Eye Movements in Language and Cognition:

A Brief Introduction

Daniel C. Richardson
Department of Psychology, Stanford University

Rick Dale
Department of Psychology, Cornell University

Michael J. Spivey
Department of Psychology, Cornell University

Send correspondence to:
Daniel C. Richardson
Psychology Department
420 Jordan Hall
Stanford, CA 94305
The role of eye movements in the visual system

The eye independently evolved over 40 times in nature (Fernald, 1997), yet strikingly all animals with developed visual systems actively control their gaze using eye or head movements (Land, 1995; Treue, 2001). Indeed, it can be argued that the most frequent behaviour of human beings is movement of the eyes (Bridgeman, 1992). This ‘ceaseless twitching’, as one early researcher described it (Stratton, 1906), is the visual system’s solution to the huge amount of available visual information and limited processing resources. The human eye covers a visual field of about 200°, but receives detailed information from only 2° (Levi, Klein, & Aitsebaomo, 1985). This region, about the size of a thumbnail at arm’s length, is called the fovea, and is jerked around at speeds of up to 500° a second, during which its sensitivity drops to near blindness levels (Matin, 1974; Thiele, Henning, Kubischik, & Hoffmann, 2002). During the 200-300 milliseconds it is at rest, however, over 30,000 densely packed photoreceptors in the fovea provide high acuity color vision. Eye movements direct this information conduit to relevant portions of the world, and are therefore fundamental to the operation of the visual system.

There are several classes of eye movements. Of most interest to the cognitive psychologist are saccades, the rapid, ballistic movements that move the eye around the visual field roughly 3-4 times a second. Other classes of eye movements serve to maintain fixation despite head, body or object motion, and to correct for muscle drift and inaccuracy. Typically, experimental psychologists measure when saccades are launched, where they land, and how long the eye stays there. The pattern of locations that saccades visit is termed the scanpath, the duration between a stimulus onset and a saccade is called...
the ‘saccade latency’, and the amount of time spent looking at a particular location is called the fixation duration. In this chapter we will show how eye movement metrics such as these have yielded a rich skein of data relating to cognitive processes, and suggest ways that they can be further mined by experimental psychologists.

**The value of eye movements to cognitive psychologists**

Eye movements are uniquely poised between perception and cognition. They are central to the function of the visual system, but for such scanning to be efficient, it cannot be simply a random sample of the visual world. To be useful, eye movements must be related to an organism’s memories, expectations and goals. Consequently, eye movements are driven equally by bottom-up perceptual properties of the world and top-down cognitive processes. This role in the perception-action cycle makes saccadic behaviour particularly informative for the experimental psychologist.

There are several specific characteristics of eye movements which also prove to be of great practical and theoretical benefit. Saccades occur roughly 3-4 times per second. During the response time of a typical experimental task then, eye-tracking data can provide a semi-continuous record of regions of the visual field that are briefly considered relevant for carrying out an experimental task. Crucially, this record provides data during the course of cognitive processing, not merely after processing is complete, as is often the case with more conventional measures. Eye-tracking data thus provide not only behavioral end products of our cognitive processes but also clues to the process through which they are achieved. Importantly, this sensitive semi-continuous measure of cognitive processing can also be used in ways that do not interrupt task processing with
requests for metacognitive reports or other overt responses. Thus eye-tracking allows for
certain degree of ecological validity in task performance, as the responses it collects are
ones that typically occur regardless of experimenters’ instructions and participants’
intent.

Moreover, eye movements exhibit a unique sensitivity to partially active representations
that may not be detectable by most other experimental measures, or even result in any
other overt behaviour. Since eye movements are extremely fast, quickly corrected, and
metabolically cheap, compared to other motor movements, they have a much lower
threshold for being triggered. Hence, briefly partially-active representations -- that might
never elicit reaching, speaking, or even internal monolog activity because they fade
before reaching those thresholds -- can nonetheless occasionally trigger an eye movement
that betrays this otherwise-latent momentary consideration of that region of the visual
display as being potentially relevant for interpretation and/or action. For example, in a
classic experiment in psychology, two speech sounds that vary continuously between
“ba” and “pa” are categorically perceived by participants, who report hearing either a
“ba” or a “pa” and respond by pressing a corresponding button (Liberman, Harris,
Hoffman, & Griffith, 1957). However, McMurray and colleagues (McMurray,
Tannenhaus, Aslin, & Spivey, 2003) showed that eye movements between the two
response buttons increased when the speech sound was near the ba/pa boundary. Thus,
the graded nature of the perceptual process, which is lost in a categorical response, is
revealed in the time course of eye movements.

Eye movements have a long and successful history as a window into perceptual and
cognitive processing. The following sections present a subset of that research that would
be of particular interest to cognitive linguists (for a broader review, see Richardson & Spivey). First, we will briefly describe how eye movements reveal psychological processes in everyday tasks of perception and memory. Then we will show how very similar cognitive processes produce similar eye-movement patterns in more ‘offline’ situations, where relevant visual stimuli may not even be present. Lastly, we shall turn to the many ways in which eye movements reveal facts about language processing, from the mechanics of reading, to the integration of visual and verbal information, to the conceptual representation of narratives and metaphors.

**Perception and action**

Visual attention is not always coincident with eye position. Posner, Snyder, and Davidson (Posner, 1980) demonstrated that participants’ covert visual attention can be dissociated from the fovea when they are explicitly instructed to not move their eyes. However, it is highly likely that spatial attention and saccade planning are closely coupled during natural unconstrained eye movement (Findlay & Gilchrist, 1998). There is behavioural evidence that covert attention directed in one direction can lead to deviations in orthogonal saccades (B. Sheliga, Riggio, Craighero, & Rizzolatti, 1995), and neuropsychological evidence from single cell recordings suggesting that they utilize overlapping neural systems (Corbetta et al., 1998). Moreover, planning a saccade toward a location improves processing at that location, regardless of whether or not the saccade is launched (Hoffman & Subramaniam, 1995; B. M. Sheliga, Riggio, & Rizzolatti, 1994; Shepherd, Findlay, & Hockey, 1986), and indeed, evidence shows that microstimulation of neurons in the frontal eye fields can cause both a saccade to a certain location (Robinson & Fuchs, 1969), and, with a lower level of simulation, an absence of eye
movement, but improved stimulus detection at that location (Moore & Armstrong, 2003; Moore & Fallah, 2001).

How are eye movements and visual attention directed around a visual stimulus? In general, when viewing a static scene, the eyes appear to be driven by both visual properties of the stimulus and top-down effects of knowledge and expectations (Henderson, 2003). For example, Buswell (1935) showed that a viewer will pay scant attention to solid regions of colour in a painting, and instead will tend to fixate regions of contrast and high spatial frequency; top-down effects will be seen in the viewers’ saccades to semantically important regions of the painting, such as faces, and the ways in which a naïve viewer will inspect the painting differently from an art expert. These findings have been recently replicated on a large scale. Wooding and colleagues installed an autonomous eye tracker in a public museum in London, and collected data from over 5000 subjects looking at works from the National Gallery (Wooding, 2002; Wooding, Muggelstone, Purdy, & Gale, 2002). They too found that only a small set of regions in a work of art were reliably fixated by viewers.

We live a world more dynamic and interactive than an art gallery, however. How are eye movements integrated with action in the course of everyday tasks? With the advent of headband-mounted eye-tracking, which allows natural movement of the entire body, this real-time measure of perceptual and cognitive processing has been applied to a number of more richly interactive, and ecologically valid, experimental tasks and paradigms. Eye movements can even reveal the everyday strategies we employ while carrying out basic tasks, such as making a sandwich. For example, Land and Hayhoe (1999) found that eye movements are tightly linked with moment-to-moment goals and sub-tasks. Task-related
Eye movements in Language and Cognition – Richardson, Dale & Spivey

fixations illuminating visual memory processes have been examined in detail using the block-copying task developed by Ballard, Hayhoe and colleagues (Ballard, Hayhoe, & Pelz, 1995; Ballard, Hayhoe, Pook, & Rao, 1997; Hayhoe, Bensinger, & Ballard, 1998).

Ballard et al. (1995) recorded participants’ eye movements during a block-pattern copying task, with a model pattern, a resource of blocks, and a workspace in which to copy the model. The participants’ hand actions were recorded, and a headband-mounted eye tracker recorded their eye movements to obtain a window on the strategy used in the task. One method participants could use is to look at the model area and memorize the pattern; each block could then be located in the resource area and placed in the workspace. A second method, which is a less memory-intensive option, would be to remember the color and location of one block from the model, collect it from the resource, place it in the workspace, and then consult the model again for the next block. The strategy used by participants, however, most often entailed the minimal possible memory demands. Participants would commonly fixate the model, then fixate and pickup a correctly colored block from the resource area, fixate the model yet again, and then place the block in the workspace. Thus, two fixations per block were made on the model -- one to extract color information, one to extract relative spatial location information.

Eye movements thus reveal a cognitive process of “indexing,” whereby the location of an object is maintained in working memory, and other properties can be “looked up” in the environment as they are needed, moment by moment, during a task (Ballard et al., 1997). For example, in a computerized, gaze-contingent version of the block-copying task, the color of a block was changed during a saccade (Hayhoe et al., 1998). The participants
rarely noticed this property change, demonstrating that they had not encoded the information, but instead relied upon the fact that an eye movement could access it when required.

**Cognition**

A general case can be made that ‘offline’ cognitive processes such as remembering, imagining and reasoning may employ many of the same mechanisms as ‘online’ perceiving and acting in the world (Barsalou, 1999; Damasio, 1989; Kosslyn, Behrmann, & Jeannerod, 1995; Martin, 2001; Ryle, 1949). Certainly, we will see here that eye movement patterns during cognitive activity bear a striking resemblance to those during the perception and manipulation of objects in the world. This continuity between perception and cognition can be exploited by psychologists, who can use overt eye-movement behaviour to investigate internal mental processes.

A clear example of this parallel exists between Ballard et al’s (1995) task where subjects moved around blocks, and a series of experiments where subjects remembered a series of verbally presented facts. Richardson and Spivey (2000) presented four talking heads in sequence, in the four quadrants of the screen, each reciting an arbitrary fact and then disappearing (e.g., “Shakespeare’s first plays were historical dramas. His last play was The Tempest.”). With the display completely blank except for the lines delineating the four empty quadrants, a voice from the computer delivered a statement concerning one of the four recited facts, and participants were instructed to verify the statement as true or false (e.g., “Shakespeare’s first play was The Tempest.”).
While formulating their answer, participants were twice as likely to fixate the quadrant that previously contained the talking head that had recited the relevant fact than any other quadrant. Despite the fact that the queried information was delivered auditorily, and therefore cannot possibly be visually accessed via a fixation, participants systematically fixated blank regions of space. This result was replicated when the talking heads were replaced by four identical spinning crosses.

Moreover, in a ‘tracking’ condition (Richardson & Kirkham, 2004), participants viewed the grid through a virtual window in the center of the screen. Behind this mask, the grid moved, bringing a quadrant to the center of the screen for fact presentation. Then, during the question phase, the mask was removed. Even in this case, when the spinning crosses had all been viewed in the center of the computer screen, requiring no eye movements, and the relative locations of the quadrants implied by translation, participants continued to treat the quadrant associated with the queried fact as conspicuously worthy of overt attention. In fact, even if the crosses appear in empty squares which move around the screen following fact delivery, participants spontaneously fixate the square that was associated with the fact in its new location (Richardson & Kirkham, 2004). The behaviour of associating events and information with a moving location, and re-fixating that location when the information is relevant has been termed ‘spatial indexing’.

Remarkably, there is evidence of spatial indexing behaviour in the eye movements of infants as young as 6 months of age (Richardson & Kirkham, 2004).

When subjects listened to pieces of semantic information, they associated them with spatial indexes, just as the participants in Ballard and colleagues’ block moving task did for the blocks they were manipulating (Richardson & Spivey, 2000; Spivey, Richardson,
& Fitneva, 2004). As many researchers have argued (Ballard et al., 1997; O’Regan, 1992; Pylyshyn, 1989, 2001), deictic pointers can be used in visuomotor routines to conserve the use of working memory. In Brooks’ (1991) words, the ‘world can be used as its own best representation’. Instead of storing all the detailed properties of an object internally, one can simply store an address, or pointer, for the object’s location in the environment, via a pattern of activation on an attentional/oculomotor salience map in parietal cortex (Duhamel, Colby, & Goldberg, 1992), along with a spatial memory salience map in prefrontal cortex (Chafee & Goldman-Rakic, 1998, 2000; Goldman-Rakic, Chafee, & Friedman, 1993). If this spatial pointer is associated with some kind of coarse semantic information, e.g., a pattern of activation in one of the language cortices, or auditory cortex, or even visual cortex, then the spatial pointer can be triggered when sensory input activates that semantic information. Such pointers allow the organism to perceptually access relevant properties of the external world when they are needed.

It actually should not be surprising that an embodied working memory system using deictic pointers would attempt to index information from events that are over and done with. The pointer doesn’t “know” that the sought-after information at its address is long gone precisely because it has offloaded that knowledge onto the environment -- it wouldn’t be a pointer otherwise. These eye movement findings demonstrate the robustness and automaticity with which spatial indices are relied upon in order to employ the body’s environment as sort of noticeboard of ‘virtual post-it notes’ that complement our internal memory.

However, many complex tasks we face on a daily basis do not necessarily involve indexing of relevant objects in a task space. For example, producing/hearing descriptions
of far away scenes or events, gossip about people who are absent, and discussions of abstract concepts, do not involve explicit reference to visible elements of the immediate situational context. An important question concerns the extent to which eye movements may be indicative of imagery processes when carrying out such tasks. How are eye movements implicated in visualizing a complex story or description? Will scanning of the visuo-spatial backdrop that is available to a listener be at all relevant during comprehension of language that refers to things that are not visually co-present with the speech?

In a headband-mounted eye-tracking experiment, Spivey and Geng (experiment 12001; see also Spivey, Tyler, Richardson, & Young, 2000) recorded participants’ eye movements while they listened to spoken descriptions of spatiotemporally dynamic scenes and faced a large white projection screen that took up most of their visual field. For example, “Imagine that you are standing across the street from a 40-story apartment building. At the bottom there is a doorman in blue. On the 10th floor, a woman is hanging her laundry out the window. On the 29th floor, two kids are sitting on the fire escape smoking cigarettes. On the very top floor, two people are screaming.” While listening to the italicized portion of this passage, participants made reliably more upward saccades than in any other direction. Corresponding biases in spontaneous saccade directions were also observed for a downward story, as well as for leftward and rightward stories. (A control story, describing a view through a telescope that zooms in closer and closer to a static scene, elicited about equal proportions of saccades in all directions). Thus, while looking at ostensibly nothing, listeners’ eyes were doing something similar to what they would have done if the scene being described were actually right there in front
of them. Instead of relying solely on an internal “visuospatial sketchpad” (Baddeley, 1986) on which to illustrate their mental model of the scene being described, participants also recruited the external environment as an additional canvas on which to depict the spatial layout of the imagined scene.

Although eye movements may not be required for vivid imagery (Hale & Simpson, 1971; Ruggieri, 1999), it does appear that they often accompany it (Antrobus & Antrobus, 1969; Brandt & Stark, 1997; Demarais & Cohen, 1998; Hebb, 1968; Laeng & Teodorescu, 2002; Neisser, 1967). Early empirical investigations found that the frequency of eye movements increases during mental imagery, particularly that of a spatial nature (Clark, 1916; Goldthwait, 1933; Perky, 1910; Stoy, 1930; Totten, 1935); and an increase in rapid fluttering of the eyes while sleeping correlates with vividness of dreams (Antrobus & Antrobus, 1969; Goodenough, Shapiro, Holden, & Steinschriber, 1959; Roffwarg, Dement, Muzio, & Fisher, 1962). But what is it that the eyes are trying to do in these circumstances? Obviously, it is not the case that the eyes themselves can actually externally record this internal information. When the eyes move upward from the imagined 10th floor of the apartment building to the imagined 29th floor, no physical mark is left behind on the external location in the environment that was proxying for that 10th floor.

In the case of Spivey and Geng’s (2001) eye movements during imagistic spoken narrative comprehension, a few pointers allocated on a blank projection screen will obviously not make reference to any external visual properties, but they can still provide perceptual-motor information about the relative spatial locations of the internal content associated with the pointers (see also Altmann & Kamide, 2004). If one is initially
thinking about $x$ (e.g., the 10th floor) and then transitions to thinking about $y$ (e.g., the 29th floor), then storing in working memory the relation above $(y,x)$ may not be necessary if the eye movements, and their allocation of spatial indices, have embodied and externalized that spatial relationship in the environment already (cf. Pylyshyn, 1989).

In this way, a “low-level” motor process, such as eye movements, can actually do some of the work involved in the “high-level” cognitive act of representing spatial relations in visual imagery elicited by linguistic input. Eye movement data thus reveal a powerful demonstration of how language about things not co-present is interfaced with perceptual-motor systems that treat the linguistic referents as if they were co-present.

It seems clear from the evidence presented here that eye movements are a rich source of information about cognitive processing, even when the relevant items are not physically present, but are recalled from memory or merely described in memory. Although some researchers today argue on the basis of null results that eye movements are not really indicative of cognitive processes at all (Anderson, Bothell, & Douglass, 2004), other researchers are demonstrating that eye movements can reveal not just which cognitive representation might be active, but how they are being manipulated. For example, eye movements appear to have a relationship to the reasoning process in mechanical problem solving (Hegarty, 1992; Hegarty & Just, 1993; Hodgson, Bajwa, Owen, & Kennard, 2000; Rozenblit, Spivey, & Wojslawowicz, 2002) and insight problem solving (Grant & Spivey, 2003; Jones, 2003; Knoblich, Ohlsson, & Raney, 2001).
Language

Language processing encompasses a spectrum of phenomena, from the largely perceptual aspects of word identification, to the largely conceptual aspects of metaphor understanding. As one might imagine from the preceding sections, eye movements can provide insight at each of these levels. The most apparent, and most studied, link between eye movements and language is in the process of reading.

The general characteristics of eye movements during reading have been studied in great depth over the past quarter century (for thorough reviews, see Rayner, 1978, 1998). This methodology has revealed a number of important facts about how people’s eyes move when they read. For example, the eyes rest in fixation for approximately 200-250 milliseconds during reading. Saccades between fixations span an average about 2 degrees of visual angle, although this is better expressed here in terms of a span of 7 to 9 letter spaces, since the number of letters covered remains largely invariant despite differences in text size or distance (Morrison & Rayner, 1981). The chances of an individual word being fixated vary according to whether it is a content word (85%) or a function word (35%) (Carpenter & Just, 1983), and in relationship to the length of the word, with 2-3 letter words being skipped 75% or the time, but 8 letter words fixated almost always (Rayner & McConkie, 1976). Eye movements also vary as a function of the syntactic and conceptual difficulty of the text (Ferreira & Clifton, 1986; Rayner, Sereno, Morris, Schmauder, & et al., 1989). Although readers typically move their eyes forward when reading, approximately 10-15% of saccades move backward, fixating previous letters or words. These regressive saccades are thought to be related to difficulties in processing an individual word, or difficulties in processing the meaning or structure of a sentence; in
these cases, readers can often accurately re-fixate the part of the text that generated confusion (Murray & Kennedy, 1988).

These features of eye movements during reading -- gaze durations, saccade lengths, occurrence of regressions, and a number of variations on these measures -- can be used to infer moment-by-moment cognitive processing of a text by the reader (Just & Carpenter, 1980; Rayner et al., 1989). Details of the cognitive processes of pronoun resolution and co-reference, word frequency, lexical ambiguity, syntactic ambiguity, as well as the influence of semantic and discourse context on these processes, can all by gleaned from analyses of eye-movement patterns (Rayner, 1998; Tanenhaus & Trueswell, 1995).

Light, headband mounted eye trackers have allowed researchers to extend the online measurement of language processing beyond reading, to the perception and understand of spoken language in a rich, naturalistic visual context. One field of research begins with the feature of eye movements, noted above, that participants will often look briefly at an object that is initially considered relevant for action, and then quickly re-fixate their eyes on another object that becomes the actual target of the action. This feature has been exploited to study many factors in the time course of speech processing and language understanding.

For example, Spivey and colleagues (Spivey & Tanenhaus, 1998) sat participants in front of a display of objects such as a candle, bag of candy, a pencil, and a spoon. The participants were then and instructed to “Pick up the candy,”. About a third of the time participants fixated the candle for a couple hundred milliseconds before looking to and reaching for the candy. Participants typically denied looking to the candle at all, and yet
their eye movements revealed a process substantially different from their conscious report and their manual action. This kind of brief interference between similar sounding object names occurs not just for cohorts but also for rhymes (Allopenna, Magnuson, & Tanenhaus, 1998), as well as for novel words from an artificial lexicon (Magnuson, Tanenhaus, Aslin, & Dahan, 2003), and even for words that sound similar across two different languages (Marian & Spivey, 2003; Spivey & Marian, 1999). It appears that the acoustic uptake of spoken input is continuously mapped onto visually relevant lexical representations, such that partial phonological matches to the names of multiple visual objects induces competition between partially active representations, in a system something like interactive processing in the TRACE connectionist model of spoken word recognition, (Elman & McClelland, 1988; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland & Elman, 1986).

A similar influence of visual context is observed with temporary ambiguities that arise across words, in the syntax of a sentence. When presented with a display containing an apple on a towel, another towel, and an empty box, and then instructed to “Put the apple on the towel in the box,” participants often looked briefly at the irrelevant lone towel near the end of the spoken instruction before returning their gaze to the apple, grasping it, and then placing it inside the box (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey Knowlton, Eberhard, & Sedivy, 1995). (With unambiguous control sentences, such as “Put the apple that’s on the towel in the box,” they almost never looked at the irrelevant lone towel.) In this case, the syntax is ambiguous as to whether the prepositional phrase “on the towel” is attached to the verb “put” (as a movement destination) or to the noun “apple” (as a modifier). Given the actions afforded by the
display, the latter syntactic structure is the correct one. However, people tend to have a
bias toward interpreting an ambiguous prepositional phrase as attached to the verb
(Rayner, Carlson, & Frazier, 1983), at least when it is an action verb like “put” (Spivey-
Knowlton & Sedivy, 1995). Thus, the brief fixation of the irrelevant lone towel indicates
a temporary partially-activated incorrect parse of the sentence. To demonstrate the
influence of visual context on this syntactic ambiguity resolution process, the display was
slightly altered to include a second apple (resting on a napkin). In this case, the visual
co-presence of the two potential referents for the phrase “the apple” should encourage the
listener to interpret the ambiguous prepositional phrase “on the towel” as a modifier (in
order to determine which apple is being referred to) rather than as a movement
destination (Altmann & Steedman, 1988; Spivey & Tanenhaus, 1998). And, indeed, with
this display, participants rarely fixated the irrelevant lone towel, indicating that visual
context had exerted an immediate influence on the incremental syntactic parsing of the
spoken sentence (Knoeferle, Crocker, Scheepers, & Pickering, 2004; Spivey et al., 2002;
Tanenhaus et al., 1995).

The word-by-word interfacing between spoken language and visual perception is also
evidenced by reference resolution with complex noun phrases. Eberhard, Spivey-
Knowlton, Sedivy, and Tanenhaus (1995) presented participants with a display of blocks
of various shapes, colors, and markings, and gave them instructions like “Touch the
starred yellow square.” When the display contained only one starred block, participants
often fixated on the target block before the head noun of the noun phrase had even been
spoken. Fixation of the target block was slightly later when the display contained another
starred block that was not yellow, and later still when the display also contained a starred
yellow block that was not a square. This result shows that even before hearing the noun that refers to the object being described, listeners are processing the pre-nominal adjectives as they are heard and mapping their meaning onto the options available in the visual context.

More recently, researchers have employed eye movement techniques to show that listeners are remarkably sensitive to subtle aspects of language, and employ that information in directing their gaze around a display. For example, Altmann and Kamide (2004) have demonstrated that participants will fixate a cake before hearing the word spoken in the sentence, ‘The boy will eat the cake’. This anticipatory saccade will not occur when subjects hear, ‘The boy will move the cake’. This evidence demonstrates participants are activating rich ‘thematic role’ knowledge (Ferretti, McRae, & Hatherell, 2001) of the verb “eat”, and fixating likely candidates for this action before the word is spoken.

Matlock and Richardson (2004) have provided further evidence of the nuances of language that drive eye movements around a scene. In previous reading time studies, Matlock (2004) found evidence that readers would mentally simulate motion when reading sentences such as ‘the fence runs along the garden’. Such use of figurative language is termed ‘fictive motion’, since although the motion verb ‘run’ is used, no literal motion takes place. Matlock and Richardson (2004) showed that listeners would look longer along the relevant path when they heard ‘the fence runs along the garden’, compared to ‘the fence is around the garden’. This suggests that fictive motion, far from being an example of a ‘dead metaphor’, elicits something like a perceptual simulation.
(e.g., Barsalou, 1999) and influences how a listener directs their attention across a picture.

Eye movements thus reveal the incremental and interactive nature of spoken language comprehension. Subjects are gradually influenced by the incremental delivery of linguistic information, and eye movements exhibit the continuous, partially active representations that arise during processing. In addition, eye movements have permitted the observation of powerful interactive effects between language and vision. It seems that this incremental process of language comprehension can be strongly constrained by appropriate visual contexts, and that moment-by-moment visual perception can be driven by subtle aspects of language such as thematic roles and figurative motion.

**Eye movement methodology**

Despite these benefits, tracking these naturalistic responses has not always involved the most naturalistic of circumstances. The facility with which eye movements are tracked has improved considerably over many decades of developing technology. Devices attached directly to eyeball (Delabarre, 1898; Huey, 1898; Yarbus, 1965) gave way to photographic techniques (Diefendorf & Dodge, 1908; Tinker, 1928) and in the last few decades, have been replaced by electronic detection of the small differences among reflective properties of the eye (Cornsweet & Crane, 1973; Merchant, Morrissette, & Porterfield, 1974; Young, 1970) that permit rapid and accurate calculation of gaze direction (for a review of the history of eye tracking methods, see Richardson & Spivey, 2004a). In particular, new headband-mounted eyetrackers point an additional “scene
camera” at the subject’s field of view from the subject’s perspective, thus allowing a point-of-regard to be superimposed on the scene camera’s image regardless of where or how the subject moves (Ballard et al., 1995; Eberhard et al., 1995; Land & Lee, 1994). This permits direct inspection of the “responses” elicited through eye movements, from the visual perspective of the subject herself, allowing analysis of the distribution of eye movements over the visual scene in an experimental task.

We hope that the reader is convinced by now that recording eye-movements does indeed provide a unique source of data for constraining one’s theories about language and cognition. The ways in which the visual environment constrains, and is used by, various linguistic and cognitive processes are becoming better understood due to the insights afforded by many findings from eye-tracking experiments. However, there are a number of safeguards and practical tips that one accumulates over years of experience with eyetracking studies that are worth considering before a newcomer dives right into collecting a large mass of eye-movement data (for a brief review, see Tanenhaus & Spivey-Knowlton, 1996).

**Display Parameters.**

Perhaps the most important thing to keep in mind when designing an eye-tracking experiment with spoken linguistic input is to plan in advance exactly how the eye-position record will be coded. It is sometimes helpful to have an initial centralized fixation point where the subject is instructed to look at the beginning of each trial; and critical relevant objects should be equidistant from this initial fixation position. In order to reduce the likelihood of subjects inferring the experimental predictions due to repeated
exposure to particular patterns or relationships in the stimuli, filler objects, as well as filler trials, are recommended. Highly complex displays, such as photographs that have objects partially occluding other objects in depth, or objects whose adjoined and abutting parts are important for separate analysis, can prove rather difficult for eye-position coding. Especially if data analysis is being performed by trained coders watching frame-by-frame videotape, but even when analysis is automated in x,y computer screen coordinates, just a small amount of noise or jitter in the eye position signal can introduce uncertainty in whether one abutting object/part or the other is actually being fixated. A common solution for this is to arrange the visual displays such that there is white space and/or a contour-based divider separating each relevant region or object by at least a couple degrees of visual angle.

**“Blind” Coding**

If the eye-position record is being analyzed by human coders via frame-by-frame videotape, it is, of course, wise for the coders to be prevented from knowing the experimental conditions and predictions for each trial. If the critical experimental manipulation is in the auditory portion of the videotape, this can sometimes be solved by simply coding the silent video portion. However, in other circumstances, trained coders who are unaware of the experimental manipulation and the theoretical predictions may need to analyze each trial with a spreadsheet that uses coded labels for conditions. A more recent solution to this problem, is to store eye-position as x,y screen coordinates and map them onto the x,y coordinates of objects on the same screen. Current headband-mounted eye-tracking systems allow this automated data analysis as long as the visual
display is presented on a computer screen and the subject makes rather minimal head and trunk movements in front of that screen.

**Participant Ease.**

Whether it is a bitebar on which the subject must make a form-fitting dental impression so that her head will be held in place during the experiment or a 3/4-pound headband that the subject is being asked to wear, eye-tracking equipment looks and feels intimidating to a newcomer. And experimental findings from uptight and uncomfortable participants may not generalize to how they behave in normal everyday settings. Therefore, a good tactic when the participant first enters the lab and begins filling out the consent form, is to chat with them a bit about the classes they’re taking etc., and then describe the basics of how the eyetracker works, while demonstrating it on yourself. This way they get to see someone else wearing the equipment with ease. Also, wearing a headband-mounted eyetracker for more than half an hour can sometimes cause a headache. Experiments that last longer than 30 minutes should probably introduce a 10-minute intermission.

**Practice.**

Every eye-tracking system has its own set of unique tricks and parameter settings that take time to learn. In particular, achieving an accurate calibration (usually a 5-10 minute process), such that the subject’s actual eye position is correctly indicated for all regions of the display, is something that requires careful attention to parameters such as centering the eye in the eye camera’s view, reducing distracting reflections on the sclera, as well as a certain amount of speed and fluidity in entering data for the calibration positions to minimize head-drift during calibration. The typical graduate student can expect to
require a couple weeks of practice before being able to complete a good calibration for the majority of their experimental subjects. And there will always be some portion of experimental subjects, 5-10%, for whom three or even four attempts at calibration simply fail to produce an accurate record of eye-position. This can be due to a variety of things, such as very light or very dark irises (with some trackers), naturally droopy eyelids, downward-pointing eyelashes that obstruct the eye camera’s view, the headband not fitting the person’s head, or even incorrigible head motion during the calibration phase.

It could be said that the actual tracking of eye movements is more of an art than a science. And there are certainly hundreds of minor tricks that a research team will learn and develop through practice with any given apparatus and set-up that cannot be anticipated in advance. However, the small handful of practical tips provided in this section (particularly tailored to headband-mounted eye-tracking) are unlikely to be found anywhere else in the literature. We hope they prove helpful.

**Conclusion**

We contend that eye movements provide an index of real-time mental activity that most other methodologies do not (but cf. event-related potentials, Coulson, this volume, and continuous kinematic properties of manual responses, Abrams & Balota, 1991, Coles, Gratton, Bashore, Eriksen, & Donchin, 1985). In its concern with the mental faculties that underlie language use, cognitive linguistics stands to gain considerably from eye movement techniques, Eye movements provide a semi-continuous record of the partially-active representations competing for a overt skeletal motor response, and therefore offer rich insight into *how cognition happens* -- not just the outcomes it produces.
References


Anderson, B. Meyer & P. Olivier (Eds.), *Diagrammatic Representation and Reasoning*. Springer.


