How the Geometry of Space controls Visual Attention during Spatial Decision Making

Jan M. Wiener (jan.wiener@cognition.uni-freiburg.de)
Christoph Hölscher (christoph.hoelscher@cognition.uni-freiburg.de)
Simon Büchner (simon.buechner@cognition.uni-freiburg.de)
Lars Konieczny (lars.konieczny@cognition.uni-freiburg.de)
Center for Cognitive Science, Freiburg University, Friedrichstr. 50, D-79098 Freiburg, Germany

Abstract

In this paper we present an eye-tracking experiment investigating the control of visual attention during spatial decision making. Participants were presented with screenshots taken at different choice points in a large complex virtual indoor environment. Each screenshot depicted two movement options. Participants had to decide between them in order to search for an object that was hidden in the environment. We demonstrate (1.) that participants reliably chose the movement option that featured the longest line of sight, (2.) a robust gaze bias towards the eventually chosen movement option, and (3.) using a bottom-up description that captures aspects of the geometry of the sceneries depicted, we were able to predict participants' fixation behavior. Taken together, results from this study shed light onto the control of visual attention during navigation and wayfinding.

Keywords: visual attention; wayfinding; navigation; gaze behavior; spatial cognition; spatial perception.

Introduction

What controls visual attention when navigating through space? In the context of navigation, eye-tracking studies so far primarily investigated the role of gaze for the control of locomotory or steering behavior (Grasso, Prevost, Ivanenko, & Berthoz, 1998; Hollands, Patla, & Vickers, 2002; Wilkie & Wann, 2003). Wayfinding, however, also includes processes such as encoding and retrieving information from spatial memory, path planning, and spatial decision making at choice points (c.f. Montello, 2001). So far, very few, if any, studies made use of eye-tracking techniques to investigate such higher level cognitive processes involved in navigation and wayfinding. For example, which information do navigators attend to and process when deciding between path alternatives? And, how does gaze behavior relate to spatial decision making at all? To approach these questions we presented participants with images of choice points and asked them to decide between two movement options while recording their eye-movements.

In non-spatial contexts, gaze behavior has been shown to reflect preferences in visual decision tasks (Glaholt & Reingold, in press). In two alternative forced choice paradigms in which participants have to judge attractiveness of faces, for example, gaze probability is initially distributed equally between alternatives. Only briefly before the decision, gaze gradually shifts towards the eventually chosen stimulus (Shimojo, Simion, Shimojo, & Scheier, 2003; Simion & Shimojo, 2007). It is an open question whether similar effects can also be observed in spatial decision making such as path choice behavior.

The features people attend to when inspecting images of scenes have been investigated in numerous studies revealing both, bottom-up (stimulus derived) as well as of top-down (e.g., task) influences (for an overview see Henderson, 2003). Already in the 60s, Yarbus (1967) demonstrated influences of the *task* on the control of visual attention: participants' gaze patterns when inspecting the same drawing systematically differed when asked to judge the ages of people depicted or when asked to estimate their material circumstances. The most widely used *bottom-up* approach is that of saliency maps (Itti & Koch, 2000, 2001). A saliency map is a representation of the stimulus in which the strength of different features (color, intensity, orientation) are coded. Several studies demonstrated that saliency maps are useful predictors of early fixations, particularly when viewing natural complex scenes (e.g., Foulsham & Underwood, 2008).

It is important to stress that bottom-up approaches usually do not explicitly account for the fact that images or pictures are two-dimensional projections of three-dimensional scenes. In other words, the geometrical properties of the scenes depicted in the images are not necessarily captured or highlighted by, for example, saliency maps. For navigation and wayfinding, however, the interpretation and understanding of the depicted three dimensional structure may be inevitable. This opens up intriguing questions: Is it possible to predict gaze behavior by analyzing geometrical properties of the sceneries depicted if the viewer is solving a navigation task? If so, can the analysis of gaze behavior be used to infer the strategies and heuristics underlying different navigation or wayfinding tasks? And, which kind of description systems of spatial form and structure captures properties of space that are relevant for the control of visual attention?

Promising candidates are isovists or viewshed polygons (Benedikt, 1979), which both describe the visible area from the perspective of the observer. Isovists are essentially depth profiles and several quantitative descriptors such as the visible area, the length of the perimeter, the number of vertices, etc., can be derived that reflect local physical properties of the corresponding space. Moreover, isovists have been shown to capture properties of the geometry of environments that are relevant for experience of the corresponding space and locomotion within the space (Wiener et al., 2007; Franz & Wiener, 2008).

The specific research questions for this study were as follows:

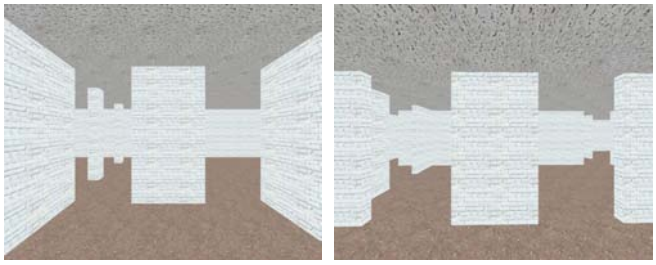

Figure 1: Two examples of decision points presented to participants (in high contrast).

1. How does gaze behavior relate to spatial decision making? Is it possible to predict participants' movement choices during navigation and wayfinding by analyzing their fixation patterns?

2. Where do navigators look when exploring unfamiliar environments? Is it possible to predict gaze behavior by analyzing geometrical properties of the spatial situations encountered?

Method

Participants

Twenty subjects (14 women, mean age: 22.45 $\pm$ 2.83 years) participated in the experiment. They were mostly university students and were paid 8 Euro per hour for participation in that study.

Stimuli

The stimuli were 30 screenshots from within large virtual architectural environments (for examples, see Figure 1). Each screenshot was taken at a decision point, depicting two path alternatives that differed with respect to their spatial form. Pilot experiments suggested that high contrast images as depicted in Figure 1, could be well comprehended parafoveally without gaze shifts. We therefore reduced the contrast of the stimuli by adjusting the colors of floor and ceiling to that of the walls. By this mean participants were forced to overtly attend to the relevant information.

Two versions of each stimulus were generated by mirroring the original stimulus along the vertical axis. Presentation of the original and the mirrored version of the stimuli were balanced between participants.

The spatial structure of the scenes were analyzed using a variant of isovist analysis (Wiener et al., 2007): for each stimulus a depth profile was calculated by contouring the edge between the ground and the walls (see Figure 2 right). The resulting contour essentially describes the distance from the observer to the walls in the stimulus. Although such depth profiles were measured in the 2d pictorial projection of the scenes and are thus compressed around the horizon, they are functionally equivalent to isovists. The angular declination of the lower border of distant walls is smaller than the declination of the lower border or walls close-by (see Figure 2). In fact, the visual system has been shown to be able to use angular declination below the horizon for distance judgments (e.g. Ooi, Wu, & He, 2001).

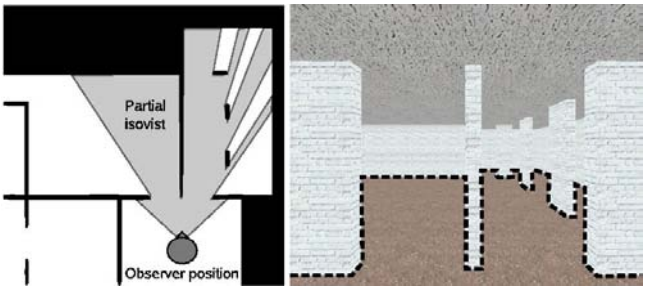

Figure 2: Left: Position in the maze from which one of the snapshots was taken. The Grey area represents the isovist (depth profile) at this position; right: corresponding view in the ego-perspective. The depth profile that is approximated by the dashed line is equivalent to the isovist displayed on the right. Note, however, that large distances are compressed in the depth profile obtained from the image as compared to the actual spatial situation captured by the isovist.

The depth profiles were used to compare spatial properties of the left and right path alternative (left and right half of the stimulus). In particular, we calculated the proportion of the length of the longest line of sight, and compared the number of vertical and horizontal edges. The latter two measures are thought to capture aspects of the spatial complexity of the path alternatives.

Procedure

Participants first read a description of the experiment along with a set of instructions stating that their task was to search for an object (a gold bar) that was placed somewhere in the environment. They would be presented with a series of single choice points at which they had to decide whether to go left or right in order to search for the object. Note that participants had no clue about where to find the target object; in other words, they either had to apply decision strategies that were independent of the stimulus (always turn right, choose randomly, etc) or they had to decide according to other stimulus-related criteria. In the latter case any such criterion would require visual attention and should be reflected in gaze patterns. Instead of actually walking through the environment they would then be presented with the next choice point they would have encountered in the environment. In order to illustrate this procedure, participants were presented with a series of snapshots taken between two choice points.

Before a novel stimulus was presented, participants were required to fixate a small cross in the center of the screen and press the 'Space' bar. Participants pressed the left or right cursor key to report their decision. Each stimulus was presented for 5 seconds, irrespective of when participants responded.

Participants movement decisions (left or right) at individual choice points did not influence which image was pre-

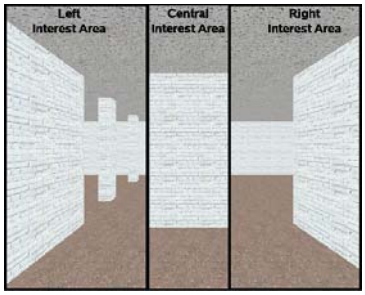

Figure 3: Left: the three interest areas superimposed on one of the stimuli.

sented next, images were presented in random order. The experiment was divided into 5 trials containing 4, 5, 6, 7, or 8 decisions. After the last decision of each trial, participants were presented with an image of a gold bar hovering in a small room.

Apparatus

The stimuli were displayed at a resolution of 1024 x 768 pixels on a 20" CRT monitor. The screen refresh rate was 100 Hz. Eye movements were recorded using a SR Research Ltd. EyeLink II eye tracker, sampling pupil position at 500 Hz. The eye tracker was calibrated using a 9-point grid. A second 9-point grid was used to calculate the accuracy of the calibration. Fixations were defined using the detection algorithm supplied by SR Research.

Analysis

Behavioral data For each stimulus presented participants' decisions (left/right) as well as the corresponding response time was recorded.

Eye movement data For each stimulus we defined three interest areas vertically dividing the image in a left part, a central part, and a right part (see Figure 3). The width of the central interest area was adjusted such as to cover the central wall. Fixations were assigned to the different interest areas. For most of the analyses conducted (unless stated otherwise), we removed the initial fixations directed towards the central interest area, because these initial fixations most likely resulted from the requirement to look at the fixation cross before the stimulus was presented.

Results

Behavioral Data

Response times for the different images ranged between 1793 ms and 2654 ms (mean: 2277 ms). Participants displayed a small yet significant tendency to choose the right over the left movement option (54.07%: T-test against chance level (50%): t(19)=2.28, p=.03) which might be related to the majority of them being right-handed (80%). An analysis of single participants' tendencies to produce stereotypical responses (i.e. to repeatedly choose left movement option or the right movement option) revealed that in 54.78% of the trials, they switched from left to right or from right to left (T-test against chance level [50%]: t(19)=1.30, p=.21). These analyses suggest that participants in fact reacted to the stimuli rather than using other search or navigation strategies such as making right or left turns only.

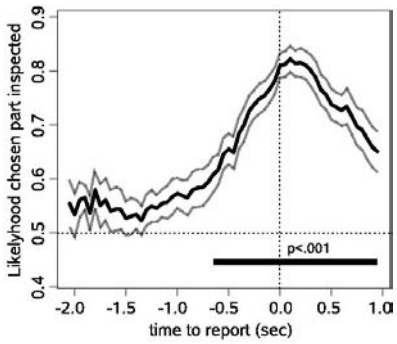

Figure 4: The likelihood that the observer's gaze was directed towards the chosen part of the image (left/right) plotted against time (synchronized at time of decision). The data represent the average across observers (n=20) and trials (n=30).

The absolute difference in the length of the longest line of sight between the left and the right part of the stimuli strongly correlated with participants relative frequency to select the left or the right movement option (r=.64, p<.001). Specifically, participants reliably chose the movement option that featured the longer line of sight.

Eye Movement Data

Fixation Duration. The mean fixation duration towards the left or right interest area before participants reported their decision was 313ms. Fixation durations significantly differed depending on whether or not the eventually chosen interest area was inspected. Fixations directed towards the chosen interest area were longer, lasting 339ms, while fixations towards the non-chosen interest area lasted 280ms (t(19)=-5.58, p<.001).

Time-Course Analyses. The likelihood that observer's gaze was directed towards the (eventually) chosen part of the stimulus changed over the time course of the trials (see Figure 4 left). Approximately 700 msec before participants pressed the button to report their decisions, the likelihood that they inspected the chosen part of the image significantly increased above chance level, reaching a maximum of 82.18% around the time of decision.

Fixation Patterns. Where did participants look when inspecting the stimuli until drawing their decisions? Figure 5 summarizes fixation patterns for the horizontal and vertical stimulus location separately. Most noticeably the distribution of fixation density along the vertical image position was sharply tuned around the horizontal center line of the images.

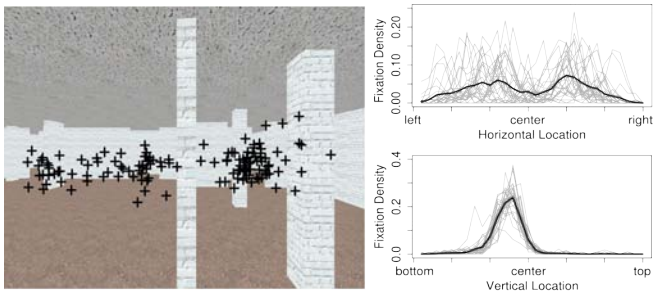

Figure 5: Left: Exemplary fixation pattern for one of the stimuli in the experiment. Single fixations are depicted as black crosses; right: fixation densities for all stimuli for the horizontal (top) and vertical (bottom) image location. Grey lines depict fixation densities for the single stimuli (areas under curve sum up to 1); the black lines reflect the average over all 30 stimuli.

Furthermore, there was very little variance in the fixation positions along the vertical position between stimuli. The distribution of fixation density along the horizontal image position, in contrast, was rather broad and there were considerable differences between the different stimuli (see Figure 5). In other words, participants scanned all spatial scenes approximately at the horizon. Differences in fixation patterns between the different scenes were primarily due to differences in the horizontal dimension. The further analysis will therefore focus on the horizontal axis.

The averaged fixation density along the horizontal image location reveals two maxima, left and right of the vertical centerline of the images. These peaks relate to the two movement options that participants had to inspect and compare in order to decide between them. Figure 6 illustrates typical fixation densities along the horizontal position for three single stimuli. A qualitative analysis of fixation behavior for these stimuli suggests that participants paid close attention to the parts of the image in which the lines of sight were particularly long (see left and right example in Figure 6). Furthermore, fixations densities for the middle image in Figure 6, in which the longest lines of sight are equivalent for both choice alternatives, suggests that fixation density was also modulated by aspects of the local complexity of spatial scene. Note that the fixation density for the left choice alternative, in which several columns are depicted, is higher than for the right choice alternative.

Taking these qualitative observations into account we will now present a tentative model of the control of visual attention in spatial decision making. The model derives its prediction for gaze behavior by analyzing geometrical features of the depicted scene.

Towards a minimalistic model of visual attention in spatial decision making

Does the three-dimensional form of a spatial situation allow predicting gaze behavior when inspecting its two-dimensional projection in an image?

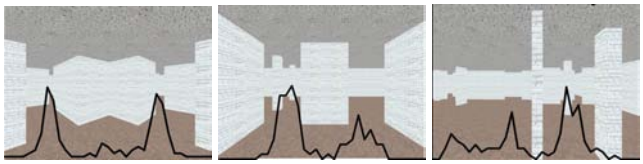

Figure 6: Exemplary fixation densities superimposed on three of the stimuli: Fixations densities (black lines) are plotted as a function of the horizontal position in the image.

The predictors. In order to derive quantitative measures of the geometry of the spatial scenes depicted in the 30 stimuli, we chose to apply a spatial analysis inspired by isovists. This was done for two reasons, (1) because isovists describe the geometry of space from the perspective of the beholder and (2), because earlier studies already demonstrated that isovist analysis captures psychologically and behaviorally relevant properties of space (Wiener et al., 2007). For each stimulus we extracted a *depth profile* directly from the image. This depth profile relates to the distances of the walls from the camera's (i.e. from the observer's) position (see Section Stimuli and Figure 2). Next, this depth profile was downsampled from 1024 bins (the images were 1024x768 pixel) to 30 bins (see Figure 7 A) and normalized such that the area under the curve summed up to 1.0. The resulting depth profile, describing the local geometry, was used as the first predictor for the model.

The depth profile was also used to generate the second predictor, the *depth-edge detector*. Starting from the vertical centerline, it progresses both to the left and to the right and detects all positions along the depth profile at which its orientation changed and exceeded 45 degrees. From these positions only those were taken into account that related to an *increase* in depth. In other words, starting from the center of the image the depth-edge detector highlights all positions at which the length of the line of sight increases sharply. We then applied a Gaussian kernel to the single edges to obtain a smoothed depth-edge profile (see Figure 7 B). Again, the resulting curve was normalized such that the total area under curve was 1.0.

To obtain a model prediction, the two predictors (*depth profile* and *depth-edge detector*) were simply added (see Figure 7).

Model evaluation. For each of the 30 stimuli we calculated the prediction of the model and correlated it with the fixation densities for each stimulus obtained in the experiment. The correlations ranged between r=.30 and r=.83. Average correlation between, the model's predictions and the empirical data was r=.67 (correlation coefficients were Fisher's Z transformed for averaging). The predictive power of the model increased when we smoothed the experimental data with a

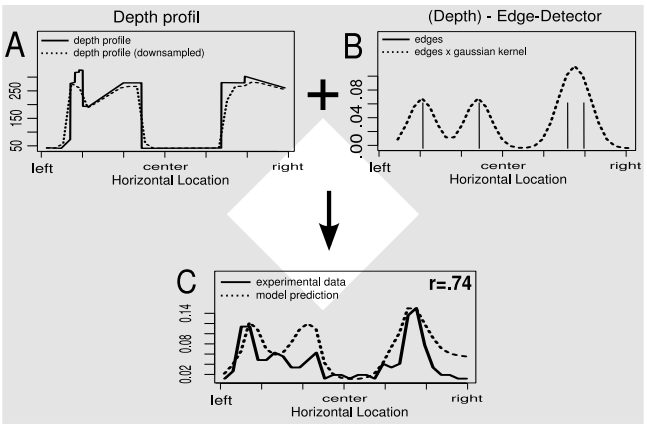

Figure 7: A tentative model of how the geometry of space influences control of visual attention in spatial decision making. (A) depth profile of the original stimulus; (B) Depth-edge detector and smoothed depth-edge profile; (C) The model's prediction and the experimental data. For this particular stimulus the correlation between the model's prediction and the experimental data was r=.74.

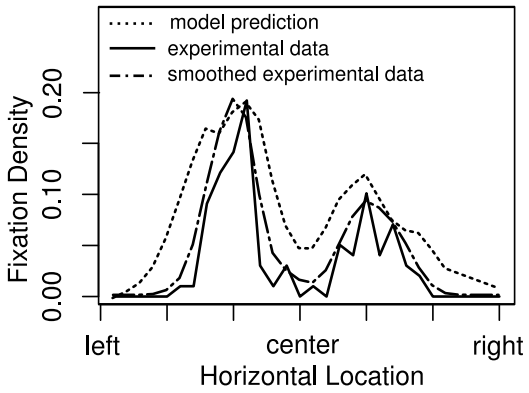

Figure 8: Model prediction, experimental data, and experimental data smoothed by a Gaussian kernel for an exemplary stimulus.

Gaussian kernel (mean correlation between model predictions and smoothed experimental data: r=.78; see Figure 8 for an example).

It should be noted at this point, that the model described above is of tentative nature for a number of reasons: (1) In its current form, the two predictors are not weighted, as if equally contributing to the control of visual attention. Possibly, better fits are obtained if the weights of the two predictors were optimized; (2) The fact that smoothing of the experimental data resulted in a noticeable increase of the predictive power of the model suggests that we might currently suffer from a sparse data problem; (3) In order to extract the predictors, we used depth profiles that were distorted: the depth profiles were extracted from the stimuli directly rather than from the corresponding floorplans. While it has been shown that the visual system can use angular declination below the horizon for distance judgments (e.g. Ooi et al., 2001), better fits may be obtained using non-distorted depth profiles.

Future versions of the model will address the points raised above.

Discussion

In this study, we investigated gaze behavior in the context of navigation and spatial decision making. Participants were presented with images of choice points displaying two different movement options and were asked to decide between them in order to search for an object that was hidden in the environment. We demonstrated that both, participants' movement decisions, as well as their gaze behavior could be predicted by certain geometrical features of the spatial scenes depicted. With respect to movement decisions, participants reliably chose the option that featured the longest line of sight. While related strategies have been demonstrated in other navigation studies (e.g., Conroy Dalton, 2003), it remains unclear why participants chose the option with the longest line of sight. A possible explanation is that the movement option with the longest line of sight promises greater information gain when traveling along than the alternative. However, further research is needed to investigate this behavior.

The analysis of gaze behavior revealed a number of interstsing results. First, gaze behavior reflected the spatial decision making process: approximately 700msec before observers reported their decisions, the likelihood that they inspected the eventually chosen movement option significantly increased above chance level. These results are in line with earlier results on visual decision tasks in non-spatial domains (e.g., Shimojo et al., 2003; Simion & Shimojo, 2007; Glaholt & Reingold, in press). Moreover, the duration of fixations was longer when inspecting the eventually chosen movement option than when inspecting the alternative.

Which parts of the scenery did participants attend to while deciding between path alternatives? Most noticeably, participants' gaze behavior was narrowly tuned along the vertical axis of the stimuli: irrespective of the specific stimulus inspected, viewers focused their fixations around the horizon. This appears to be a sensible viewing strategy in a spatial context, because (1.) information about the geometry of space is most dense around the horizon, and (2.) because by scanning a scenery along the horizon one makes sure that all behaviorally relevant geometrical information is perceived (at least in architectural spaces as used in this study). This suggests that participants were not merely responding to areas with high visual complexity, but were actually analyzing the spatial structure. Fixation densities along the *horizontal* axis systematically differed between stimuli, demonstrating that participants directed their attention to specific features in the environment.

To account for these differences in gaze behavior between different scenes we developed a tentative, minimalistic model of the control of visual attention during spatial decision making. Inspired by isovist analysis, the model extracts a depth

profile describing the visible geometry of the scene and calculates salient geometrical features from that profile. Specifically, starting from the center line and progressing to the edges, the model detects spatial situations in which the line of sight suddenly increases in length. We refer to this as the depth-edge detector. By a simply (unweighted) additive model using the depth profile the depth-edge detector, we obtained quite strong correlations between the model's predictions and the experimental data (r=.67; this correlation even increased when smoothing the experimental data). In other words, by analyzing certain features of the geometry of the depicted scenes – the depth profile, and local changes in the depth profile – we are able to predict where viewers look when deciding which of two movement options to select.

Conclusion

Taken together, results from this study provide evidence that participants did interpret the presented stimuli as three dimensional scenes rather than as flat pictures. While this appears trivial at first glance, it strongly suggests that the geometry of scenes is a relevant factor contributing to the control of visual attention when inspecting corresponding images (at least when faced with spatial tasks such as navigation or wayfinding). Earlier bottom up approaches such as the widely used saliency maps (e.g., Itti & Koch, 2001) as well as recent models combining bottom-up saliency, scene context, and top down influences (Torralba, Oliva, Castelhano, & Henderson, 2006), do not explicitly analyze the spatial structure of the inspected scenes but concentrate on features in the two dimensional projection of the scene. Here we presented a novel bottom-up model that could contribute to a more comprehensive understanding of the control of visual attention. The model specifically analyzes the spatial structure of the scene presented and highlights situations in which the line of sight or the depth profile, respectively, suddenly changes. Apparently these spatial features attract visual attention when visually exploring unfamiliar environments.

Overall, the results suggest that the integrated analysis of navigation behavior and gaze behavior can play a key role in the investigation of the information processing mechanisms and the cognitive strategies underlying human wayfinding behavior.

Acknowledgments

This work was supported by the Volkswagen Foundation and the SFB/TR8 'Spatial Cognition'. Special thanks to J. Wendler, J. Henschel, and A. Günther for their help in carrying out the experiment and analyzing the data.

References

Benedikt, M. L. (1979). To take hold of space: Isovists and isovist fields. *Environment and Planning B*, *6*, 47-65.

Conroy Dalton, R. (2003). The secret is to follow your nose: Route path selection and angularity. *Environment & Behavior*, *35*(1), 107–131.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*, 1–17.

Franz, G., & Wiener, J. (2008). From space syntax to space semantics: a behaviorally and perceptually oriented methodology for the efficient description of the geometry and topology of environments. *Environment & Planning B: Planning and Design*, *35*(4), 574-592.

Glaholt, M. G., & Reingold, E. M. (in press). The time course of gaze bias in visual decision tasks. *Visual Cognition*.

Grasso, R., Prevost, P., Ivanenko, Y., & Berthoz, A. (1998). Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. *Neuroscience Letters*, *253*, 115–118.

Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Hollands, M. A., Patla, A. E., & Vickers, J. N. (2002, Mar). "Look where you're going!": gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental Brain Research*, *143*, 221–230.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.*, *40*, 1489–1506.

Itti, L., & Koch, C. (2001, Mar). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.

Montello, D. R. (2001). Spatial cognition. In *International encyclopedia of the social & behavioral sciences* (p. 14771-14775). Oxford: Pergamon Press.

Ooi, T., Wu, B., & He, Z. (2001, Nov). Distance determined by the angular declination below the horizon. *Nature*, *414*, 197–200.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003, Dec). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*, 1317–1322.

Simion, C., & Shimojo, S. (2007). Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, *69*(4), 591-595.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006, Oct). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113*, 766–786.

Wiener, J., Franz, G., Rossmanith, N., Reichelt, A., Mallot, H., & Bülthoff, H. (2007). Isovist analysis captures properties of space relevant for locomotion and experience. *Perception*, *36*(7), 1066–1083.

Wilkie, R., & Wann, J. (2003). Eye-movements aid the control of locomotion. *Journal of Vision*, *3*, 677–684.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum.