The integration of figurative language and static depictions: An eye movement study of fictive motion

Daniel Richardson *

Psychology Department, University of California, Santa Cruz

Teenie Matlock *

Social and Cognitive Sciences, University of California, Merced

Contact information:

Daniel Richardson
Psychology Department
273 Social Sciences 2
Santa Cruz, CA 95064
Phone: (831) 459-2002
Email: dcr@ucsc.edu

* Authors note: This research was carried out while both authors were postdoctoral researchers in the Stanford Psychology Department.

MAIN TEXT: 2970 words

Abstract

Do we view the world differently if it is described to us in figurative rather than literal terms? An answer to this question would reveal something about both the conceptual representation of figurative language and the scope of top down influences on scene perception. Previous work has shown that participants will look longer at a path region of a picture when it is described with a type of figurative language called fictive motion (*The road goes through the desert*) rather than without (*The road is in the desert*). The current experiment provided evidence that such fictive motion descriptions affect eye movements by evoking mental representations of motion. If participants heard contextual information that would hinder actual motion, it influenced how they viewed a picture when it was described with fictive motion. Inspection times and eye movements scanning along the path increased during fictive motion descriptions when the terrain was first described as difficult (*The desert is hilly*) as compared to easy (*The desert is flat*); there were no such effects for descriptions without fictive motion. It is argued that fictive motion evokes a mental simulation of motion that is immediately integrated with visual processing, and hence figurative language can have a distinct effect on perception.

Introduction

Our comprehension of a picture is more than the sum of its pixels; our comprehension of a sentence is more than the sum of its words. Both words and pictures need interpretation. When spoken words describe what we see in front of us, we must integrate these interpretations on the fly. How do these visual and verbal processes interact? Since Cooper (1974) demonstrated that eye movements are often directed towards objects referred to in speech, research has revealed a close integration of visual and linguistic processing (see Henderson & Ferreira, 2004; Trueswell & Tanenhaus, 2005). For example, visual processes are engaged during processing syntactic structure (Tanenhaus, Spivey Knowlton, Eberhard, & Sedivy, 1995), differentiating semantic roles (Altmann & Kamide, 1999) and resolving anaphoric reference (Runner, Sussman, & Tanenhaus, 2003), and the degree to which listeners' eye movements are coupled to speakers' reflects levels of comprehension (Richardson & Dale, 2005).

Yet studies of verbal and visual integration have focused on literal language. Even though figurative expressions are pervasive in everyday language and exist in all cultures (Gibbs, 1994; Lakoff, 1987), research has not addressed how figurative language affects the process through which we perceive the world. In the current experiment, we investigated how a scene would be perceived when it was described by forms of literal and figurative language that are reported to have equivalent meaning. If the mental representation of a figurative expression is identical to that of a literal expression, then there would be no difference between eye movement patterns. Similarly, if the mental representation of a figurative expression does not interact with visual processes, then there would be no difference between eye movement patterns. Therefore, any differences that are present in eye movement patterns can tell us about both the distinct mental representations that are evoked by figurative language, and the scope of the integration between visual and verbal processing.

Fictive motion

We chose to study a class of figurative spatial descriptions known as *fictive motion* (FM) sentences. Two examples are shown in (1a) and (1b).

(1a) *The road goes through the desert*

(1b) *The fence follows the coastline*

Pervasive in English and many other languages, including Swedish, Finnish, Italian, Chinese, and Japanese, the descriptions are figurative because they contain a motion verb but describe no motion (Huumo, 2005; Matlock, 2004a; Matsumoto, 1996) They highlight the spatial relation between a path or linear entity and a landmark (Talmy, 2000), for instance, the road and the desert in (1a) and the fence and the coastline in (1b). In this way, these fictive motion descriptions are equivalent to literal spatial descriptions, or non-fictive motion sentences (non-FM) such as those in (2a) and (2b).

(2a) *The road is in the desert*

(2b) *The fence is next to the coastline*

Experimental evidence supports the idea that simulated motion is evoked by fictive motion sentences such as (1a) and (1b). In a study by Matlock, Ramscar, and

Boroditsky (2005) it was shown that thinking about the meaning of fictive motion sentences affected how people would conceptualize time spatially. Participants in the study were primed with FM sentences (e.g., *The tattoo runs along his spine*) or non-FM sentences (e.g., *The tattoo is next to his spine*) before answering this ambiguous question about time: "Next Wednesday's meeting has been moved forward two days. What day is the meeting now that it has been re-scheduled?" The expression "move forward" is ambiguous because both Monday and Friday are possible answers. When primed with descriptions with fictive motion, participants in Matlock et al. (2005) were encouraged to take an ego-moving perspective and more likely to say Friday (versus Monday), but when primed with non-FM descriptions they were split between Monday and Friday. Similarly, fictive motion direction (either away or toward, as in *The road goes all the way to New York* or *The road comes all the way from New York*) affected how participants conceptualized of time, namely, more Fridays with going away and more Mondays with coming toward. Together, the results of Matlock et al. (2005) parallel those of other studies on time, space, and motion (Boroditsky, 2000; Boroditsky & Ramscar, 2002; Ramscar, Matlock, & Boroditsky, 2005), suggesting that thinking about motion (fictive or actual) induces an ego-moving perspective when thinking about time.

Figurative language and visual processing

We have found suggestive evidence that fictive motion descriptions can have an immediate and distinct effect on visual processing. Matlock and Richardson (2004) presented participants with simple drawings of paths such as roads, rivers and pipelines. They heard either FM or non-FM descriptions of these paths while their gaze was tracked. The FM descriptions caused participants to spend more time inspecting the region of the path. These gaze differences did not merely result from minor differences in sentence length. Nor did they result from different semantic content, for FM and non-FM sentences were judged as having similar meanings, to be equally semantic sensible, and to be equally good descriptions of the pictures.

Why might fictive motion descriptions have influenced eye movements in this way? One possibility is that participants simply found the FM descriptions to be more interesting, and so viewers paid more attention to the paths. Another possibility is that comprehending fictive motion descriptions evokes mental representations of motion (Matlock, 2004a, 2004b; Matlock et al., 2005; Talmy, 2000), and that these motion representations result in more visual attention being directed to the path. The first goal of the current experiment was to distinguish between these two possibilities. The second goal was to learn more about the eye movements produced by fictive motion descriptions. Is it simply that the whole path attracts more visual attention, or do fictive motion descriptions also evoke a pattern of eye movements that is related to motion along a path? We addressed these goals by introducing an additional experimental factor and an additional dependent variable.

In Matlock's (2004b) reading time studies, participants read stories about protagonists travelling through spatial domains (e.g., valley), followed by target sentences with fictive motion (e.g., *The road goes through the valley*). In general, participants were quicker to process fictive motion target sentences after reading about terrains that were easy to traverse (e.g., *The valley was flat and smooth*) versus terrains that were difficult to traverse (e.g., *The valley was bumpy and uneven*). Critically, there was no difference for comparable literal target sentences without fictive motion (e.g., *The road is in the valley*). These results suggest that the comprehension of descriptions of fictive motion across a domain is influenced by factors that would affect actual motion across the domain. Following that logic, in the current experiment we presented participants with descriptions of easy and difficult terrains and then FM sentences or non-FM sentences. If terrain information modulated looking behavior with FM sentences, it would show that it was not merely something generally eye catching about the combination of non-literal motion verb and path preposition (e.g., *runs along*, *goes through*) that influenced the looking times in Matlock and Richardson (2004), but rather, the engagement of contextually appropriate simulated motion.

We hypothesized that fictive motion descriptions would activate representations of motion. If so, then perhaps we would see not only longer looking times to the path, but also sequences of eye movements that correspond to motion. Spivey and colleagues found that as participants listened to a narrative and looked at blank screen (Spivey & Geng, 2001) or closed their eyes (Spivey, Tyler, Richardson, & Young, 2000), they tended to make eye movements that corresponded to spatial content in the stories. For example, more vertical eye movements were made when hearing about someone repelling down a canyon wall, and more horizontal eye movements were made when hearing about a train pull out of a station. Eye movements were increased along a specific axis of motion, rather than sequentially in a particular direction. We adapted this idea to our experiment, and counted the number of occasions that participants made *path scanning* eye movements, in which one region of the path was fixated immediately after any other path region. In addition to looking time differences, we predicted that participants would make more path scanning looks along the path during a fictive motion description when they had previous heard a description of a difficult rather than easy terrain, but there would be no such difference for non-fictive motion descriptions.

Method

Participants. Sixty-three Stanford University psychology students with normal or corrected vision participated. Data from six participants were discarded because a successful calibration was not achieved.

Stimuli. The visual stimuli consisted of 32 pictures of spatial scenes. All of these pictures were matched on luminance, and all were created with a Microsoft drawing program. Of the 32 pictures,16 were experimental and 16 were fillers. All experimental pictures contained two paths, one represented vertically in the picture plane, and the other horizontally (see Figure 1). These paths were traversable objects, such as roads or trails, or linearly extended objects, such fences or rows of trees.

The verbal stimuli consisted of 64 sentences recorded in 16 blocks of four sentences. Each block contained two pairs of descriptions. One pair described the vertical path, and the other described the horizontal path. Each pair contained two experiment sentences: a fictive motion (FM) sentence and a comparable non fictive motion (non-FM) sentence, such as *The road runs through the valley* and *The road is in the valley*. The experiment was designed such that each participant would hear one sentence from each of the 16 blocks in addition to 16 sentences for the filler pictures. Norming studies reported in Matlock and Richardson (2004) showed that these FM and non-FM sentences were judged to be equal in semantic content and semantic sensibility, and to be equally good descriptions of the scenes.

We recorded two terrain descriptions to precede each experimental sentence. Each terrain description referred to a region in which movement could be conceptualized as easy or difficult, for example, *The valley was flat and smooth* (easy), and *The valley was full of potholes* described (difficult). We did a norming study to ensure that all sentences would in fact be equally compatible with the scenes they described. The participants were told to judge how well the sentences go with the scenes in the pictures. Using a scale that ranged from 1 for “not at all” to 7 for “very well“, 10 Stanford undergraduates judged all pairs to be well-matched. The means were FM + slow-terrain 5.72, FM + fast-terrain 5.62, non-FM + slow-terrain 5.74, non-FM + fast-terrain 5.73. No combination of terrain description and experimental sentence was any better than the other, $F(3, 124) = .4$, $p > .1$, suggesting that all sentence-picture combinations were plausible pairings. In addition to the primary stimuli, we created filler descriptions for all filler sentences.

Apparatus. An ASL 504 remote eye tracking camera was positioned at the base of a 17” LCD stimulus display that was set to 800x600 resolution. Participants were unrestrained and sat about 30” from the screen. The stimuli were 560 pixels square, which subtended approximately 18º square of visual angle. The camera detected pupil and corneal reflection position from the right eye, and the eye-tracking PC calculated point-of-gaze in terms of coordinates on the stimulus display. This information was passed to a PowerMac G4, which controlled the stimulus presentation and collected gaze duration data. Prior to the experiment proper, participants went through a 9 point calibration routine that took one to three minutes.

Procedure. After establishing a successful eye track, participants were told: "Look at the pictures and listen to the sentences." Participants were first presented with 4 practice trials and then a random sequence of 16 filler trials and 16 experimental trials. At the beginning of every trial, they first saw a gray square that was the same size and luminance as the pictures. Next they heard a terrain sentence or a filler sentence. After 500ms, they saw a new picture and after a further 1000ms, they heard a FM sentence, a non-FM sentence, or a filler sentence. The picture remained on screen for a total of 6000ms. The trial ended with a 2000ms inter-stimulus interval.

Coding. Eye movements were recorded for the 6000ms that the picture was on the screen. The eye movement data consisted of which regions-of-interest were fixated at $1/30^{th}$ of a second intervals. The path region-of-interest was a strip 80 pixels wide that extended vertically or horizontally across the image. This path was further divided into seven equally sized, square regions-of-interest.

Results

Participants' eye movement data were parsed into two dependent variables: the total *looking time* in the region of the path, and the frequency of *path scanning* fixations, in which participants fixated one path region followed immediately by another. Analyses were performed by participants (F_1) and items (F_2). Though we intended for all paths in the visual images to have symmetrical arrangements, the path in one image was erroneously asymmetric; it contained an anomaly on one end (water coming out of a garden hose). As additional evidence of this image being inappropriate for our purposes, it elicited unusually long looking times to the bottom region of the vertical path, regardless of fictive or terrain condition. For this reason, that item was removed from all analyses.

The listeners' eye movements were influenced by a combination of terrain descriptions and fictive motion language, as shown in Figure 2. As predicted, *looking times* to the path were affected by an interaction of sentence type and terrain description, ($F_1(1,56) = 11.78$, $p <. 001$; $F_2(1,14) = 15.25$, $p <. 001$). Critically, with FM sentences, participants spent more time inspecting paths after difficult terrain descriptions (M = 2014ms) than after easy terrain descriptions (1621ms) (Tukey's LSD, $p < .05$), but for non-FM, there was no reliable difference (1681ms and 1847ms, respectively). There were no main effects of terrain (F_1 (1,56)=2.30; $F_2(1,14) = 0.10$) or sentence type (F_1 (1,56)=0.45; $F_2(1,14) = 1.21$) for looking times.

This pattern of results was echoed by analysis of the *path scanning* data. There was a significant interaction between sentence type and terrain description ($F_1(1,56)$ = 6.87, $p <. 02$; $F_2(1,14) = 4.77$, $p <. 05$). Participants made more path scanning fixations after hearing a FM sentence preceded by a difficult (M = 3.6) rather than an easy terrain description (M = 2.8) (Tukey's LSD, $p < .05$), but there was no reliable difference for non-FM sentences (2.86 and 3.16, respectively). There were no main effects of terrain (F_1 (1,56)=1.57; $F_2(1,14) = 0.16$) or sentence type (F_1 (1,56)=1.02; $F_2(1,14) = 0.98$).

Discussion

Figurative language can have an immediate effect on how we look at the world. Our results suggest that this is because of the distinct spatial representations that figurative descriptions can evoke that their literal counterparts do not. The way participants inspected paths was affected by information about the terrain and the figurative language that described the path. Critically, eye movements were not influenced by descriptions of difficult or easy terrain by themselves. They were influenced *only* when the terrain descriptions were paired with fictive motion sentences. A plausible explanation for the interaction between fictive motion language and terrain information, we argue, is that comprehending a fictive motion sentence involves a mental representation of motion along a path (Langacker, 1987; Matlock, 2004b; Talmy, 2000), and that the representation incorporates information about terrain. Consequently, difficult terrain would result in slow motion, for example, and the resulting representation is shown by the longer amount of time participants looked at a path and the increased number of fixations scanning along its length.

Our interpretation of these results is congruent with perceptual simulation theories (Barsalou, 1999; Glenberg, 1997; Zwaan, 2004), which hold that language comprehension is a process of generating perceptual-motor representations. Comprehension of fictive motion descriptions led to eye movements along the depicted path that mirrored an internal simulation of movement. More generally, simulated motion is known to figure into a broad range of cognitive processes, such as inferring motion from static images (Freyd, 1983; Kourtzi & Kanwisher, 2000), comprehending descriptions of actual motion (Zwaan, Madden, Yaxley, & Aveyard, 2004), and solving everyday physics problems (Schwartz & Black, 1999).

Our fictive motion experiments are an interesting test case for perceptual simulation theories for two reasons. First, previous experiments compared different scenes, such as *the nail was hammered into the floor* versus *into the wall* (Stanfield & Zwaan, 2001), or concepts, such as *a watermelon* versus *half a watermelon* (Solomon & Barsalou, 2001), and found evidence for differing perceptual-motor activation. In contrast, we are comparing literal and figurative spatial descriptions of the *same* scene. Though the descriptions are equivalent in objective terms, they have *different* interactions with perceptual mechanisms. Therefore, we can distinguish between the identical semantic commitments of the sentences and their differing perceptual simulations. Second, previous experiments have been forced to infer the involvement of perceptual-motor representations in language comprehension from reaction time differences in concurrent tasks, such as sensibility judgements, picture matching or visual discriminations (Glenberg & Kaschak, 2002; Richardson, Spivey, McRae, & Barsalou, 2003; Zwaan, Stanfield, & Yaxley, 2002). In contrast to these studies, our eye movement paradigm allows us to directly measure the effect of figurative language on perceptual mechanisms that are unconstrained by any task other than looking and listening.

In this experiment all we manipulated was the presence of figurative language, a change that did not alter the literal meaning or truth conditions of the sentence. Nevertheless this change appeared to alter visual processing. We argue that eye movements were affected because fictive motion language evokes a dynamic mental simulation which interacts with the ways in which the visual system interprets and inspects the world. Our findings, which have consequences for both the linguistic accounts of figurative language and the scope of top-down influences in visual perception, help illuminate the ways in which verbal and visual processes are intertwined.

Acknowledgements

The authors are indebted to Herbert Clark, Natasha Kirkham, Paul Maglio, Michael Ramscar, Michael Spivey, and our anonymous reviewers for helpful discussions and comments.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22(4)*, 577-660.

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition, 7*, 1-28.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13*(2), 185-189.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84-107.

Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33*(6), 575-581.

Gibbs, R. W., Jr. (1994). *The poetics of mind: Figurative thought, language, and understanding*. New York, NY: Cambridge University Press.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*(1), 1-55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.

Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The integration of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Huumo, T. (2005). How fictive dynamicity motivates aspect marking: The riddle of the Finnish quasi-resultative construction. *Cognitive Linguistics, 16*, 113-144.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12*(1), 48-55.

Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago, IL: The University of Chicago Press.

Langacker, R. W. (1987). *Theoretical Prerequisites* (Vol. 1). Stanford, CA: Stanford University Press.

Matlock, T. (2004a). The conceptual motivation of fictive motion. In G. Radden & K. Panther (Eds.), *Studies in linguistic motivation [Cognitive Linguistics Research]*. New York and Berlin: Mouton de Gruyter.

Matlock, T. (2004b). Fictive motion as cognitive simulation. *Memory & Cognition, 32*, 1389-1400.

Matlock, T., Ramscar, M., & Boroditsky, L. (2005). The experiential link between spatial and temporal language. *Cognitive Science, 29*, 655-664.

Matlock, T., & Richardson, D. C. (2004). *Do eye movements go with fictive motion?* Paper presented at the 26th Annual Cognitive Science Society, Chicago.

Matsumoto, Y. (1996). Subjective motion and English and Japanese verbs. *Cognitive Linguistics, 7*, 183-226.

Ramscar, M., Matlock, T., & Boroditsky, L. (2005). The experiential basis of abstract language comprehension. *Manuscript under review.*

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*, 1045-1060.

Richardson, D. C., Spivey, M. J., McRae, K., & Barsalou, L. W. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767-780.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition, 89*(1), B1-B13.

Schwartz, D., & Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning Memory and Cognition, 25*(1), 116-136.

Solomon, K. O., & Barsalou, L. W. (2001). Representing Properties Locally. *Cognitive Psychology, 43*(2), 129-169.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research/Psychologische Forschung, 65*(4), 235-241.

Spivey, M. J., Tyler, M., Richardson, D. C., & Young, E. (2000). *Eye movements during comprehension of spoken scene descriptions*. Paper presented at the 22nd Annual Conference of the Cognitive Science Society, Philadelphia.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*, 153-156.

Talmy, L. (2000). *Toward a cognitive semantics (Volume 1: Concept structuring systems)*. Cambridge, MA, US: The MIT Press.

Tanenhaus, M. K., Spivey Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Trueswell, J. C., & Tanenhaus, M. K. (Eds.). (2005). *Approaches to studying world-situated language use: bridging the language-as-product and language-as-action traditions*. Cambridge: MIT Press.

Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 44, pp. 35-62). New York: Academic Press.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic mental representations in language comprehension. *Cognitive Science, 28*, 611-619.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science, 13*, 168-171.

Captions

Figure 1. Example stimuli

Figure 2. Total looking time and frequency of path scanning fixations

Figure 1

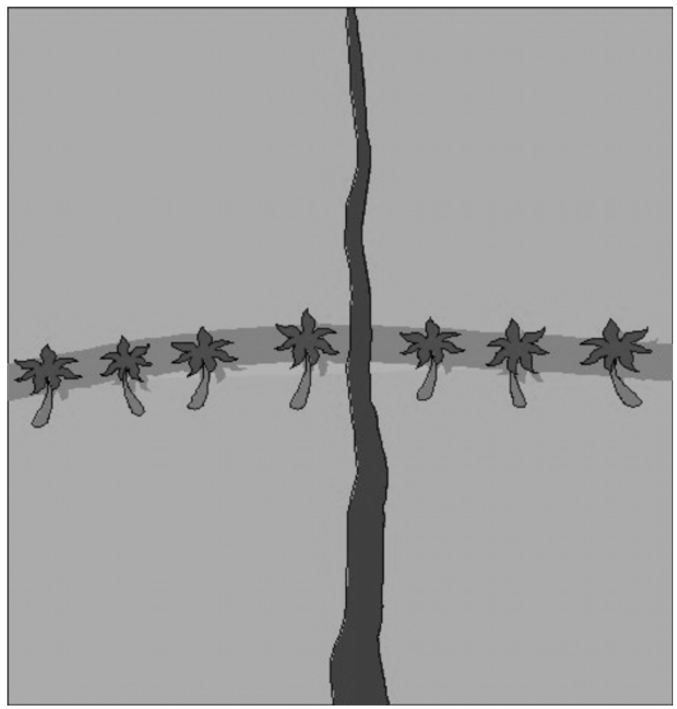
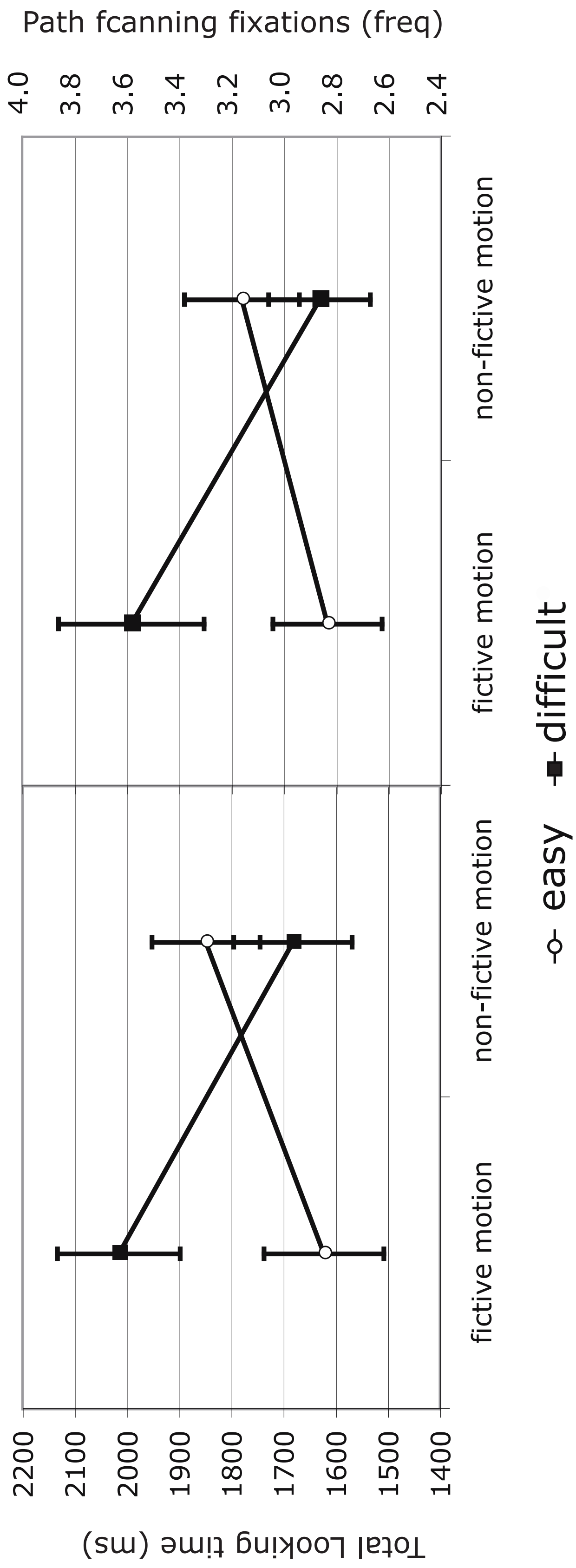
PRIME SENTENCE	PICTURE	TARGET SENTENCE
Easy Terrain The valley is covered with dust Hard Terrain The valley is covered with ruts		Vertical Path Non-Fictive: The road is in the valley Fictive: The road runs through the valley Horiztontal Path Non-Fictive: The palm trees are in the valley Fictive: The palm trees run across the valley

Figure 2


Path fcanning fixations (freq)
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
fictive motion
non-fictive motion
fictive motion
non-fictive motion
2200
2100
2000
1900
1800
1700
1600
1500
1400
Total Looking time (ms)
easy
difficult