Published paper

Lessels, S. and Ruddle, R.A (2006) *Three levels of metric for evaluating wayfinding.* Presence: Teleoperators and Virtual Environments, 15 (6). pp. 637-654. http://dx.doi.org/10.1162/pres.15.6.637

Three Levels of Metric for Evaluating Wayfinding

Roy A. Ruddle and Simon Lessels

School of Computing, University of Leeds, Leeds, LS2 9JT, UK.

email: royr@comp.leeds.ac.uk

Tel: +44 (0)113 343 5430

Fax: +44 (0)113 343 5468

Ruddle, R. A., & Lessels, S. (2006). Three levels of metric for evaluating wayfinding. *Presence: Teleoperators and Virtual Environments, 15*, 637-654.

Abstract

Three levels of virtual environment (VE) metric are proposed, based on: (1) users' task performance (time taken, distance traveled and number of errors made), (2) physical behavior (locomotion, looking around, and time and error classification), and (3) decision making (i.e., cognitive) rationale (think aloud, interview and questionnaire). Examples of the use of these metrics are drawn from a detailed review of research into VE wayfinding. A case study from research into the fidelity that is required for efficient VE wayfinding is presented, showing the unsuitability in some circumstances of common metrics of task performance such as time and distance, and the benefits to be gained by making fine-grained analyses of users' behavior. Taken as a whole, the article highlights the range of techniques that have been successfully used to evaluate wayfinding and explains in detail how some of these techniques may be applied.

1 Introduction

Navigation is one of the most common types of operation that people perform in virtual environments (VEs), and is a broad term that incorporates many different types of task, from wayfinding (the act of traveling between places), to planning routes and giving directions. Research into navigation has investigated all of these, in addition to component-level tasks that are involved in the process of navigation such as path integration, distance perception, estimates of direction and orientation specificity (e.g., Chance, Gaunet, Beall, & Loomis, 1998; Richardson, Montello, & Hegarty, 1999; Riecke, van Veen, & Bülthoff, 2002; Waller, 2000; Wilson, Foreman, & Tlauka, 1997).

This article focuses on wayfinding. Studies have investigated the effect on wayfinding of various aspects of VE systems design, for example, the content of the visual scene, movement interface, field of view (FOV) and wayfinding aids, but major limitations exist with the methods used for evaluation. In particular, commonly used metrics: (a) are sometimes insufficiently fine grained to detect the differences that exist between different configurations of system (e.g., Lessels & Ruddle, 2004), (b) tend to treat extremes of behavior as a nuisance factor, even though it is these extremes that sometimes identify major problems in usability, and (c) are poor at providing a concrete explanation for any differences in wayfinding performance that are observed.

To help overcome these limitations, and in common with frameworks for the evaluation of human-computer interaction (Sweeney, Maguire, & Shackel, 1993), we propose the use of three distinct levels of metric to evaluate wayfinding. At the bottom level are direct measures of task performance. At a secondary level are metrics that measure users' physical behavior while they navigated, and at the third level are metrics that provide an explanation for this behavior by trying to establish users' underpinning cognitive rationale (or lack of it!). The remainder of this article is divided into two major sections. The first reviews metrics in each of the three levels that have been used to evaluate wayfinding in a variety of user studies, and identifies certain deficiencies. The second major section then shows how some of these metrics have been applied to help evaluate wayfinding in a suite of the authors' own studies. The overall goal of the article is to make a substantial improvement in the techniques used to evaluate wayfinding, and to communicate these techniques to a wider audience. Over a period of time this should assist research into wayfinding and promote closer comparisons between the studies carried out by different researchers than was previously possible.

2 Three levels of metric

The overall purpose of a set of metrics is to assimilate raw data gathered during an evaluation, and to allow key features of these data to be presented in a concise form that allows the meaning of the data to be readily comprehended. Metrics should provide valid data that supports the conclusions drawn in a written report or presentation, and be sufficiently reliable for the data to be recreated in future evaluations.

As we progress from one level of metric to the next the data presented is of finer granularity, providing explanations for differences apparent in the levels above. At the bottom level (Level 1) are metrics that measure how well a user performs a task, which for wayfinding involves a user in finding a particular place. Physical behavior metrics (Level 2) provide information about what a user was doing during a given task, not just how long they took or how accurately they performed. Rationale metrics (Level 3) operate at the highest of the three levels, and can provide an explanation for why users exhibit given behaviors.

Even though, at the end of the day, it is users' performance in wayfinding that counts (Durlach et al., 2000), this performance is driven by the users' decision making processes and subsequent behavior. It is users' rationale that dictates their physical behavior and subtle changes in behavior can have a substantial effect on performance, as is well known within the general field of human-computer interaction (e.g., Gray & Boehm-Davis, 2000; Gray & Fu, 2001). Rationale and behavior metrics facilitate detailed evaluation of wayfinding and can often identify important changes in usability caused by different designs of VE systems, even if those differences have yet to become apparent at the level of task performance. In fact, Rationale and behavior metrics may be used in this way to identify the optimal components of each of several different designs, leading to an improved design that, overall, is superior. Another situation where cognitive metrics have successfully been used in VEs is to compare fidelity with the real world (Mania, Troscianko, Hawkes, & Chalmers, 2003).

The next sections review the metrics in the three levels. Details of the metrics used in each of the published studies that were included in the review are contained in Appendix A.

2.1 Task performance metrics

The most common method of evaluating wayfinding is to directly measure task performance. The key characteristic of any given task performance metric is that it reports a single measure that provides an overview of users' performance during a task, and this can be used as a basis for the statistical analysis of the effect of different factors in a VE's design (e.g., interface or scene fidelity), or to compare users' performance against a previously established norm.

Most evaluations of VE wayfinding use at least one direct measure of task performance, with the most common being the time taken to complete the task, distance traveled, or number of errors made (see Figure 1). Time and distance are well suited to situations where users may navigate freely, whereas counting errors or the number of correct turns that are made are used to measure the accuracy of users' wayfinding when they are not allowed to deviate from a given path.

Care is needed to make sure that the performance metrics chosen are suited to the evaluation being performed, and in this three particular issues should be highlighted. First, consider two users performing a wayfinding task that involves searching for a set of locations. One of the users spends a considerable amount of time planning where to travel and trying to memorize where they have traveled, whereas the other user searches in a near-arbitrary manner (the word “random” is deliberately avoided, because wayfinding movements in VEs are rarely random; users typically travel forwards much more often than they turn left or right; Ruddle, Howes, Payne, & Jones, 2000). In terms of distance traveled the first user has a better wayfinding performance, but in terms of time it is the second user who may well have performed better. This issue arises not just in terms of individual differences, but is also caused by interfaces that impose different speed limits on the rate of travel, as has been shown for menus in a graphical user interface (O’Hara & Payne, 1998) and travel in VEs with a hyperlink interface that allowed users to jump between locations instead of walking (Ruddle et al., 2000).

The second issue also refers to a trade off between the motor and cognitive costs of wayfinding. When searching a spatial layout, users may choose to adopt certain strategies such as repeatedly traveling back to a given, known location (Elvins, Nadeau, Schuk, & Kirsh, 2001; Ruddle et al., 2000; Ruddle & Péruch, 2004). Although this increases the distance traveled, compared with always traveling somewhere new, backtracking makes it easier to remember where one has searched, and so lowers the risk of leaving some sections of the environment completely untouched. Performance metrics sometimes need to take account of actions that were deliberately performed because, even if those actions could be considered to be inefficient at the motor level, they may lower the workload involved in wayfinding and, therefore, actually boost efficiency from a cognitive standpoint. A similar trade off was observed in a computerized version of the game Tetris (Kirsch & Maglio, 1994). Expert players rotated the pieces substantially more often than was necessary, and it was surmised that this occurred because pieces that entered the screen in certain orientations were first rotated to an orientation that made it easier to mentally solve the question of where a piece should be placed (epistemic action) and then rotated and moved to that place (pragmatic action).

The third issue is particularly common in situations where users are assessed in terms of their performance when traveling to several different places. If overall performance is measured (e.g., total distance traveled) then very poor performance finding one of those places may mask the ease with which every other place was found. An example of a performance metric that was developed to overcome this issue is the perfect search metric used in the case study (see §3.2.1 Perfect search metric). This metric measures how close a user was to performing a wayfinding task in a perfect manner, so small mistakes such as narrowly missing a target do not result in a large penalty (e.g., traveling a large additional distance because the whole environment had to be searched again).

2.2 Physical behavior metrics

These metrics measure what users were physically doing, but are much less widely used (22 of the 40 studies in Appendix A) than task performance metrics so there is less of a consensus about the most appropriate data to gather or the ways in which it should be presented. The overall pattern of users' movement encompasses a rich variety of behavior, including the path that is followed, periods of time spent in motion or at rest, head movements and body orientation (Winkel & Sasanoff, 1970). In the present paper, physical behavior metrics are divided into the following main categories: (i) physical actions (subdivided into locomotion, looking around and general observation), (ii) classification of the time spent performing different types of action, and (iii) classification of any errors made (as opposed to the performance metric of simply counting the number of errors made). Details of these are summarized in Table 1 and discussed in the next sections.

2.2.1 Physical actions: Locomotion

The first aspect of users' physical actions that needs to be addressed is metrics to measure where users travel (locomotion) during their wayfinding. The most challenging problem lies in developing metrics that can either combine the movements of different users in a meaningful way or find patterns in the movements made by different users that may, at first glance, appear dissimilar.

One approach is to count the number of times each part of an environment is visited or searched. If the environment is structured like a road network or a building then it can be approximated as a graph (nodes and links) and a count made of the number of traversals of each link or visits to each node. A user's locomotion can then be expressed as either this number (Elvins et al., 2001) or the percentage of the environment that was visited more than a given number of times (Ruddle & Péruch, 2004). This approach is also used to analyze hits on web pages (e.g., Tauscher & Greenberg, 1997) and often presented as a histogram. If an environment has an open structure (e.g., virtual seascapes, campuses, and rooms) then it is necessary to divide it into zones, instead of nodes and links. Once this is done the number of times a user's path enters each zone (Darken & Sibert, 1996; Ruddle & Jones, 2001), or each zone falls within a user's field of view, may be counted. For both types of environmental structure, number of visits data may be superimposed on a plan view of the environment using color, grayscale or line thickness to indicate the relative frequency of visits (Darken & Sibert, 1996; Elvins et al., 2001) and these techniques may be extended to provide time-dependent visualizations that take account of the speed of users' movements (Chittaro & Ieronutti, 2004).

A second approach is to break down users' locomotion into a sequence of events (movements in a straight line or along curves) interspersed by periods when the users pause (Thiel, 1970). Once this has been done it is possible to characterize the nature of users' locomotive movements in terms of the average or profile of distances traveled (or time taken) between pauses, the frequency with which movements are made in different directions (e.g., forward vs. backward), the number or pattern of inputs made to the relevant interface devices (Gamberini, Cottone, Spagnolli, Varotto, & Mantovani, 2003; Sas, O'Hare, & Reilly, 2004; Strommen, 1994), the frequency with which collisions are made with the fabric of the environment (Gamberini et al., 2003; Waller, Hunt, & Knapp, 1998) or use biomechanics techniques to derive motion profiles (Whitton et al., 2005). In the case study below, users' locomotion was characterized by the profile of the distance traveled during each period of discrete movement (see §3.2.3 Movement breakdown; Figure 5).

A third approach is taken by metrics that provide information about wayfinding paths. Plan views have been used to great effect to show the actual path taken by a typical user (Gamberini et al., 2003) or illustrate the extreme difficulty some users have wayfinding in VEs (Ruddle & Jones, 2001). The paths taken by different users may be combined onto a composite map to highlight commonalities in the actual paths, with separate maps sometimes being created for users in different categories, for example, users initially turning left versus right on entering an environment (Winkel & Sasanoff, 1970).

Path information may also be captured using a comprehensive notation for describing users' physical actions during wayfinding (Thiel, 1970). Different channels of information (time, distance, rate, direction and ascent) are presented in parallel, and these may be expressed in terms of movement to, along or across districts, nodes, paths, edges and landmarks, the five basic elements of cities proposed by Lynch (1960). However, the authors are not aware of research that has adopted this notation.

Finally, path information may be expressed by identifying the macro or micro heuristics that users have used. Macro heuristics characterize a user's movement through an environment as a whole. Heuristics that have been identified in VEs include moving directly to a place versus traveling via somewhere else (relevant when revisiting a place), and following the edge (perimeter) of major environmental features, searching from the center of an environment outwards, or following a lawnmower-type pattern (applicable to revisiting places or making an exhaustive search of an environment; Darken & Sibert, 1996; Ruddle, Payne, & Jones, 1999a; Ruddle & Péruch, 2004). In all of these studies the heuristics were assessed subjectively, in some cases by two human raters who worked independently. However, it is recommended that a more objective approach is used, whereby heuristics are described either in concrete terms or as a mathematical equation and tested for goodness of fit with users' actual paths. An example of this approach is shown in the case study, where searches were classified as perimeter-first or lawnmower (see §3.2.2 Macro-level search heuristic; Figure 4), and another comes from research into the paths drawn by participants who were asked to solve the traveling salesman problem for six, 10 or 18 points that were displayed on a piece of paper (Hirtle & Gärling, 1992). Four heuristics were analyzed (nearest neighbor, straight-line (a set of points will be taken in order along the line), zig-zag (lawnmower) and cluster (locations are partitioned into clusters, which are then connected using one of the other heuristics)). Micro heuristics have generally been assessed within the context of an environment's graph (node and link) structure, with the most commonly identified heuristics being backtracking and loops (Elvins et al., 2001; Ruddle et al., 2000; Ruddle & Péruch, 2004; Ruddle, 2005).

2.2.2 Physical actions: Looking around

The second aspect of users' physical actions is how users look around while they travel. For many years, researchers have investigated the effect that the freedom of movement provided by head-tracked rather than stationary displays has on performance in spatial tasks such as targeting radiotherapy treatment beams, target detection and maintaining one's orientation (e.g., Bakker, Werkhoven, & Passenier, 1999; Chung, 1992; Pausch, Shackelford,

& Proffitt, 1993). These studies highlight the benefits of head-tracking in terms of speed and accuracy, but few studies have actually reported behavioral data that relate to how people look around in VEs.

Looking around during wayfinding takes place in three primary frames of reference: (i) view-referenced, (ii) body-referenced, and (iii) world-referenced. In a view reference frame, looking around relates to the movements a user makes to attend to different parts of a display or aspects of a scene that are shown on a display at any given moment in time. Users often have to make specific inputs (e.g., a key press) to view parts of a display such as a map (Darken & Banker, 1998) or a guidebook (Elvins et al., 2001), so it is straightforward for VE software to record the times for which each part was being displayed. Determining the aspects of a virtual scene that a user is attending to, and for how long, involves eye tracking. The use of eye tracking to investigate navigation within hypertext and the WWW is becoming commonplace and, for data analysis and presentation, the path of a user's eye fixation may be overlaid onto a view of the environment, annotated with time stamps and accompanied by a transcript of the user's physical actions (e.g., Card et al., 2001). An identical approach could be applied to VEs.

Looking around within a body reference frame concerns the direction of a user's view in a VE relative to the orientation of their virtual body, whereas looking around in a world reference frame concerns changes in the global direction of the user's view. For a user who is "walking" through a VE, the body and world reference frames are identical for changes of direction in pitch and roll. Changes of heading (yaw) are also identical in the two reference frames if movement is made using the most commonly implemented walking metaphor, which is view-direction movement (also known as gaze-directed movement; Bowman, Koller, & Hodges, 1997). However, changes of heading are different in body and world reference frames if the user’s view and body directions are decoupled, although the basic methods used to quantify the changes remain the same.

One of these methods involves annotating a plan view of the path taken by a user with either lines or view cones that show how the view heading changed during wayfinding (Zanbaka et al., 2004). Animation of the view heading is useful for detailed analysis of users’ actions and snapshots may be used to illustrate behavioral phenomena (see §3.2.5 Looking around; Figure 8). A second method determines the amount of time for which each part of an environment was displayed, taking into account occlusions caused by walls and objects in a VE (Chittaro & Ieronutti, 2004). A third method is to identify patterns of behavior that can be summarized verbally, for example, users systematically looking in every direction at junctions (Strommen, 1994). A fourth method is born out of the fact that it is often desirable to be able to quantify the differences that occur between one design of VE system and another, although examples of this are rare. Useful data that may be reported include the mean rate of direction changes, which have showed that participants look around twice as much when navigating using a head-mounted display (HMD) than a desktop display (Ruddle, Payne, & Jones, 1999b), the total angle users turn by and the number of separate movements in which this is achieved (Sas et al., 2004), the cumulative angle of all direction changes made during a given task, and the profile of those changes (see §3.2.5 Looking around; Figure 7). The profile and rate of change of pitch viewing angles have also been used to assess the causes of different levels of VE sickness in wayfinding and other tasks (Ruddle, 2004).

2.2.3 Physical actions: Observation

Most studies of VE wayfinding have only evaluated what users did within a VE, but occasionally studies have formally recorded and related users’ on and off screen activities (Gamberini et al., 2003; Murray, Bowers, West, Pettifer, & Gibson 2000). The former adapted existing forms of notation to combine information about each user’s physical real-world actions, verbally-reported comments, movements in the VE and events beyond the user’s control that happened in the VE (e.g., a virtual fire starting) into a single transcription. These transcriptions were then analyzed in conjunction with other behavioral and performance data to provide quantitative and qualitative information about the VEs that were being evaluated. By contrast, Murray et al. made a qualitative assessment of similarities in their observations of different users and, drawing on reporting techniques used in areas such as conversation analysis (Sacks, 1992), provided quotes from users that were exemplars of the observed behaviors.

2.2.4 Time classification

These metrics classify behavior in terms of the amount or proportion of time users spend performing different types of action. In wayfinding the primary candidates for classification are: (a) whether or not a user is locomoting (stationary vs. traveling), and (b) whether or not a user is changing the direction in which they are looking within a given frame of reference (static vs. looking around). Some studies have only discriminated between the time users spent stationary versus traveling (Bowman, Johnson, & Hodges, 2001; Ruddle & Jones, 2001) but others include all four combinations of locomoting and changing direction (Elvins et al., 2001; Gamberini et al., 2003). The latter is recommended, and easily implemented if a user’s position and view direction are recorded at the same rate as the VE’s frame rate and each time step coded using bit-wise ORing. The metrics of looking around that were described in the previous section can then be calculated for the whole of a wayfinding task, or for periods when locomotion is either taking place or not (see §3.2.5 Looking around; Figure 7).

Sometimes finer grained classifications of time are performed to distinguish between the amount of time spent traveling in each direction allowed by an interface (e.g., forward vs.

backward; Gamberini et al., 2003). Other studies have adopted different classifications for particular purposes with examples being the amount of time spent viewing environments from different altitudes (Witmer, Sadawski, & Finkelstein, 2002), looking down at one's virtual feet to see the obstacles with which one has collided (Ruddle & Jones, 2001) and traveling along wide versus narrow pathways (Zhai, Kandogan, Smith, & Selker, 1999).

2.2.5 Error classification

Although the number of errors that users make during wayfinding is frequently used as a measure of performance (see above) it is rare that the errors are classified, even though this can highlight some of the root causes of difficulty users have navigating in a given VE. The type of error that is most commonly identified in wayfinding is a miss, which occurs when a user travels within sight of a given location without turning to look at it (Ruddle, Payne, & Jones, 1998; Ruddle & Péruch, 2004). The most likely causes, of course, are the impoverished FOV that is provided by most VE systems and restrictions on the speed at which users can look around, the equivalent of glancing momentarily over one's shoulder in the real world usually being impossible to achieve.

For VEs such as virtual buildings, which are structured like a graph, errors can be classified in terms of where along a path each error was made (e.g., the first vs. a subsequent decision point; Ruddle et al. 1998). VEs that have an open structure can make use of Delaunay triangulation or Voronoi regions (see §3.2.4 Error classification; Figure 6).

2.3 Cognitive rationale metrics

A small number of wayfinding studies (6 of the 40 studies in Appendix A) have attempted to gain direct insights into users' decision-making processes. The best-known technique that provides data while wayfinding is actually taking place is for a user to think aloud. This has successfully been used in studies involving wayfinding / navigation in VEs

(Gamberini et al., 2003; Grammenos, Filou, Papadakos, & Stephanidis, 2003; Murray et al., 2000), the WWW (Card et al., 2001) and the real-world (Hayes-Roth & Hayes-Roth, 1979). Other techniques that provide information about users' decision-making are interviews and questionnaires (Elvins et al., 2001; Whitton et al., 2005; Zhai et al., 1999), and these usually take place after wayfinding has finished. Think aloud can provide detailed information during the process of wayfinding, albeit with the risk that the requirement to think aloud will impede a user's wayfinding performance. Interviews and questionnaires do not affect the process of wayfinding but rely more on the user having an accurate memory, although in an interview the user could be prompted by an observer using notes made during the wayfinding. All of these rationale metrics rely to a certain extent on users being able to elucidate the reasons for the decisions that they make, but wider adoption of the metrics for the evaluation of VE wayfinding would be a positive development.

3 Case study: Investigating VE fidelity

It is well known that people frequently experience great difficulty wayfinding in VEs, and typically take much longer to learn the layout of virtual spaces than equivalent real-world spaces (Richardson et al., 1999; Witmer, Bailey, Knerr, & Parsons, 1996). The authors' current research is investigating what is required if wayfinding tasks are to be performed as easily in a VE as they are in the real world, focusing on the effects of visual fidelity versus FOV (Lessels & Ruddle, 2004) and visual versus movement fidelity (Lessels & Ruddle, 2005). In the research, participants' task is to find targets that are placed in a small number of explicitly identified, possible locations in a 10 x 10 m virtual room. Eight of the locations contain a target and the other eight are decoys (see Figures 1 and 2). Looking once in each location guarantees success. By using room-sized VEs, the effects of fidelity can be studied in the absence of the additional complexity caused by wayfinding in a large-scale space such as a building. Ultimately, however, it is hoped to extend the research findings to large-scale VEs.

The philosophy that guides the research is as follows. If it was possible to build a “virtual reality”, which in all respects was indistinguishable from its real-world counterpart, then it follows that tasks would be as easy to perform in the virtual version of the environment as they were in the physical version. Given the task we are using in our experimental studies is straightforward to perform in the real world (Lessels & Ruddle, 2005), it would also be straightforward in a virtual reality. Unfortunately, this is not possible given the limitations of current VE technology and so users often find the virtual version of the task difficult to perform.

3.1 Overview of experiments

The behavioral experiments completed so far have all used desktop VEs. This section provides an overview of the method. Section 3.2 then explains the problems that were encountered when interpreting the data and how these problems were overcome by applying some of the performance and behavior metrics described in Section 2. For full details of the method and results of the experiments, readers are referred to Lessels & Ruddle (2004; 2005).

The interfaces have allowed participants to travel through the VEs by holding down keys that allow movement to be made either: (a) only in a forward direction, or (b) forward / backward / sideways / diagonally. With all of the interfaces participants’ direction of view was the same as the orientation of their virtual body, and the same method was used to control changes of direction. The amount a participant looked up / down was directly mapped to the vertical position of the cursor on the screen (zero order control), and participants could turn left / right at a rate that increased with the cursor’s horizontal offset from the center of the screen (first-order control). The maximum rate of turning was 135 degrees/second and the view direction did not change at all if the cursor was within the middle 10% of the screen. A slip collision response algorithm was implemented, which automatically guided participants around the cylindrical obstacles in the environment and along the walls.

The VE application was written in C++ and OpenGL Performer^{TM} and ran on an SGI Onyx 3400 with a frame rate of 60 Hz, giving an overall system latency of approximately 30 ms. Six different visual scenes have been utilized. All of the scenes used the same visual representation of the cylinders and target / decoy boxes, but differed substantially in terms of the texture maps used to represent the environment's walls and floor (see Figure 2). Rendering used directional light sources and Gouraud shading. Participants either viewed one computer monitor (48 x 39 degrees; a normal FOV for a VE) or three monitors that were arranged in an arc (144 x 39 degrees; a wide FOV).

A between participants design was used, with different participants taking part in each experiment. After a period spent practicing the interface, each participant performed two practice trials that allowed them to become familiar with the search task, and then completed four test trials. Each trial began at the starting point in the boundary recess (see Figure 1) and participants searched until they had found and selected all eight targets. Participants were informed that the targets were always in the boxes that were on top of the cylinders, but that the position of the boxes and targets changed between trials.

3.2 Application of performance and behavioral metrics

The primary problem faced in interpreting the experimental data was as follows. When participants performed the task in a real-world version of the environment they completed 93% of the trials by checking each target / decoy box only once, which, in terms of the number of places checked, represents a perfect search. In each experimental condition used in the VEs, participants sometimes completed the task perfectly, but on other occasions experienced great difficulty and revisited a substantial proportion of the environment.

Conventional task performance metrics such as the time taken or distance traveled proved to be insufficiently fine-grained to detect differences that did exist in the way that participants searched VEs that were constructed using different combinations of the visual scenes, FOVs and interfaces outlined above. As a result, the data were analyzed using a selection of metrics, namely:

1. Task performance
    a. Proximity to conducting a perfect search (related to the number of errors)
2. Behavior
    a. Locomotion: Macro-level search heuristic
    b. Locomotion: Movement breakdown (start-stop frequency, and time classification)
    c. Error classification
    d. Looking around (quantity, profile and impact on errors made)

3.2.1 Perfect search metric

Overall, participants completed 47% of the VE trials by checking each target / decoy box only once. Of the remaining trials, participants completed some in a near-perfect manner, for example, double checking just one target or decoy, but on other occasions searched large parts of the VE several times. Sometimes this was a direct consequence of having previously traveled past a target without stopping to check it so, once the participant had completed the initial phase of their search, they had no idea of the location of the missing target because they had already traveled through the whole environment (see Figure 3a). When this occurred, some participants searched the environment again in a systematic manner, but others changed their strategy and appeared choose the boxes to recheck on an arbitrary basis. On other occasions the general region that contained a target was left untraveled during the initial search (see Figure 3b).

To distinguish between the above cases, a performance metric was developed that measured how close each trial came to being completed in a perfect manner (checking each target / decoy only once). This metric was calculated by summing the proximity score for each target up until the point when the first target or decoy was revisited. Targets that had been found scored zero, as did targets that the participant had bumped into but not checked. For other targets the score was the closest distance to the target so far. An analysis of variance showed that participants came significantly closer ($p < .05$) to conducting a perfect search when a high fidelity visual scene was combined with a wide FOV than with other combinations of scene and FOV, even though there was no significant difference in terms of total distance traveled or time taken (see Lessels & Ruddle, 2004). Group means for the perfect search ranged from 1.2m (high fidelity / wide FOV) to 5.3m for searches conducted in VEs, compared with 0.05m for those conducted in the real world.

3.2.2 Macro-level search heuristic

In the real world study, 93% of trials used a heuristic that, by dividing the VE into quadrants and writing down the sequence in which they were entered, could unambiguously be classified as either perimeter-first or lawnmower (see Figure 4). Perimeter-first searches visited the quadrants in the order 1-2-3-4-1 (clockwise) or 1-4-3-2-1 (anticlockwise). Lawnmower searches involved a sequence of passes that crossed the VE’s centerline. The same two heuristics were used in 97% of the VE trials but, as has already been explained, only half of these were conducted in a perfect manner. Thus, although real-world and VE behavior was similar at a macro-level, that did not translate to a similarity in search performance. This identified the need to analyze participants’ behavior at a finer level of detail, in terms of their movement, the errors made, and the extent to which they looked around.

3.2.3 Movement breakdown

Two aspects of movement were analyzed: the extent to which participants traveled in one continuous movement from one target / decoy to the next, and the percentage of time spent pausing in between targets and decoys. On average in the VEs participants paused once en-route to every target / decoy in each trial, but never paused between targets and decoys when the task was performed in the real world. The effect of the pauses on the distance participants traveled during each period of continuous movement is highlighted in Figure 5. Overall, these pauses accounted for approximately one third of the time participants took to perform the trials in the VEs and, although a significantly smaller proportion of time was spent pausing when a wide rather than a normal FOV was used ($p < .05$), indicating a shift toward a real-world pattern of behavior, the magnitude of the difference was small (30% vs. 38%) (see Lessels & Ruddle, 2004).

Taken together, these data highlight a clear difference between the smooth, continuous movements participants made when conducting the task in the real world, and the start-stop nature of their movements in a VE. The use of a view-direction walking metaphor for movement in the VE prevented participants from looking around while they traveled in a straight line and that is one factor that is likely to have caused them to pause. However, similar behavior was found in an earlier study that used an HMD and de-coupled the view and body directions, which, in theory at least, made it straightforward for participants to glance to the side while they traveled (Ruddle & Jones, 2001). Therefore, it seems likely that the stop-start behavior arises from more fundamental limitations imposed by VE technology, with candidates being the abstract mechanism (key presses) that is used to control translationary movements even in most immersive VEs, and the impoverished FOV that prevents users from seeing much of their immediate surroundings at one moment in time and, therefore, impedes the planning of movements that are going to be made in the short- to medium-term future. The chunking together of path segments makes routes much easier to learn than if a long sequence of individual cues and wayfinding decisions that make up that route have to be memorized (Golledge, 1992). The changing of movement from continuous (real-world) to discrete segments (VE) is likely to have reduced the extent to which participants were able to chunk together sequences of targets and decoys to make it easier to remember the parts of the environment that had been searched.

3.2.4 Error classification

The perfect search metric highlighted the fact that participants sometimes passed near a target without checking it and on other occasions left substantial areas of a VE untraveled until the rest of the VE had been repeatedly searched. This led to the division of the VE into regions using Delaunay triangulation and the classification of three types of error (<u>miss</u>, <u>local neglect</u> and <u>global neglect</u>; Lessels & Ruddle, 2004) to describe the targets that were not found during the initial stage of a participant's search. A miss was recorded if the participant had previously touched the cylinder on which the target's box was located. Local neglect was recorded if the participant had previously traveled through any of the Delaunay triangles connected to the target's cylinder. Global neglect was recorded for all other errors, indicating that the participant had not been in the target's immediate vicinity (see Figure 6). On average, 1.6 targets were not found during the initial search of each trial and, of these, misses accounted for 5%, local neglect 29% and global neglect 66% (see Lessels & Ruddle, 2004; 2005). The percentage of errors in each category was similar for each combination of visual fidelity, FOV and movement interface, but categorization of the errors allowed the effect of FOV to be assessed in detail (see below).

3.2.5 Looking around

For each trial, the cumulative amount by which the view heading was changed was calculated for periods when a participant was: (i) standing in one place, or (ii) traveling through the VE. The majority (73%) of the change took place when participants had stopped traveling. Due to the fact that all of the VE interfaces used a view-direction walking metaphor for movement, changes of view heading while traveling took place whenever participants changed their direction of travel. When standing in one place some changes of view direction occurred when participants were looking around and others when they stopped, chose a new direction in which to travel and then recommenced movement. It was predicted that participants would look around less when they were provided with a wide FOV, because that allowed them to view a large amount of the environment by scanning the monitor displays rather than turning their virtual head, and this would be exhibited by a change in the profile of the angles by which participants changed their view heading when standing in one place. However, there was little difference in the profiles for wide and normal FOV (see Figure 7).

A finer-grained analysis of looking around was performed to relate changes in participants' view heading to the types of error that were identified above. For each trial where a miss or local neglect occurred, an animation of participants' FOV was overlaid on the path they followed and used to investigate where they were looking during the time when each error was made. In 94% of the misses that participants made with a normal FOV they looked directly at the target concerned when they were standing beside it (Lessels & Ruddle, 2004). With a wide FOV, the target was displayed on the central monitor for 80% of the misses and on either the left or right monitor for the remaining 20%. However, there was a different story for the local neglect (see Figure 8). With a normal FOV, 98% of the targets that were neglected never appeared in participants view when they were in the immediate vicinity of a target (i.e., within the Delaunay triangles that defined the local region). On the other hand, with a wide FOV 56% of these targets were visible on either the left or right monitor when participants were in a target's immediate vicinity, with the other 44% never being visible.

4 Summary

Wayfinding is one of the most often performed and frequently studied tasks in VEs. However, although the vast majority of studies measure users' performance in at least one way, there is a tendency to rely on hypotheses and anecdotal observation to explain any differences that occur. This paper reviews metrics that may be used to measure various aspects of wayfinding behavior and the decision making process that accompanies it. By highlighting the techniques used by some researchers, this paper aims to stimulate the wider adoption of metrics of behavior and rationale, leading to greater insights into the whole process of wayfinding and allowing easier comparison between the results of the studies carried out by different research groups.

To illustrate the practical use of some of the metrics the authors' current research into the effects of visual fidelity, FOV, and movement fidelity on VE wayfinding was used as a case study. These metrics showed that participants used the same general heuristics (perimeter or lawnmower search pattern) to guide their wayfinding in both real world and VE versions of the task, but this did not translate into a similarity in terms of performance. Therefore, it was necessary to make a more detailed analysis of behavior. Breaking down participants' wayfinding into periods when they were either traveling or stationary highlighted the contrast between the continuous nature of participants' movements between targets and decoys in the real world, with the start-stop nature of their movement in VEs. This, and its impact on the ease with which participants could chunk together segments of their path through an environment, provide one explanation for the general difficulty that participants experienced performing the task in a VE. A major challenge for the developers of

VE systems and technology is to provide interfaces that allow free-flowing movement during wayfinding.

The role of FOV was highlighted by a fine-grained analysis of the errors that participants made in conjunction with the extent to which they looked around. Every target that was missed was visible within a participant's FOV when they were standing beside it, so an important question is why did they not check those targets? Did the participant think that they had already checked them, or did they miss them because they attending to other aspects of the visual scene? Use of a think aloud methodology would help to provide an answer.

Analysis of participants' looking around also showed that, with a normal FOV, almost every time a local target was neglected it was simply because the target was outside the FOV. One way of resolving this would be to substantially reduce the time cost of looking around and allowing participants to perform the real-world equivalent of glances to one side while they traveled through a VE. This could be achieved by mapping the view direction directly to the orientation of a sensor (e.g., an Intersense InertiaCube2) instead of using a steering metaphor whereby participants rotated their view direction at a rate that depended on the offset of a cursor from the center of the display. This would increase the speed of precise direction changes, as would be expected with zero- rather than first-order control. Even when a wide FOV was provided, participants frequently neglected targets that were visible on the sides of the combined display. This highlights the need for investigations into different methods of displaying a wide FOV, including distorting the geometric FOV relative to the physical FOV, and the use of eye tracking to determine the parts of a display to which participants attended.

Finally, a large proportion of participants' errors occurred because parts of the VEs were completely neglected during the initial period of wayfinding in each trial. Future use of a think aloud methodology may help explain why this neglect occurred.

Software download

The software used to process data for a variety of the metrics described in this article may be downloaded from http://www.comp.leeds.ac.uk/royr/publications/.

Acknowledgements

This research was supported by grant GR/R55818/01 from the EPSRC.

References

Bakker, N. H., Werkhoven, P. J., & Passenier, P. O. (1999). The effects of proprioceptive and visual feedback on geographical orientation in virtual environments. Presence: Teleoperators and Virtual Environments, 8, 36-53.

Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence: Teleoperators and Virtual Environments, 6, 73-86.

Bowman, D. A., Johnson, D. B., & Hodges, L. F. (2001). Testbed evaluation of virtual environment interaction techniques. Presence: Teleoperators and Virtual Environments, 10, 75-95.

Bowman, D. A., Koller, D., & Hodges, L. F. (1997). Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. Proceedings of the Virtual Reality Annual International Symposium (VRAIS'97, pp. 45-52). Los Alamitos, CA: IEEE.

Card, S. K., Pirolli, P., van der Wege, M., Morrison, J. B., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001). Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability. Proceedings of the Computer Human Interfaces Conference (CHI'01, pp. 498-505). New York: ACM.

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.

Chen, J., Pyla, P. S., Bowman, D. A. (2004). Testbed evaluation of navigation and text display techniques in an information-rich virtual environment. Proceedings of the IEEE Virtual Reality Conference (VR'04, pp. 181-188). Los Alamitos, CA: IEEE.

Chittaro, L., & Ieronutti, L. (2004). A visual tool for tracing users' behavior in virtual environments. Proceedings of the Working Conference on Advanced Visual Interfaces (AVI'04, pp. 40-47). New York: ACM.

Chung, J. C. (1992). A comparison of head-tracked and non-head-tracked steering modes in the targeting oh radiotherapy treatment beams. Proceedings of the 1992 Symposium on Interactive 3D Graphics (pp. 193-196). New York: ACM.

Czerwinski, M., Tan, D.S., & Robertson, G.G. (2002). Women take a wider view. Proceedings of Computer Human Interfaces (CHI'02, pp. 195-202). New York: ACM.

Darken, R. P., & Banker, W. P. (1998). Navigating in natural environments: A virtual environment training transfer study. Proceedings of the Virtual Reality Annual International Symposium (VRAIS 98, pp. 12-19). Los Alamitos, CA: IEEE.

Darken, R. P., & Sibert, J. L. (1996). Navigating large virtual spaces. International Journal of Human-Computer Interaction, 8, 49-71.

Durlach, N., Allen, G., Darken, R., Garnett, R. L., Loomis, J., Templeman, J., & von Wiegand, T. E. (2000). Virtual environments and the enhancement of spatial behavior: Towards a comprehensive research aganda. Presence: Teleoperators and Virtual Environments, 9, 593-615.

Elvins, T. T., Nadeau, D. R., Schuk, R., & Kirsh, D. (2001). Worldlets: 3-D Thumbnails for Wayfinding in Large Virtual Worlds. Presence: Teleoperators and Virtual Environments, 10, 565-582.

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.

Gamberini, L., Cottone, P., Spagnolli, A., Varotto, D., & Mantovani, G. (2003). Responding to a fire emergency in a virtual environment: different patterns of action for different situations. Ergonomics, 46, 842-858.

Golledge, R. G. (1992). Place recognition and wayfinding: Making sense of space. Geoforum, 23, 199-214.

Grammenos, G., Filou, M., Papadakos, P., & Stephanidis, C. (2002). Virtual prints: Leaving trails in virtual environments. Proceedings of the Eighth Eurographics Workshop on Virtual Reality (pp. 131-138). Aire-la-Ville, Switzerland: Eurographics Association.

Grant, S. C., & Magee, L. E. (1998). Contributions of proprioception to navigation in virtual environments. Human Factors, 40, 489-497.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. Journal of Experimental Psychology: Applied, 6, 322-335.

Gray, W. D., & Fu, W. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. Proceedings of the Computer Human Interfaces Conference (CHI’01, pp. 112-119). New York: ACM.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. Cognitive Science, 3, 275-310.

Hirtle, S. C., & Gärling, T. (1992). Heuristic rules for sequential spatial decisions. Geoforum, 23, 227-238.

Howes, A., Miles, G. E., Payne, S. J., Davies, A. J., & Mitchell, C. D. (2001). Incidental memory and navigation in panoramic virtual reality for electronic commerce. Human Factors, 43, 239-254.

Kirsch, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. Cognitive Science, 18, 513-549.

Lessels, S., & Ruddle, R. A. (2004). Changes in navigational behaviour produced by a wide field of view and a high fidelity visual scene. Proceedings of the 10th Eurographics Symposium on Virtual Environments (EGVE'04, pp. 71-78). Aire-la-Ville, Switzerland: Eurographics Association.

Lessels, S., & Ruddle, R. A. (2005). Movement around real and virtual cluttered environments. Presence: Teleoperators and Virtual Environments, 14, 580-596.

Lynch, K. (1960). The image of the city. Cambridge: MIT.

Mania, K., Troscianko, T., Hawkes, R., & Chalmers, A. (2003). Fidelity metrics for virtual environment simulations based on spatial memory awareness states. Presence: Teleoperators and Virtual Environments, 12, 296-310.

Murray, C. D., Bowers, J. M., West, A. J., Pettifer, S., & Gibson, S. (2000). Navigation, wayfinding, and place experience within a virtual city. Presence: Teleoperators and Virtual Environments, 9, 435-447.

O’Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. Cognitive Psychology, 35, 34-70.

O’Neill, M. J. (1992). Effects of familiarity and plan compexity on wayfinding in simulated buildings. Journal of Environmental Psychology, 12, 319-327.

Parush, A., & Berman, D. (2004). Navigation and orientation in 3D user interfaces: The impact of navigation aids and landmarks. International Journal of Human-Computer Studies, 61, 375-395.

Pausch, R., Shackelford, M. A., & Proffitt, D. (1993). A user study comparing head-mounted and stationary displays. Proceedings of IEEE Symposium on Research Frontiers in Virtual Reality (pp. 41-45). Los Alamitos, CA: IEEE.

Pierce, J. S., & Pausch, R. (2004). Navigation with place representations and visible landmarks. Proceedings of the IEEE Virtual Reality Conference (VR'04, pp. 173-180). Los Alamitos, CA: IEEE.

Regian, J. W., Shebilske, W. L., & Monk, J. M. (1992). Virtual reality: An instructional medium for visual-spatial tasks. Journal of Communication, 42, 136-149.

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. Memory and Cognition, 27, 741-750.

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.

Ruddle, R. A. (2004). The effect of environment characteristics and user interaction on levels of virtual environment sickness. Proceedings of the IEEE Virtual Reality Conference (VR'04, pp. 141-148). Los Alamitos, CA: IEEE.

Ruddle, R. A. (2005). The effect of trails on first-time and subsequent navigation in a virtual environment. Proceedings of the IEEE Virtual Reality Conference (VR'05, pp. 115-122). Los Alamitos, CA: IEEE.

Ruddle, R. A., Howes, A., Payne, S. J., & Jones, D. M. (2000). The effects of hyperlinks on navigation in virtual environments. International Journal of Human-Computer Studies, 53, 551-581.

Ruddle, R. A., & Jones, D. M. (2001). Movement in cluttered virtual environments. Presence: Teleoperators and Virtual Environments, 10, 511-524.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings in "desk-top" virtual environments: Experimental investigations using extended navigational experience. Journal of Experimental Psychology: Applied, 3, 143-159.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1998). Navigating large-scale “desk-top” virtual buildings: Effects of orientation aids and familiarity. Presence: Teleoperators and Virtual Environments, 7, 179-192.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999a). The effects of maps on navigation and search strategies in very-large-scale virtual environments, Journal of Experimental Psychology: Applied, 5, 54-75.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999b). 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.

Sacks, H. (1992). Lectures on conversation (Ed. G. Jefferson). Oxford: Blackwell.

Sas, C., O’Hare, G., & Reilly, R. (2004). A performance analysis of movement patterns. In M. Bubak, G. D. Albada, P. M. A. Sloot, & J. Dongarra (Eds.), Lecture notes in computer science, 3038, 954-961. Berlin: Springer-Verlag.

Steck, S. D., & Mallot, H. A. (2000). The Role of Global and Local Landmarks in Virtual Environment Navigation. Presence: Teleoperators and Virtual Environments, 9, 69-83.

Strommen, E. (1994). Children’s use of mouse-based interfaces to control virtual travel. Proceedings of Computer Human Interfaces (CHI’94, pp. 405-410). New York: ACM.

Swan II, J. E., Gabbard, J. L., Hix, D., Schulman, R. S. & Kim, K. P. (2003). A comparative study of user performance in a map-based virtual environment. Proceedings of the IEEE Virtual Reality Conference (VR’03, pp. 259-266). Los Alamitos, CA: IEEE.

Sweeney, M., Maguire, M., & Shackel, B. (1993). Evaluating user-computer interaction: A framework. International Journal of Man-Machine Studies, 38, 689-711.

Tan, D. S., Robertson, G. G., & Czerwinski, M. (2001). Exploring 3D navigation: Combining speed-coupled flying with orbiting. Proceedings of Computer Human Interfaces (CHI'01, pp. 418-425). New York: ACM.

Tate, D. L., Sibert, L., & King, T. (1997). Virtual environments for shipboard firefighting training. Proceedings of the Virtual Reality Annual International Symposium (VRAIS'97, pp. 61-68). Los Alamitos, CA: IEEE.

Tauscher, L., & Greenberg, S. (1997). How people revisit web pages: Empirical findings and implications for the design of history systems. International Journal of Human-Computer Studies, 47, 97-137.

Theil, P. (1970). Notes on the description, scaling, notation, and scoring of some perceptual and cognitive attributes of the physical environment. In H. M. Proshansky, W. H. Ittelson, & L. G. Rivlin (Eds.), Environmental psychology: Man and his physical setting (pp. 593-619). New York: Holt, Rinehart and Winston.

Tlauka, M., & Wilson, P. N. (1994). The effects of landmarks on route-learning in a computer-simulated environment. Journal of Environmental Psychology, 14, 305-313.

Waller, D. (2000). Individual differences in spatial learning from computer-simulated environments. Journal of Experimental Psychology: Applied, 6, 307-321.

Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. Presence: Teleoperators and Virtual Environments, 7, 129-143.

Whitton, M. C., Cohn, J. V., Feasel, J., Zimmons, P., Razzaque, S., Poulton, S. J., McLeod, B., & Brooks Jr., F. P. (2005). Comparing VE locomotion interfaces. Proceedings of the IEEE Virtual Reality Conference (VR'05, pp. 123-130). Los Alamitos, CA: IEEE.

Wilson, P. N., Foreman, N., & Tlauka, M. (1997). Transfer of spatial information from a virtual to a real environment. Human Factors, 39, 526-531.

Winkel, G. H., & Sasanoff, R. (1970). An approach to an objective analysis of behavior in architectural space. In H. M. Proshansky, W. H. Ittelson, & L. G. Rivlin (Eds.), Environmental psychology: Man and his physical setting (pp. 619-631). New York: Holt, Rinehart and Winston.

Witmer, B. G., Bailey, J. H., Knerr, B. W., & Parsons, K. C. (1996). Virtual spaces and real-world places: Transfer of route knowledge. International Journal of Human-Computer Studies, 45, 413-428.

Witmer, B. G., Sadawski, W. J., & Finkelstein, N. M. (2002). VE-based training strategies for acquiring survey knowledge. Presence: Teleoperators and Virtual Environments, 11, 1-19.

Zanbaka, C., Lok, B., Babu, S., Xiao, D., Ulikski, A., & Hodges, L. F. (2004). Effects of travel technique on cognition in virtual environments. Proceedings of the IEEE Virtual Reality Conference (VR'04, pp. 149-156). Los Alamitos, CA: IEEE.

Zhai, S., Kandogan, E., Smith, B. A., & Selker, T. (1999). In search of the "Magic Carpet": Design and experimentation of a bimanual 3D navigation interface. Journal of Visual Languages and Computing, 10, 3-17.

Physical actions: Locomotion	No. of visits to each part of VE Characteristics of movement (e.g., movement frequency, and mean distance traveled in each step) Path followed through VE (macro and micro heuristics)
Physical actions: Looking around	View-referenced: Looking at different parts of display (e.g., VE scene vs. map vs. guidebook) View-referenced: Attending to different aspects of scene (eye tracking) Body/world-referenced: View direction changes (annotation of direction changes onto plan view of path traveled; rate, profile and cumulative magnitude of changes)
Physical actions: Observation	Transcription notation that combines VE and real-world actions Exemplars of observed behavior
Time classification	Locomotion: Stationary vs. traveling View direction: Static vs. looking around
Error classification	Node-based ("miss" or decision point that relates to VE's graph structure) Region-based (Delaunay triangulation or Voronoi region; open virtual spaces)

Table 1. Metrics of physical behavior used to evaluate VE wayfinding.

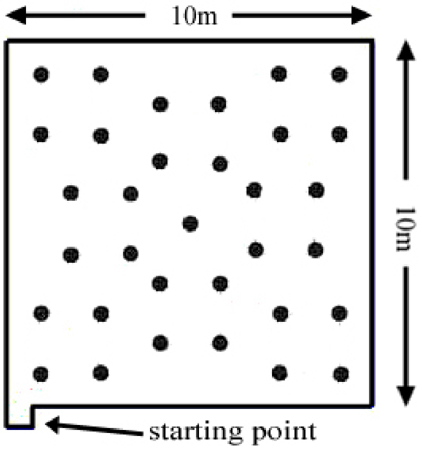
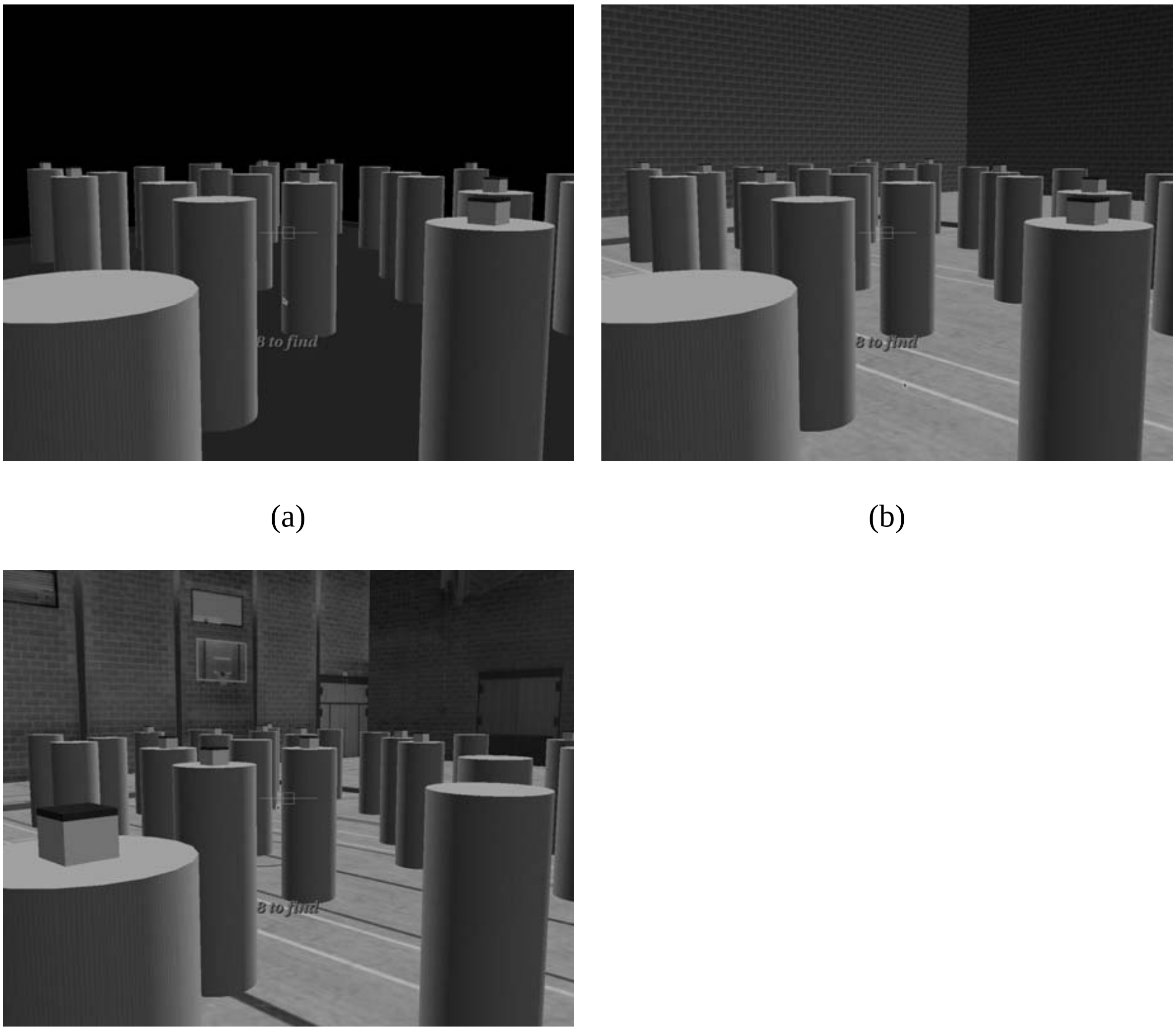


Figure 1. Plan view of the virtual room. Targets and decoys are placed on 16 of the 33 cylinders.



(a) (b)

(c)

Figure 2. Views inside three of the visual scenes using in the studies: (a) lowest fidelity (similar to Ruddle & Jones, 2001), (b) intermediate fidelity (repeated brick texture superimposed on the walls of the virtual sports hall; another version did not have the lines on the floor), and (c) highest fidelity (digital photographs of the sports hall used to texture map the four different walls; other versions had fewer (see (b)) or no lines on the floor).

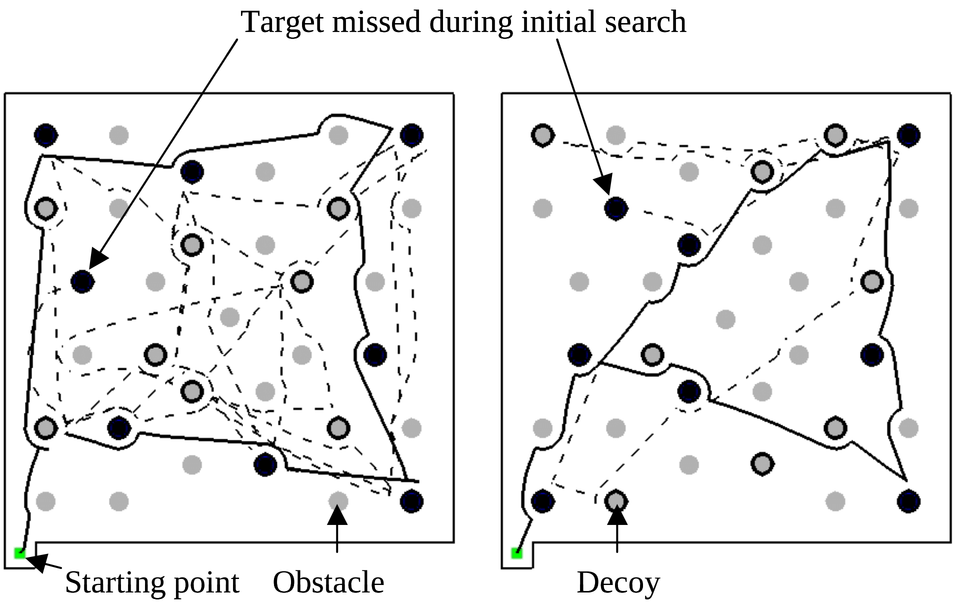


Figure 3. Path followed in trial where participant traveled past one target without checking it before re-searching a large amount of the environment (left), and left the general region containing a target untraveled during the initial part of their search (right). In both cases the solid line shows the path traveled up until the first target or decoy was revisited, and the dashed line shows the path that was subsequently traveled.

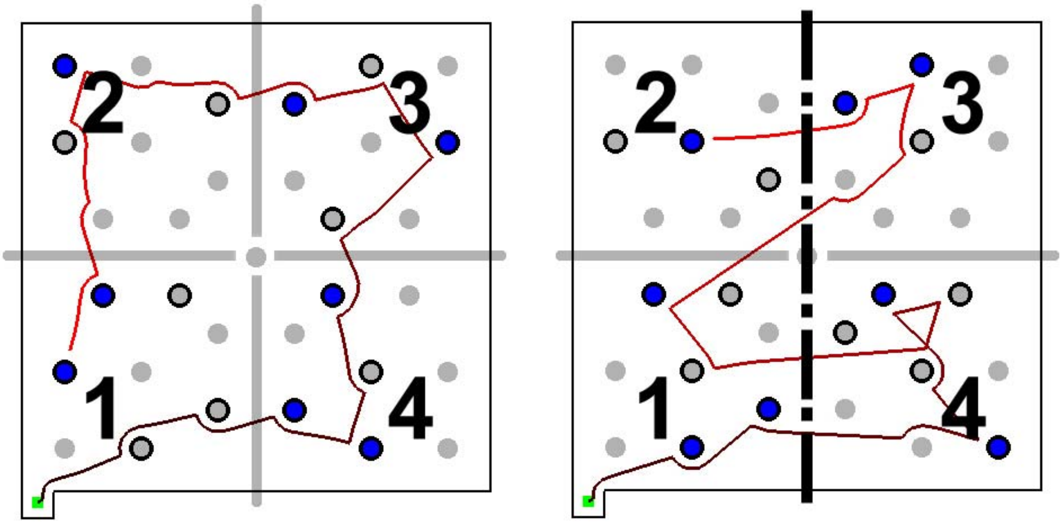


Figure 4. Examples of participants’ search paths in the real-world study, showing a perimeter search (left), and a lawnmower search (right). Both examples show the quadrants that were used to classify the heuristics.

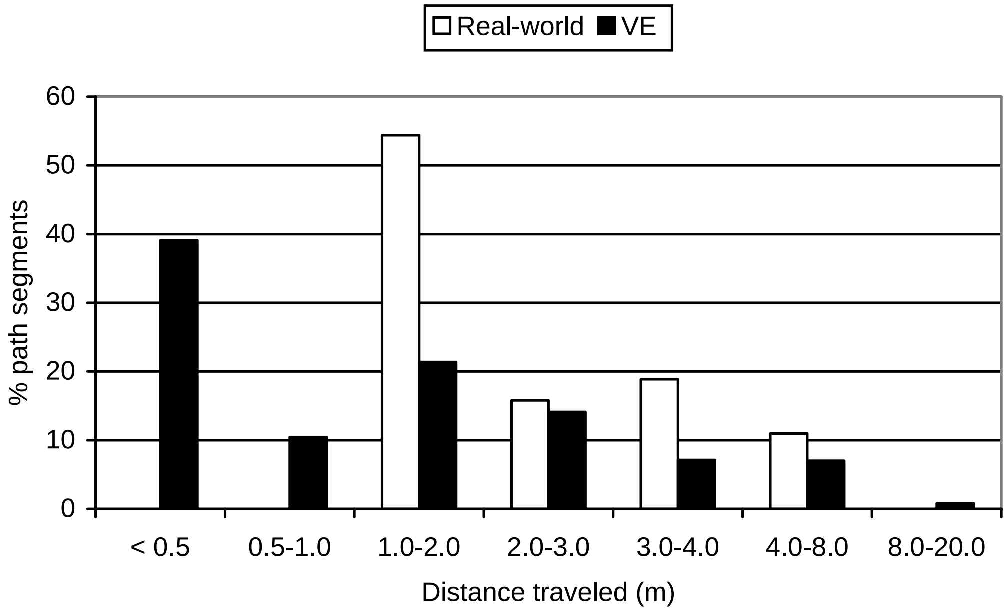
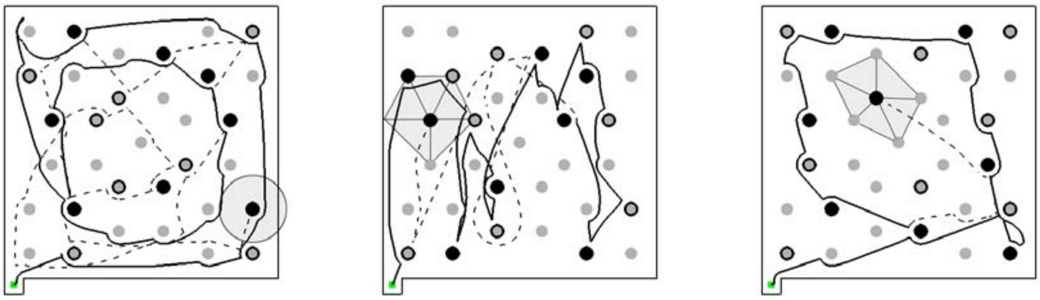


Figure 5. Profile of the length of participants' path segments in the real world and VE studies. One segment stopped and the next started each time a participant paused during their travels.



(a) Miss (b) Local neglect (c) Global neglect

Figure 6. Examples of the errors made by participants. Solid line shows path up to the point that the first target/decoy was revisited, and the dashed line shows the participant's path for the remainder of the trial. In (a) the participant's path was deflected by the cylinder with the missed target (see shaded circle). In (b) and (c) the Delaunay triangulation (shaded) defines the region through which the participant had to pass if the error was to be defined as local neglect.

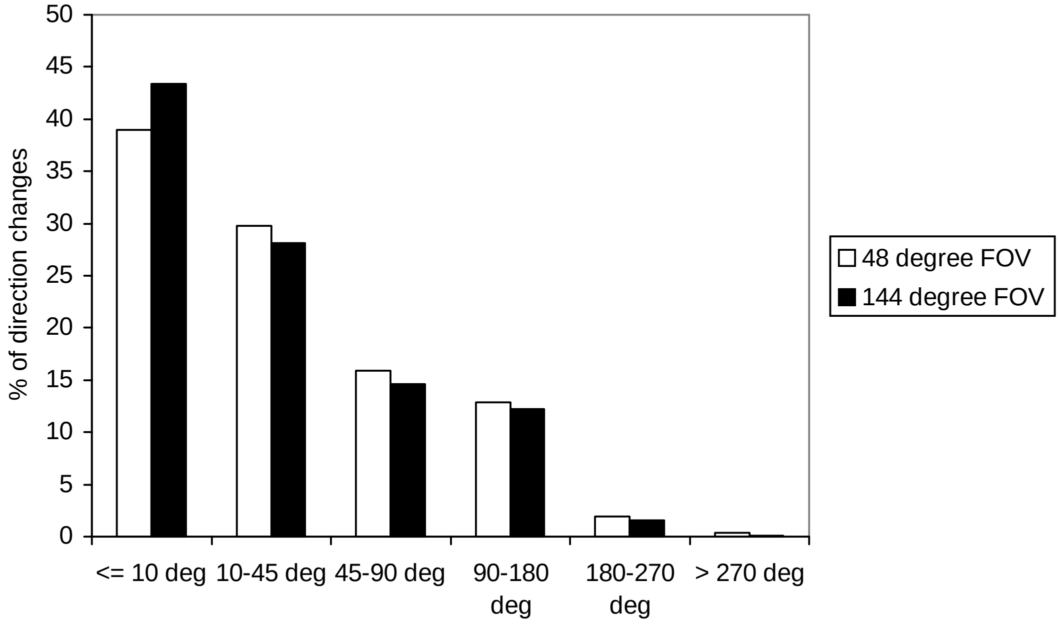
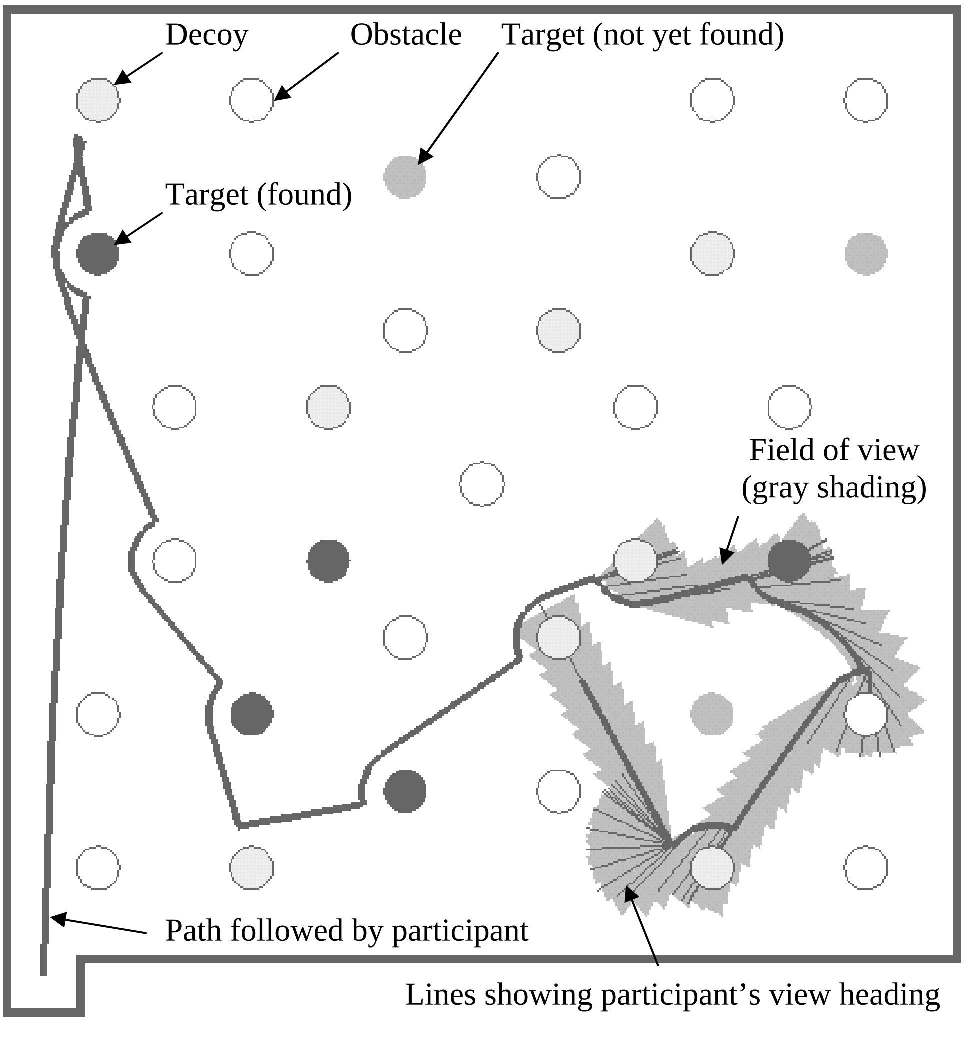


Figure 7. Profile of view heading changes that participants made while standing in one place (i.e., not traveling through the VE).

Figure 8. Plan view showing a participant's path when they traveled completely around a target without it ever appearing within their FOV. The FOV was 48 degrees (horizontal).



Appendix A: Metrics used in studies of VE wayfinding

This appendix summarizes the metrics used in each of the wayfinding studies that was reviewed for the present paper. Studies that used metrics such as direction and distance estimates to investigate aspects of navigation other than wayfinding are not included. The primary insights provided by physical behavior and cognitive rationale metrics in each study are summarized at the end.

Study	Performance			Physical behavior					Cognitive rationale		
	Time taken	Distance traveled/no. of rooms entered	No. of errors or correct turns	Physical actions: Locomotion	Physical actions: Looking around	Physical actions: Observation	Time classification	Error classification	Think aloud	Interview	Questionnaire
Bliss et al. (1997)	*		*								
Bowman et al. (2001)	*						*				
Chen et al. (2004)	*										
Chittaro & Ieronutti (2004)				*	*						
Czerwinski et al. (2002)	*	*									
Darken & Banker (1998)			*		*						
Darken & Sibert (1996)	*			*							
Elvins et al. (2001)	*	*		*	*		*			*	*

Study	Performance			Physical behavior					Cognitive rationale		
	Time taken	Distance traveled/no. of rooms entered	No. of errors or correct turns	Physical actions: Locomotion	Physical actions: Looking around	Physical actions: Observation	Time classification	Error classification	Think aloud	Interview	Questionnaire
Farrell et al. (2003)	*		*								
Gamberini et al. (2003)	*			*		*	*		*		
Grammenos et al. (2002)									*		
Grant & Magee (1998)	*	*									
Howes et al. (2001)	*	*									
Lessels & Ruddle (2004)	*			*	*		*	*			
Lessels & Ruddle (2005)	*	*		*				*			
Murray et al. (2000)						*			*		
O’Neill (1992)	*		*								
Parush & Berman (2004)	*	*									
Pierce & Pausch (2004)	*										
Regian et al. (1992)		*									
Ruddle (2005)		*		*				*			
Ruddle & Jones (2001)	*	*		*			*				
Ruddle, & Péruch (2004)		*		*				*			

Study	Performance			Physical behavior					Cognitive rationale		
	Time taken	Distance traveled/no. of rooms entered	No. of errors or correct turns	Physical actions: Locomotion	Physical actions: Looking around	Physical actions: Observation	Time classification	Error classification	Think aloud	Interview	Questionnaire
Ruddle et al. (1997)		*									
Ruddle et al. (1998)		*						*			
Ruddle et al. (1999a)		*		*							
Ruddle et al. (1999b)	*	*			*						
Ruddle et al. (2000)	*	*		*							
Sas et al. (2004)	*			*	*						
Steck & Mallot (2000)			*								
Strommen (1994)				*	*						
Swan II et al. (2003)	*										
Tan et al. (2001)	*	*									
Tate et al. (1997)	*		*								
Tlauka & Wilson (1994)			*								
Waller et al. (1998)	*			*							
Whitton et al. (2005)				*							*
Witmer et al. (1996)	*		*								

Study	Performance			Physical behavior					Cognitive rationale		
	Time taken	Distance traveled/no. of rooms entered	No. of errors or correct turns	Physical actions: Locomotion	Physical actions: Looking around	Physical actions: Observation	Time classification	Error classification	Think aloud	Interview	Questionnaire
Witmer et al. (2002)	*	*					*				
Zhai et al. (1999)	*						*			*	*

Notable insights provided by the physical behavior metrics

Physical actions: Locomotion	Macro heuristics similar in real world and VE wayfinding, but performance much worse in latter (Lessels & Ruddle, 2005) Two micro heuristics explained 92% of users' movements when navigation was aided by a trail (Ruddle, 2005). Provided critical data for choosing interface in children's game (Strommen, 1994).
Physical actions: Looking around	Users look around twice as much with HMD than desktop display (Ruddle et al., 1999b).
Combination of locomotion & looking around	Identify (un)popular parts and exhibits of virtual museum, and users' visiting styles (Chittaro & Ieronutti, 2004). Movement patterns of inefficient vs. efficient navigators (Sas et al., 2004).
Physical actions: Observation	Users needed time to understand what was happening and decide course of action when fire occurred in VE. Actions used to derive specific types of behavioral response (Gamberini et al., 2003).
Time classification	Usage of human's-eye vs. arial views during exploration and targeted search (Witmer et al., 2002).
Error classification	Targets shown toward sides of wide FOV were often neglected (Lessels & Ruddle, 2004)

Notable insights provided by the cognitive rationale metrics

Think aloud	Users walked along virtual pavement because that's what they do in a real city. Traveled down wide streets instead of narrow alleyways because interface was difficult to control (Murray et al., 2000).
Interview & Questionnaire	Participants' own assessment of how easy different interfaces were to learn and use (Elvins et al., 2001; Whitton et al., 2005; Zhai et al., 1999)