Evaluating Spatial Memory in Two and Three Dimensions

A. COCKBURN[†] AND B. MCKENZIE

Human-Computer Interaction Lab, Department of Computer Science, Private Bag 4800. University of Canterbury, Christchurch, New Zealand

Tel: +64 3 364 2987 x7768

Fax: +64 3 364 2569

[†] *Author for correspondence.*

Abstract

Prior research has shown that the efficient use of graphical user interfaces is strongly dependent on human capabilities for spatial cognition. One facet of spatial cognition is the ability to quickly and accurately recall and access the location of objects in a spatial arrangement. This paper describes a series of experiments aimed at determining whether three-dimensional user interfaces better support spatial memory than their more traditional two-dimensional counterparts. The experiments are conducted using both computer-supported systems and physical models that vary the depth and perspective cues in spatial arrangements of interface items. The physical models were used to escape some of the dimensional ambiguities that are hard to control using computer displays. Results strongly suggest that adding a third dimension to computer displays does not aid users' spatial memory. Although there were no significant differences between the effectiveness of spatial memory when using two- and three-dimensional computer interfaces, participants' memory for the location of cards representing web-pages was reliably better when using a two-dimensional physical model than when using an equivalent three-dimensional physical model.

KEYWORDS: Spatial memory, 3D user interfaces, information visualisation, evaluation.

Evaluating Spatial Memory in Two and Three Dimensions

A. COCKBURN[†] AND B. MCKENZIE

Human-Computer Interaction Lab, Department of Computer Science, Private Bag 4800. University of Canterbury, Christchurch, New Zealand

Tel: +64 3 364 2987 x7768

Fax: +64 3 364 2569

[†] *Author for correspondence.*

Abstract

Prior research has shown that the efficient use of graphical user interfaces is strongly dependent on human capabilities for spatial cognition. One facet of spatial cognition is the ability to quickly and accurately recall and access the location of objects in a spatial arrangement. This paper describes a series of experiments aimed at determining whether three-dimensional user interfaces better support spatial memory than their more traditional two-dimensional counterparts. The experiments are conducted using both computer-supported systems and physical models that vary the depth and perspective cues in spatial arrangements of interface items. The physical models were used to escape some of the dimensional ambiguities that are hard to control using computer displays. Results strongly suggest that adding a third dimension to computer displays does not aid users' spatial memory. Although there were no significant differences between the effectiveness of spatial memory when using two- and three-dimensional computer interfaces, participants' memory for the location of cards representing web-pages was reliably better when using a two-dimensional physical model than when using an equivalent three-dimensional physical model.

KEYWORDS: Spatial memory, 3D user interfaces, information visualisation, evaluation.

1. Introduction

The efficient use of graphical user interfaces relies heavily on human capabilities for spatial cognition. Several research projects, summarised in the following section, have shown that measures of spatial cognition are strongly correlated with performance in a variety of user interface tasks. This correlation raises the question "what can be done to better exploit human spatial capabilities in user interfaces?"

The powerful 3D graphics hardware available in desktop computers provides an attractive opportunity for enhancing interaction. It may be possible to leverage human spatial capabilities by providing computer generated 3D scenes that better reflect the way we perceive our natural environment. Systems such the 'Data Mountain' (Robertson et al. 1998), the 'Task Gallery' (Robertson et al. 2000), and Win3D[1] all work towards this goal by providing 3D alternatives to the 'flat' desktop metaphor.

Spatial memory is one component of spatial cognition that is readily supported by user interfaces. There is some evidence supporting the argument that 3D improves spatial memory. Robertson et al. (1998) showed that task times and error rates were lower when retrieving web pages using their 3D Data Mountain than when using the standard 2D 'Favorites' mechanism of Internet Explorer. Their studies with the Data Mountain provide a variety of insights into the power and robustness of spatial memory, and they conclude that "3D visualization techniques such as those described in this paper can lead to improved user memory" (Czerwinski et al. 1999). Tavanti and Lind (2001) compared the effectiveness of spatial memory in computer generated 2D and 3D displays. They found that spatial memory was much better in the 3D condition, and like Czerwinski et al., they conclude that "a realistic 3D display better supports a specific spatial memory task, namely learning the place of an object". These results are important for user interface design. Spatial cognition is an important predictor of user interface efficiency, so if 3D interfaces improve spatial memory (an important facet of spatial cognition), then 3D interfaces are likely to improve user performance.

The experiments reported in this paper replicate Tavanti and Lind's study and use physical models to imitate and evaluate the functionality of systems based on the Data Mountain. Although rarely done in HCI, replication of important research results is a pivotal component of scientific endeavour. Our experiments revisit the earlier studies, and aim to directly compare interfaces that vary only in their use of depth cues.

The following section details related work on using depth cues to provide a sense of 3D, and reviews prior work on the importance of spatial cognition in effective use of user interfaces. It also describes prior research comparing 2D and 3D interfaces. Two experiments are then described, the first being a variant of Tavanti and Lind's study, and the second using physical emulations of Robertson et al.'s Data Mountain.

---

[1] www.clockwise3d.com

2. Background

In order to clarify our use of the terms "2D" and "3D", this section begins by describing the range of cues that can be used to give a sense of depth and perspective in user interfaces. We then review two areas of related work: that showing a relationship between users' spatial capabilities and their performance with user interfaces; and that comparing the effectiveness of 2D and 3D user interfaces.

2.1 DISTINGUISHING '2D' FROM '3D'

Although desktop environments such as Microsoft Windows or MacOs X are often thought of as 2D environments they use a variety of 3D depth cues. Microsoft Windows, for example, uses shadow casting to perceptually raise the cursor above the background; widgets such as buttons and text entry regions use shape-from-shading (see below) to make items appear raised or recessed; and occlusion is heavily used with overlapping windows, icons, and so on. Despite these depth cues, few would argue that standard desktop environments provide a rich three dimensional experience. Modern graphical user interfaces and '3D' environments exist on a spectrum between two and three dimensions—they are all are essentially 2.xD, where x varies between giving a largely 'flat' impression to providing a richly immersive 3D experience.

Ware (2000) and Goldstein (1989) provide excellent summaries of the depth cues that combine to help the user comprehend a 3D scene. In monocular static displays (the focus of our studies), the more important depth cues are as follows:

- Linear perspective — parallel objects appear to converge on a 'vanishing point' as they recede into the distance.

- Texture gradient — related to linear perspective, the elements of a uniformly textured surface appear closer and smaller with distance.

- Size constancy — familiar objects help to disambiguate the depth and size of other unfamiliar objects in a scene.

- Distance from the horizon — when objects are resting on a surface that recedes from the viewer, the vertical separation between objects and the horizon decreases with distance.

- Occlusion/interposition — near objects occlude more distant ones.

- Depth from focus — scenes can reveal that a set of items are at a similar depth by revealing them in sharp focus while all others are blurred.

- Aerial perspective — all light scatters over distance, and blue light scatters more than other colours. Consequently, distant items are slightly blurred and bluer than near ones.

- Cast shadows and shape-from-shading — shadows are very useful for aiding interpretation of object shape and altitude. They also communicate depth because when one object casts a shadow on another it must be closer to the light source.

- Ground intercept — airborne and subterranean objects are subject to a severe depth/altitude ambiguity in monocular static displays. One artificial display mechanism to overcome this problem is to reveal ground-intercept information by connecting a flying object to the ground with a line.

The relative importance of these factors is an open research question. Ware (2000) provides a comprehensive review of studies and competing theories of the additive effects of depth cues.

2.2 Spatial Cognition and User Interface Performance

Performance with user interfaces is strongly predicted by spatial aptitude. This result has been confirmed in many separate experiments and with varied interface types. Egan and Gomez (1985) showed that measures of spatial memory and age provided the best predictors of how well participants learned to use a text editor. Gagnon (1985) reported the surprising result that computer game scores were not correlated with measures of hand-eye coordination, but were correlated with scores on a spatial memory test. Vicente et al. (1987) and Leitheiser and Munro (1995) concur that measures of spatial ability predict performance in hierarchical file browsing tasks and in a variety of file management tasks.

As mentioned earlier, the Data Mountain's spatial arrangement of webpage thumbnail images allowed more rapid and accurate page retrieval than the 'Favorites' mechanism in Microsoft Internet Explorer (Robertson et al. 1998). Impressively, a follow-up evaluation showed that participants were able to rapidly retrieve pages from their spatial arrangements four months after creating them (Czerwinski et al. 1999). The strength of the spatial cue is dramatically demonstrated by the fact that retrieval times did not significantly worsen when the thumbnail images were replaced with blank outlines.

Ehret (2002) provides interesting insights into how users learn the location of items in a user interface. His experiment shows that users learn locations more effectively when targets poorly represent their function. In other words, the higher the 'evaluation cost' (the degree of effort the user must put into finding the function of an item), the better the location is learned. Ehret presents a theory that predicts how well users learn the location of interface items as evaluation cost varies.

Jones and Dumais (1986) provide some cautions on over-reliance on spatial organisation. Their evaluation indicates that semantic labels provide stronger retrieval cues than spatial organisation alone, but indicate that combinations of semantic and spatial organisation enhance performance.

There has been a great deal of prior work comparing the general effectiveness of 2D and 3D user interfaces, particularly in the military and aviation domains. Many of the findings are dependent on the precise tasks under analysis. Wickens et al. (1995), for example, examined navigation on an aircraft landing approach with 2D and 3D displays. Their results were mixed, showing that their 3D interface better supported navigation on the lateral axis, but at a substantial cost to performance on the vertical axis. Tests of their participants' terrain awareness revealed slightly better performance in 2D, contradicting those of St. John et al. (2000), who found better understanding of terrain shape using 3D. 3D visualisations of airborne or subterranean objects are susceptible to a depth/altitude ambiguity (Wickens et al. 1994). Delucia (1995) shows that 'ground-intercept' information—consisting of a line between objects and the ground—can ease this ambiguity, and Barfield and Rosenberg (1995) show that stereoscopic views can also help.
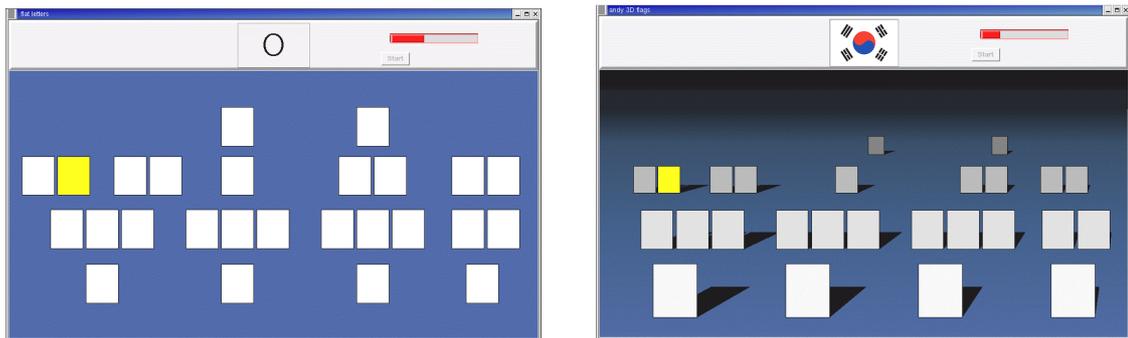
Outside aviation and military research there have been several evaluations comparing 2D and 3D, with varying results. Risden et al. (2000) compared 2D and 3D visualisations of web content, finding no significant differences in user performance or satisfaction with the visualisations. Hicks et al. (2003) compared user performance using 2D and 3D visualisations of telecommunication traffic, finding that the 3D interfaces negatively affected user performance, but that they received higher subjective satisfaction ratings (probably due to their novelty value). Ware and Franck (1996) conducted a comprehensive comparison of 2D and various modes of 3D in tasks involving comprehension of 3D graphs. The 3D conditions reliably outperformed 2D, with the ability to move or rotate the graph proving a valuable feature in 3D. Finally, Smallman et al. (2000) showed that 2D symbolic representations of military targets allowed faster and more accurate identification than 3D icons.

Wickens et al. (1997) provide a fitting summary for prior work on 2D versus 3D evaluations: "whether the benefits of 3D displays outweigh their costs turns out to be a complex issue, depending upon the particular 3D rendering chosen, the nature of the task, and the structure of the information to be displayed."

The specific question addressed in this paper is "do 3D interfaces result in better spatial memory than 2D ones"? The studies reported in this paper are replications of prior studies showing spatial memory advantages for 3D, but with attempts to isolate dimensionality as a factor.

3. Experiment One: Replicating Tavanti and Lind's Study

Tavanti and Lind's study (2001) compared the effectiveness of spatial memory in computer generated 2D and 3D displays. Their tasks involved recalling the location of letters of the alphabet hidden behind 'cards' depicted in hierarchical 2D and 3D displays. Their results

(a) 2D: revealing a letter.   (b) 3D: revealing a flag

FIGURE 1: The 2D and 3D interfaces used in the experiment.

showed much better recall in the 3D condition. There were, however, several potential confounding factors in their experiment. These include the vertical versus horizontal orientation of the windows, the use of letters as search items, the incorrect use of shadowing effects (their shadows receded towards a common vanishing point), and the sizes and separation of the individual icons. Another important difference between the two interfaces is that their 2D version effectively provided a largely linear arrangement with no two icons overlapping on the y-axis.

Our evaluation is essentially a replication of Tavanti and Lind's, with changes to the 2D interface and to the items recalled. The 3D interface (Figure 1b) is a faithful re-implementation of Tavanti and Lind's, including the unusual shadows (discussed in Section 3.2), but our 2D interface (Figure 1a) diverges from theirs by using a horizontal arrangement. Other than perspective effects, both interfaces behaved identically, and both were sized at 1000×510 pixels.

Tavanti and Lind evaluated recall of the twenty-seven letters of the Swedish alphabet, whereas we used the twenty-six letters of the Roman alphabet. We additionally evaluated recall of national flags because preliminary trials revealed that participants used mnemonic aids to construct words, sounds, or word sequences from letters. For example, if the bottom row of letters revealed 'K', 'D', 'O', 'Q', the participant might form the mnemonic 'Klingons Don't Order Quietly'. Mnemonics such as these confound the intended measurement of spatial capabilities. The national flags used in the experiment were those of the twenty-six most heavily populated countries.

Participants completed memorisation and recall tasks with either the 2D or 3D interface (randomly assigned). During memorisation, the users tried to learn the location of letters/flags by pressing the mouse over one of the 'cards', which would highlight and reveal the letter or flag 'hidden behind' it in a display area at the top of the window (see Figure 1). When the mouse button was released, the letter/flag and highlighting would disappear. Tavanti and Lind's systems behaved identically. This separation between selection location and display location is

unusual in user interfaces, but the task remains focused on spatial memorisation because the card's location is the sole cue to the associated letter/flag. Software automatically logged the time, location, and item displayed for all mouse-button presses.

During recall tasks the users tried to select the card associated with the letter/flag shown in the display area. The interface did not reveal whether the selection was correct or not. Software logged the time and location of all mouse-button presses, whether the card was the correct one, and, if not, the pixel distance between the correct one and the one selected (on both the x and y-axes). It also administered questionnaires that recorded a variety of subjective measures and comments.

Participants carried out three separate memorise/recall tasks, with the first being a training task used to familiarise the participants with the experimental procedure. The participants were given 20 seconds to memorise the location of four characters ('!', '@', '#', and '$'), and 30 seconds to find them. Data from the training exercise were discarded. The second and third tasks both involved memorising 26 items (flags or letters). The cards always appeared in the locations shown in Figure 1. Participants had three minutes for memorisation (the complete time period was always used) and five minutes for recollection (usually only one minute was required). A time-bar showed the remaining time in both the memorisation and recall activities. The order in which the participants used letters or flags was randomly assigned, as was the relationship between individual letters/flags and the cards that hid them.

3.1 PARTICIPANT AND EQUIPMENT DETAILS

Forty-four (forty male, four female) Computer Science undergraduate students participated in the experiment. Seventy-eight of the participants were of European origin, and twenty-two percent were of Asian descent. They were randomly assigned (gender-balanced 20-2) to either the 2D or 3D condition. Participation in the experiment lasted approximately thirty-five minutes and was rewarded with a $5 shopping voucher. Forty percent of the 2D participants and 36% of the 3D participants stated that they regularly played 3D computer games. Only one of the 2D and two of the 3D participants stated that they had never played 3D computer games.

The experiment was run on a set of Pentium III computers with 17inch displays running at 1600×1200 resolution.

3.2 RESULTS

As mentioned above, we faithfully replicated Tavanti and Lind's 3D interface, including the inaccurate use of shadows. We did not notice the shadow errors until after conducting the experiment, raising the concern that our results (and those of Tavanti and Lind) are confounded by inaccurate representation of 3D. There is, however, strong evidence that the 3D condition successfully depicted depth. All participants responded to a five-point Likert-scale question (1
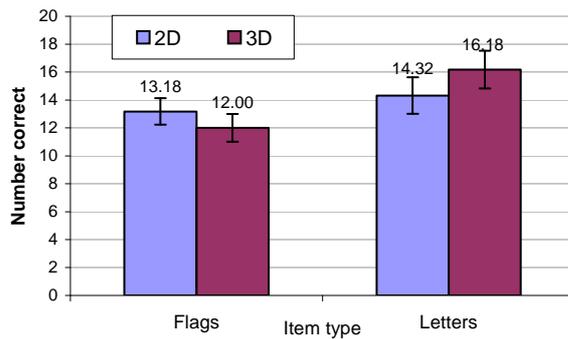
FIGURE 2: Mean number of correctly recalled items per condition. Error bars ±1 standard error.

disagree, 5 agree) "The display provides a sense of depth (some icons seem further away than others)". Responses were low for 2D (mean 1.3, s.d. 0.7), but dramatically higher for 3D (mean 4.2, s.d. 1.1). This is a statistically significant difference Mann-Whitney U=19.5, p<.001.

The primary dependent measure in the experiment was the number of pages correctly recalled. The data is analysed in a 2×2 mixed-factors analysis of variance (ANOVA) for factors interface-type (between-subjects, 2D and 3D) and item-type (within-subjects, letters and flags).

The mean number of items correctly recalled across both interface types and both item types was 13.9 (s.d. 5.6). The range of correctly recalled items extends from a minimum of three by a participant in the 3D flags condition to a maximum of all 26 by three participants (two 3D, one 2D) all using letters. The mean miss distance across all conditions was 128 pixels (s.d. 71).

The means for the 2D and 3D conditions were very similar at 13.8 (s.d. 5.3) and 14.1 (s.d. 5.9), yielding no significant difference: $F_{1,42}$=.06, p=0.8. This result disagrees with Tavanti and Lind. Figure 2 shows the mean recall counts in our experiment for each of the four conditions. The mean absolute miss distance also showed no significant differences between dimensions with 2D and 3D means of 128 (s.d. 63) and 128 (s.d. 78) pixels: $F_{1,42}$=0.001, p=0.9.

There was a significant difference between the number of items correctly recalled when using letters (mean 15.3, s.d. 6.2) and flags (mean 12.6, s.d. 4.5): $F_{1,42}$=8.2, p<0.01. Comments from the participants supported our conjecture that the recall of letters is readily aided by mnemonics:

TABLE 1: Percentage of correct recalls for letters and flags. Note the wide distribution for flags (19 to 92%) compared to letters (49 to 78%).

Percent	Flags	Letters
11-20%	Egypt (19)	
21-30%	Ethiopia (27) Indonesia(30) Iran (30) Mexico (30) Philipines(30)	
31-40%	Bolivia (32) Russia (32) Turkey (35) India (38)	
41-50%	France (41) Italy (41) Pakistan(41) China (46)	**M N T** 49 49 49
51-60%	Congo (51) Thailand(51) B'ladesh (54) Ukraine (54) Vietnam (54) Nigeria (57)	**H R V O F J** 51 51 51 54 57 57 **S D K Y Z** 57 59 59 59 59
61-70%	Germany(68) Brazil (70)	**C E G P Q U** 62 62 62 62 62 62 **W B I L X** 62 65 65 65 68
71-80%	Japan (73) USA (76)	**A** 78
81-90%	UK (81)	
91-100%	S.Korea (92)	

> "Letters were easy. For example, the second row contained NUL HBI TOY PM. They're all words, more or less, except for the HBI bit. Flags were a nightmare. I had to remember where each flag was. It was easier when I recognised the flag, like the UK was right there (points) and America there, but I didn't know most of the flags."

There was no significant interaction between factors interface-type and item-type ($F_{1,42}$=2.7, p=0.11), indicating no reliable difference between the ways that letters and flags were memorised in 2D and 3D.

Table 1 shows a dramatic difference between the ways that letters and flags were memorised. There is a narrow 29% range between the least and most frequently recalled letter: from letters M, N, and T which were correctly found in 49% of tasks, to letter A which was correctly found in 78% of tasks. All other letters were in a narrow range between 51% and 68% of tasks. Successful recall of flag locations was much more varied, with a 73% range: the Egyptian flag was recalled in only 19% of tasks, but the South Korean flag was recalled in 92% of tasks. Twenty two percent of the participants were of Asian descent. The participants mentioned two factors that affected whether they recalled flags: first, whether they knew the country associated

with the flag; and second how visually distinct it was from others. These observations are consistent with psychological research showing that recall improves when items are meaningful, concrete and form an image in the mind (Paivio et al. 1968; Paivio et al. 1968), and when images are simple rather than complex (Attneave 1955).

4. Experiment Two: The Physical Data Mountain

Robertson et al.'s Data Mountain (1998) allowed users to arrange 'thumbnail' images of web pages on an inclined plane that recedes into the distance. Their evaluation showed that the 3D Data Mountain improved retrieval times and reduced error rates in comparison to Internet Explorer's Favorites. Despite the Data Mountain's impressive performance, their experiment did not reveal whether the performance gains were due to 3D or to some other factor such as the need to scroll when using Favorites. As Robertson et al. state: "We would like to understand the relative contributions to this successful study of the various components (3D versus 2D…)" To this end in our prior work we compared 2D and 3D re-implementations of the Data Mountain, and found no significant difference between 2D and 3D (Cockburn and McKenzie 2001).

There remains a risk, however, that our results, and those of any computer generated depiction of 3D, are confounded by visual subtleties that are hard or impossible to replicate using computers (human focal depth, for instance, will remain at a short distance when using a computer screen). For this reason, we conducted a study to investigate spatial memory in tasks similar to those of the Data Mountain evaluation, but using physical models rather than computer systems (Cockburn and McKenzie 2002). Although the experiment originally used both physical and virtual systems, only the physical systems are reported here (results of the comparison between virtual systems showed no significant difference, supporting the result of Experiment One).

4.1 EXPERIMENTAL METHOD

The experimental design was a 2×3 mixed factors analysis of variance (ANOVA). The between-subjects factor 'dimension' had levels 2D and 3D, with participants randomly assigned to the conditions. The within-subjects factor 'density' had levels sparse, medium and dense, which were measured when the interface contained 33, 66 and 99 pages.

(a) 2D   (b) 3D

FIGURE 3: The two physical interfaces used in Experiment Two.

Apparatus. The web page 'thumbnails' in the interfaces were 90×90mm photo-quality printed images of web pages rendered in Netscape Navigator. The images were mounted on stiff cardboard and covered in clear plastic for protection. The title information for each page was overlaid on top of the Netscape window banner information, allowing a clearly identifiable text title at a distance of approximately two metres. Clips on the back of each card allowed them to be inserted into the physical interfaces. The experimenters positioned the cards within the interfaces under the participants' instructions.

The 2D interface (Figure 3a) was constructed from chipboard and horizontal lines of taut fishing-line separated by 2cm in a single vertical plane. String marked the 900×710mm page placement boundaries. In all interfaces the participants were allowed to request that cards overlap on another, provided the page titles remained visible.

The 3D interface (Figure 3b) was constructed from painted steel rods and horizontal lines of taut fishing-line placed at 5cm intervals vertically and horizontally. On every second horizontal plane of lines, each second line was removed to create a 'tunnel' of lines to allow the experimenter to place cards within the structure. The 3D structure allowed pages to be placed in a 900×900×750mm x, y and z space. The larger size of the 3D interface on the y-axis (900mm versus 710mm for the 2D interface) was intended to partially compensate for the coarse granularity of placement alternatives on the vertical axis (5 or 10cm versus 2cm in the 2D interface).

In both interfaces the fishing-line caused minimal occlusion of pages.

The participants sat on a height-adjustable chair set approximately 50cm from the front-edge of the interface. This gave an angle at the eye of approximately 84° between the left and right front edge of each interface and approximately 40° at the back of the 3D interface. Head positions

were normally approximately mid-height in the 2D interface and one-third height in the 3D interface.

A laser pointer was used to identify target pages during retrieval tasks.

Procedure. The evaluation consisted of repeating storage and retrieval exercises for the sparse, medium and dense conditions. Starting with an empty interface, the participants added 33 pages, one at a time, with freedom to relocate any pages already in the display. The pages were presented in random order in a separate computer-based cueing interface. The cueing condition for each page consisted of a magnified thumbnail of the page, the page URL and its title. Participants were asked to read the title aloud and to ask for clarification if they did not understand the page topic. The experimenter quickly accessed the correct page-card and showed it to the participant who told the experimenter where to place the card, through a combination of pointing, gestures and comments such as "left a bit", "back a bit", "overlapping Greenpeace, there", etc.

The same set of 99 web pages was used for all participants. Pages included media providers such as cnn.com, major international commercial organisations such as coke.com, local retail companies, and a variety of university sites.

The retrieval tasks consisted of finding, as quickly as possible, ten randomly selected pages from the display, one at a time. The cueing condition for each page was identical to that used for storage: the user was shown the magnified image, its URL and title. For each page retrieval the participants were given a preparatory "three, two, one" countdown prior to displaying the page to be found. A clock on the computer running the cueing interface started as soon as the page to be found was displayed. The experimenter stopped the clock by pressing the space bar as soon as the subject illuminated the target page using a laser pointer. To help the experimenter stop the clock at the correct moment, subjects were encouraged to clearly state when they had located the page with utterances such as "There!" or "Got it".

On completing the retrieval tasks in the sparse condition (33 pages), the participants proceeded to the medium condition, adding a further 33 pages to the display. They then retrieved ten randomly selected pages from the 66 pages shown. Finally, participants repeated the tasks for the last 33 pages in the dense condition.

The twenty participants (all Computer Science students) were randomly assigned to one of the two interfaces. Each evaluation session lasted approximately one hour. Training typically lasted ten minutes and consisted of organising and retrieving a set of eight training pages (not included in the evaluation set) using the cueing interface. The participants were advised that the best way to organise pages was to cluster them into groups of related pages. They were given no hints
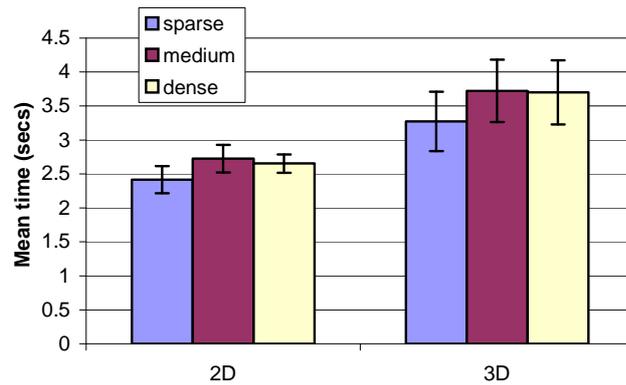
FIGURE 4: Mean page retrieval times using the 2D and 3D physical interfaces across three levels of density in Experiment Two. Error bars show one standard error above and below the mean.

about what groupings would appear. They were also told that they could ask to modify page locations at any time.

4.2 RESULTS

Across the 600 trials in the two interfaces, the mean time to retrieve pages was low at 3.39 seconds (s.d. 1.4). The 100 second time-limit was exceeded in only one trial, in the 3D-dense condition. Many participants commented that they were much faster at retrieving pages than they expected, indicating that their spatial memory was effective, but not trusted. Several participants found the visual effect of the 3D condition delightful, expressing statements like "Wow… The pages are just floating in front of my eyes".

The means for the three densities were not significantly different ($F_{2,36}$=1.6, p=0.2) at 2.8 (s.d. 1.1), 3.2 (s.d. 1.2) and 3.2 (s.d. 1.2) seconds for the sparse, medium and dense conditions. Although surprising, our observations of the participants provided a clear explanation for the similarity of results across density. Early in the evaluation (during the sparse condition), the participants were still in the process of forming their 'chunks' of related web pages. In many cases, the categories remained unclear until more pages arrived to clarify the boundaries between chunks—for example, Air New Zealand and IBM might initially have formed part of a 'large corporations' chunk, but after adding pages for other airlines and computer companies, the need for separate and more precise chunks became clear.

There was a reliable difference between mean retrieval times with the 2D (2.6, s.d. 0.6 seconds) and 3D (3.6, s.d. 1.4 seconds): $F_{1,18}$=5.5, p<.05. This is interesting because the 3D interface allowed the participants to organise images on a 2D plane that was slightly larger than that available in the 2D interface—in effect, it allowed many 2D planes, each slightly larger that the 2D interface. From our observations, it appeared that the additional flexibility in placement location provided by the third dimension caused an undesirable lack of constraint. Many times we observed 3D participants uttering statements to the effect of "where did I put that group" and subsequently visually searching across multiple dimensions: up/down, left/right, and front/back.

14

Equivalent statements were much less common in 2D, although participants frequently ran visual scans over the two dimensions available. We conjecture that the visual search was additionally difficult in 3D because users needed to refocus between near and distant objects.

Prior to the evaluation we had suspected that occlusion might inhibit efficient use of the 3D physical interface, but the participants were careful to arrange items in such a way that occlusion was rarely an issue.

5. Discussion

The results of Experiment One showed no difference between retrieval using 2D and 3D, and that the participants believed the 3D interface provided a strong visual sense of depth. This leads to the question, why are our results different to Tavanti and Lind's? We suspect that the vertical orientation of Tavanti and Lind's 2D display made the formation of effective letter mnemonics more difficult than the horizontal 3D layout—words and word combinations normally run horizontally left to right (in English and Swedish). Certainly, our participants reported that they made extensive use of mnemonics with letters, which can partially explain the dramatically better recall rates for letters than flags.

There remains a risk, however, that the incorrect use of shadows by Tavanti and Lind (replicated in our study) may have confounded the results, despite the participants reporting that the 3D interface provided a strong sense of depth. In an attempt to bypass the unavoidable differences between computer-generated 3D and 'real' 3D, we compared spatial memory using physical models in Experiment Two, with tasks similar to those of the Data Mountain. Results in this experiment showed worse retrieval times in 3D than 2D, with the presence of depth appearing to complicate the formation of chunks and the visual search of the data-space.

Naturally, there remain experimental concerns about the physical interfaces. Firstly, although the fishing-line was successful at minimising occlusion, it affected the way the participants used the systems by affording linear page arrangements along the line. It also created discrete placement locations on the y and z axes. In the 2D environment, cards could not be placed less than 2cm vertically apart, and in the 3D environment they could not be closer than 5cm on the y and z axes. We doubt that this had a major impact on the results because when vertical arrangements were used in the 2D environment the participants almost all requested a vertical separation of two lines (4cm). Finally, there is a risk that the participant's 3D arrangements were affected by the default location of the experimenter, who normally stood on the participants' right. The participants may have been biased toward right-hand placements in order to avoid making the experimenter walk around the 3D structure repeatedly. Again, we doubt that this will have affected the results.

6. Conclusions

Prior research has shown that spatial organisations of information can allow users to access data items surprisingly quickly. Furthermore, there have been research claims that monocular-static 3D visualisations can enhance the effectiveness of spatial memory.

This paper described two experiments, one using computer-generated interfaces and one using physical models that emulate prior computer-supported systems, to investigate the contribution of the third dimension in supporting effective spatial memory. The experiments are heavily based on prior studies.

Results showed no significant difference between the effectiveness of spatial memory when using 2D and 3D computer-supported systems, but significantly better performance with 2D than 3D when using the physical systems. It seemed that the 3D physical systems under-constrained participants in their freedom to organise data items, and that the differing focal depths of data items made visual search more difficult.

Spatial memory clearly provides an effective aid to information retrieval, but we are sceptical of the role that 3D plays in aiding retrieval from static-perspective spatial organisations.

References

Attneave, F. (1955). "Symmetry, Information, and Memory for Patterns." American Journal of Psychology **68**: 209-222.

Barfield, W. and C. Rosenberg (1995). "Judgements of Azimuth and Elevation as a Function of Monoscopic and Binocular Depth Cues Using a Perspective Display." Human Factors **37**(1): 173--181.

Cockburn, A. and B. McKenzie (2001). 3D or Not 3D? Evaluating the Effect of the Third Dimension in a Document Management System. Proceedings of CHI'2001 Conference on Human Factors in Computing Systems, Seattle, Washington, March 31--April 6. 434--441.

Cockburn, A. and B. McKenzie (2002). Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. Proceedings of CHI'2002 Conference on Human Factors in Computing Systems, Minneapolis, Minnesota, 20--25 April. 203--210.

Czerwinski, M., M. vanDantzich, G. Robertson and H. Hoffman (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3D. Proceedings of Interact '99, Edinburgh, Scotland, IOS press. 163--170.

Delucia, P. (1995). "Effects of Pictorial Relative Size and Ground-Intercept Information on Judgements about Potential Collisions in Perspective Displays." Human Factors **37**(3): 528--538.

Egan, D. and M. Gomez (1985). Assaying, Isolating, and Accomodating Individual Differences in Learning a Complex Skill. Individual Differences in Cognition. R. Dillon**:** 173--217.

Ehret, B. (2002). Learning Where to Look: Location Learning in Graphical User Interfaces. Proceedings of CHI'2002 Conference on Human Factors in Computing Systems. Minneapolis, Minnesota, 20--25 April**:** 211--218.

Gagnon, D. (1985). "Videogames and Spatial Skills: An Exploratory Study." Educational Communication and Technology **33**(4): 263--275.

Goldstein, E. (1989). Sensation and Perception, Belmont, CA: Wadsworth.

Hicks, M., C. O'Malley, S. Nichols and B. Anderson (2003). "Comparison of 2D and 3D Representations for Visualising Telecommunication Usage." Behavior and Information Technology **22**(3): 185--201.

Jones, W. and S. Dumais (1986). "The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location versus Name." ACM Transactions on Office Information Systems **4**(1): 42--63.

Leitheiser, B. and D. Munro (1995). An Experimental Study of the Relationship Between Spatial Ability and the Learning of a Graphical User Interface. Proceedings of The Inaugural Americas Conference on Information Systems, Pittsburgh, PA, August 25--27.

Paivio, A., T. Rogers and P. Smythe (1968). "Why are Pictures Easier to Recall than Words?" Psychonomic Science **11**: 137-138.

Paivio, A., J. Yuille and S. Madigan (1968). "Concreteness, Imagery, and Meaningfulness Values for 925 Nouns." Journal of Experimental Psychology **76**(1, part 2.).

Risden, K., M. Czerwinski, T. Munzner and D. Cook (2000). "An Initial Examination of Ease of Use for 2D and 3D Information Visualizations of Web Content." International Journal of Human Computer Studies **53**: 695--714.

Robertson, G., M. Czerwinski, K. Larson, D. Robbins, D. Thiel and M. v. Dantzich (1998). Data Mountain: Using Spatial Memory for Document Management. Proceedings of the 1998 ACM Conference on User Interface Software and Technology, November 1--4. San Francisco, California.**:** 153--162.

Robertson, G., M. Czerwinski, K. Larson, D. Robbins, D. Thiel and M. vanDantzich (1998). Data Mountain: Using Spatial Memory for Document Management. Proceedings of UIST'98 ACM Conference on User Interface Software and Technology, November 1--4., San Francisco, California. 153--162.

Robertson, G., M. vanDantzich, M. Czerwinski, K. Hinckley, D. Thiel, D. Robbins, K. Risden and V. Gorokhovsky (2000). The Task Gallery: A 3D Window Manager. Proceedings of CHI'2000 Conference on Human Factors in Computing Systems. The Hague, The Netherlands, April 1--6**:** 494--501.

Smallman, H., M. StJohn, H. Oonk and M. Cowen (2000). Track Recognition Using Two-Dimensional Symbols or Three-Dimensional Realistic Icons, SPAWAR Systems Center, San Diego.

StJohn, M., H. Oonk and M. Cowen (2000). Using Two-Dimensional and Perspective Views of Terrain, SPAWAR Systems Center, San Diego.

Tavanti, M. and M. Lind (2001). 2D vs 3D, Implications on Spatial Memory. Proceedings of IEEE InfoVis 2001 Symposium on Information Visualization, San Diego, 22--23 October. 139--145.

Vicente, K., B. Hayes and R. Williges (1987). "Assaying and Isolating Individual Differences in Searching a Hierarchical File System." Human Factors **29**(3): 349--359.

Ware, C. (2000). Information Visualization: Perception for Design, Morgan Kaufmann.

Ware, C. and G. Franck (1996). "Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions." ACM Transactions on Graphics **15**(2): 121--139.

Wickens, C., C. Liang, T. Prevett and O. Olmos (1995). Egocentric and Exocentric Dipslays for Terminal Area Navigation. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting. 16--20.

Wickens, C., S. Miller and M. Tham (1994). The Implications of Data-Link for Representing Pilot Request Information on 2D and 3D Air Traffic Control Displays. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting. 61--65.

Wickens, C., O. Olmos, A. Chudy and C. Davenport (1997). "Aviation Display Support For Situation Awareness." (ARL-97-10/LOGICON-97-2).