Proceedings of the Annual Meeting of the Cognitive Science Society

Strategies for Orientation: The Role of 3D Landmark Salience and Map Alignment

Permalink
https://escholarship.org/uc/item/5d79337m

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Davies, Clare
Peebles, David

Publication Date
2007

Strategies for Orientation: The Role of 3D Landmark Salience and Map Alignment

Clare Davies (clare.davies@ordnancesurvey.co.uk)
Research Labs, Ordnance Survey,
Romsey Road, Southampton, SO16 4GU, UK.

David Peebles (D.Peebles@hud.ac.uk)
Department of Behavioural Sciences, University of Huddersfield,
Queensgate, Huddersfield, HD1 3DH, UK.

Abstract

An experiment and eye movement study investigated the strategies people use to orientate themselves in urban settings using a streetmap. Previous studies have suggested that the preferred strategy involves choosing salient landmarks to match between the scene and the map. We presented stimuli for which single-landmark matching was not the optimal strategy; the only unambiguous information available for matching was the map's 2D geometry which could also be abstracted from the scene. However, most participants still chose a landmark-based strategy. We discuss the implications for cognitive models, for understanding individual differences, and for potentially improving map designs to aid orientation.*

Keywords: orientation; spatial cognition; individual differences; mental rotation; landmarks

Introduction

Orienting oneself in an environment with the aid of a map is a common problem carried out in a variety of real-world contexts. One must match a direction within the visible scene on the ground with a specific direction on the map, assuming that one already knows where one is located†. It is generally assumed to depend upon mental rotation, to match the 2D and 3D object representations, and as such is assumed to be subject to the individual differences in ability and training which are well known for mental rotation tasks (Dror, 1992; Shepard & Hurwitz, 1984).

One obvious strategy for performing this task involves studying the geometry of the scene in front of the observer, and deriving from it a mental representation of the 2D shapes of the ground layout (as would be seen if viewed from above, i.e. from the map's perspective). However, a short-cut strategy could be employed in many situations, if some salient cue or landmark can be identified in both the scene and the map. The observer could then use the cue as an orientation indicator and would be able to match other items in the scene according to their position relative to it, rather like having a huge *north* arrow marked on the ground.

Previous studies of orientation tasks (e.g., Gunzelmann & Anderson, 2006; Pick, Heinrichs, Montello, Smith, & Sullivan, 1995; Warren, Rossano, & Wear, 1990) have suggested that people do tend to use this landmark-based strategy, picking a salient object or visible feature to match rather than abstracting the overall geometry. This appears to contradict the conclusions of some recent studies of reorientation tasks (e.g., Hermer & Spelke, 1994) which have been concerned with establishing geometric reasoning as a basic cognitive module. In these tasks, after having been disoriented within a room-sized space, animals and children (and sometimes adults) seem to be strongly dependent on the geometry of the space rather than its contents when re-establishing an orientation heading.

One factor in this difference may be that in previous orienting-with-map studies the landmark-based strategy was both possible and appropriate: obvious features in the 3D scene also tended to be obvious on the map. An exception is the recent study of Gunzelmann and Anderson (2006), but this involved highly simplified artificial environments and a task that focused on identifying a specific target among identical geometric shapes within an unstructured circular scene (viewed from outside a circle by the observer), rather than orienting from within a scene (and map) which is the more common task in everyday life. The other previous studies have also focused on environments which were either extremely sparse, or limited to a view of a single building. For this study we wished to extend the paradigm to real-world urban landscapes, and at the same time to test users' strategies and orientation capabilities in a scenario where single-landmark matching would not be so easy. In addition, we wished to check previous findings that the degree of alignment of the map with the space affects performance (even when participants are apparently performing single-landmark matching).

Urban environments provide a rich and varied set of building shapes, road patterns and architectural features. This is especially true for European and other 'evolved' cities, as opposed to New World or other planned settlements (Hillier, Penn, & Dalton, 1992). Very few cities in the UK, for example, are based on a grid or block pattern, but have evolved organically over time so that a wide range of architectural and development styles often coexist within a local area. This means that the 2D geometry is almost always uniquely specified from any point in a typical urban space: it is rarely completely symmetrical and hence tends not to be ambiguous in terms of orientation. This factor has often been surmised to

---

†For the purposes of this study we have assumed a scenario where one knows one's location but not direction. This can occur, e.g., when emerging from a subway station, when viewing a 'you-are-here' map signboard, or on reaching a decision point in navigation having managed without a map until that point.

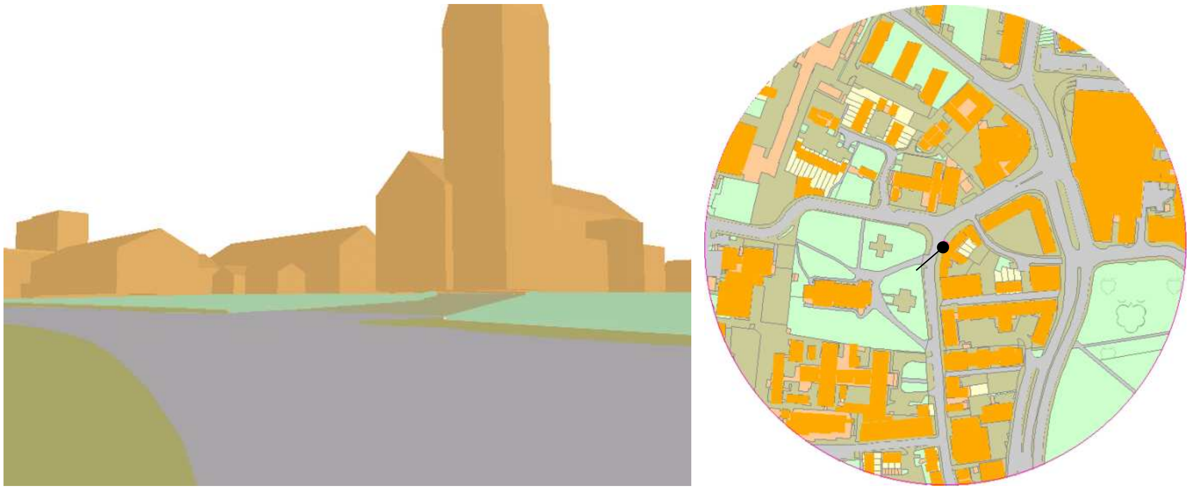
(a) (b)

Figure 1: Scene (a) and corresponding map (b), stimulus 18. © Crown copyright 2007. Reproduced by permission of Ordnance Survey.

have an impact on spatial cognitive processing (e.g., Freundschuh, 1991; Montello, 1991).

We therefore decided to investigate orientation strategies where the scene people viewed was an image taken from a 3D model of a UK city, Southampton, with only the 2D ground layout and the 3D building shapes being shown. The scenes were shorn of irrelevant detail that did not appear on the map and thus could not be used for the task, and the map in turn contained no name labels or other indicators to differentiate buildings and other objects. For items remaining within the scene, the only remaining visual cues were overall size (both in terms of ground area and height), shape (again in terms of both roof line and ground layout), and (one single) colour.

Rather than depicting buildings in their actual colours, the same colour scheme was used for both the scene and map, to emphasise the similarity of their 2D geometry and to facilitate its use in matching. Therefore, choosing a single item based on salient 3D cues (e.g., height), and attempting to match it to the map was unlikely to be successful, since its 2D geometry would probably not be sufficiently unambiguous on its own (but only when combined with other ground layout cues or relative object positions).

Figures 1 and 2 show typical scene-map pairings from our stimuli and Figure 3 shows the actual real-world streets corresponding to the experiment scenes. In Figure 1 two large and distinctive 3D objects can be seen within the scene, revealed by Figure 3a to be a church with a steeple in real life. Neither of these objects can be unambiguously identified on the map however. We hypothesised therefore that people would learn within a few trials to use aspects of the 2D geometry (such as roadside shape or relative object locations) to solve the orientation problem, rather than focusing on these visually salient but task-irrelevant 3D objects. If instead they were distracted by the latter, then performance would be worse for scenes such as Figure 1 than for those such as Figure 2 where heights and shapes were less variable, although 2D ground layout was often just as complex.

Experiment

Method

Design and Participants Forty-nine students and members of staff from the University of Huddersfield took part in the experiment. All participants saw the entire set of stimuli in random order. An additional five participants carried out the experiment while having their eye movements and verbal protocols recorded to enable qualitative assessment of their apparent strategies in solving the task. The other 49 participants were encouraged to perform the task as quickly and accurately as possible.

Materials The experiment was carried out using PC computers with 17 inch displays. The eye movement and verbal protocol study was conducted using a Tobii 1750 remote desktop eye tracker with a 17 inch display. The stimuli were 25 scenes and corresponding maps from various locations in the city of Southampton, UK. The scene images were generated using a buildings-only 3D model overlaid on OS MasterMap® Topography Layer and draped on an OS Land-Form PROFILE® terrain model to provide a realistic and accurate representation of height information (see e.g., Figures 1a and 2a). The maps were circular sections of OS MasterMap® Topography Layer at 1:1250 scale. A black dot in the centre of the map indicated the location of the observer. When the mouse cursor was moved over the map, a short black line of fixed length was drawn from the centre of the dot toward the tip of the cursor (see e.g., Figures 1b and 2b). This rotated around the dot as the mouse was moved around the map so that it always pointed toward the mouse cursor. Scenes and maps were selected to represent a wide range of

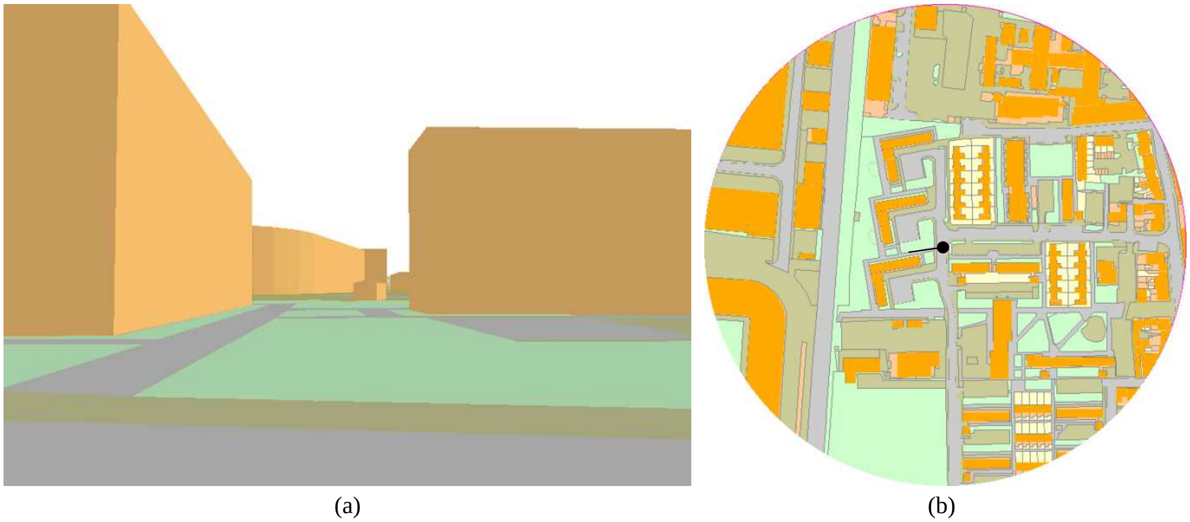
(a) (b)

Figure 2: Scene (a) and corresponding map (b), stimulus 7. © Crown copyright 2007. Reproduced by permission of Ordnance Survey.

building shapes, degrees of salience and distinctiveness, together with a range of urban features such as green spaces and road patterns. The stimuli were also controlled for alignment so that the correct response ranged across the full 0–360 degree circle and there were roughly equal numbers of roughly north- and south-facing scenes.

Procedure Participants were introduced to the experiment through the following scenario: "Imagine that you are standing in the street in an unfamiliar town, holding a map. You know where on the map you are standing, but you need to find out which way you are facing". They were then shown an example scene/map pair and told that their task was to work out in which direction they must be facing on the map in order to see the scene. Participants were instructed how to make a response, asked to respond as rapidly and as accurately as possible, and told that the maps were all the same scale and that they should avoid the natural assumption that the 'upwards' direction on the map indicates 'forward' in the environment (cf. Shepard & Hurwitz, 1984).

There were five practice trials and 20 experiment trials in total. Participants initiated each trial by clicking a button at the top of the screen. On each trial, a scene and corresponding map were presented on the screen as shown in Figures 1 and 2. Participants were able to take as long as necessary to make a judgement. When the participant responded by clicking on the map, the angular degree of the response was recorded, from 0° pointing directly to the top of the map to 180° pointing directly to the bottom, with the sign of the angle indicating left (negative) or right (positive). The duration between the onset of the stimulus to the participant's response was also recorded. Participants in the eye movement and verbal protocol study were asked to talk through each trial as they attempted to solve the problem, in particular to say what they were looking at, how they were thinking through the problem, and why and how they were choosing a particular direction.

Results

Solution strategies In line with previous studies (e.g., Warren et al., 1990) we scored a response as correct if the angle of the centre line fell within 15 degrees of the true angle in either direction (i.e. within the 30 degree range that it bisected), at the point when the participant clicked the mouse. Given that the scenes tended to subtend about 60 degrees of visual angle in total, which is also typical of a photograph taken with a normal camera, this meant that the participants had got within half a scene of the exact line.

In order to test whether performance was influenced by the presence of salient 3D landmarks, the scenes were coded according to the presence or absence of such a landmark. Ten scenes included at least one. Similarly, scenes were also coded according to the presence of distinctive 2D ground layout information in the foreground of the scene, which would facilitate identification on the grounds of 2D layout. Nine of the 20 scenes included such 2D features, e.g., an extensive and irregularly shaped strip of lawn or pavement in the foreground. For example Figure 1 shows a scene that includes both a salient 3D landmark and distinctive ground layout cues, whereas Figure 2 is typical of a scene with predominantly 2D layout cues.

The mean response time and percentage of correct responses for the stimuli categorised by the presence or absence of 2D and 3D cues are presented in Table 1. It should be noted that with the more complex scenes and maps of this experiment, these were typically around 30-40 seconds rather than the few seconds recorded in previous studies (e.g., Gunzelmann & Anderson, 2006). Separate 2×2 repeated-measures ANOVAs were performed on participants' error rates and response times, with presence or absence of 2D and 3D cues as the two within-subjects factors. For errors, there was a significant effect of the presence of salient 3D landmarks, $F(1,48) = 40.35$, $p < 0.0001$, and the presence of distinctive 2D ground layout, $F(1,48) = 5.47$, $p < 0.05$. There was also a significant interaction between them, $F(1,48) = 5.26$, $p < 0.05$. The directions of these effects showed that while presence of an obvious 2D cue was able to decrease error rates, this was only in the absence of a salient 3D cue which always greatly increased them.

The analysis of response times, however, showed that both 3D, $F(1,48) = 29.7$, $p < 0.0001$, and 2D, $F(1,48) = 9.28$, $p < 0.005$, cues seemed to slow participants down. There was again a mild interaction, $F(1,48) = 4.37$, $p < 0.05$, which indicated that the presence of both a 2D and a 3D cue had the most marked effect of all on response times; the presence of a 2D landmark made only a small difference except when a 3D landmark was also present.

Some caution should be expressed with the above analyses since both the response time and error data showed minor deviations from normality; however, the main effects were also checked using non-parametric Wilcoxon signed-rank tests, which showed the same significance patterns (but could not, of course, test the interaction effects).

This finding was independently confirmed by qualitative verbal protocol and eye movement analysis of the 5 additional participants. By far the most commonly reported feature used for solving the problem was ‘buildings’, and the eye movement patterns in the scenes with the most salient 3D landmarks (e.g., large skyscrapers or church steeples) tended to strongly focus around those landmarks.

Table 1: Mean RT (s) and Percentage of Erroneous Responses for stimuli categorised by the presence or absence of 2D and 3D cues

	3D Landmark	
2D layout cue	Present	Absent
Mean RT		
Present	49.0	36.3
Absent	40.7	34.6
% Error		
Present	49.0	26.5
Absent	49.8	39.1

Map alignment Previous studies where a map is matched to a scene have tended to find a distinctive ‘M shape’ pattern in the effect of map alignment with observer position (e.g., Gunzelmann & Anderson, 2006; Hintzman, O’Dell, & Arndt, 1981). Performance typically is better not only at 0 degrees (where ‘up’ on the map exactly corresponds to the forward direction within the scene), but also at 90, 180 and 270 (i.e. -90) degrees. It seems that mental rotation to angles at or close to these cardinal directions is easier than with more oblique angles. In the current study however, these patterns were considerably less clear, as shown in Figure 4. In Figure 4, the response times from our experiment are plotted as a function of map alignment and compared with the M-shaped curve found in Gunzelmann and Anderson (2006), Experiment 1. In order to compare the two data sets on the same axis, the RTs from the Gunzelmann and Anderson study were scaled by a factor of 12. Although the M shape is also partly visible in the RTs from our study, many scenes appear to violate it: indeed, the alignment angles for the three fastest scenes were -53, 76 and -17 degrees. Potential reasons for these findings are discussed below.

Discussion

The persistence of a landmark-based strategy by participants in this study, even when geometry was a far more reliable cue, could be seen to be at odds with the recent assumed pre-eminence of geometry as the primary source of orientation information for both humans and other animals (e.g., Hermer & Spelke, 1994; Cheng & Newcombe, 2005). Generally the question that is posed is whether geometry is so fundamental to cognition that a specific area of the brain has a specially evolved module for it. As summarised by Cheng and Newcombe (2005), this has been followed by the idea that perhaps the ability of adult humans (but possibly not all young children) to use landmarks as well as geometry is due to a relatively late-developing facility aided by language — or at least that it is secondary to a fundamental use of geometric information above all else.

Certainly, with respect to the linguistic angle, it is easy to imagine a participant (and we sometimes saw this) muttering “OK…so find the church” to themselves as they moved from viewing the scene to studying the map. However Cheng and Newcombe (2005) reviewed various evidence that some non-linguistic animals, and sometimes young children, can also use landmarks to help them reorientate in a space (although it should be noted that the reorientation task generally used for these studies is different in a number of respects from our map-matching task). They also argue that other findings that would support a linguistic explanation, based on the apparent disruption of orientation with landmarks by adding a verbal shadowing dual-task paradigm but not a rhythmic one (Hermer-Vazquez, Spelke, & Katsnelson, 1999), may only mean that the linguistic task was harder, since it also disrupted orientation without a landmark.

Whether or not landmark use could involve linguistic processes, or at least those that were in some sense propositional more than spatial, Cheng and Newcombe’s review pointed out that spatial information could be used either approximately, “to tell broadly which direction is which” or precisely to “pinpoint a target location exactly” (p.15). We would argue that studies on location memory (e.g., Lansdale, 1998) have suggested a similarly dual means of encoding spatial location - one precise and easily disrupted (arguably geometric), and the other vague and landmark-related (and again, perhaps more descriptive or propositional than purely spatial) but apparently more robust and persistent in memory. This presents a different and more flexible view of spatial encoding than the

(a)

(b)

Figure 3: Street locations for scenes shown in Figure 1a (a) and Figure 2a (b). © Crown copyright 2007. Reproduced by permission of Ordnance Survey.

one encouraged by the recent focus on geometric modularity.

Cheng and Newcombe (2005) concluded that the evidence to date does not allow us to distinguish between a model of spatial processing that integrates features and geometry into a single representation, and the modular view that geometry is primary and that under different circumstances features may or may not be added to it. While our study was not intended to help distinguish between these two models, the consistent finding by our and other studies in the map-matching paradigm that landmarks are often used *in preference to* geometry suggests that the modularity proponents and opponents may need to take account of a broader range of evidence about spatial cognition. This point was also implied by Cheng and Newcombe's inclusion of studies of people learning spaces from different kinds of perspective, as well as the pure reorientation paradigm, in their review of relevant evidence. Further reviews across the spatial cognition domain might help to resolve these apparent contradictions between task paradigms.

Previous studies have demonstrated people's tendency to match a single salient landmark between a 2D and 3D representation of a scene, and particularly to pick on a landmark with a distinctive 3D (but not 2D) shape despite the absence of 3D cues in the 2D map. The present results indicate that this continues to be a preferred strategy when available, even when not only inappropriate but also discouraged by the nature of the stimuli. In the scenes used in the present study, as in the studies by Gunzelmann and Anderson (2006), the 2D shapes and colours were directly matchable between the scene and the map (though they would not be in real-world scenes or photographs), and all distracting salient cues were removed other than the 3D geometry. Yet participants still made errors through attending to the latter rather than the more reliable 2D geometry. This, along with the slower response times where a 2D ground layout cue was provided (which would still need some spatial transformation to be matched accurately to the map), implies that participants may find it quite difficult to abstract a 2D overhead layout from the 3D scene.

As well as the obvious implications for cognitive modelling of human cognition of large-scale spaces, this may also help to explain the public popularity of bird's eye urban maps that show the buildings from an oblique angle rather than from overhead (e.g., Gombrich, 1982). It also implies that if large-scale maps were to be designed explicitly to aid their use in orientation, it would help to include specific landmarks that could be easily matched to the scene around the traveller or viewer. However, it would not be sufficient merely to include orienting landmarks at places where the 2D geometry was an ambiguous cue, since it may not be used efficiently even when unambiguous.

The disruption of the usual 'M shape' effect of map alignment shown in Figure 4 indicates that map alignment alone (implying a strong role of mental rotation in the task) is not the only factor influencing orientation performance. The scenes which had unexpectedly good performance despite their alignment angle were apparently those where it was relatively easy to match an unambiguous cue to the map, regardless of its angle from the map's upward (north) direction.

The graph therefore show that with a more realistic sampling of typical real-world urban scenes, other factors beyond map alignment must be considered if we are to effectively model the cognitive processes of orientation tasks.

Acknowledgements

The authors wish to thank Claire Cannon and Jon Gould for their help in analysing and interpreting the data, Isabel Sargent, Jon Horgan and Dave Capstick (creators of the 3D building model), Guy Heathcote and Tim Martin for help in using the model and mapping to create the stimuli, the experiment participants for their time and cooperation, Glenn Gunzelmann and Glen Hart for insights and inspiration, and Ordnance Survey of Great Britain for funding and supporting this work.

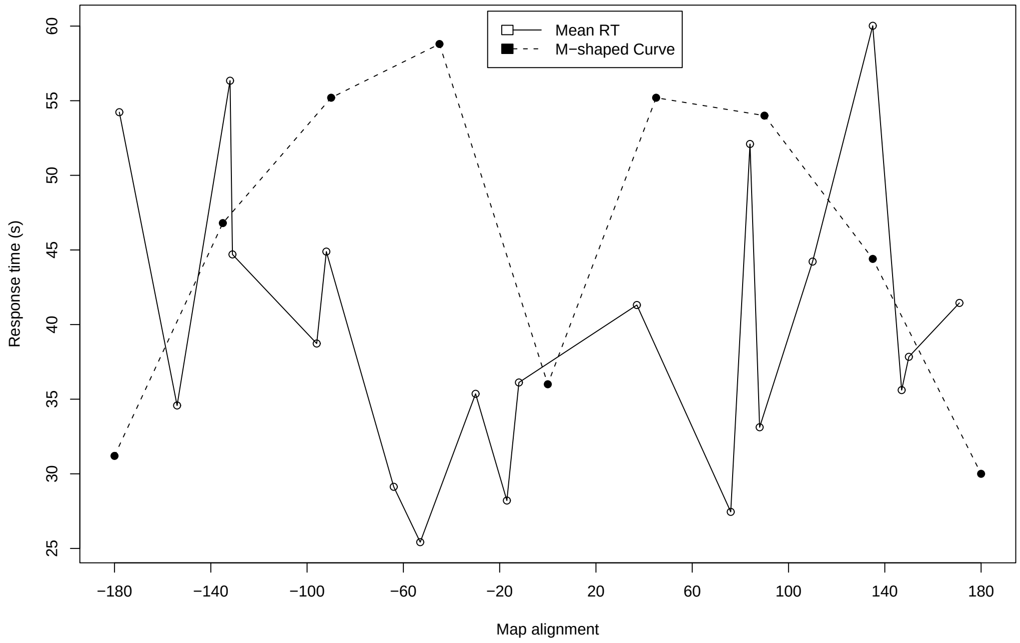

Figure 4: Response latencies plotted as a function of the map's alignment with observer position and compared with the M-shaped RT curve rescaled from Gunzelmann and Anderson (2006), Experiment 1.

References

Cheng, K., & Newcombe, N. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review*, *12*, 1–23.

Dror, I. (1992). Abilities and individual differences in complex skill acquisition. In *Innovations for interactions: Proceedings of the Human Factors Society 36th annual meeting* (Vol. 2, pp. 1368–1372). Santa Monica CA: Human Factors Society.

Freundschuh, S. (1991). The effect of the pattern of the environment on spatial knowledge acquisition. In D. M. Mark & A. U. Frank (Eds.), *Cognitive and linguistic aspects of geographic space* (pp. 167–183). Dordrecht, Netherlands: Kluwer.

Gombrich, E. H. (1982). *The image and the eye: Further studies in the psychology of pictorial representation*. London: Phaidon Press.

Gunzelmann, G., & Anderson, J. R. (2006). Location matters: Why target location impacts performance in orientation tasks. *Memory and Cognition*, *34*, 41–59.

Hermer, L., & Spelke, E. (1994). A geometric process for spatial representation in young children. *Nature*, *370*, 57–59.

Hermer-Vazquez, L., Spelke, E., & Katsnelson, A. (1999). Source of flexibility in human cognition: Dual task studies of space and language. *Cognitive Psychology*, *39*, 3–36.

Hillier, B., Penn, A., & Dalton, N. (1992). Look back to London. *Architects' Journal*, *195 (15)*, 42–45.

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*, 149–206.

Lansdale, M. W. (1998). Modeling memory for absolute location. *Psychological Review*, *105*, 351–378.

Montello, D. R. (1991). Spatial orientation and the angularity of urban routes: a field study. *Environment and Behavior*, *23*, 47–69.

Pick, H. L., Heinrichs, M. R., Montello, D. R., Smith, K., & Sullivan, C. N. (1995). Topographic map reading. In P. A. Hancock, J. Flach, J. Caird, & K. Vicente (Eds.), *Local applications of the ecological approach to human-machine systems* (Vol. 2, pp. 255–284). Hillsdale, NJ: Lawrence Erlbaum.

Shepard, R. N., & Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition*, *18*, 161–193.

Warren, D. H., Rossano, M. J., & Wear, T. D. (1990). Perception of map-environment correspondence: The roles of features and alignment. *Ecological Psychology*, *2*, 131–150.