Daniel C. Richardson

Research Statement

Overview

I study the relationship between perception and action in complex behaviors such as language-use and memory. When two people are discussing a scene in front of them, such as a painting or a mathematical proof, how do they assimilate their understanding of what they see with their understanding of what is said? How do they keep track of relevant parts of the image, so that they can refer to them verbally and attend to them visually? How are these visual and linguistic processes coordinated, and how do they influence each other? I investigate the ways in which these interactions between perception and action drive visual attention and play out in the eye movements of people perceiving the world, acting upon it and communicating with each other.

My theoretical questions are based on a broad interest in how cognition is contiguous with the representations and processes of perception and action. In various lines of research I have revealed a number of cases – semantic memory, visual memory, verb comprehension, conceptual representation – where cognitive processes draw on the spatial formats inherent in perceptual and motor representations. I have pursued these notions in developmental studies, and found that infants as young as six months of age demonstrate a remarkable ability to spatially index the locations of rich multi-modal events. These results are integrated into a theoretical framework in which 'higher' cognitive functions emerge from rich, integrative perceptual and motor representations that serve to guide action in the world.

My studies pursue several theoretical goals. First, they address traditional questions in cognitive psychology in a new way. For example, how a person views a picture that is being described can reveal the processes and representations that underlie language comprehension. In one case, I have found that even subtle aspects of figurative language alter how the eyes move across an image. Second, my studies broaden the scope of questions asked from a vision science perspective. In contrast to the static top down influences on visual attention that are typically studied, such as the beliefs or expertise of an individual, I am investigating a host of online top-down influences generated by memory and linguistic information. For example, I have demonstrated that comprehending a narrative or recalling semantic information can draw subjects' eyes around an entirely blank display. Third, my research addresses a novel set of questions regarding the precise integration of linguistic and perceptual processing in the service of complex behaviour. For example, how is the moment-by-moment understanding of two people engaged in a conversation reflected in the relationship between their eye movement patterns?

How do people keep track of the location of information?

Various perceptual faculties exist to help us keep track of the location of objects in a dynamic environment. In a series of experiments, I have argued that this spatial indexing behavior extends to keeping track of the location of verbal semantic information. In our first set of experiments (Richardson & Spivey, 2000) participants listened to pieces of factual information (e.g. *The capital of Australia is Canberra*) while they looked at events (e.g. a talking face delivering the spoken sentence) occurring in one of four grid regions on a computer screen. In front of a blank grid, they were then asked a question relating to one of those facts. While answering, participants made significantly more saccades to the empty region of space where the semantic information had been previously presented. Importantly, the accuracy of their answers was not related to whether or not they fixated this region of space. Therefore, although location was explicitly irrelevant to the task and remembering locations did not improve performance, participants nevertheless systematically encoded the spatial location of auditory information, and re-fixated the location when recalling the information.

We extended this paradigm to investigate the existence and flexibility of spatial indexing in infants (Richardson & Kirkham, 2004). Six-month-olds were shown two different animated objects on a computer screen. Each object had a distinct sound, and appeared in a certain location in one of two square frames in the top and bottom halves of the screen. After six presentations, the two square frames smoothly rotated into a horizontal alignment on the left and right sides of the screen. In two test trials, the sound associated with one of the objects was played while the infants looked at the frames. The infants looked significantly longer at the empty frame associated with the test sound, even though it occupied a new location in space. These results are particularly surprising since they show that a propensity for spatial indexing is not just a feature of the mature adult visual system, but emerges by 6 months along with some of the first uses of adult-like spatial reference frames.

How do people coordinate their visual attention when they talk to each other?

The Eye Chat project investigates how two people inspect a common visual scene while they are engaged in a conversation. Previously, speech production and comprehension have been studied in separate eye tracking paradigms, using simple, single sentences. It is unclear, therefore, how well those results would generalize to spontaneous speech exchanged between conversants. When discussing a joint visual scene – a painting, diagram, map or movie – conversants must simultaneously plan utterances, comprehend speech, coordinating turn taking and locate and integrate visual information. Clark (1996) has argued that language use can be best characterized as a form of 'joint activity' that is grounded in the shared visual context. I am using eye-tracking technology to answer a number of simple, but important questions that have not been addressed before. When people talk about something in front of them are their eye movements linked in some way? Does the degree to which conversants' eye movements are linked relate to how well they understand each other? When people have problems communicating - for example, when their speech is disfluent, they interrupt each other or seek clarification - is this in part because they are looking at different things?

In the first Eye Chat experiments (Richardson & Dale, in press) we investigated the coupling between speakers' and listeners' eye movements. Some participants talked extemporaneously about a TV show whose cast members they were viewing on a screen. Later, other participants listened to these monologues while viewing the same information. Eye movements were recorded for all speakers and listeners. According to cross-recurrence analysis, the listener's eye movements most closely matched the speaker's eye movements at a delay of about two seconds. Indeed, the more closely a listener's eye movements were coupled with the speaker's, the better the listener did on a comprehension test. In a second experiment, low-level visual cues were used to manipulate the listeners' eye movements, and these, in turn, influenced their latencies to comprehension questions. Just as eye movements reflect the mental state of an individual, we found that the coupling between a speaker's and a listener's eye movements reflects the success of their communication.

In current Eye Chat experiments, we are tracking the gaze of two subjects engaged in live discussions about politics, art, tangrams and TV shows while looking at relevant visual stimuli. We are testing a number of hypotheses linking the moment-by-moment eye-movement coupling with their turn-taking, speech disfluencies, interruptions, and their degree of agreement, understanding and efficiency.

Can figurative language influence visual perception?

Our comprehension of a picture is more than just its pixels; our comprehension of a sentence is more than the sum of its words. Both visual and verbal stimuli need interpretation. Though speech processing and scene perception are typically studied in isolation, when spoken words describe what we see in front of us, these interpretations must be integrated. Everyday language is suffused with figurative and non-literal expressions, yet their relationship to perception is rarely studied. Using a novel eye-tracking and language paradigm I developed at Stanford, we investigated how figurative language might shape scene perception (Matlock & Richardson, 2004). In one experiment, we studied pervasive fictive motion descriptions such as *The road runs through the valley*, and *The fence goes along the cliff*. Both sentences use a motion verb (*run*/*go*) but no motion is actually expressed. Even though these sentences convey roughly the same information as literal counterparts (*the road is in the valley*, *The fence is next to the cliff)*, previous work suggests that they evoke mentally simulated motion (Matlock, in press). In the eye-tracking study, we found that gaze durations along a path in a picture differed according to whether subjects heard fictive motion or literal descriptions. In another experiment, prior descriptions of terrain (*the field was hilly / flat*) modulated gaze durations when fictive motion descriptions were used (*the path goes through the field*). These results suggest that fictive motion evokes a mental simulation of motion that is immediately integrated with visual processes. In future work, we will continue to investigate similar forms of implicit spatial information that is conveyed not just in figurative language, but also in progressive versus perfective tense, forms of Spanish subjective progressives, and conceptual metaphors.

How is spatial information represented in abstract verbs and narratives?

Cognitive psychology and cognitive linguistics have produced provocative evidence that a substantial portion of language is encoded in the form of spatial representations that are grounded in perception and action. We demonstrated that naïve participants display a high level of agreement when asked to choose or draw schematic representations of concrete and abstract verbs (Richardson, Spivey, Edelman & Naples, 2001). For example, participants tended to ascribe a horizontal schema to the verb “push”, and a vertical schema to the verb “respect.” A further set of experiments showed that this spatial information is activated during real-time language processing (Richardson, Spivey, Barsalou & McRae, 2003). We predicted that if comprehending a verb activates a spatial representation that is extended along a particular horizontal or vertical axis, it will affect other forms of spatial processing along that axis. Participants listened to sentences containing concrete and abstract verbs while engaged in a visual discrimination task and a picture memory task. In both cases, reaction times showed an interaction between the horizontal/vertical nature of the verb’s image schema, and the horizontal/vertical position of the visual stimuli. Similarly, our eye-tracking research has shown that when subjects look at a blank screen (or even have their eyes closed) and listen to a story that is spatially arrayed along a certain axis (e.g. a description of a train going by), their eye movements are extended along that axis (Spivey, Tyler, Richardson & Young, 2000). These results suggest that even relatively high-level cognitive representations that are generated from linguistic input can contain elements of perceptual and motor representations

Implications

My research aims to uncover the interrelationships among perception, memory, and language, in the service of cognition and action. I tackle these theoretical questions with novel methodologies as well as conventional ones. For example, spontaneous conversation has previously been studied by video recording participants and coarsely coding their behaviour by hand. For the first time, my research employs the fine-grained temporal and spatial resolution of eye-tracking technology to examine the moment-by-moment coupling of conversants' visual attention. Traditionally, eye-tracking technology has been used to study the perceptual processes involved with either reading text or viewing a scene in silence. Studies of spoken language and eye movements are typically limited to single, literal sentences produced or comprehended by individual subjects. My research, in contrast, provides people with rich visual contexts and presents figurative, spontaneous and interactive speech. Although these conditions are rarely present in most laboratories, these are the features of cognition and behaviour in everyday life.

These theoretical and methodological qualities generate several practical applications for my work. Interfaces for human-computer interaction can be informed by my studies of visual and verbal processing, and my work on the coupling of eye movements in conversation has implications for web-conferencing and interactive workspaces. My research is currently featured in the Media X portfolio of Stanford research projects that are of interest to industry partners. In addition, my work on the development of spatial indexing has potential clinical benefits. Burgeoning theories suggest that infants at risk for autism spectrum disorders have trouble integrating information across modalities. My multimodal spatial indexing paradigms are currently being used by a group at UC Davis' Mind Institute investigating infant markers for autism.