance, and attention prioritization in real-world scenes. *Psychonomic Bulletin & Review, 12*(6), 1061–1067.

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: effects of object–scene consistency. *Visual Cognition, 16*(2–3), 375–390.

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

Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: an empirical evaluation. *Knowledge-Based Systems, 20*(6), 557–574. doi:10.1016/j.knosys.2007.04.010.

Conati, C., Aleven, V., & Mitrovic, A. (2013). Eye-tracking for student modelling in intelligent tutoring systems. In R. Sottilare, A. Graesser, X. Hu, & H. Holden (Eds.), *Design recommendations for intelligent tutoring systems—Volume 1: Learner modeling* (pp. 227–236). Orlando: Army Research Laboratory.

Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Attention, Perception, & Psychophysics, 62*(4), 673–683.

D'Mello, S. K., Olney, A. M., Blanchard, N., Samei, B., Sun, X., Ward, B., & Kelly, S. (2015). *Multimodal capture of teacher-student interactions for automated dialogic analysis in live classrooms.* Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI 2015), New York.

Damrad-Frye, R., & Laird, J. D. (1989). The experience of boredom: the role of the self-perception of attention. *Journal of Personality and Social Psychology, 57*(2), 315.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research, 36*(12), 1827–1837.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: a gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*(5), 377–398.

Drummond, J., & Litman, D. (2010). In the zone: Towards detecting student zoning out using supervised machine learning. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems. (Vol. 6095)* (pp. 306–308). Berlin: Springer-Verlag.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology, 48*(1), 269–297.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777.

Forbes-Riley, K., & Litman, D. (2011). When does disengagement correlate with learning in spoken dialog computer tutoring? In S. Bull & G. Biswas (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 81–89). Berlin: Springer.

Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin & Review, 18*(5), 992–997.

Gluck, K. A., Anderson, J. R., & Douglass, S. A. (2000). Broader bandwidth in student modeling: What if ITS were "eye" TS? In C. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th international conference on intelligent tutoring systems* (pp. 504–513). Berlin: Springer.

Graesser, A., Lu, S., Olde, B., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: comprehending illustrated texts on devices when the devices break down. *Memory and Cognition, 33*, 1235–1247. doi:10.3758/BF03193225.

Hegarty, M., & Just, M. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717–742.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Attention, Perception, & Psychophysics, 57*(6), 787–795.

Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). *Predicting affect from gaze data during interaction with an intelligent tutoring system.* Paper presented at the Intelligent Tutoring Systems.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Kardan, S., & Conati, C. (2012). Exploring gaze data for determining user learning with an interactive simulation. In S. Carberry, S. Weibelzahl, A. Micarelli, & G. Semeraro (Eds.), *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization (UMAP 2012)* (pp. 126–138). Berlin: Springer.

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology, 43*, 711–743.

Knoblich, G., Öllinger, M., & Spivey, M. J. (2005). Tracking the eyes to obtain insight into insight problem solving. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 355–375). Oxford University Press.

Kopp, K., D'Mello, S., & Mills, C. (2015). Influencing the occurrence of mind wandering while reading. *Consciousness and Cognition, 34*(1), 52–62.

Mathews, M., Mitrovic, A., Lin, B., Holland, J., & Churcher, N. (2012). Do your eyes give it away? Using eye tracking data to understand students' attitudes towards open student model representations. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K.-K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 422–427). Berlin: Springer.

Mills, C., & D'Mello, S. K. (2015). Toward a real-time (day) dreamcatcher: Detecting mind wandering episodes during online reading. In C. Romero, M. Pechenizkiy, J. Boticario & O. Santos (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*. International Educational Data Mining Society.

Moss, J., Schunn, C. D., Schneider, W., & McNamara, D. S. (2013). The nature of mind wandering during reading varies with the cognitive control demands of the reading strategy. *Brain Research, 1539*, 48–60.

Muir, M., & Conati, C. (2012). An analysis of attention to student–Adaptive hints in an educational game. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 112–122). Berlin: Springer.

Olney, A., D'Mello, A., Person, N., Cade, W., Hays, P., Williams, C., . . . Graesser, A. (2012). Guru: A computer tutor that models expert human tutors. In S. Cerri, W. Clancey, G. Papadourakis & K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 256–261). Berlin/Heidelberg: Springer-Verlag.

Olney, A., Risko, E. F., D'Mello, S. K., & Graesser, A. C. (2015). Attention in educational contexts: The role of the learning task in guiding attention. In J. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 623–642). Cambridge: MIT Press.

Ponce, H. R., & Mayer, R. E. (2014). Qualitatively different cognitive processing during online reading primed by different study activities. *Computers in Human Behavior, 30*(1), 121–130.

Raca, M., Kidzinski, L., & Dillenbourg, P. (2015). *Translating head motion into attention-towards processing of student's body-language.* Paper presented at the Proceedings of the 8th International Conference on Educational Data Mining.

Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: a theory-driven meta-analysis of attention regulation. *Psychological Bulletin, 140*(6), 1411–1431.

Rapp, D. N. (2006). The value of attention aware systems in educational settings. *Computers in Human Behavior, 22*(4), 603–614.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: comparisons to other models. *Behavioral and Brain Sciences, 26*(4), 445–476.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science, 21*(9), 1300.

Roda, C., & Thomas, J. (2006). Attention aware systems: theories, applications, and research agenda. *Computers in Human Behavior, 22*(4), 557–587. doi:10.1016/j.chb.2005.12.005.

Scheiter, K., & Eitel, A. (2010). The effects of signals on learning from text and diagrams: How looking at diagrams earlier and more frequently improves understanding. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *6th International Conference on Diagrammatic Representation and Inference* (pp. 264–270). Heidelberg: Springer.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). New York: Cambridge University Press.

Sewell, W., & Komogortsev, O. (2010). *Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network.* Paper presented at the Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, Austin, TX.

Sibert, J. L., Gokturk, M., & Lavine, R. A. (2000). The reading assistant: eye gaze triggered auditory prompting for reading remediation. *Proceedings of the 13th annual ACM symposium on User interface software and technology* (pp. 101–107). New York: ACM.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*, 487–518.

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010). Out of mind, out of sight: eye blinking as indicator and embodiment of mind wandering. *Psychological Science, 21*(6), 786–789.

Sottilare, R., Graesser, A., Hu, X., & Holden, H. K. (Eds.). (2013). *Design recommendations for intelligent tutoring systems: Volume 1: Learner modeling*. Orlando: U.S. Army Research Laboratory.

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering: phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica, 136*(3), 370–381.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110*(16), 6313–6317.

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99.

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25*(3), 785–791. doi:10.1016/j.chb.2009.02.007.

Volkmann, F. C. (1986). Human visual suppression. *Vision Research, 26*(9), 1401–1416.

Wang, H., Chignell, M., & Ishizuka, M. (2006). Empathic tutoring software agents using real-time eye tracking. *Proceedings of the 2006 symposium on Eye tracking research &applications* (pp. 73–78). New York: ACM.