ERGONOMICS, 1997, VOL. 40, NO. 1, 69–77

The perception of spatial layout in real and virtual worlds

E. J. ARTHUR, P. A. HANCOCK† and S. T. CHRYSLER

Human Factors Research Laboratory, University of Minnesota, 141 Mariucci Arena, 1901 Fourth St. SE, Minneapolis MN 55455, USA

Keywords: Virtual environments; Spatial orientation; Gender differences.

As human-machine interfaces grow more immersive and graphically-oriented, virtual environment systems become more prominent as the medium for human-machine communication. Often, virtual environments (VE) are built to provide exact metrical representations of existing or proposed physical spaces. However, it is not known how individuals develop representational models of these spaces in which they are immersed and how those models may be distorted with respect to both the virtual and real-world equivalents. To evaluate the process of model development, the present experiment examined participant's ability to reproduce a complex spatial layout of objects having experienced them previously under different viewing conditions. The layout consisted of nine common objects arranged on a flat plane. These objects could be viewed in a free binocular virtual condition, a free binocular real-world condition, and in a static monocular view of the real world. The first two allowed active exploration of the environment while the latter condition allowed the participant only a passive opportunity to observe from a single viewpoint. Viewing conditions were a between-subject variable with 10 participants randomly assigned to each condition. Performance was assessed using mapping accuracy and triadic comparisons of relative inter-object distances. Mapping results showed a significant effect of viewing condition where, interestingly, the static monocular condition was superior to both the active virtual and real binocular conditions. Results for the triadic comparisons showed a significant interaction for gender by viewing condition in which males were more accurate than females. These results suggest that the situation model resulting from interaction with a virtual environment was indistinguishable from interaction with real objects at least within the constraints of the present procedure.

1. Introduction

In different forms, virtual representations have a surprisingly long history (Ellis 1991). However there has been a recent and highly visible impetus in the development and promotion of virtual systems (Rheingold 1991, Kruger 1991, Durlach and Mavor 1995). The evolution of graphical computational ability and comparable progress in light-weight, head-mounted visual displays have meant that virtual systems are now accessible to users beyond developmental research laboratories. Virtual environment systems also form part of the obvious evolution of the human-machine interface (Hancock 1996a, b). From early card reading systems through alpha-numeric designators to direct manipulation (Shneiderman 1983) and graphic user interfaces, the manner in which users interact with computer systems has obviously progressed toward the way that individuals interact with the everyday real-world environment (Helander 1989). Although human beings can imagine different dimensional worlds (Abbott 1926), we have evolved and become skilful with our four-dimensional

†Corresponding author.

behavioural environment. Consequently, it is unsurprising that design has progressed in this direction and virtual environments represent the next step along this progression (Furness, 1988, Sheridan 1992, Wells 1992).

The transformation of the interface into a three-dimensional virtual environment does not obviate all problems of interaction. Indeed, there are specific problems in transferring interaction from three- to four-dimensional worlds and unique problems that arise in generating surrogate worlds as linkages to real-world systems. In the present work, we address one of these transition problems, which is the question of the accuracy of the situation model. A crucial question in the adoption of virtual environments for everyday use is the accuracy of the mental representation that comes from interaction with such virtual environments (Peruch and Lapin 1993, Peruch *et al.* 1995). At present, there is limited research that attests to this supposition. Henry and Furness (1993) evaluated virtual and real spaces by comparing estimated room dimensions and object orientation judgements in a real world, with a monitor only, using a head mounted display (HMD) without head tracking, and finally, using an HMD with head tracking. Their results suggested that virtual environment users underestimated the size of environments compared to participants in the real-world conditions. The explanation given by Henry and Furness was that HMDs offered a limited field of view and the edges of HMDs distort the image more than the centre. These factors tended to make the environment appear systematically smaller than it actually was.

Caird and Hancock (1991) studied perceived distances as a function of experience of using a driving simulator. Half of the participants were given 30 mins experience of using the simulator driving around the environment that they were to judge, the other half had no such experience. Participants were then asked to judge the absolute distance from their stationary location to nine locations in the environment and to judge the relative distances between pairs of locations. The results showed that participants in the visual experience group were more accurate in absolute and relative distance judgements than participants who had no exploratory experience. However, it was the case that even the experienced group showed systematic distortions from actual distances, especially for longer inter-location distances. An alternative body of applicable research has used real environments, texts describing environments, and maps to construct and test spatial representations (McNamara 1986, McNamara and LeSueur 1989, McNamara *et al.* 1984, Presson and Hazelrigg 1984, Presson *et al.* 1989, Taylor and Tversky 1992a, b, Tversky 1981). A majority of these studies have shown a distortion of the environment to reflect the routes taken during navigation. For example McNamara *et al.* (1984) presented subjects with a map showing cities and routes. Some cities were connected by routes, others were not. Despite the fact that there were many situations in which cities were the same Euclidean distance apart, participants showed a greater priming effect for cities which were connected over those which were not. Objects that are closer in route space also tend to be recalled with other objects from the same locale. Another well-known phenomenon is that mental representations of space often contain 90° angles where the actual angle observed in the environment is clearly different. Global coherence has a large effect on mental representations of space, for example, Tversky (1981) found that participants were very likely to judge Santiago, Chile as being west of New York City when in fact it is east of NYC. The central point is that mental representations are not veridical with respect to the external space, people distort perceived environments in systematic ways.

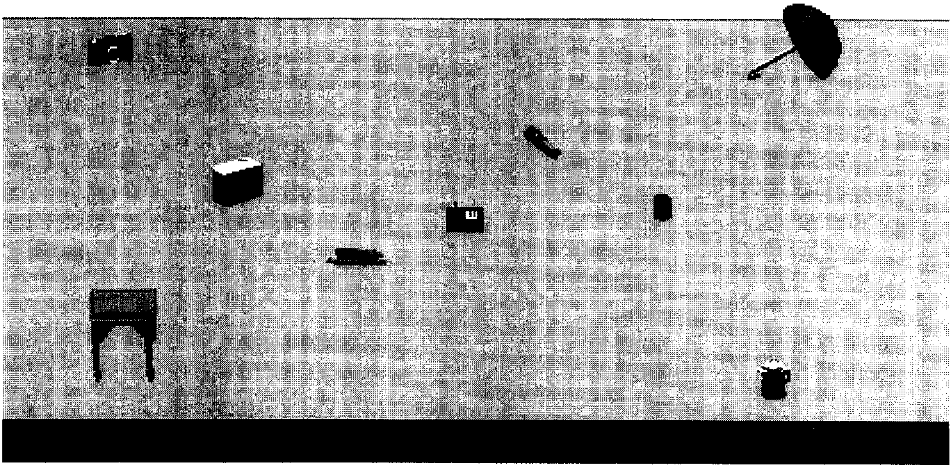

Figure 1. Layout of objects in the room. From left to right: table, camera, toaster, stapler, telephone, flashlight, mug, umbrella, and coffee pot.

Consequently a central issue for the use of virtual environments both as an interface and as a training tool is how users mentally represent virtual space. To the present there has been little research concerning cognitive processes such as navigation, orientation, and memory in virtual environments. The question addressed here is how individuals can map virtual worlds and how their recognition and memory of such worlds correspond to actual conditions. The utility of virtual environments for any application for which they are being proposed, is predicated upon the accuracy of the spatial representation formed in the virtual environment. Therefore, the present experiment examines the utility of virtual environments as a tool for the investigation of tasks requiring a spatial representation. Two binocular viewing conditions were used that were virtual and real-world. There was also a control condition that used one monocular viewing condition, single fixed viewpoint. Participants were allowed unrestricted locomotion in the binocular conditions. The independent variable was the viewing condition, virtual environment (VE), real room (RR), or single fixed viewpoint (SFV), and the dependent variables were accuracy of object placement in a map drawing task and triadic comparison of relative inter-object distances. The hypothesized outcomes were that the real-world condition would be superior to the virtual world condition, and that the virtual world would be superior to the single fixed viewpoint in terms of spatial accuracy. Therefore, accuracy of the inter-object distance should be the highest in the real room condition, next highest in the virtual room, and lowest in the single fixed viewpoint.

2. Experimental method

2.1. *Experimental participants*

Twelve male and eighteen female undergraduates from the University of Minnesota participated in the experiment. They received credit points in exchange for their time. Participants were right-handed and had normal or corrected-to-normal vision.

2.2. *Experimental apparatus*

Performance assessment in the real room and single eye conditions were conducted in a 16 × 18 ft room where the walls, floor, and ceiling were painted matt black. Nine objects: a camera, end table, toaster, umbrella, coffee-pot, stapler, flashlight, telephone, and mug were placed on the floor. The scaled arrangement of these objects is shown in figure 1. One part of each object was covered by purple construction paper, and participants were instructed to use that part of the object as the reference point for subsequent responses. A false wall was constructed and a 1 in. diameter hole was cut at 68 in. as the viewpoint for the single eye condition. The virtual environment contained a three-dimensional graphic representation of the same nine objects. Objects were scaled to appear exactly as they did in the real room. A homeomorphic physical mapping was used so that the same part of the object that was covered by purple paper in the real room was also coloured purple in the VE condition and this surface represented the reference point for all judgements. The objects in the VE condition were situated on a cyan plane, which was the same relative size as the room. The VE system used was a VPL RB2 system using VPLs Isaac and Body Electric software and VPL Dataglove and Eyephone LX (VPL, Redwood City, CA, USA).

2.3. *Experimental procedure*

Participants were randomly assigned to one of the viewing conditions; virtual environment, real room, and single fixed viewpoint. Viewing condition was thus a between-subjects factor. Participants in the virtual environments condition were first given 20 min. experience in a different virtual environment, to familiarize them with locomotion and orientation procedures in VE. After the VE training, participants were offered the opportunity to rest. Participants in the real room and VE conditions were then allowed unrestricted locomotion in the environment.

Instructions for all groups were the same. Participants were requested to study the spatial layout of the objects so that they were subsequently able to draw a map of the objects' position. Participants were also informed that the object names would be provided in the map drawing phase so that they should not be concerned about memorizing the names of objects, only their location. Participants were given as much time as they required to view the environment. When they indicated their readiness, subjects were taken from the room or the virtual world for map drawing and triadic comparisons. They were given a map with two of the objects filled in to provide them with the scale and orientation, and a list of the other seven objects that were to be located. Following the map completion task, participants were asked to make relative distance judgements between object pairs in all of the possible 84 triads (Kosslyn *et al.* 1974). This task required subjects to rank order the distance between the three possible pairs within any triad combination. Participants were debriefed following the completion of the triad comparisons.

3. Experimental results

3.1. *Triadic comparisons*

The data for the triad comparisons were analysed by summing the rank orders across the seven instances of each pair for each participant creating a similarity matrix. This sum was then correlated with the sum of the correct responses for each pair in the questionnaire to assess the accuracy of the subject's rank orderings. Pearson correlations were converted to Z scores using Fischer's R to Z formula yielding a single data point for each subject. The resulting Z scores were analysed by analysis of variance (ANOVA). This analysis is comparable to a two-dimensional multidimensional scaling (MDS) analysis and when the MDS analysis was conducted the result was the same (Davison 1992). While the data analysis was done on Z scores, the results are illustrated in terms of mean correlations as they are more comprehensible. There were

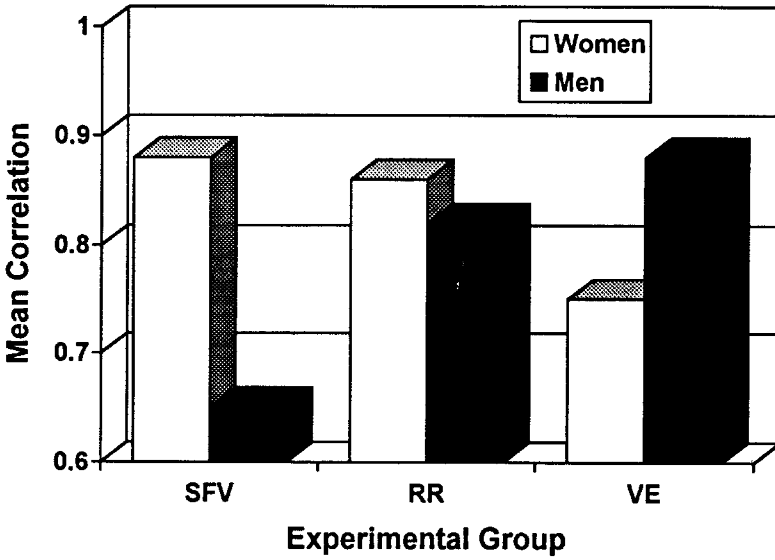

Figure 2. Interaction of Gender by Group for the triadic comparisons

no significant main effects of either group or gender for the triadic comparisons, $F < 1$. The interaction of group and gender was significant. $F[2, 24] = 8{\cdot}13$, $p < 0{\cdot}01$. As can be seen in figure 2, the women performed significantly more effectively than the men in the single eye condition. *Post hoc* analysis on the interaction, using Turkey's HSD, showed that only the difference between men and women in the single eye condition was reliable, $MSe = 0{\cdot}08$, $p < 0{\cdot}05$.

3.2. *Map comparisons*
The data for the maps were prepared in the same way as the scaling questionnaire, except that there was only one observation for each object pair. The dependent measure of interest was the physical distance between points on the maps. Six participants were excluded from this analysis, four because they had forgotten to include one point on their maps, one because pictures of the objects were drawn rather than a dot placed in the object location, and one because the Z score was more than 3 standard deviations from the mean. There was a main effect of group, $F[2, 18] = 3{\cdot}40$, $p < 0{\cdot}05$, and *post hoc* analysis using Tukey's HSD revealed that the single eye condition tended to be more accurate than either the virtual environment or real room, $MSe = 0{\cdot}03$, $p < 0{\cdot}09$, the Z-score means were SFV (1·42), VE (1·25), and RM (1·20). The effect of gender approached significance $F[1, 18] = 4{\cdot}04$, $p < 0{\cdot}06$, with the Z-score means of males (1·39) being more accurate than those of females (1·22). The interaction between group and gender was not significant, $F < 1$.

3.3. *Object orientation*
Franklin and Tversky (1990) have found that observers are faster in verifying the presence of objects that are either directly in front or behind rather than placed laterally relative to the observer and this effect is called the spatial orientation hypothesis. This effect has been observed with situation models constructed by reading spatial descriptions and navigation through environments. Consequently the authors examined whether this effect was present in the current data. The assumption of the analysis was predicated on the map task having fixed the

orientation of the spatial representation and the lack of physical locomotion or visual input during the triadic comparisons had no effect on the orientation of the mental representation, thus the representation was in the same orientation as expressed in the last sketch map. These analyses were limited to objects that were directly in front of each other or beside each other relative to the SFV and the mapping task (figure 1). Objects from the display that met these criteria were selected for further analyses. Five instances of each type (front-back or left-right) were examined. The five instances of each type were correlated with the correct responses, converted to Z scores, and then subjected to ANOVA. Direction was significant for both the triadic comparison, $F(1, 48) = 10{\cdot}61$, $p < 0{\cdot}01$, $MSe = 0{\cdot}36$, and the map construction task, $F(1, 40) = 7{\cdot}02$, $p < 0{\cdot}05$, $MSe = 0{\cdot}03$. The front-back estimates were more accurate than the left-right for the triadic comparisons, where the front-back mean was 0·87 and the left-right mean was 0·70. For the map construction task, the front-back mean was 0·94 and the left-right mean was 0·83.

To analyse whether there were any systematic distortions in object placement, a repeated measures ANOVA was performed on the difference between the participant's estimates of inter-object distances as expressed in their mapping response and the true inter-object distances as expressed in their mapping response and the true inter-object distances between the locations of pairs of objects. There was a main effect of gender, $F(1, 20) = 4{\cdot}65$, $p < 0{\cdot}05$, $MSe = 39{\cdot}85$, where males (3·86 cm) produced less error than females (4·38 cm). The main effect of triad pair was also significant, $F(35,700) = 46{\cdot}38$, $p < 0{\cdot}001$, $MSe = 3{\cdot}31$, and the interaction of viewing condition and object pairs was significant, $F(70, 700) = 1{\cdot}67$, $p < 0{\cdot}001$, $MSe = 3{\cdot}31$. Pairwise comparisons of all of the pairs was conducted using the Tukey A test. Each pairwise comparison contained four objects and there were 630 such comparisons. The occurrence of each object in each significant comparison was tallied. Visual examination of the distribution of significant pairwise comparisons led to the hypothesis that objects on the edges of the maps were involved in a greater number of significant pairwise comparisons than objects that were more centrally located, and further that this difference was only in the horizontal and not vertical dimensions. To test this hypothesis, the authors calculated mean absolute deviations from the centre of the maps for each object and created a mean deviation for each of the viewing conditions. This mean deviation was correlated with the number of significant pairwise comparisons yielding Pearson R values of 0·83 for the binocular real world, 0·46 for the virtual environment, and 0·51 for the single fixed viewpoint. In a similar analysis using constant error, no significant differences were found, thus participants were not consistently overestimating or underestimating distance on the map task by viewing condition or gender.

4. Discussion

The present results suggest that representations formed from the experience of virtual objects does not differ significantly from that of the actual objects. This would support the contention that VE can be used to effectively simulate spatial relations, and that the mental representation resulting from experience in a VE does not differ fundamentally from a mental representation resulting from experience with real objects. While the present result accepts the null hypothesis, both measures utilized in this study converge on the same result. This finding is useful for many types of applications and argues for the utility of VE for human-machine interfaces (Hancock 1996b, Kozak *et al.* 1993).

The finding that the single eye condition was superior to both the virtual and real conditions is of particular interest. The finding that maps were more accurately drawn by participants in the single fixed viewpoint condition on the surface looks surprising but in retrospect others (Thorndyke and Hayes-Roth 1982) have found that individuals who learned spatial layouts from maps were much more accurate in judging Euclidean distances than those who had learned the same environment via navigation. The SFV could be considered very similar to viewing a map because of the single vantage point with which the environment was viewed. The superiority of the single eye condition could have been due to the fact that participants viewed and mapped the objects from that single orientation. Both the VE and RR conditions were viewed from multiple angles and it is uncertain that participants ever saw the environment from the angle from which they were asked to map it. Another possible explanation is that participants in the single eye condition were more constrained in the strategies available to them and the VE and RR conditions were less constrained. Participants could also have been less efficient in the encoding process in the VE and RR conditions. Less efficient here means devoting less attention to the spatial layout and focusing on the objects themselves rather than where they were located in space.

The design of applications directed to the use of virtual environments should recognize the single eye condition superiority as a preliminary result. The task used was primarily assessing the memory component of spatial relations, but dynamic performance in virtual and real environments often has vastly different characteristics in which memory load minimization is a critical design principle (Norman 1992). The increase in accuracy for the spatial orientation hypothesis (front-back versus left-right) that was observed in both the map and the triadic comparison data could be explained by combining the work of Tversky and her associates together with the work of Rieser and his associates (Franklin and Tversky 1990, Rieser 1989). Franklin and Tversky (1990) have shown a speed advantage for objects that are in front or behind a person relative to objects that are either to the left or the right of a person. The effect persists regardless of whether an egocentric or allocentric frame of reference is used. The accuracy finding could be an extension of this effect. The persistence of the effect into the triadic comparisons is not what would be expected, unless the mental representation of the space was still in the orientation specified by the mapping task. Rieser (1989) has shown that the transformation of a mental representation of a space is much slower when the space is rotated mentally, e.g. turn to the left, than when physical locomotion is employed, actually turning to the left. Consequently, it might be posited that the representation is not rotated unless the constraints of the task demand rotation and even then the rotation is costly in terms of mental workload.

The finding that objects on the edge of the maps were placed less accurately than objects at the centre of the map is interesting. It could be that participants perceived that objects in the centre have more constraints on their position or that the constraints are more salient even though objects have the same number of constraints regardless of their centrality in the image. Ferguson and Hegarty (1994) found that landmarks were placed correctly more frequently with greater accuracy than details. One could use this paradigm by making a cluster of objects in the centre and at the edges of the map and examining the accuracy of placement within the cluster by map position.

5. Summary and conclusions

Does the homeomorphic mapping of the physical world with the virtual world translate into a direct perceptual mapping? This is a critical question for the future use of VE that is the overwhelming approach used by programmers at present. This is because direct physical replication contains the implicit assumption that a metrically structured representational environment implies the same perception-action constraints as occur in its real world equivalent. There is a potential fallacy that permeates all simulation, that being the more the simulated world appears like the real world, the more useful it is as a surrogate. Over a century of work in psychophysics has confirmed that the perceiver's appreciation of the real world itself is systematically distorted. However, psychophysics parses perception from action. When this link is re-established, the *average* perception across individuals of the real world accords with its metrical structure. Therefore, it is vital to establish whether actions in virtual environments, which are specifically constructed to involve presence and immersion, are the same as in the real world. Evidence from earlier research in VE suggests that there are systematic distortions (Caird and Hancock 1991). However, these distortions apparently affect the edges of virtual environments or extended distances in surrogate, computer-generated worlds. Results obtained from the present experiment provide evidence that the spatial representation resulting from interactions with small-scale virtual environments is comparable to real-world experience. However, the present work is a first step in understanding spatial representation in VE and clearly additional experimentation is needed in this new area of interface possibilities.

Acknowledgements

The authors would like to thank Dennis Foderberg Director of the Intelligent Transportation Systems Institute of the Center for Transportation Studies at the University of Minnesota, Mn/DOT and FHWA for supporting this work. Susan Chrysler is now with 3M Corporation in St Paul, MN.

References

Abbott, E. A. 1926, *Flatland*, (Oxford University Press, Oxford).

Caird, J. K. and Hancock, P. A. 1991, Perceived spatial layout of a simulated scene as a function of experience, *Proceedings of the Human Factors Society*, **35**, 1447–1450.

Davison, M. L. 1992, Personal communication.

Durlach, N. I. and Mavor, A. S. 1995 (Eds), *Virtual reality: Scientific and Technological Challenges*. National Academy Press, National Research Council, Washington, DC.

Ellis, S. 1991, Nature origins of virtual environments: a bibliographical essay, *Computer Systems in Engineering*, **2**, 321.

Ferguson, E. L. and Hegarty, M. 1994, Properties of cognitive maps constructed from texts, *Memory & Cognition*, **22**, 455–473.

Franklin, N. and Tversky, B. 1990, Searching imagined environments, *Journal of Experimental Psychology: General*, **119**, 63–76.

Furness, T. A. 1988, *Harnessing virtual space.* Human Interface Technology Laboratory Report HITL-M-88-1.

Hancock, P. A. 1996a, Effects of control order, augmented feedback, input device, and practice on tracking performance and perceived workload, *Ergonomics*, **39**, 1146–1162.

Hancock, P. A. 1996b, On convergent technological evolution, *Ergonomics in Design*, **4**(1), 22–29.

Helander, M. (Ed.) 1989, *Handbook of Human-Computer Interaction.* (North-Holland, Amsterdam).

Henry, D. and Furness, T. 1993, *Spatial perception in virtual environments: evaluating an architectural application*, Technical Report No. FU-20. Human Interface Technology Laboratory, University of Washington, Seattle, WA.

Kosslyn, S. M., Pick, H. L., Jr. and Fariello, G. R. 1974, Cognitive maps in children and men, *Child Development*, **45**, 707–716.

Kozak, J. J., Hancock, P. A., Arthur, E. J. and Chrysler, S. T. 1993, Transfer of training from virtual reality, *Ergomonics*, **36**, 777–784.

Kruger, M. 1991, *Artificial Reality II*, (Addison-Wesley, Reading, MA).

McNamara, T. P. 1986, Mental representations of spatial relations, *Cognitive Psychology*, **18**, 87–121.

McNamara, T. P. and LeSeur, L. L. 1989, Mental representations of spatial and nonspatial relations, *Quarterly Journal of Experimental Psychology*, **41A**, 215–233.

McNamara, T. P., Hardy, J. K. and Hirtle, S. C. 1989, Subjective hierarchies in spatial memory, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 211–227.

McNamara, T. P., Ratcliff, R. and McKoon, G. 1984, The mental representation of knowledge acquired from maps, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 723–732.

Norman, D. A. 1992, *Turn Signals are the Facial Expressions of Automobiles*, (Addison-Wesley, Reading, MA).

Peruch, P and Lapin, E. A. 1993, Route knowledge in different spatial frames of reference, *Acta Psychologica*, **84**, 253–269.

Peruch, P., Vercher, J. and Gauthier, G. M. 1995, Acquisition of spatial knowledge through visual exploration of simulated environments, *Ecological Psychology*, **7**(1), 1–20.

Presson, C. C. and Hazelrigg, M. D. 1984, Building spatial representations through primary and secondary learning, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 716–722.

Presson, C. C., DeLange, N. and Hazelrigg, M. D. 1989, Orientation specificity in spatial memory: what makes a path different from a map of the path?, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 887–897.

Rheingold, H. 1991, *Virtual Reality* (Summit Books, New York).

Rieser, J. 1989, Access to knowledge of spatial structure at novel points of observation, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 1157–1165.

Sheridan, T. B. 1992, *Telerobotics, Automation, and Human Supervisory Control* (MIT Press, Cambridge, MA).

Shneiderman, B. 1983, Direct manipulation: a step beyond programming languages, *IEEE Computer*, **16**, 57–69.

Taylor, H. A. and Tversky, B. 1992a, Descriptions and depictions of environments, *Memory & Cognition*, **20**, 483–496.

Taylor, H. A. and Tversky, B. 1992b, Spatial mental models derived from survey and route descriptions, *Journal of Memory and Language*, **31**, 261–292.

Thorndyke, P. W. and Hayes-Roth, B. 1982, Differences in spatial knowledge acquired from maps and navigation, *Cognitive Psychology*, **14**, 560–589.

Tversky, B. 1981, Distortions in memory for maps, *Cognitive Psychology*, **13**, 407–433.

Wells, M. J. 1992, Virtual reality: technology, experience, assumptions, *Human Factors Society Bulletin*, **35**(9), 1–3.

Arthur, E.A., Hancock, P.A., & Chrysler, S.T. (1996). Perception of spatial orientation in real and virtual environments. Ergonomics, 40 (1), 69-77.