For efficient navigational search,

humans require full physical movement but not a rich visual scene

Roy A. Ruddle and Simon Lessels

University of Leeds, UK

Word count: 2457

Corresponding author:

Roy A. Ruddle

School of Computing

University of Leeds

Leeds

LS2 9JT

UK

email: royr@comp.leeds.ac.uk

Tel: +44 (0)113 343 1711

Fax: +44 (0)113 343 5468

This is a preprint of an Article accepted for publication in *Psychological Science* © 2003 American Psychological Society

For efficient navigational search,

humans require full physical movement but not a rich visual scene

ABSTRACT

During navigation, humans combine visual information from their surroundings with body-based information from the translational and rotational components of movement. Theories of navigation focus on the role of visual and rotational body-based information, even though experimental evidence shows they are not sufficient for complex spatial tasks. To investigate the contribution of all three sources of information, we asked participants to search a computer generated "virtual" room for targets. Participants were provided with either only visual information, or visual supplemented with body-based information for all movement (walk group) or rotational movement (rotate group). The walk group performed the task with near-perfect efficiency, irrespective of whether a rich or impoverished visual scene was provided. The visual-only and rotate groups were significantly less efficient, and frequently searched parts of the room at least twice. This suggests full physical movement plays a critical role in navigational search, but only moderate visual detail is required.

During navigation we update knowledge of our position and orientation (spatial updating) to prevent ourselves from becoming lost. This process involves combining body-based information about our translational and rotational movements with other sensory information, principally vision. Theories of navigation focus on the role of visual information and the rotational component of movement (e.g., Gopal, Klatzky, & Smith, 1989; Mou & McNamara, 2002), but experimental evidence highlights many unknowns and suggests translational body-based information is also critical. The objective of the present study was to determine the contribution all three sources of information make to our ability to efficiently perform a navigational search task[1].

The environments people navigate on an everyday basis contain visual cues that act as landmarks (Janzen & van Turennout, 2004) and provide optic flow (Warren, Kay, Zosh, Duchon, & Sahuc, 2001). Studies using virtual environments (VEs) show humans rely on landmarks when they are available (Foo, Warren, Duchon, & Tarr, 2005) and, in rich visual scenes, basic spatial tasks such as path integration may be accurately performed even if no body-based information is provided (Riecke, van Veen, & Bülthoff, 2002). However, visual information alone is not sufficient for cognitively demanding tasks such as learning the layout of a building, as witnessed by the difficulty participants frequently have navigating VEs displayed on a desktop monitor (Ruddle, 2001).

Previous research into the relative importance of translational versus rotational body-based information has been inconclusive. Studies conducted using basic spatial tasks imply that the rotational component of movement is critical, with examples coming from inter-object pointing, path integration and exhaustive search.

[1] I.e., where a person has to travel through a space to search it. By contrast, visual search generally involves eye movements and a single display, and gaze-based search involves head and eye movements from a fixed position.

Participants pointed more accurately and quicker between objects in a room if they physically turned, rather than imagined they turned. However, there was no significant difference between physical and imagined translationary movements (Rieser, 1989; Presson & Montello, 1994; see also Mou, McNamara, Valiquette, & Rump, 2004).

Path integration was performed accurately in an immersive VE[2] that provided optic flow for all movement but body-based information only for rotational movement. By contrast, large errors were made when participants were provided with no body-based information in a VE, only a verbal description, or observed someone else walking the path (Klatzky, Loomis, Beall, Chance, & Golledge, 1998; see also Avraamides, Klatzky, Loomis, & Golledge; 2004).

Participants took substantially longer to exhaustively search a room from a fixed position (gaze-based search) if the direction of view was controlled by hand rather than head movements (Pausch, Proffitt, & Williams, 1997). The cause was attributed to parts of the room being searched more than once with hand movements. In everyday life the use of head musculature to look around is well practiced.

In more complex spatial tasks, full (i.e., translational and rotational) body-based information appears to hold advantages over rotational information on its own, with evidence coming from studies in which participants estimated the direction to targets along a route. In one study (Chance, Gaunet, Beall, & Loomis, 1998), participants were divided into three groups that all used an immersive VE but differed in terms of the body-based information that was provided: (i) none (visual-only), (ii) rotation (participants physically turned but controlled forward speed using a joystick), and (iii) rotation and translation (participants literally walked through the VE while

[2] An immersive VE is one in which a participant has (almost) no view of the outside world. This is most commonly achieved by presenting the VE on a head-mounted display (HMD), which obscures the outside world and leaves the participant visually "immersed" in the VE.

physically situated in a large empty room). Participants who walked estimated directions significantly more accurately than the visual-only group. Rotation-only was not significantly different from either of the other groups. In another study participants either walked a route while viewing video images displayed on an HMD, or viewed recorded video while physically stationary in the laboratory (Waller, Loomis, & Haun, 2004). Again, participants who walked estimated directions significantly more accurately than those who were provided with no body-based information.

Further evidence concerning the minimal contribution made by rotational body-based information comes from studies in which participants learned the layout of a large-scale environment (Ruddle, Payne, & Jones, 1999; Ruddle, & Péruch, 2004). Participants navigated one environment in an immersive VE (rotational body-based information provided) and another in a desktop VE (visual-only), but there was no difference in route knowledge accuracy (distance traveled between specific targets), or any consistent difference across the studies for survey knowledge (estimates of direction and relative straight-line distance).

To investigate the importance of visual information, and rotational and translational body-based information in complex spatial tasks we performed an experiment in which participants searched a room-sized space for eight targets that were randomly placed in 16 explicitly identified, possible locations.

MAIN EXPERIMENT

The experiment was conducted within a photorealistic virtual model of our laboratory. A between participants design was used, with each participant randomly assigned to one of three groups that differed in terms of the type of body-based information that was provided and the visual display (see Table 1).

Method

Participants

Thirty individuals (14 female) with a mean age of 24 years (*SD* = 3.4) took part. All gave informed consent and were paid an honorarium for their participation. The study was approved by the Ethics Committee, Institute of Psychological Sciences, University of Leeds.

Materials

The photorealistic VE model was constructed using measurements of the laboratory's geometry (see Figure 1a) and photographs of the interior. Added to the model were 33 identical cylinders and 16 identical boxes (see Figure 1b) that, in each trial, were placed on top of a cylinder chosen at random. Half of the boxes contained a red target and the others were decoys. In each trial, participants were asked to travel around the VE until they had found the eight targets, pressing either a button on a 3D mouse (walk and rotate groups) or a key on a keyboard (visual-only group) to raise/lower a box's lid to see whether a target was inside. The VE software prevented more than one box lid from being raised at any given moment in time. Another button/key was pressed to indicate a target had been found, causing it to turn blue. The VE was rendered by an SGI Onyx4 graphics workstation at 60 frames/sec, with overall system latency of approximately 50 ms.

Participants in the *walk* group physically walked around the laboratory while viewing the corresponding virtual model on an HMD (48º × 36º field of view; 100% binocular overlap; see Figure 1d). Participants in the *rotate* group stood in one place, viewed the VE on the HMD and achieved movement by physically rotating but holding down a button on the 3D mouse to translate. Thus, the setups of these two

groups were similar to Chance et al. (1998). Participants in the *visual-only* group viewed the VE on a 21-inch monitor and used the mouse and keyboard to change position and orientation. The graphical field of view (48º × 39º) was similar to the angle subtended by the monitor from a normal viewing distance (600mm).

Procedure

Each participant in the visual-only group performed four practice trials to allow familiarization with the interface controls and search task, and then performed four test trials. Participants in the walk and rotate groups did two practice trials using the same system as the visual-only group, and then two more practice trials and the four test trials using the type of movement relevant to their group (walk or rotate). This allowed participants' initial familiarization with the task to take place while sitting in front of a monitor, rather than wearing an HMD that obscured the experimenter.

Results and Discussion

Our interest centered on the efficiency of participants' searches, as defined by the amount of the environment that was visited twice (or more) before a trial was successfully completed. The dependent variable used to measure search efficiency was the number of target and decoy boxes that were checked more than once during a trial. A "perfect" search was one in which no boxes were re-checked.

The rotate and visual-only groups performed 45% and 43%, respectively, of their trials perfectly, and in 10% of trials re-checked at least half of the boxes. The walk group performed 90% of trials perfectly, comparable to participants in an earlier study who performed a similar task in the real world with either a normal field of view (93% trials perfect) or wore goggles that limited field of view to 20º × 16º (87%) (Lessels & Ruddle, in press).

The distribution of the search efficiency data was normalized using a square root transformation. A 3 × 2 × 4 (movement × gender × trial) mixed factorial analysis of variance (ANOVA) showed a significant effect of movement on search efficiency, $F(2, 24) = 9.74$, p-rep = .99, $\eta_p^2 = .45$ (see Figure 2). Bonferroni post-hocs showed that participants who walked re-checked significantly fewer boxes than those in the rotate (p-rep = .97) and visual-only groups (p-rep = .99), these latter two groups being equivalent. The main effects of gender and trial were not significant, and there were no significant interactions.

These results show that both translational and rotational body-based information were necessary for participants to efficiently search a room-sized space for targets. If translational information was not provided then performance was similar to when participants had to search using just visual information.

In previous research, participants pointed to targets along a route significantly more accurately when full body-based information was added to visual information (Chance et al., 1998; Waller et al., 2004). However, never before has experimental evidence demonstrated the importance of the translational component of body-based information, over and above rotational component. In doing so, our findings help explain why participants learned the layouts of buildings at a similar rate both with and without rotational body-based information (Ruddle et al., 1999; Ruddle, & Péruch, 2004)

The visual environment used in the current experiment contained a rather homogenous region of cylinders that was searched, together with many salient surrounding features (e.g., door, cupboards and computers; see Figure 1) that may have helped participants maintain their orientation and, therefore, identify the parts of the cylinder region that had (not) been searched. To investigate whether rich visual

information, as well as full body-based information, was required we conducted a supplementary experiment using an impoverished VE model.

SUPPLEMENTARY EXPERIMENT

For the supplementary experiment we replaced the photorealistic VE model with one that just contained the cylinders, boxes, targets and four gray walls (see Figure 1c). This impoverished environment contained far less visual information for a participant to use. Twenty new participants (12 female) with a mean age of 22 years (*SD* = 4.0) were recruited and randomly assigned to two groups. Half of these participants walked around the impoverished VE and the others moved using mouse and keyboard (visual-only).

Once again, search efficiency was measured by counting the number of target/decoy boxes that were checked more than once during a trial, and the percentage of perfect trials was similar to the main experiment (45% for the visual-only group; 90% for the walk group). The distribution of the search efficiency data was normalized using a square root transformation. A 2 × 2 × 4 (movement × gender × trial) ANOVA showed that the walk group re-checked significantly fewer boxes than the visual-only group, $F(1, 16) = 15.66$, p-rep = .99, $\eta_p^2 = .49$ (see Figure 2). A second ANOVA showed no difference between participants who used the walking interface in the impoverished and photorealistic environments, $F(1, 16) < 0.01$, p-rep = .50, $\eta_p^2 < .01$. No other main effects or interactions were significant in either analysis.

This supplementary experiment showed that rich visual information was not required for efficient searching if full body-based information was provided.

GENERAL DISCUSSION

Our results demonstrate the critical role that body-based information from full physical movement (translation and rotation) plays in navigational search. This is in marked contrast to basic spatial tasks, for which rotational body-based information is sufficient.

A likely explanation is the cognitive demands of a task. In path integration, inter-object pointing and route following, participants were instructed to make particular movements, so they could devote their cognitive resources to updating their position relative to objects in the environment. Our task was a form of foraging with simultaneous target encounters (Stephens & Krebs, 1986). Participants had to plan where to travel, detect every target in their vicinity as they moved, and remember where they had been. Full physical movement allowed detection and position updating to be largely automated, so the information necessary for ongoing planning during a search ("embodied cognition"; see Wilson, 2002) was made available at minimal cognitive cost.

Our results also show that if full body-based information is provided then a rich visual scene is not necessary for efficient searching. This extends to a more complex setting the findings from path integration (Kearns, Warren, Duchon, & Tarr, 2002) and obstacle avoidance (Loomis, Beall, Macuga, Kelly, & Smith, in press).

The present research raises important issues in three distinct areas. First, theoretical models of human navigation and spatial memory tend to focus on the rotational aspects of movement concentrating, for example, on the role of rotation in defining the frames of reference used to accomplish spatial tasks (e.g., Mou & McNamara, 2002). It is now clear that these theories also need to take account of the role of body translation in spatial updating.

Second, concerns have previously been raised that many VEs used to research navigation lack the visual complexity and richness of a real environment and, therefore, are not ecologically valid (Spiers & Maguire, 2004). However, we suggest a far greater concern is the widespread use of desktop environments to study navigation because these provide none of the body-based information that has been shown to be essential.

Third, this study reports the most complex navigational task to date where performance in a VE was comparable to the real world. This represents a notable step toward the creation of a virtual "reality", and highlights the need for renewed efforts to develop effective technologies that allow people to "walk" though large virtual spaces (e.g., Iwata, Yano, Fukushima, & Noma, 2005). Success would have widespread impact on applications in such as training (Farrell et al., 2003), as well as the use of VEs as simulators for studying navigation in realistic settings (e.g., Tarr & Warren, 2002).

Acknowledgements

This work was supported by Grant GR/R55818/01 from the EPSRC. We also thank J. Loomis, A. Ruppertsberg, J. Cutting and the anonymous reviewer for insightful comments about drafts of this article.

REFERENCES

Avraamides, M. N., Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (2004). Use of cognitive versus perceptual heading during imagined locomotion depends on the response mode. *Psychological Science, 15*, 403-408.

Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of

vestibular and proprioceptive inputs to path integration. *Presence: Teleoperators and Virtual Environments, 7*, 168-178.

Farrell, M. J., Arnold, P., Pettifer, S., Adams, J., Graham, T., & MacManamon, M. (2003). Transfer of route learning from virtual to real environments. *Journal of Experimental Psychology: Applied, 9*, 219-227.

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 195-215.

Gopal, S., Klatzky, R. L., & Smith, T. R. (1989). NAVIGATOR: A psychologically based model of environmental learning through navigation. *Journal of Environmental Psychology, 9*, 309-331.

Iwata, H., Yano, H., Fukushima, H., & Noma, H. (2005). CirculaFloor: A locomotion interface using circulation of movable tiles. *Proceedings of the IEEE Virtual Reality Conference (VR'05*, pp. 223-230). Los Alamitos, CA: IEEE.

Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience, 7*, 673-677.

Kearns, M. J., Warren, W. H., Duchon, A. P., & Tarr, M. J. (2002). Path integration from optic flow and body senses in a homing task. *Perception, 31*, 349-374.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science, 9*, 293-298.

Lessels, S., & Ruddle, R. A. (in press). Movement around real and virtual cluttered environments. *Presence: Teleoperators and Virtual Environments*.

Loomis, J. M., Beall, A. C., Macuga, K. L., Kelly, J. W., & Smith, R. S. (in press). Visual control of action without retinal optic flow. *Psychological Science*.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 162-170.

Mou, W., & McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 142-157.

Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques ((SIGGRAPH '97*, 13-18). New York: ACM.

Presson, C. C., & Montello, D. R. (1994). Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception, 23*, 1447-1455.

Riecke, B. E., van Veen, H. A. H. C., & Bülthoff, H. H. (2002). Visual homing is possible without landmarks: A path integration study in virtual reality. *Presence: Teleoperators and Virtual Environments, 11*, 443-473.

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

Ruddle, R. A. (2001). Navigation: Am I really lost or virtually there? In D. Harris (Ed.) *Engineering psychology and cognitive ergonomics (volume 6)*, 135-142. Burlington, VT: Ashgate.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999). Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? *Presence: Teleoperators and Virtual Environments, 8*, 157-168.

Ruddle, R. A., & Péruch, P. (2004). Effects of proprioceptive feedback and environmental characteristics on spatial learning in virtual environments. *International Journal of Human-Computer Studies, 60*, 299-326.

Spiers, H. J., & Maguire, E. A. (2004). A 'landmark' study on the neural basis of navigation. *Nature Neuroscience, 7*, 572-574.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton NJ: Princeton University.

Tarr, M. J. & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience, 5*, 1089-1092.

Waller, D., Loomis, J. M., & Haun, D. B. M. (2004). Body-based senses enhance knowledge of directions in large-scale environments. *Psychonomic Bulletin & Review, 11*, 157-163.

Warren, W. H., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience, 4*, 213-216.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625-636.

TABLE 1

Body-based and visual information provided to each group of participants

Group name	Body-based information		Visual information
	Translation	Rotation	
Walk	yes	yes	Stereo HMD
Rotate	no	yes	Stereo HMD
Visual-only	no	no	Monitor (not stereo)

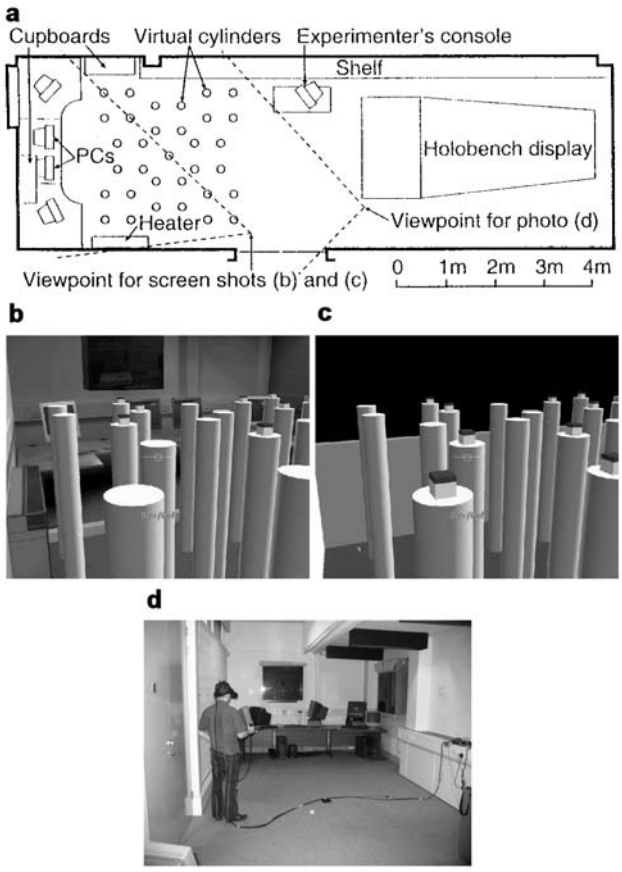

Fig. 1. Experimental setup: (a) Plan view of the physical laboratory, showing location of the virtual cylinders, (b) Photorealistic VE used in the main experiment, (c) Visually impoverished VE used in the supplementary experiment, and (d) Person standing in the position used to generate views (b) and (c), wearing the HMD.

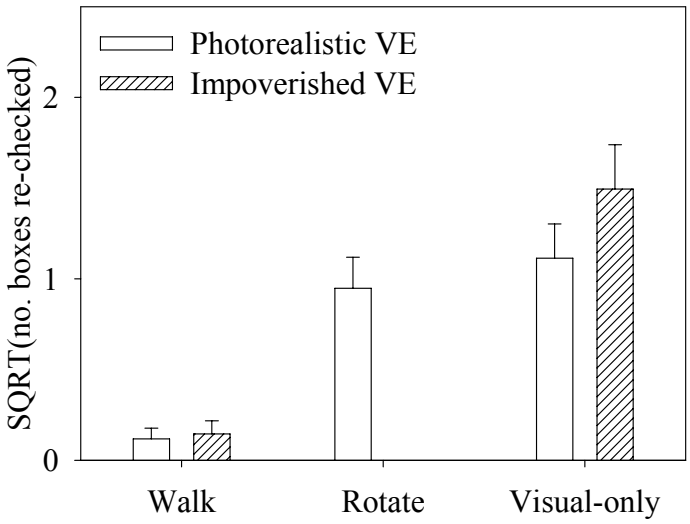

Fig. 2. Search efficiency, defined as mean of √(number of target and decoy boxes re-checked in each trial). Error bars indicate the standard error.