(2000) Proceedings of the Twenty-second Annual Meeting of the Cognitive Science Society, (pp,487-492), Erlbaum: Mawhah, NJ

Eye Movements During Comprehension of Spoken Scene Descriptions

Michael J. Spivey
(spivey@cornell.edu)

Daniel C. Richardson
(dcr18@cornell.edu)

Melinda J. Tyler
(mjt15@cornell.edu)

Ezekiel E. Young
(eey1@cornell.edu)

Department of Psychology
Cornell University
Ithaca, NY 14853 USA

Abstract

A recent eyetracking experiment has indicated that, while staring at a blank white display, participants engaged in imagery tend to make eye movements that mimic the directionality of spatial expressions in the speech stream (Spivey & Geng, 2000). This result is consistent with a spatial mental models account of language comprehension (e.g., Johnson-Laird, 1983), adds a *motor* component to evidence for activation of *perceptual* mechanisms during visual imagery (e.g., Kosslyn, Thompson, Kim, & Alpert, 1995), and fits with claims regarding the embodiment of cognition (e.g., Varela, Thompson, & Rosch, 1991). However, some methodological concerns remain. We report some preliminary observations, and a controlled experiment, in which these methodological concerns are resolved. We demonstrate that, even when the speech includes no instructions to imagine anything, and even when participants' eyes are closed, participants tend to make eye movements in the same direction (and especially along the same axis) as the described scene when listening to a spatially extended scene description..

Introduction

More than three decades ago, Donald O. Hebb (1968) suggested that the very same eye-movement scanpaths associated with *viewing* an object may be automatically triggered (via transcortical cell assemblies) when a person is *imagining* that object -- and some empirical support for this claim has recently been reported. When viewing a blank screen and being instructed to imagine a previously-viewed block pattern, observers produced scanpaths that bore some resemblance to the scanpaths elicited during original viewing of the actual block pattern (Brandt & Stark, 1997).

Such oculomotor behavior in the absence of visual input is consistent with the notion that, when imagining or remembering an object or event, we often develop a mental representation of that object or event that has a distinctly spatial structure to it This spatial format of representation is thus able to take advantage of properties inherent to Cartesian space, such as topography and metric relationships. During the construction and interrogation of such spatial mental models (e.g., Bower & Morrow, 1990; Bryant, 1997; Johnson-Laird, 1983, 1996), cognition often uses linguistic input to activate memory representations, and imagery may then use those memory representations to partially activate perceptual representations (e.g., Farah, 1995; Finke, 1986; Kosslyn et al., 1995).

The present study demonstrates that, even in the absence of any visual stimulus at all, such "perceptual simulations" (Barsalou, 1999) often trigger corresponding oculomotor responses. In a sense, one might say that *thinking* of something often involves *pretending to look at it*. This finding contributes to the developing "embodied" view of the mind (e.g., Ballard, Hayhoe, Pook, & Rao, 1997; Brooks, 1995; Varela, Thompson, & Rosch, 1991), in which an adequate characterization of cognition requires special attention to the repertoire of actions available to the organism or agent.

Looking at Objects that Aren't There

In a recent study, Spivey and Geng (2000, Experiment 1) had participants simply listen to pre-recorded instructions to imagine visual scenes while looking at a blank white projection screen and wearing a headband-mounted eyetracker. Each of the descriptions had a specific directionality (rightward, leftward, upward, and downward) to the manner in which new objects or events were introduced in the scene. In addition, a control scene description was presented, in which no particular directionality was present.

Pilot results with this methodology produced eye movement patterns very much in accordance with the directionality of the scene description, however most participants developed rather accurate suspicions of our experimental hypothesis. Although eye movements are relatively automatic, and usually not very susceptible voluntary control, the concern remained that participants may not have produced such behavior if they hadn't known that their eye movements were being recorded.

To avoid potential strategy effects, we introduced a sham task (of following instructions to move objects around on a table), and referred to the imagery session as a break from the experiment during which the eyetracker would be turned off ("but don't take off the headband because then we'd have to recalibrate the tracker when we return to the experiment"), Although two participants suspected that their eyes were still being tracked, and two participants closed their eyes during the imagery session, the remaining six participants produced eye movement patterns that were remarkably consistent with the directionality of the scene descriptions. Figure 1 shows example data from the Control (left panel) and Rightward (right panel) scene descriptions.

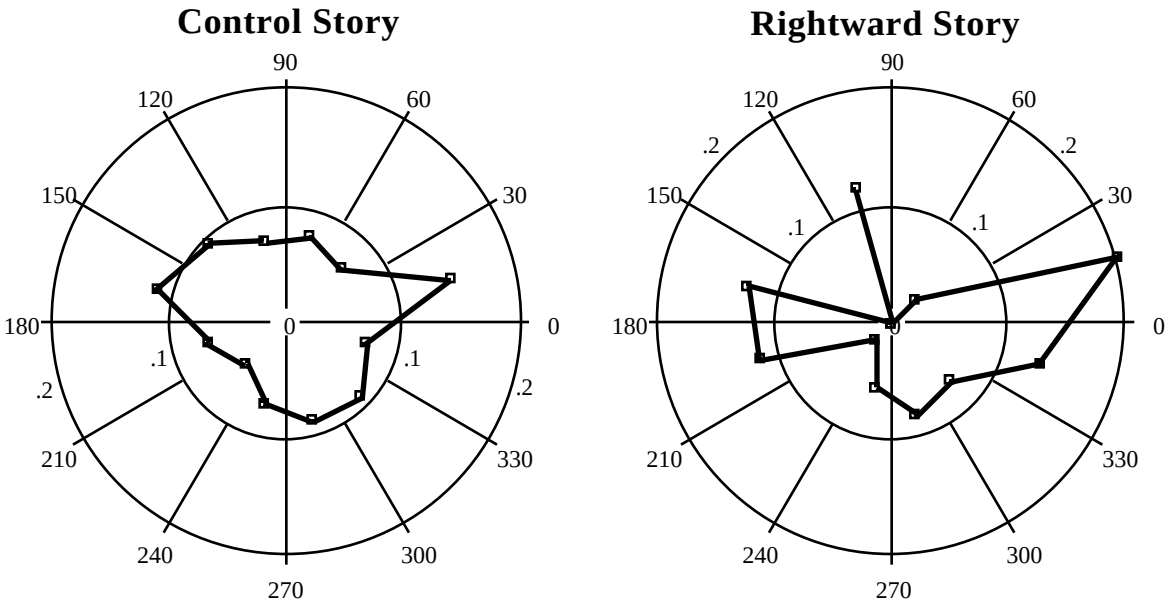

Figure 1: Polar coordinate plots of the average proportion of saccades in various directions while listening to stories. In the control story, participants made an approximately equal proportion of eye movements in all directions. In the rightward story, .2 and .14 of all saccades were in the two rightward directions (15˚ and 345˚, respectively).

Despite remarkable direction-selectivity in saccades during the imagery instructions (for all four directionally-biased scene descriptions), some concerns about this methodology remained. To begin with, the scene descriptions began with an explicit instruction to "imagine" the scene, which may be importantly different from normal language use. Moreover, because they knew that eye movements were involved in the rest of the experiment, participants' eye movement behavior may still have been somehow unnatural. Ideally, these findings needed to be replicated under circumstances where there were no explicit instructions to imagine something, and where the participant had no idea that his/her eye movements were being recorded.

Observation

We have been developing a methodology for recording a participant's eye movements with an ISCAN, Inc. remote eyetracking camera without the participant's knowledge. (The extreme difficulty in getting a precise calibration has alleviated, for us, any worries about the potential unethical uses of such a methodology.) For our purposes, the deception involves telling participants that the camera directed at them is recording subtle thermal changes in the face as a result of emotional arousal. They are encouraged to sit as still as possible during the experiment in order to allow an accurate thermal image.

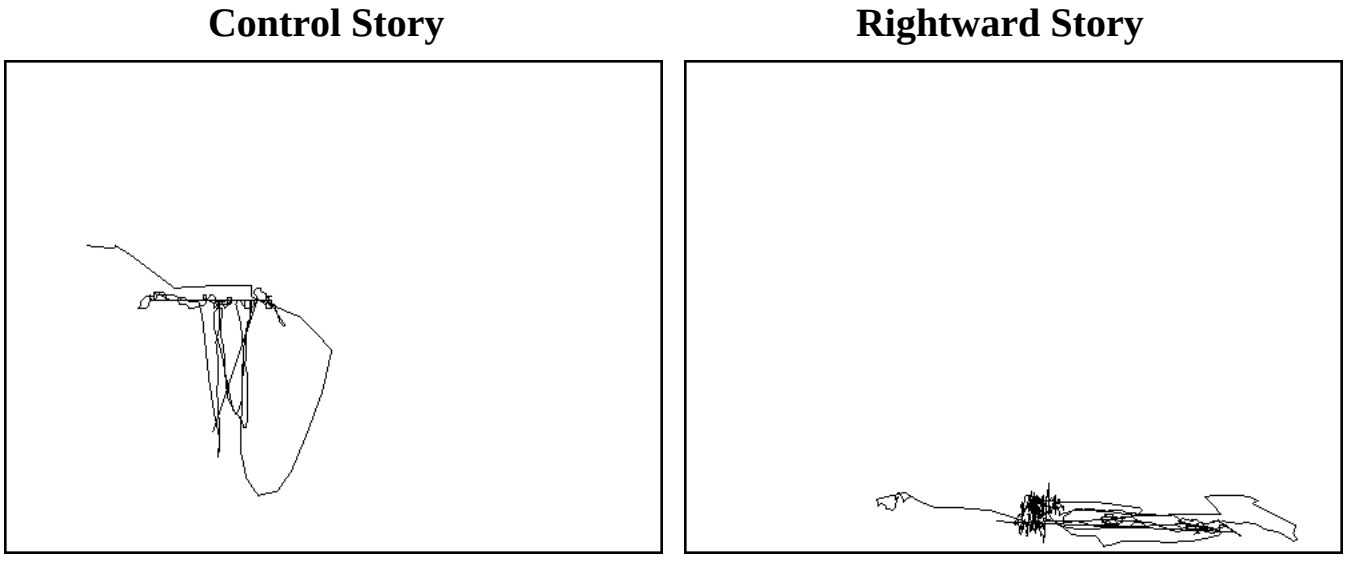

Figure 2: Example scanpaths from a participant while listening to the Control and Rightward scene descriptions.

For calibration, we use a poster of the painting 'Farbstudie Quadrate' by Kandinsky, spanning 20 X 30 degrees of visual angle, mounted on black foam core board. With the cover story being that we must first collect some baseline thermal readings, participants are asked to report the colors in the centers of various concentric circles in the painting, in the four corners and near the center, during which time we record those five eye positions as calibration points. In order to reduce the amount of visual information in the environment that participants might fixate during the time that they listen to the pre-recorded stories, the poster is then flipped over to display the black-colored board.

Table 1: Pre-recorded scene descriptions

CONTROL STORY
"You are on a hill looking at a city through a telescope. Pressing a single button zooms a specific block into view. Another button brings a gray apartment building into focus. Finally a third button zooms in on a single window. Inside you see a family having breakfast together. A puppy appears and begs for a piece of French toast."
RIGHTWARD STORY
There is a fishing boat floating on the ocean. It's facing leftward from your perspective. At the back of the boat is a fisherman with a fishing pole. The pole extends about 10 feet to the right beyond the edge of the boat. And from the end of the pole, the fishing line extends another 50 feet off to the right before finally dipping into the water."
DOWNWARD STORY
"You are standing at the top of a canyon. Several people are preparing to repel down the far canyon wall across from you. The first person descends 10 feet before she is brought back to the wall. She jumps again and falls 12 feet. She jumps another 15 feet. And the last jump, of 8 feet, takes her to the canyon floor."
LEFTWARD STORY
"There is a train extending outwards to the left. It is pointed to the right, and you are facing the side of the engine. It is not moving. Five cars down is a cargo holder with pink graffiti sprayed on its side. Another six cars down is a flat car. The train begins to move. Further down the train you see the caboose coming around a corner."
UPWARD STORY
"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."

The pre-recorded auditory stimuli consist of ten short stories, each story lasting approximately thirty seconds. Half of the stories are filler stories which contain no directional cues, but which are intended to be slightly emotionally-engaging, in order to divert the participants from suspecting the actual experimental hypothesis. The five test stories are derived from Spivey & Geng (2000), but do not begin with an instruction to "Imagine...". The test stories contain systematic directional cues (see Table 1), and are not emotional in content. The order of the ten stories is pseudo-randomized, with one filler story interleaved between each pair of test stories. After about five minutes for calibration, the listening session begins with a filler story, and lasts another five minutes.

At the end of the session, participants are debriefed and thoroughly questioned as to their beliefs about the nature of the study. With the thermal camera cover story, none of the participants have correctly guessed the experimental hypothesis, and generally all are surprised that their eye movements were being examined rather than patterns of thermal change on the face.

In developing this methodology, we have encountered numerous complications in acquiring and testing an accurate calibration of the eyetracking system without betraying the deception. Due to blinks and other movements, an accurate calibration with the remote eyetracking camera typically requires multiple fixations of each the five calibration points. As the calibration period drags on with more and more inventive questions regarding the four corners and center of the painting, participants can become suspicious of the cover story. Moreover, following an acceptable calibration, head movements during the listening phase of the experiment can make it difficult for the software, or a human controller, to maintain a centralized camera image of the eye.

Despite the complications, some preliminary observations from this methodology are available and worth reporting. It is already clear that listeners make a greater number of saccades during the directionally-biased stories compared to the Control Story. Although these eye movements are not always in the specific direction of the directionally-biased story, they do tend to be limited to the appropriate axis of orientation; see Figure 2. This should be expected if one considers the fact that a listener simply couldn't make rightward or upward eye movements indefinitely as the story continues to add rightward or upward expressions. At some point, an eye movement in the opposite direction is necessary to "re-center" the imagined scene in head-centered coordinates. (It is possible that this might occur less frequently if participants were allowed to turn their heads.) In any case, there is nothing stopping the listener from voluntarily "examining", or "looking back to", previously described elements of the scene.

Further development of this methodology, with more accurate tracks from additional naive participants, will allow averaging of saccades in polar coordinates (as shown in Figure 1). This work appears likely to confirm the findings of Spivey and Geng (2000), under circumstances where there are no explicit instructions to imagine anything, and participants are completely unaware that their eye movements are being recorded.

Experiment

The described complications with the above methodology (as well as concerns about participants looking at objects in the visual field beyond the 20˚ X 30˚ black display board) point to a methodology with which it is easier to collect data from more reliably naive participants, but which produces data that is admittedly more difficult to analyze. In this experiment, participants were instructed to close their eyes while they listened to the same ten stories; the stories in Table 1 as well as the five filler stories. A standard video camera was directed at the participant's face, and the camera's image of the participants' closed eyes was later used to estimate incidence and direction of eye movements. (One could, in principle, accurately record movements of closed eyes with a search coil [a contact lens with a copper wire and loop, the position and orientation of which is precisely determined via an electromagnetic field in which the participant sits]. However, it might be difficult to convince someone wearing a search coil that their eye movements were not being recorded.)

Method

Participants Eleven Cornell University undergraduates participated in the study for extra credit in Psychology courses. None of them had previously participated in an eyetracking experiment.

Stimuli and Apparatus This experiment used the same pre-recorded scene descriptions as described in the previous methodology. A standard video camera was positioned in front of the participant, with a black curtain as background. Stories were presented in a pseudo-randomized order, with each test story preceded by a filler story.

Procedure Participants were asked to relax in their seat, but to remain still and to close their eyes while listening to the ten short stories. They were informed that we would be examining their facial expressions as related to story content. Participants were told that their shoulders and face were being videotaped while the stories were being played. Upon achieving a well-focused image of the participant's closed eyes through the video camera, recording began. The stories were played from a cassette player. Each session lasted approximately five minutes. At the end of each session, participants were debriefed and thoroughly questioned as to their beliefs about the nature of the study. None of the participants correctly guessed the study hypothesis, and all were surprised that their eyes movements had been of interest.

Coding Coding was made easier by focusing on the movement of a round spot of luminescence on each of the participant's eye lids. The spot of light reflectance on the eye lids corresponded to the protuberant round area of the cornea directly over the pupil sitting beneath the lid. Movements of the eye were considered to be clearly visible shifts in the spot of reflectance on each lid (as opposed to ambiguous twitches in the lid, or jitters in the positioning of the spot of reflectance that were too small in distance and too short in duration to be easily interpreted.) See Figure 3, where dark lines are added to indicate the points of inflection due to the corneal bulge. Such eye movements are much easier to discern when multiple frames are seen in real-time (http://node15.psych.cornell.edu/home/eyesclosed.html). A definite movement of the eye ball was considered to be a movement of the spot of reflectance lasting for 3 video frames or longer and of such a distance that the direction of movement was unambiguous to 2 independent trained coders. Assessment of start time and direction of eye movements involved pinpointing the movements on video tape by playing approximately 10 frames of tape at a time, then comparing the previous position of the spot of reflectance with the new position. It was frequently necessary to rewind or forward the position of the tape one frame at a time to specify as precisely as possible when each eye movement took place. Inter-rater reliability for the two coders was high, as measured by a Pearson correlation; r=.84. It is difficult to estimate how large a saccade (in degrees of visual angle) is detectable with this coding method. However, it is likely that many small eye movements are being made in this task, the direction of which cannot be determined with the present method.

Results

For each scene description, proportion of detectable eye movements in each of eight directions was averaged across all eleven participants. Two of the eleven participants made no detectable eye movements during the entire experiment.

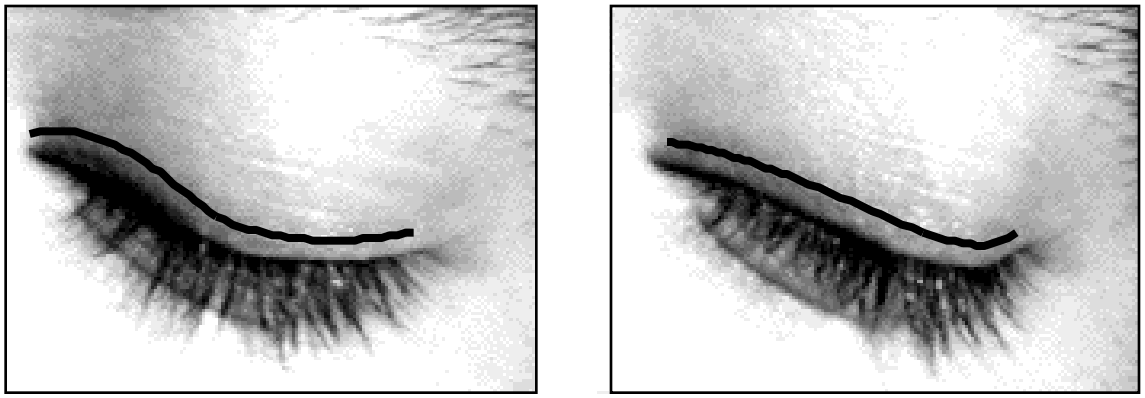

Figure 3: Examples of a detectable rightward eye position (left panel) and leftward eye position (right panel).

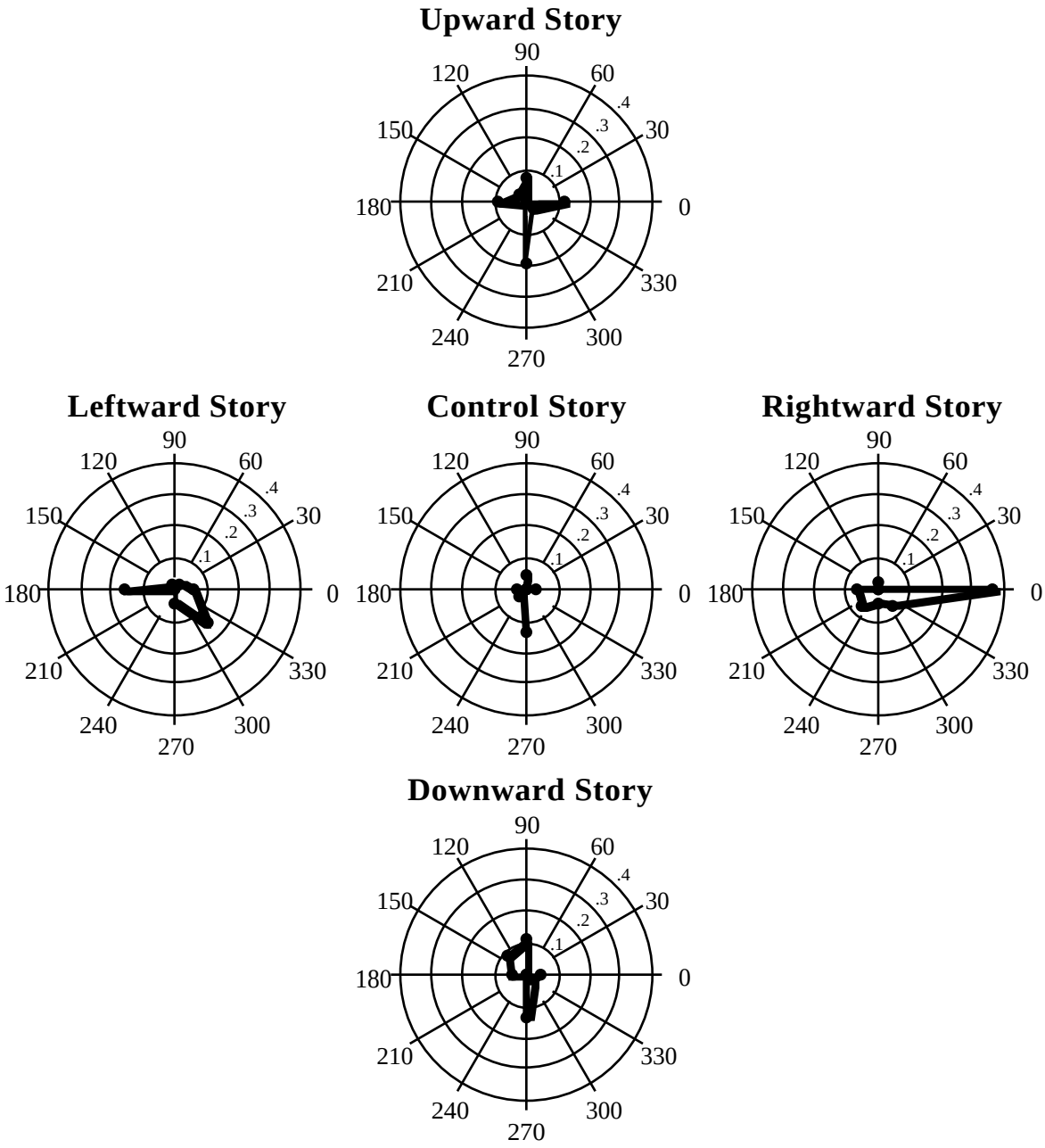

Figure 4: Polar coordinate plots of the average proportion of detectable eye movements in eight directions while participants listened to scene descriptions with their eyes closed.

Consistent with the previous observations, only three of the eleven participants made any detectable eye movements during the Control Story. Unfortunately, this contributes to a rather unreliable estimate of the direction-selectivity profile for that story. Figure 4 shows polar coordinate plots of the eye-movement direction-selectivity profiles for the five scene descriptions. It is noteworthy that when the detectable eye movements were in the unpreferred direction, they were frequently in the exact opposite direction. Thus, even when the eye movements do not follow the specific directionality in the scene description, they nonetheless tend to be limited to the appropriate axis of orientation (horizontal or vertical). In a paired t-test, the average proportion of eye movements in a preferred direction was significantly greater than the average proportion of eye movements in the unpreferred directions; $t(10) = 4.49$, $p<.01$.

Discussion

These results demonstrate that, even when participants' eyes are closed, they tend to move their eyes in directions that accord with the directionality of the scene being described. Although comprehension of these scene descriptions may involve some of the same mechanisms involved in imagery tasks (e.g., Kosslyn et al., 1995), the present results do not

rely on explicit instructions to imagine anything. We suggest that comprehension of scene descriptions employs a decidedly spatial format of representation, and that oculomotor coordinates may be an important component of that representation.

Conclusions

With no visual information available, participants constructing mental models of complex scenes tend to make eye movements that mimic the kinds of eye movements that would be made when viewing that actual scene. In a similar vein, Spivey and Geng (2000, Experiment 2) and Richardson and Spivey (2000) report experiments in which participants attempting to recall visual or auditory information tend to make eye movements to the region of space (now empty) where that information was first encoded. In conjunction with results in visuomotor coordination by Ballard et al. (1997) and in attention by Pylyshyn (1994), this work suggests that the mind/brain exploits useful properties inherent to spatial formats of representation by relying on oculomotor pointers to spatial indexes out in the world or, in the present case, perceptual simulations of them. This perspective may point to a pivotal role for space, independent of perceptual modality, in mental representations (cf. Bryant, 1997).

One could possibly interpret our findings as a tangential endorsement of a sort of 'Cartesian Theater' account of the mind (cf. Dennett, 1991): that when we imagine something, it is like viewing it "in our mind's eye", and that (perhaps epiphemonenally) our real eyes simply echo the motion of our internal spectating. However, it is well known that this kind of interpretation too easily falls into the infinite regress of a homunculus inside the mind. Therefore, we prefer to place these results in the light of Ryle's (1949) comment that "[A person picturing his nursery in his mind's eye] ... is not being a spectator of a resemblance of his nursery, but he is resembling a spectator of his nursery." That is to say, we would argue that our data are indicative of an embodied system that naturally activates 'lower level' motor actions to accompany 'higher level' cognitive processes because, rather than being separate functions that are triggered by a mental state, motor actions are fundamental components of the mental state.

In such an embodied view of the mind, action determines cognition as much as perception does. Indeed, a number of researchers have suggested that an important aspect of *perceiving* an environment is knowing *how to interact* with it (e.g., Brooks, 1995; Gibson, 1979; Milner & Goodale, 1994; Turvey & Carello, 1996). Perhaps we can add to this embodied view of perception that part of *perceiving* a scene is also knowing *how to look at it* -- even when it's not there.

Acknowledgments

We are grateful to Michael Tanenhaus, Mary Hayhoe, and Dan Earley for helpful discussions of this work, and to Jessica Evett-Miller and Koji Park for assistance with data collection and analysis. Supported by a Sloan Foundation Fellowship in Neuroscience to MJS.

References

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723-767.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Bower, G. H. & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247*, 44-48.

Brandt, S. A. and Stark, L. W. (1997) Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*, 27-38.

Brooks, R. (1995). Intelligence without reason. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents*. Hillsdale, NJ: Erlbaum.

Bryant, D. J. (1997). Representing space in language and perception. *Mind and Language, 12*, 239-264.

Dennett, D. (1991). *Consciousness explained*. Boston: : Little, Brown and Co.

Farah, M. J. (1995). Current issues in the neuropsychology of image generation. *Neuropsychologia, 33*, 1455-1471.

Finke, R. A. (1986). Mental imagery and the visual system. *Scientific American, 254*, 88-95.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, Massachusetts: Houghton Mifflin.

Hebb, D. O. (1968). Concerning imagery. *Psychological Review, 75*, 466-477.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Cambridge University Press.

Johnson-Laird, P. N. (1996). Space to think. In P. Bloom & M. Peterson (Eds.), *Language and space.* Cambridge, MA: MIT Press.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995) Topographical representations of mental images in primary visual cortex. *Nature, 378*, 496-498.

Milner, A. D. & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*, 363-384

Richardson, D. C. & Spivey, M. J. (2000). *Representation, space, and Hollywood Squares: Looking at things that aren't there anymore*. Manuscript submitted for publication.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Spivey, M. J. & Geng, J. J. (2000). *Oculomotor mechanisms triggered by imagery and memory: Spontaneous eye movements to objects that aren't there*. Manuscript submitted for publication.

Turvey, M. & Carello, C. (1995). Some dynamical themes in perception and action. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Varela, F. J., Thompson, E., & Rosch E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.