Metadata of the chapter that will be visualized online

Chapter Title	Human-Centric Chronographics: Making Historical Time Memorable	
Copyright Year	2013	
Copyright Holder	Springer Science+Business Media New York	
Corresponding Author	Family Name	**Korallo**
	Particle	
	Given Name	**Liliya**
	Suffix	
	Division	Psychology Department
	Organization	Middlesex University
	Address	The Burroughs, Hendon, London, NW4 4BT, UK
	Email	lkorallo@yahoo.co.uk
Author	Family Name	**Davis**
	Particle	
	Given Name	**Stephen Boyd**
	Suffix	
	Division	School of Design
	Organization	Royal College of Art
	Address	Kensington Gore, London, SW7 2EU, UK
	Email	stephen.boyd-davis@rca.ac.uk
Author	Family Name	**Foreman**
	Particle	
	Given Name	**Nigel**
	Suffix	
	Division	Psychology Department
	Organization	Middlesex University
	Address	The Burroughs, Hendon, London, NW4 4BT, UK
	Email	n.foreman@mdx.ac.uk
Author	Family Name	**Moar**
	Particle	
	Given Name	**Magnus**
	Suffix	
	Division	School of Art and Design
	Organization	Middlesex University
	Address	The Burroughs, Hendon, London, NW4 4BT, UK

Email m.moar@mdx.ac.uk

Abstract

A series of experiments is described, evaluating user recall of visualisations of historical chronology. Such visualisations are widely created but have not hitherto been evaluated. Users were tested on their ability to learn a sequence of historical events presented in a virtual environment (VE) flythrough visualisation, compared with the learning of equivalent material in other formats that are sequential but lack the 3D spatial aspect. Memorability is a particularly important function of visualisation in education. The measures used during evaluation are enumerated and discussed. The majority of the experiments reported compared three conditions, one using a virtual environment visualisation with a significant spatial element, one using a serial on-screen presentation in PowerPoint, and one using serial presentation on paper. Some aspects were trialled with groups having contrasting prior experience of computers, in the UK and Ukraine. Evidence suggests that a more complex environment including animations and sounds or music, intended to engage users and reinforce memorability, were in fact distracting. Findings are reported in relation to the age of the participants, suggesting that children at 11–14 years benefit less from, or are even disadvantaged by, VE visualisations when compared with 7–9 year olds or undergraduates. Finally, results suggest that VE visualisations offering a 'landscape' of information are more memorable than those based on a linear model.

Human-Centric Chronographics: Making Historical Time Memorable

Liliya Korallo, Stephen Boyd Davis, Nigel Foreman, and Magnus Moar

Abstract A series of experiments is described, evaluating user recall of visualisations of historical chronology. Such visualisations are widely created but have not hitherto been evaluated. Users were tested on their ability to learn a sequence of historical events presented in a virtual environment (VE) flythrough visualisation, compared with the learning of equivalent material in other formats that are sequential but lack the 3D spatial aspect. Memorability is a particularly important function of visualisation in education. The measures used during evaluation are enumerated and discussed. The majority of the experiments reported compared three conditions, one using a virtual environment visualisation with a significant spatial element, one using a serial on-screen presentation in PowerPoint, and one using serial presentation on paper. Some aspects were trialled with groups having contrasting prior experience of computers, in the UK and Ukraine. Evidence suggests that a more complex environment including animations and sounds or music, intended to engage users and reinforce memorability, were in fact distracting. Findings are reported in relation to the age of the participants, suggesting that children at 11–14 years benefit less from, or are even disadvantaged by, VE visualisations when compared with 7–9 year olds or undergraduates. Finally, results suggest that VE visualisations offering a 'landscape' of information are more memorable than those based on a linear model.

L. Korallo (✉) • N. Foreman
Psychology Department, Middlesex University, The Burroughs, Hendon, London NW4 4BT, UK
e-mail: lkorallo@yahoo.co.uk; n.foreman@mdx.ac.uk

S.B. Davis
School of Design, Royal College of Art, Kensington Gore, London SW7 2EU, UK
e-mail: stephen.boyd-davis@rca.ac.uk

M. Moar
School of Art and Design, Middlesex University, The Burroughs, Hendon, London NW4 4BT, UK
e-mail: m.moar@mdx.ac.uk

W. Huang (ed.), *Handbook of Human Centric Visualization*,
DOI 10.1007/978-1-4614-7485-2_19,

1 Introduction

Our work is concerned with *chronographics*—the visualisation of chronology, especially that of history. The approach is human-centric in two respects. We have undertaken extensive user-testing, comprising 12 experiments involving a cumulative total of 512 participants, the results of which are summarised and discussed in this chapter. While many chronographic visualisations have been created in recent years, none has been evaluated experimentally until now. Our investigation focuses in particular on questions of memorability. The second human-centric aspect of the work is that the user is literally placed at the centre of our visualisations using virtual environment (VE) technologies, positioned so as to take egocentric views on time past, to undertake personal explorations of 'history-space' looking through time and, in our most recent work, looking 'across' time too, rather as though exploring a landscape. We hoped that the use of such an embedded, spatialised user view would produce particular benefits.

Although our application was the learning of history and especially the recall of chronology, our findings have broad relevance. We report on surprising differences in the effectiveness of VE visualisations for different age groups, on some effects of multimedia and other components which are not strictly functional in expressing chronological information, and in our most recent work, suggestions that exploitation of two dimensional 'landscapes' of information are more effect than those that are effectively linear.

1.1 Chronographic Visualisation: The Timeline

In what follows we use the word *timeline* frequently, denoting a graphic layout where time is mapped to a surface or space. The word first appears in its modern sense in William James' Principles of Psychology of 1890 [1], in relation to recording experimental data against time. More than a century earlier there had been a shift from typographic, tabular layouts of historical events to truly graphical time-maps inspired by the ideas of Descartes and Newton [2]. For centuries prior to that, historical events had only been organised into lists and tables. In the mid-eighteenth century, French and then English pioneers began instead to map events in a linear, graphical way. One example was a printed paper chart 16.5 m (54 ft) long, attempting to encompass all history since the biblical Creation. The idea of situating the user within a dynamic representation of historical time was already claimed as a benefit: the timeline was described by its author Barbeu-Dubourg as 'a moving, living tableau, through which pass in review all the ages of the world [. . .] where the rise and fall of Empires are acted out in visible form' [3], and in fact this particular example was available in a 'machine' where time could be scrolled back and forth by turning handles [4], a surprising anticipation of modern digital approaches to navigating history.

Many of the early aspirations for chronographic visualisations are still with us now. A recurrent theme was memorability, the focus of the present chapter. Le Sage, for example, asked, 'Why is it that an object in geography communicates an idea that is so precise and so specific, and leaves such lasting traces, while a moment in history, by contrast, sinks into nothingness, leaving behind nothing but fleeting impressions? [. . .]: simply that the knowledge of geography is engraved in our mind by *images*, while that of history is only arrived at by *words*.' [5] (original emphasis). These alleged advantages of visualisation were of course based on intuition and assumptions, not experimental evidence. There was no way to judge whether one visualisation was more successful for the user than another.

Currently digital timelines proliferate, especially on the Web. Often the term is used just to mean a time-ordered list, but many truly graphical examples also exist, plotting time horizontally, vertically or in virtual depth. Sometimes events are attached to a single line, as in most of our examples discussed below, or to multiple lines or a time 'surface'. Different degrees of interactivity are made available, above all scrolling and zooming and related forms of navigation. But again any form of user-centric evaluation is noticeable by its absence.

1.2 A Problem and a Possible Solution: Adopting VE Technologies

We originally set out to address a problem in the learning of history, particularly within school education. An important aspect of historical knowledge is the *framework* of events: both sequence in time, and synchronism of contemporaneous events, perhaps in different fields. History only 'makes sense' when events can be fitted into a framework of this kind. Yet historical time and sequences of historical events are difficult concepts for children to acquire and comprehend. In schools, children usually learn about such abstract concepts by relying on semantic information most often provided on printed worksheets. To learn dates of events, for example, children have no option but to memorize them laboriously, which imparts little understanding of meaningful historical relations. Responding to a questionnaire conducted by the present authors, history teachers reported having used timelines to make history 'less kaleidoscopical and more coherent' [6]. The timeline is the most popular classroom tool to assist children in understanding chronology [7–9].

We wanted to know whether locating the user *within* such a visualisation, using Virtual Environment (VE) technologies to construct a three-dimension time-space which the user could navigate, would make a difference in particular to the memorability of the information it contained. No timeline visualisations have previously been subjected to this kind of research. Our findings do not offer an unequivocal answer, but our most recent experiments suggest the most promising routes to follow.

It is important to note that VE technologies have been extensively applied to history, but generally with the aim of recreating historical sites and artefacts. Our work instead visualises historical time itself, positioning visual markers such as paintings, photographs or objects representing events in a three-dimensional space, of which one dimension represents time. One of the most striking uses of VE technologies for a three-dimensional timeline is Kullberg's 1995 M.Sc. project representing the history of photography [10]. The user could navigate among photographs attached to lines representing the lives of individual photographers, travelling in different directions, and had a choice of either obtaining further information about a selected photograph (by clicking on a relevant icon) or moving on in time to further items. It offered an overview of the environment (from an elevated virtual viewpoint) making it potentially easier for the user to establish spatial relationships—to establish an effective cognitive 'map' [11] amongst places/images—that may subsequently improve recall of the information. However, Kullberg's project included no user-evaluation.

1.3 The Rationale for Using Virtual Environments

One might expect that VE presentations of historical data would have all the standard benefits of visualisation when compared with memorising lists of names and dates. In addition, by situating the user in a time-space we hoped to harness spatial memory rather than semantic memory, in particular since spatial memory is not obviously limited in terms of capacity [12]. Although participants could in principle remember a simple verbal nine-item list, it was hoped that spatial memory would be employed preferentially. In previous studies, for example in which participants experienced rows of shops in a VE, they quickly acquired a good spatial memory for the layout of the shopping mall and for positions of individual shops [13, 14]. After a short period exploring a VE, a participant can make spatial judgments that could only be made using a cognitive "map", such as pointing in the direction of currently-not-visible landmarks [15, 16]. Ours is the most recent incarnation of a long tradition of using physical spaces as mnemonic aids, often referred to as the 'Theatre of Memory', for example in Yates's seminal study The Art of Memory [17].

2 Overview of the Series of Experiments

In all, twelve experiments took place. We do not describe each experiment in detail but rather focus on illustrative examples and on the accumulated findings and discussion. The reader is referred to our other publications for more detailed accounts [12, 18–20]. The purpose here is to give sufficient information to indicate the general nature of the investigation, the characteristics of the different participant groups, and to indicate some firm and some more tentative findings relevant to user-centric visualisation. The aggregated findings on gender effects over all the experiments were inconclusive, so this aspect is omitted.

In our studies, except where otherwise noted, nine historical events were presented as images in a chronological sequence in three conditions, each condition experienced by independent experimental groups.

In many of the experiments described below, two screen-based conditions were evaluated. One was a VE visualisation, in which the user navigated a simple 3D space, so that it seems as though the user travelled in both space and historical time. The other used PowerPoint to sequence a series of images and associated text. In the VE condition, we did not take advantage of the immersive effects of head-mounted displays and stereoscopic vision, principally because our target users were mainly school-children for whom such facilities would currently remain inaccessible. Our use of VE technologies was therefore limited to the construction and delivery of time-spaces which were subsequently displayed and navigated on conventional computer displays.

In the simplest format used for most experiments, pictures or virtual objects representing events were positioned along a line in the virtual space, with successive images appearing alternately on the left or right of the axial line representing time. The user navigated along this timeline sagittally, that is orthogonally to the surface of the screen (for a discussion of the use of the three cardinal dimensions for time, see [21]). Clearly in the case of both screen conditions, the image is in reality two dimensional; however the design of the VE condition using perspectival cues and movement creates an impression that the user moves through a time-space rather than simply seeing a sequence of images.

In the case of classroom studies, efforts were made with the help of teachers to ensure that the comparison groups were equally capable in terms of their previous classroom performance in history lessons, as reflected in standard classroom assessments.

Some aspects of the experiments were modified in the light of experience. Early experiments simply exposed participants to the timeline material. As this produced generally poor results, an element of challenge was introduced into the exploration. These changes are described in more detail below. Other differences between experiments occurred through adaptation to local circumstances in the United Kingdom and in the Ukraine.

The size of the groups used in the experiments ranges from 10 to 20 participants per condition. From a practical point of view, conducting experiments using VEs in schools, it is difficult to access larger populations. The size of the groups was equivalent to those in previous studies of spatial learning conducted using VEs [13, 22–24].

Virtools Dev 3.0 educational version software (www.virtools.com) was used to create the virtual fly-throughs as a Virtools Player File. This was run in the Virtools Player in a standard browser on desktop computers with graphics cards sufficient to deliver smooth full-screen animation and, where necessary, synchronised audio.

Participating schools and teachers were told that the purpose of the research was to attempt to discover means to assist history teaching and learning, so that some benefits might accrue to the school (and other schools) in the medium term. After completion of the studies, children were presented with a simplified version of the results, and teachers were also debriefed. Staff were told their assistance would be acknowledged in any publications. No other incentives were offered. Consent forms were signed and returned by parents in conformity with ethical requirements. With regard to the studies conducted in schools in Ukraine, two separate ethical approvals were obtained: from Middlesex University and from local education authorities in Ukraine.

2.1 Scoring Methods for Experiments

In all 12 experiments, a score was allocated to the degree of error per item in each participant's performance when attempting to place items in the correct sequence, and to the number of correct answers in allocating the items to sequenced slots. A number of other measures were used depending on the focus of each experiment. These are summarised in Table 1.

1. REM score (i.e., "REMOVED" score—how far a picture was placed from its correct position in the sequence; see [12]). For instance, for a particular picture that ought to be placed in position 3, but was placed in position 6, the REM score would be $6-3=3$. Correspondingly, a correctly placed picture would obtain a score of 0. Each list constructed by a child was given an overall REM score by totalling the REM scores for each of the nine items in the list. This measure was used in all experiments.
2. REM1 the same score as REM, but analysed after a delay period (variously 2–6 weeks). This measure was used in Experiments three, four, six, seven, eight and nine.
3. REM2 was calculated by subtracting from the total Removed Score, the score that was ascribed to the highest-scoring picture. In a nutshell, the difference between REM and REM2 lies in the fact that the former indicates overall accuracy of ordering the pictures, whereas the latter avoids very high scores due to the very bad placement of a single item, despite the overall sequences of the nine pictures being generally well remembered (perhaps all otherwise correctly remembered). This measure was used in Experiments two, three.
4. Correct Order measurement (Corr) indicated how many of the nine pictures were placed correctly in their true list positions in the initial testing phase; participants were given nine slots on a page, as successive dotted lines and labelled 1–9; they therefore placed as many items as possible in the correct numbered slot. This measure was used in all experiments.
5. Correct Order 1 (Corr1), the same as Correct order but measured after delays. This measure was used in Experiments six, seven, eight, nine.

Table 1 Summary of the measures applied in each experiment. See text for an explanation of the abbreviations. The accumulated N for all experiments is 512. Experiments reported in detail elsewhere are indicated in the right-hand column (F07: [12]; K12a: [18]; K12b: [19])

Exp	REM	REM1	REM2	Corr	Corr1	SPE	Qs	Tries	TotErr	Location	Ages	N	Pub'd
1	•			•						UK	U/grad	45	
Environment complexity and 'decoration' has no effect on recall for undergraduate students													
2	•		•	•		•	•			UK	18-22	39	F07:1
VE visualisation enhanced recall compared with two paper-based conditions for undergraduate students													
3	•	•	•	•		•	•			UK	11-14	62	F07:2
VE produced no benefits in recall compared with paper-based or PowerPoint conditions for middle-school children													
4	•	•		•						UK	7-9	72	F07:3
VE impeded recall for primary school children; multimedia effects seemed counter-productive													
5	•			•		•	•			UK	18-27	36	
VE with integrated challenge enhanced recall compared with paper-based and PowerPoint conditions for u-grad students													
6	•	•		•	•	•		•	•	UK	8	52	K12a:1
VE produced no benefits in recall compared with paper-based or PowerPoint conditions for primary school children													
7	•	•		•	•	•		•	•	UK	8-9	45	K12a:2
VE enhanced recall compared with paper-based or PowerPoint conditions for primary school children, given more extensive exposure to the material													
8	•	•		•	•			•	•	Ukraine	7-8	30	K12a:3
VE enhanced recall (but not long-term) compared with paper-based or PowerPoint conditions for primary school children, as Exp7, in an alternative context.													
9	•	•		•	•			•	•	Ukraine	mean 12	30	
VE produced no benefits in recall compared with paper-based or PowerPoint conditions for middle-school children, as Exp3													
10	•			•				•	•	UK	Middle School	49	
VE did not benefit recall compared with paper-based or PowerPoint conditions for middle-schoolchildren, despite the use of perhaps more engaging material													
11	•			•				•	•	UK	mean 25	25	
The addition of music synchronised to events located in the VE seemed to be counterproductive													
12	•			•				•	•		U/grad	27	K12b:1
The use of three parallel timelines to create a VE 'landscape' benefitted recall in undergraduate students													

6. SPE: serial position effects. It was of interest to know whether, after experiencing a series of locations laid out in a sequence in space, information would be remembered best (or selectively lost) at the start (primacy) or middle, or end (recency) of the list. The number of items correctly remembered and placed in list positions 1–3, 4–6 and 7–9 was therefore recorded. This measure was used in Experiments two, three, five, six and seven.
7. Qs: Use of a set of questions in the form "Did X come before Y?" Not all studies were designed to explore this variable. Used in Experiments two, three and five.

Measures eight and nine were used when challenge was introduced into the protocol, as described below (Experiment five onwards):

8. Tries: The number of passes through the experiment that participants required to meet the researcher's criterion of two successive passes without error in the training phase. This measure was used in Experiments six, seven, eight, nine, ten, eleven and twelve.
9. TotErr: A total error score, i.e., how many errors were made throughout all passes prior to reaching criterion in the training phase. This measure was used in Experiments six, seven, eight, nine, ten, eleven and twelve.
10. In Experiment twelve, where multiple timelines were used in parallel, additional variables were introduced.

3 The Experiments

We present a sequence of 12 experiments, each of which contributes to one or more of our main findings overall. Two interim discussions are offered, while overall conclusions and discussion end the chapter.

3.1 Experiment One: A Comparison of Historical Chronological Learning from Three Complexities of VE

We describe this experiment in some detail in order to indicate the kinds of VE visualisation created and the experimental methods used. The specific question in the first experiment was whether, in order to be effective for recall, an environment should include non-functional environmental features, imparting some sense of visual and experiential realism, or whether simpler 'diagrammatic' characteristics should be preferred.

Forty-five participants took part in the experiment (9M, 36F). The participants were selected pseudorandomly from within a university student population. The subject domain was the history of art. All participants confirmed that they had no formal art education and were unfamiliar with most art works presented to them during testing. It was established that they were unaware of the chronological ordering of the paintings or the specific year when any one was painted.

Nine paintings were included in the timeline. Within the environment, each painting was inscribed with its title, author and date. Participants were pseudorandomly allocated to one of the three conditions: high, medium and low VE complexity: one (basic or low complexity) was a featureless corridor, one (medium complexity; Fig. 9) modelled a real corridor with windows and other features, and a third (high complexity) allowed user manoeuvres, i.e., using a lift between floors and going upstairs and downstairs (Fig. 1).

 Participants could move at a constant velocity forward through the virtual space by depressing a key. Other movements were disabled—we had discovered during

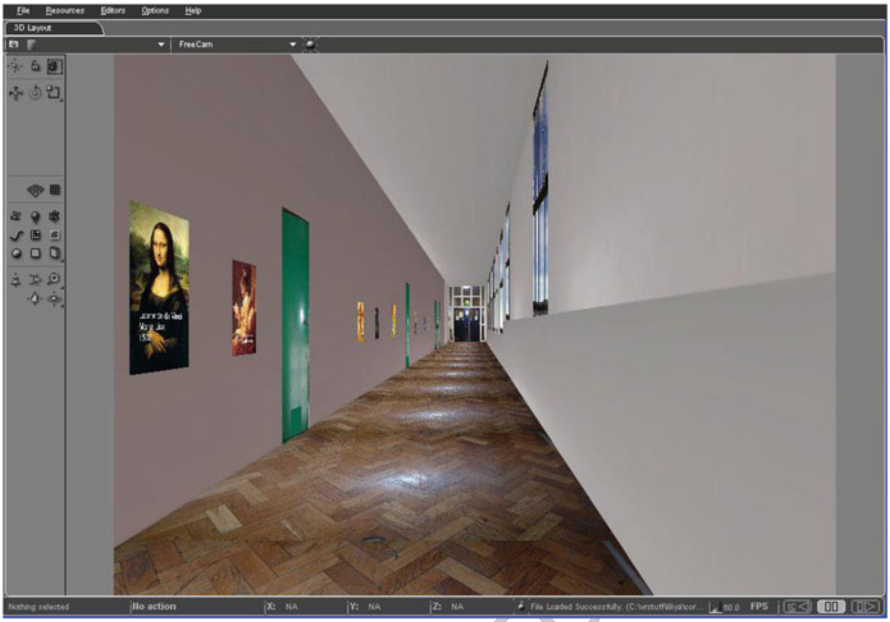

Fig. 1 Art History represented in a medium-complexity VE. Nine pictures from the history of art were located in a virtual corridor

a previous pilot phase that users could become disoriented and travel backwards in time while believing themselves to be travelling forwards. The speed of movement gave participants time to read the name of the artist, title, and year of each painting. Participants could also pause in their journey. Passing through the VE took typically 5–6 min. Participants passed through five times, after which they moved on to testing, being given a set of the nine images that they had seen in the environment (minus the inscribed text), printed on A4 sheets of paper. They were asked to place these in the order in which they had seen them on the computer. When they had completed the task, the order of their placed pictures was recorded and scored.

Two dependent variables were analysed: the number of pictures placed in their correct positions in the sequence (Number Correct), and REM (removed scores) using a one way independent ANOVA. The result showed that there was no significant difference obtained between the three conditions on either the REM scores or the Number Correct variable, $F(2,42) = .388$, $p > .05$ and $F(2,42) = .691$, $p > .05$, respectively.

The results showed that the effectiveness of the environment did not depend on its complexity or the inclusion of potentially distracting details. Statistical analysis revealed that participants retained the same amount of information irrespective of the complexity of the environment they experienced. However, later experiments (Four, Eleven) cast additional light on the possible effects of VE complexity, particularly when these use multiple media.

3.2 Experiment Two: Memory for Imaginary Historical Information Acquired from a VE, a 'Washing-Line', Text Alone

In this study, undergraduate students were tested using a nine-item series of historical events that depicted the chronological history of an imaginary planet. A 'washing-line' condition, described below, was introduced because this is a popular way of conveying chronology in school class rooms [7, 9, 25]. A verbal/text protocol was used as the control condition, its presentation using only semantic information being familiar from conventional teaching without visualisations.

A group of 39 undergraduate students (15M and 24F, aged 18–22 years), was pseudorandomly chosen from among the university student population and was pseudorandomly allocated to one of three groups, no specific attention being paid to their prowess in history classes in school. None was a history specialist.

A set of nine images comprising pictures and dates was created, each representing an event in the history of the imaginary planet. These were positioned as successive objects in a VE timeline. Participants could fly through the environment using forward movement only but with full control over their velocity. For the 'washing-line' control condition, the same pictures (with captions and dates) were printed on nine A4 sheets which were then pegged along a string across one wall of the room. For the printed verbal/text control condition, the procedure was the same except that the nine images plus event name and dates were printed, three per page, on three successive A4 sheets in portrait orientation.

The participants were allowed to spend as much time as they wished in each pass-through of the VE (the total time required at maximum velocity being 67 s). After each fly-through, an on-screen dialog prompted them to return to the beginning of the sequence.

In the washing-line condition, participants were asked to scan slowly along the line from left to right. In the verbal/text condition they were asked to look at the three A4 sheets. In all three conditions, the participants were asked to attempt to memorise the history of the planet represented.

All participants, in all three conditions, passed through the materials five times, taking roughly the same length of time to complete the exercise.

The test had two parts: a questionnaire that posed nine questions of the form "Did X come before Y?" requiring true/false responses; and a task to place the nine pictures in their correct chronological order. No time limit was imposed but on average, participants did not spend any longer than 4–6 min doing this.

The following measurements were taken: (1) "Correct number" was the number of pictures placed in their original places in the one to nine sequence; (2) the second was the number of questions correctly answered (out of nine) on the questionnaire; (3) the REM or "Removed" score assessed how far each picture was placed from its correct position; an additional score, Removed2 or REM2 was used, when testing was repeated after an interval. In order to examine serial position effects in the data (SPEs), the number of items placed correctly in list positions 1–3, 4–6, and 7–9

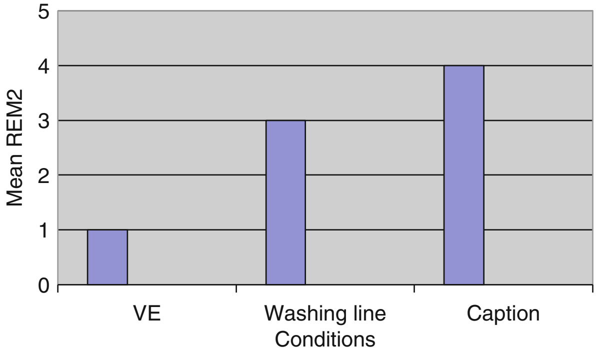

Fig. 2 Mean REM2 scores for the three groups (Experiment Two). The lower the score, the better the performance

were recorded separately for each participant. All data were normally distributed, allowing the use of parametric testing. Post-hoc group comparisons were made using the Least Significant Differences test (with two-tailed probabilities, unless otherwise specified) following the main analysis, when effects were found to be significant. There was a group significant difference in placing the pictures in their correct position $F(2,33) = 4.41$, $p < .05$. Participants in the VE group performed significantly better than participants in the two other groups (either washing-line, $p < .05$, or verbal/text, $p < .05$, groups). Further analysis showed that there was no significant difference found between washing-line and verbal/text groups, $p > .05$. For the number of questions answered correctly, the analysis revealed no significant difference, though the result bordered on significance, $F(2,33) = 2.99$, $p = .06$. Mean scores for the VE group indicated that the number of errors committed in this condition was arithmetically less than the numbers of errors made in the other two groups.

The third variable investigated was the Removed scores. There was a highly significant difference among groups, $F(2,33) = 5.95$, $p < .05$. The participants in the VE condition performed significantly better than those in the other two groups, washing-line and text (p's < .05 and .003 respectively). No significance was found between washing-line and text groups, $p = .19$. An additional variable was investigated, Removed2 scores, which revealed the same tendency (Fig. 2), ANOVA indicating that the three groups differed, $F = (2, 33) = 4.64$, $p < .05$. The VE group performed significantly better than the washing-line and caption/paper groups, p's < .05 and .005 respectively but no significance emerged between the latter groups, $p > .05$.

Since data variances were not homogeneous for the SPE measure, this was analysed by employing a non-parametric test, the Kruskal-Wallis test, to conduct a one way independent groups' analysis on each successive serial block. Group differences were then examined using the Mann–Whitney U-test. The result showed that there was a group difference in the middle block only (position block 4–6),

$X2(2) = 5.91$, $p < .05$. The VE group achieved higher scores compared to the washing-line and text/paper groups, $U(13,13) = 42$, $p < .05$; the latter two groups failed to differ significantly.

This study revealed significant differences on three out of four measures, and almost reached significance on the fourth, the number of questions answered correctly. Notably, participants who used the VE made fewer errors than those in the other two groups. However, we were aware that undergraduate students might be a special group, with more experience of working with computers than school-age children, and a fuller conceptualization of time and history [26]. With this in mind, a VE was used in the next study to assess whether middle school pupils would benefit from the use of VEs in learning about medieval history as required by the UK national curriculum.

3.3 *Experiment Three: The Use of VE Fly-Throughs as Adjuncts to National Curriculum History at Middle School Level*

Sixty-two children in two North London schools (29M, 33F, age 11–14 years) were pseudorandomly allocated by class teachers into two groups in one school and three groups in the other (Paper: $N = 24$, 17F, 7M; VE: $N = 26$, 9F, 17M; PowerPoint: $N = 13$, 7F,6M). The data were subsequently combined. The teachers were asked to equally distribute their children across the groups, taking into account their classroom performance in history.

The material, this time on medieval history, was presented in a similar manner to Experiment Two. An innovation was to introduce into Sub-study two ($N = 13$; 6M, 7F) a PowerPoint condition as a second control variable. The visual materials used in PowerPoint were identical to those used in the other two conditions. In order to move on to the next image, a key was pressed. At the end of a session of nine images, an additional screen would appear to invite the participants either to continue with the training task by returning to the starting point (as in the VE) or to proceed to a testing stage. The time taken to pass through the nine items was paced such that it was similar to that in the VE condition. The Paper Condition ($N = 24$; 7M, 17F) involved the children looking through the images presented by the researcher. The pictures would appear in the predetermined correct order, the time taken to pass through all nine being similar to that in the VE condition. As usual, the sequence of nine events was passed through five times in all three conditions.

Each participant was tested individually. The interval between exposure and testing was 48 h. The researcher explained the task by showing nine images presented in an A4 format and asking the children to place the nine images in the correct order. After this, the children were asked to complete a questionnaire, in which they were required to answer questions of the form “Did X come before Y?” To test the hypothesis that VE materials have a greater durability compared with the

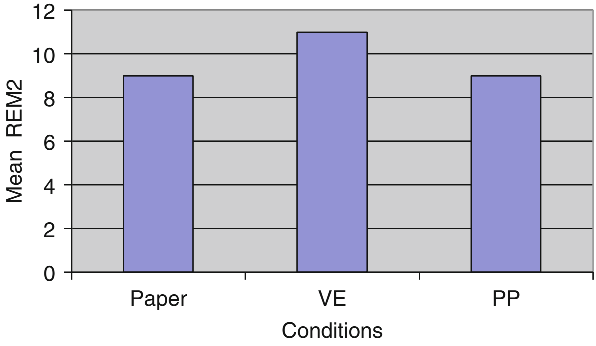

Fig. 3 The mean scores for REM2 for the three conditions (Experiment Three)

materials used in other conditions, a further test session was carried out, comparing a sample of the Paper Group (13 participants) with a sample of the VE Group (13 participants), two months after the original training and testing was completed.

Data were analysed in the same manner as in the previous experiments. In terms of picture ordering, there was no significant effect of condition; $F(2,57) = 1.12$, $p > .05$. When the number of questions answered correctly was analysed, the same pattern emerged, there being no significant differences found. Removed and Removed 2 Scores (Fig. 3) also failed to show any significant result. There was no significant result observed between groups in terms of primacy, middle or recency position blocks, X2(2) = 1.03, 1.18 and 1.53 respectively; p's > .05.

After a two month interval, there was no difference obtained when two subgroups were compared. Children in the VE condition were not better able to remember the items than those in the Paper condition. Further analysis revealed that there was a high correlation between the picture ordering score in the first round of testing and in testing after the delay, $r(24) = .7$, $p < .001$, suggesting that the measure used was sensitive and reliable.

The results from this experiment were disappointing: the VE presentation was not successful in promoting good scores as seen with undergraduate participants in the previous study. Indeed, participants showed no benefit on any measure from using the VE format in learning the sequence of historical events, and there was no benefit of using a VE in terms of the longevity of memory.

3.4 *Experiment Four: The Use of VEs in the Teaching of Primary Level History*

The next study involved younger children (primary school participants) who worked with material that had not yet been taught to them in the classroom.

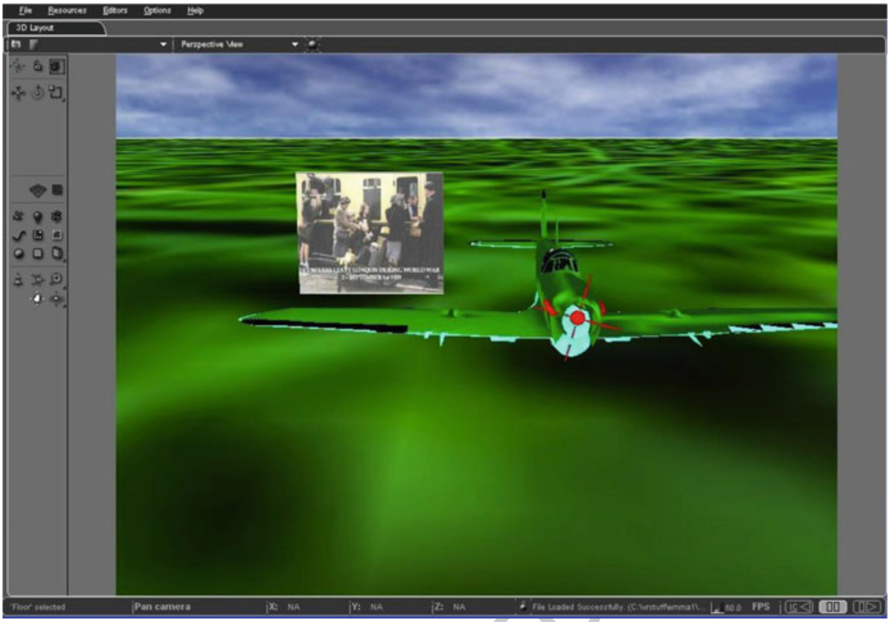

Fig. 4 The multimedia VE with animation and sound used in Experiment Four

Seventy two primary school children took place in this experiment (39M, 33F), 35 children in year 3 (19M, 16F, 7–8 years) and 37 children in year 4 (20M, 17F, 8–9 years). All children had at least some regular classroom experience of operating a computer keyboard (Fig. 4).

A set of nine images was used as in the previous studies. A new, multimedia VE format was used, incorporating some animations and sounds such as a French battle cry accompanying the depiction of the battle of Hastings, a rolling Viking boat, and a noisy Hurricane aircraft flying over a depiction of evacuees in World War II. Movement through the environment was controlled by depressing the space bar. The PowerPoint condition materials were presented as sequences of slides, without any auditory material, using the same computer as the VE condition. The same nine images were used in the Paper condition.

The participants were divided into three separate groups with equivalent numbers of boys and girls and ability range in each condition (VE condition N = 24; 13M, 11F; PowerPoint N = 23; 12 M, 11F; Paper N = 25; 14M, 11M). Testing took place two days after exposure. Each participant was tested individually and spent about 57 min completing the testing task, placing the nine images in order. Subsequently nine yes/no questions were posed in the form "Did X come before Y?"

With regard to the task in which the participants had to place pictures correctly, ANOVA revealed that the three conditions failed to show any difference, $F(2,66) = 1.38, p > .05$.

For the number of questions answered correctly, the three conditions differed significantly, $F(2,66) = 3.86$, $p < .05$. The Paper condition was significantly better compared to two other conditions, PowerPoint and VE, p's < .05 respectively, these latter groups failing to show any significant difference from one another. Teachers' prior ratings of ability correlated significantly with the questionnaire performance (Spearman's rho[N = 72] = .22; p [one-tailed] = 03).

When the difference between Removed and Removed1 was analysed statistically, AQ2 there was no significant difference, $F(2,66) = 1.8$ and 1.4; p's > .05 (Fig. 15). The teachers' ratings of ability were significantly correlated with the participants' performances on both Removed and Removed1 scores, rho[N = 72] = .19 and .20, p's [one-tailed] .05.

Serial order effects were analysed. The Kruskal-Wallis test showed that there was no significant difference when comparisons were made within individual serial position blocks. When data from the first two blocks were combined, however, placement accuracy in these list positions (1–6) was significantly better in the Paper group than in the two computer groups combined, $U(25,47) = 423$, $p < .05$.

The results showed a *dis*advantage of using a VE. The detrimental effect was especially evident when scores for items in early/intermediate positions were analysed.

4 Interim Discussion: Introducing Challenge and Pre-Training VE Experience

At this point in our research, it was clear that our VE chronological visualisations were not universally useful, and could even be counter-productive. Although undergraduate students seemed to benefit from using the VE format, other age groups did not. Middle school children failed to recall more than from other media. Moreover, primary school children actually performed worse compared to control conditions, though they were perhaps distracted by the animations and sounds used in the version of the VE visualisation they experienced.

Other issues might include a lack of engagement with the environment which could affect how much information participants could retrieve when tested since they had experienced it only passively: the only activities available to them were to look and to move their position, far less then, for example, when playing a computer game [16].

In light of out generally disappointing results from simple navigation of a VE visualisation, we next experimented with a more game-like format, in which successive representations of events (paintings, representing epochs of art history) had to be memorized and anticipated during use. As in a computer game, participants' scores were displayed in the upper right corner of the computer screen.

In the interests of brevity, details of the statistical analysis are omitted for the remaining experiments, the conclusions being summarised. Further details of Experiments Six, Seven, Eight and Twelve are available in [12, 18, 19]

4.1 *Experiment Five: Introducing Challenge Into the Interaction*

Thirty six undergraduate students (18M, 18F) were pseudorandomly drawn from an undergraduate population. They were aged 18–27 years.

The environment used was as in the studies above. The nine pictures were displayed as successive objects in the space with the title, name of the artist, and date of the painting displayed in the upper right corner of each picture. The viewpoint was held stationary while participants guessed what the up-coming image would be. The PowerPoint and Paper conditions were as previously.

All participants were trained individually. For the VE condition, the participants were instructed to observe the environment carefully while depressing the forward arrow key to move through the environment. They were told to look at the pictures and try to remember the order of the pictures, if necessary using terms such as "blue flowers" as descriptors. No attempt was made to draw their attention to specific elements depicted within each picture. The same initial procedure as in Experiment Four was applied. However, on the second fly-through, at the point when the next picture became visible, it was always blank (Fig. 5) and the viewpoint was held stationary by the experimenter. The participant had to describe the still invisible picture; if the answer was correct, the experimenter would click on the screen to display the hidden picture, after which the participant was free to move forward to repeat the procedure with the next image, and the score would increase by one. If the participant described the picture incorrectly, he/she was asked to choose again and an error was recorded. At the beginning of each pass through the environment, the screen counter was reset to zero. The experimenter noted all errors and continued until the participant achieved two successive error-free fly-throughs. In the PowerPoint condition, the same images were displayed as in the VE condition, using full screen images. For the training procedure with challenge a blank screen was displayed and replaced when the image was correctly anticipated. For the Verbal/Text condition, participants were tested with semantic information provided on each plain sheet of paper only (the artist's name, as text, the picture's name and the date it was painted). Following training, after an interval of 5 min participants were assessed using three tests:

1. The numbers one to nine were listed vertically down a test sheet and users were asked to fill in as much information as they could recall about the nine successive pictures, if possible providing the painter's name and the picture's title and date. Then the sheet was removed.

Fig. 5 The VE timeline with following images blanked, waiting for the user to attempt to recall what the next image will be (Experiment Five)

2. A list of nine questions about picture order, of the form: "Did Kandinsky come before Matisse?" was then posed, answerable with "yes" or "no".
3. Finally, a set of nine pictures were placed pseudorandomly (without names or dates) and participants were asked to order them correctly, i.e., to reproduce the order in which they were shown in the training stage. No time limit was imposed, though on average 8 min were spent completing the three tasks.

The dependent measures were: a) During use of the VE visualisation: (1) the number of passes required, excluding the first, for users to achieve two successive error-free passes; (2) the total number of errors made before criterion was reached. b) During post-use testing: (1) the amount of information provided correctly in the first test (nine pictures each having three items of information: painter, picture title, date) so a possible maximum of 27 items; (2) the number of questions answered correctly out of a total of nine; (3) the number of pictures placed in correct order, calculated using a REM score procedure as previously.

Analysis of error positions was conducted by totalling the number of errors made by each participant in training within three successive list position blocks, representing list position blocks 1–3, 4–6 and 7–9, respectively. Note that the VE condition showed almost error free learning, and therefore median scores for all

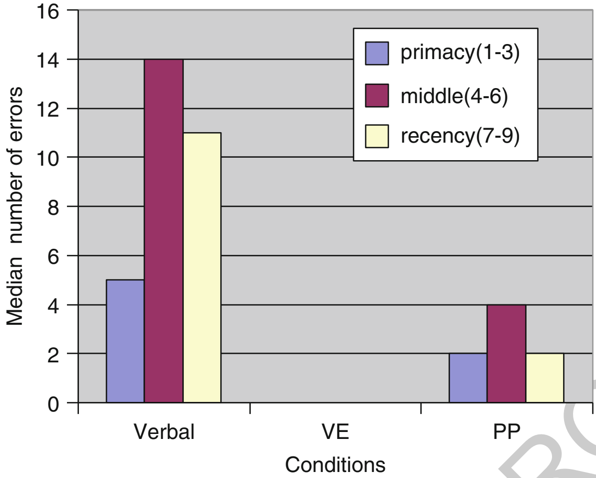

Fig. 6 Median number of errors made in the three conditions (Experiment Five)

blocks were zero, while the PowerPoint group made errors most frequently in the middle list position, and for controls a large number occurred in middle list positions (Fig. 6).

When participants were assessed for their ability to remember information about the pictures, the VE group differed significantly both from the PowerPoint and Verbal/Text controls. When REM score (reflecting the ability to place the pictures in correct chronological order) was assessed, the VE group's performance was entirely error free. As for the PowerPoint condition, two participants made two errors each, while 11 out of the 12 controls made between 2 and 4 errors (overall group mean was 2.5 errors).

This study showed a strong advantage of using a VE format compared with the PowerPoint condition (cf. [27, 28]. Participants in the verbal control group performed especially poorly. During the training procedure, it was evident that participants from the VE condition learned more, and more quickly, compared to the two other conditions; the latter two groups also showed poor retention when tested afterwards. This accords with the study by Hartley et al. [29], who claimed that the spatial relationship between objects is durable and can remain stable over a long time.

It appears that the verbal control group concentrated more on particular items (the picture name) while the experience of each picture with its accompanying textual information enabled the VE participants to absorb more of all kinds of information provided in the environment (spatial sequential and associated verbal). Interestingly, although the amount of information recalled (out of a maximum of 27 items) far exceeded the 7+/−2 items associated with short term memory [30], suggesting that participants were using a memory store with a limit greater than that traditionally regarded as the short term memory store used for the learning of simple lists of items. On the other hand, the VE group was far from perfect, and their results revealed that they could remember only half of the total information presented.

4.2 Experiment Six: The Use of Challenge in Enhancing Learning in Primary History Teaching

In an earlier study with primary children (Experiment Four, above), nine sequential images were presented chronologically in a VE, depicting eras of history from ancient Greece to World War II. It was found that children in this primary group did not benefit from exploring historical events in the VE format. In fact, they performed significantly poorer in the VE condition than pupils given the same information sequentially on paper (Paper condition) or as a non-spatial sequence displayed sequentially on a computer monitor (PowerPoint condition). The present experiment was designed to improve upon the earlier study by encouraging children to anticipate what was going to appear next, at each sequential choice point. When they anticipated correctly they scored a point (their score being displayed on the screen). This format, therefore, involved more active participation of children in the task and moderate challenge, rather like many computer games. Besides, children were asked to think carefully about historical events presented to them. This adapted protocol might also help to overcome another disadvantage. In the previous experiment (Experiment Four above), children were apparently overexcited by the animations used in the environment and perhaps concentrated less on the main task as a result. By introducing challenge (requiring anticipation, and displaying their score on the screen), children were arguably more concentrated on the main task in this experiment. It was hypothesised that children in this study would perform considerably better than those in the earlier study. Further, the environment itself was designed not to feature any elements that could be considered distracting.

Participants were 52 children (32M, 20F) drawn from a primary school in North-East London, UK. The children were from a single class, the average age being 8 years and 6 months at the time of testing.

A set of nine pictures was used, representing historical epochs, the same set as used in the previous study with primary age children (Experiment Four above). Each picture was dated. A brief description of the picture was added to each; for instance, to represent the ancient Egyptian era, a picture was used which depicted pyramids, with label and date added conspicuously in white lettering to the upper part of the figure. Features were 3D-modelled and introduced into the VE to help to make the child feel "located" in space rather than just viewing a picture. For example, models of Egyptian pyramids were located around the picture of pyramids (Fig. 7), and a virtual Hurricane aircraft flew overhead as the participant approached the evacuation picture. As before, participants could proceed in a forward direction only, achieved by depressing the space bar.

Fig. 7 Images in the VE are supplemented by relevant 3D objects to create a sense of inhabiting the space (Experiment Six)

For the PowerPoint condition during the training phase, the pictures were separated from each other by using a blank PowerPoint screen, displayed for approximately the same length of time (8 s) as participants took to move from one picture to the next in the VE.

In the control Paper condition, the same material was used as above, the pictures being printed on A4 sheets and presented to the child in landscape orientation with text added as in other conditions. Intervening blank pages were shown for 8 s each.

Children were pseudorandomly divided into three groups on the advice of teaching staff, to encompass a similar range of ability in history in each group. These were a Paper group (N = 16; 8M, 8F), a PowerPoint group (N = 18; 12M, 6F), and a VE group (N = 20; 14M, 6F). As in Experiment Five, participants were introduced to the VE with all pictures initially visible, followed by additional passes in which the anticipation (challenge) element was introduced. Scoring was as before. On average, participants required four fly-throughs to achieve criterion. In the Text condition, the nine pictures were presented to the participants by the experimenter. The same anticipation routine was used in all three conditions.

When training was complete, the participant was taken to an adjacent set of desks on which were placed the nine test items, in pseudorandom order. The participant had to place these in the correct chronological order.

Five scores were obtained. Two were during initial training: (1) the number of passes to criterion and (2) a total error score, summing all errors committed prior to reaching criterion. Two further measures were obtained from the initial post-criterion testing: (3) REM score, and (4) Correct Order. A further two scores were obtained when testing was repeated 3 weeks after the original training and testing phases: (5) REM1, and (6) Correct Order1 scores, measures (5) and (6) being calculated in the same ways as (3) and (4).

The results showed that there was no significant difference between groups on any measure, though on total errors, a group effect approached significance. Participants trained in the Text condition were found to have made fewer errors than the PowerPoint group, a result that approached significance. The data showed that even the introduction of challenge into a VE visualisation of historical chronology is not sufficient to ensure effective recall. Indeed, in terms of total errors committed prior to achieving criterion, the Text condition made arithmetically fewer errors than those in the PowerPoint condition to an extent approaching significance, but there was no hint of significance between VE and PowerPoint conditions. Other measures showed no significant differences. The results reinforced earlier findings, that PowerPoint seems to be an especially ineffective medium for conveying chronologically sequenced information [31], and indicated that children of primary school age appear not to experience the kind of benefit from using VEs in the history context that characterizes an older, undergraduate sample.

Some children commented that they did not have computers at home and that they found the task rather difficult to perform, so the poor results might have arisen partly from participants' lack of familiarity in using computers generally. Therefore in Experiment Seven, the same basic protocol was used as in the first study, but children were given extra experience with the environment and input device control before full training commenced.

4.3 Experiment Seven: Challenge and Pre-Training Experience in the Use of VEs to Teach Historical Chronology at Primary Level

This experiment was as the previous one, but the children were given more time to explore the VE. It was hypothesized that by making this modification, ensuring adequate familiarity with the medium, error free learning would be achieved.

Forty-five primary school children (32M, 13F) were pseudorandomly drawn from the population of a school in north-east London by the teaching staff. The numbers for three conditions were: Paper N = 15; 11M, 4F; VE N = 15; 10M, 5F; PowerPoint condition N = 15; 10M, 5F. The children in this study had approximately 5–10 min of extra exposure compared with Experiment Six. Children in PowerPoint and Paper conditions were also given an extra pass through the materials which was adjusted to take approximately the same time as the extra VE training. All other procedures were followed as in Experiment Six.

The same six measures were taken as in Experiment Six. When the number of rounds/passes to criterion was analysed the result was highly significant, showing substantial differences between groups. Post-hoc tests revealed that participants in the VE condition required fewer trials to meet the criterion than in Paper and PowerPoint conditions while there was no significant difference between Paper and PowerPoint conditions. The Kruskal-Wallis Test was used to compare the three conditions, VE, PowerPoint, and Text. The result obtained was significant, $X2(2) = 6.2$, $p < .05$. The Mann–Whitney U-test showed that the VE group was significantly superior to the Paper group on the REM variable, $U(N1 = N2 = 15) = 64.00$, $p < .05$ (two-tailed) and that VE trained participants performed better than PowerPoint participants $U\ (N1 = N2 = 15) = 70.5$, $p < .05$ (two-tailed) while there was no significant difference found between Paper and PowerPoint groups, $U(N1 = N2 = 15) = 108.5$, $p > .05$ (two-tailed). Clearly from these results, PowerPoint presentation was not as ineffective as suggested by the results of the earlier studies (above). Post-hoc tests failed to reveal significant differences between VE and Paper or VE and PP groups. Other measures showed no significant group differences.

The results were compared with the previous findings from primary school children who did not have challenge incorporated in the protocol, and who performed particularly poorly (Experiment Four, above). Comparability between schools is complicated by differences that may arise from differences in curriculum, computer use and teaching strategies, though this of course applies equally to all of the experimental conditions in which the children were tested. Control groups (both Paper and PowerPoint) did not differ between the experiments but those who used VEs did perform significantly better on the REM and Correct Order variables than those using VEs previously.

Compared with the previous study, the addition of extra pre-training for VE participants clearly improved retention of the historical materials. Significantly better learning was reflected in the lower number of trials needed to reach performance criterion in training and by significantly better Correct Order and REM scores at test. Indeed, performance was error free for all VE participants and thus substantially better than for VEs in the previous Experiment Six and in earlier work.

5 Can VEs Benefit Children's Learning of Historical Chronology in a Culture Where Computer Experience Is More Limited?

The studies reported in above chapters were all conducted in schools in a culture, the UK, where most pupils reported using computers on a regular basis. This might influence results in at least two ways: computer familiarity might make it easier for pupils to use VEs, and navigate more naturally and freely, leading to good retention of historical materials. On the other hand there was evidence from one study

(Experiment Four) that primary children (with limited knowledge of computers, on account of their age) were apparently overawed by the computer experience, leading to especially poor retention. Therefore two studies were conducted, in primary and secondary schools in Sumy, Eastern Ukraine, to examine the effects of using VEs in a country where children have much lower levels of computer familiarity. Challenge was incorporated, as above, by having the children anticipate up-coming images, plus prolonged pre-training, since this combination proved effective in the present experiment. The same comparisons were made among conditions as in UK samples.

5.1 *Experiment Eight: Use of a VE to Enhance the Learning of Ukrainian History in a Primary School in Eastern Ukraine*

Thirty pupils (14M, 16F, aged 7–8 years old) from school number N.23 in Sumy in the Eastern part of Ukraine took part in the experiment. Children were randomly selected and equally divided into three conditions by the teachers: PowerPoint (N = 10, 4M, 6F), VE (N = 10, 5M, 5F) and Paper (N = 10, 5M, 5F). Teachers asked pupils for details of their typical daily computer use, which was found to be an average of 2.5 h per week. This compares with 10.5 h per week in Experiment Six and 13.8 h per week for Experiment Seven, both conducted in the UK. Unfortunately the VE visualisations used in Ukraine were not identical in form to those used in the UK because of the preferences of the teachers concerning the design.

Nine pictures representing significant events in Ukrainian history were selected with the assistance of teachers, based on the materials used to teach history in the classroom to this age group and representing events considered important for children to remember chronologically. A new VE format was designed (based on teachers' requests) that consisted of four gallery rooms located on two floors in a virtual gallery, similar to those that pupils might visit on school excursions. Each floor consisted of two rooms of the same size. On level one a first room contained two pictures, on opposite walls, while another had two on adjacent walls. The same room layout was replicated on a second floor, in which three pictures were placed in one room and two pictures placed in the final room. In order to move from the first to the second floor a child was required to call a lift, from which the participant was required to go along a short corridor, leading to the first of the level two rooms, after which they could pass across the corridor to the final room. In the training stage, all pictures were dated and named. The PowerPoint condition was conducted using the same materials but as a succession of single screen displays with dates and text; the Paper/Text condition used A4 pictures with dates and text, so replicating the conditions used in Experiments Six and Seven.

As before, teachers were asked to select the children randomly and children were assigned to the VE, PowerPoint or Text conditions. Children in the VE group were asked to look at the VE visualisation together with the researcher, who explained how the environment worked. As in the preceding UK experiment, the experimenter went through the environment with the participant reading and explaining all information depicted on each picture. They were told to try to remember the order of the pictures, plus dates and titles. Participants were then invited to explore the environment by themselves until they were comfortable to move to another stage of the training phase. At this point challenge was introduced (as in Experiments Six and Seven, above) so that participants had to guess which picture would be displayed next, using the same protocol as previously. On average, children required two to three passes to reach criterion. The same was conducted with the other conditions, moving between PowerPoint slides or between successive sheets of A4 paper with printed images, in all cases having the same labels and dates as displayed in the VE condition. After reaching criterion, all children were required to place the images (provided on A4 sheets, but without dates) in correct chronological order. On average 3–4 min were spent completing this task. Overall, children spent 7–10 min carrying out the whole experimental procedure. After a 2 week interval, the testing was repeated.

Six dependent variables (as in Experiments Six and Seven) were analysed. Post-hoc Tukey analyses revealed that for Total Errors, the computer groups (VE and PowerPoint) made more errors than the Paper group, p's $< .05$. On the REM variable the VE group performed much better than the PowerPoint group, $p < .05$, but there was no significant difference between VE and Text groups, $p > .05$. On the Correct order variable, the VE group answered more questions correctly than PowerPoint, $p < .05$ but there was no significant difference between VE and Text groups, $p > .05$. On the Correct1 variable (2 weeks after initial training and testing), the VE group gave fewer correct answers compared to the Text group, p .05.

The main result from this study showed that even among pupils who do not use computers as often as those in the UK, and do not have the same degree of computer familiarity, when challenge is incorporated there is some benefit in using a VE visualisation to acquire historical chronological information. This further reinforces the conclusions from previous studies, showing the benefits of active involvement [22, 32, 33]. Interestingly, however, children in the VE condition here answered fewer questions correctly than in the Paper condition when they were retested after a 2 week interval, which suggests that there was no benefit of using VE presentations in terms of the longer-term retention of information.

Another controversial aspect of the findings was that participants in the VE group during the training phase made more errors in the course of the trials required to meet the "two successive correct passes" criterion, compared to the Paper condition. This is not consistent with the findings from previous experiments (Experiment Six and Seven), in which VE participants made fewer errors in the course of training trials. It is unclear why children did better in the VE group when tested straight after the training phase, but failed to show any significant effect after 2 weeks. This is in conflict with the finding [29] that spatial memory remains stable over a long period of time.

It is important to reiterate that the VE used in this study was different from the environments employed in the research described above. Despite the complexity of the environment that required additional mental effort (using left/right turns, manoeuvring up and down the lift) primary school children did benefit significantly from the VE experiences, although this advantage was no longer evident at follow-up testing, and so there was no lasting effect.

5.2 *Experiment Nine: Use of a VE to Enhance the Learning of Ukrainian History in a Middle School, Eastern Ukraine*

Having achieved generally disappointing results from the middle school in the UK, but a significant benefit of using VEs with challenge when children were adequately pre-exposed to the medium, the aim of this study was to see whether this would apply equally to a group of children of the same age in the Ukraine, but having much less experience of computer use. Challenge was again incorporated during training, and children were introduced to the VE format individually by the experimenter and given time to explore the environment, to familiarize themselves with the medium prior to beginning the experiment per se.

Thirty (15M, 15F) pupils from a Ukrainian middle school were tested in the experiment. The group was a year group, the average age being 12 years. Typical daily computer use was found to be an average of 1.5 h per week. Ten out of 30 participants did not have any access to computers.

The same materials were used in the experiment as described in the previous study with primary children in Ukraine. The same VE layout was employed. However, three new pictures were added to the existing environment to match the learning material covered by teachers in classroom lessons. All pictures were named and dated as previously.

Children were in three groups: a Paper N = 10; 4M, 6F; PowerPoint group N = 10; 5M, 5F; and a VE N = 10; 5M, 5F, with a similar range of ability in history in each group. The protocol followed was as before. On average, participants required four fly-throughs to achieve criterion.

As in the UK sample (Experiment Three) middle school children showed no benefit from VE training. Most of the variables explored did not show any significant differences. Participants from the PowerPoint group required more trials in order to remember all historical events.

5.3 *Experiment Ten: Use of VEs to Enhance Historical Understanding Amongst Middle School Children in the UK*

In this experiment, a second exposure to the VE visualisation was included, separated by a period of time from the first. While no immediate beneficial effect of using VE visualisations with Middle School pupils had been found (Experiment Three),

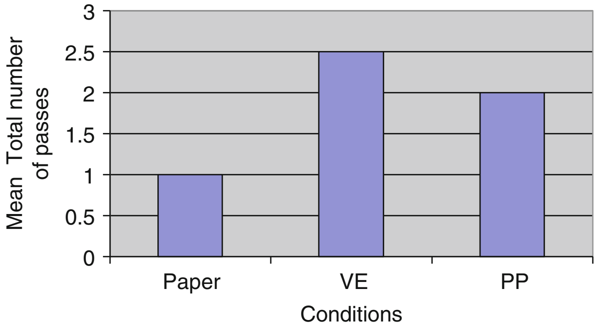

Fig. 8 Mean Total number of passes to meet criteria (Experiment Ten, Part 1)

such might become evident were participants to revisit the same environment after an interval, relearning the same materials and perhaps reinforcing retention.

It was hoped that the introduction of new materials, selected by teaching staff, in this experiment might also encourage children to be more engaged with the environment. In the previous experiment (Experiment Three) with the same age group, where performance was rather poor, it was speculated that this may have occurred because participants were asked to learn medieval materials that they had also been taught about in the classroom, which might have affected their enthusiasm for the experiment.

Forty-nine middle school pupils from North London were randomly selected by teaching staff (26M, 23F). Typical computer use 6.5 h per week.

As usual pictures and labels were the same in all conditions. Images were selected with the assistance of a teacher, who advocated using a horrific image of the victims of the Holocaust with the intention of evoking sympathy and engaging the 11–14 year-old pupils with the content. Other images represented discoveries regarded by history staff as being especially significant. Items were thus selected on the basis of the teacher's assessment of their apparent interest, rather on UK National Curriculum requirements.

The same procedure was applied as in Experiment Eight. Children were randomly divided into the same three conditions, a Paper condition (N = 16; 9M, 7F), a PowerPoint condition (N = 15; 7M, 8F) and a VE condition (N = 18; 10M, 8F).

As usual, initial exposure to the full set of materials was followed by a challenge in which users had to anticipate the next, invisible, item. The difference here was that after a 1 month interval the participants were asked to undergo the same experiment, in which they were asked to go through the training followed by the testing stages, exactly the same procedure being applied as in Part 1 of the experiment.

Five variables were analysed in the initial phase: Total number of passes required to meet criterion, Errors to criterion, REM, Correct order and Serial Order. The Total number of trials differed highly significantly $F(2,46) = 10.35$, $p < .001$ (Fig. 8). A further post-hoc test revealed that participants trained in the VE and PowerPoint conditions required more passes to meet criterion than in the Paper condition (both p's < .001).

The additional measures taken at retraining and retesting were Total errors1, Total number of trials1, REM1, Correct order1 and Serial order effect1. Participants who were trained in the PowerPoint condition tended to place more items correctly than the participants trained in the VE condition. The other three variables did not yield any statistical differences.

We found that VE participants, far from benefiting, required more passes through the environment to meet the experimental criterion. The Correct order 2 variable showed an interesting feature, insofar as participants who were trained in the Power Point condition (contrary to the previous findings above, in which the PowerPoint failed to deliver effective learning) placed significantly more items correctly than VE participants. On the other hand, the Serial Order Effect, when further analysed, showed that the participants who were trained in the VE condition placed more items correctly in the early list positions 1–3 than their counterparts in the PowerPoint condition. Despite the fact that the participants were exposed to the same training and testing stages twice, so that there was plenty of opportunity for any benefits of VEs to emerge, the results did not show any such effect. Throughout all of the above studies with middle school pupils, using different materials, different formats and with different nationalities, the absence of any advantage from using VEs (with or without challenge) was consistent and repeated, in contrast to the benefits that were observed with other age groups when equivalent training procedures were adopted.

6 Second Interim Discussion

The foregoing studies produced interesting data insofar as they demonstrated that VEs might not be effective as memorable media for chronological materials for all age groups, and especially not with middle school children. Despite the fact that other age groups profited from the use of VEs once challenge and familiarity with the medium were incorporated (see Experiments Seven and Eight), children aged 11–13 years old were found consistently not to benefit. In the second study the participants were allowed to explore the environment longer by being trained and tested twice after a short interval. The same strategy has been employed in classroom for children using 2-D timelines [25]. Still, the result showed that even extra exposure did not provoke the participants to perform better in the VE. An additional variable was tested, exploring the lasting effect of the use of VEs.

Although most of the results were non-significant, Experiment Ten demonstrated that children in the second part of the experiment showed a better grasp of materials learnt in a PowerPoint format. The present findings are therefore in disagreement with previous results consistently demonstrating the ineffectiveness of PowerPoint learning (Experiments Six, Seven and Eight).

There are several possible explanations for the fact that middle school children consistently failed to benefit from VE use. First, as suggested above, they may suffer an overload of information, which could be related to rapid biological/hormonal changes that may indirectly affect their ability to concentrate on a task or remember any new materials.

The changes that children experience in their lives at this age should not be underestimated; they experience novel activities that require independent thinking and are encouraged to take full responsibility for their actions (such as travelling to school independently, and learning new routes and strategies). This may reflect changes taking place in their cognitive styles and skills. Studies (see [34–36]) have previously argued that this is a stage at which children's spatial memory is undergoing important changes. In the context of the present studies this is important, since it means that children approach the test situation with immature structures and strategies that might be expected to make high demands on working memory. In other words, they perhaps have greater difficulty than other age groups, in employing the necessary strategies to encode materials in chronological time-space.

From the previous studies, it is evident that the use of a VE format to present sequential historical material for retention might not be beneficial for all ages, especially for middle school children. Undergraduate participants did retain more historical information from VE exposure compared to the other conditions. This can be explained in terms of their familiarity with computers and computer games, though it could also reflect better developed spatial capacities. When challenge and pre-training experience were introduced, undergraduates showed virtually error-free learning, but children at primary level also substantially improved their performance in retaining historical chronological materials. It seems that a computer "game" format might be effective in the teaching of historical chronology when using a VE as it allowed active participation and engagement, and introduced challenge that encouraged participants to be more motivated and try harder. In addition, most of the studies showed that a PowerPoint presentation might not be effective; participants tended to retain less historical information after PowerPoimt experience when compared to the other two conditions.

6.1 Experiment Eleven: Use of VEs in Conveying Parallel Timelines: Art and Music

Following the earlier success in studies carried out with undergraduate students working with VE visualisation, a new study was designed in a similar fashion, using the same paradigm as previously with the same age group, but including an additional domain of information, combining art and music. The number of items presented in each timeline was again nine. While our experiments have dealt with nine-item timelines using a single line, we also want to know whether this number can be exceeded using a more complex VE visualisation. If spatial memory is harnessed in the recall of VE visualisations, we can take advantage of the high capacity of human spatial recall. This should allow us to far exceed Miller's 7+/−2 [30], but if the short term memory buffer is the limiting factor, and it becomes overloaded as successive items are remembered, art information will be dislodged by musical information, so that the total items remembered from the display may total nine but will not exceed it.

In an initial study designed to investigate the storage capacity for materials learned from a VE, the new study used a single timeline but with both art and musical materials presented simultaneously. In this case, a single timeline was used but it incorporated two domains of information—music was played as a line of successive pictures was viewed. The situation replicated what is sometimes reported anecdotally: that a particular piece of music can help spatial recall of a place, or that returning to a place might evoke a memory of music previously heard there. Examples of evocative paintings would seem particularly appropriate to this purpose. The use of spatial memory would be indicated were the amount of information recalled from this timeline greatly exceed nine. For both art and music events, the name of the picture and the tune, the name of the artist and the composer, and the year/period in which they were both created were presented in combination, so that a total of 45 items of information were presented in the course of a participant's passing from the start to the end of the VE visualisation. It was hypothesized that (1) after several successive exploratory trips through the VE, the total information remembered would exceed nine, and (2) a greater proportion of these 45 items of information would be recalled after exploring the timeline in a VE format than by either hearing the extracts of music while viewing the linked pictures as PowerPoint screens or while viewing them conventionally printed on sheets of paper.

Twenty-five undergraduates (9M, 16F) took part in the experiment: VE ($N = 7$), Power Point ($N = 11$) and Paper ($N = 7$). The average age was 25 years.

The nine images were placed in correct chronological order with the title of the paintings, the name of the artist and the year in which the painting was produced superimposed on the picture. A text adjacent to the picture gave details of the concurrent music (name of composer, title of the extract, and year in which it was composed). The music and paintings were selected and paired in such a way that they were chronologically matched.

The extract was programmed to begin playing as the corresponding picture was approached. Challenge was introduced into the three environments. A pair of headphones was used to allow participants to listen to the music excerpts in the VE condition. In a second condition PowerPoint was used, the same protocol being used as in the VE condition, such that participants viewed the same paintings along with the music excerpts. Similarly, the music details as well as the paintings' details were also shown on the screen. In the third, Paper condition, the painting was provided on a plain piece of paper with the name of the artist and the title of the painting. In contrast to the other two conditions, the music was not played at the same time as the pictures were shown, but the details of the music were displayed.

Individual training was provided for each participant in the VE condition. Participants were told that their task was to remember the order and details of each painting as it appeared on the screen. Participants could only move forward as in previous experiments. At the same time as the painting was displayed the music was played, matching the duration of time with the painting displayed. After this, participants were told to move forward to reach another painting; the same protocol was used throughout the environment. The participants were instructed to look at each painting along with the details of the music. Also they were told that they would be later asked to anticipate which painting was going to appear next. To meet a criterion, the participants had to guess nine paintings correctly twice in a succession. After completing the first fly-through, they were asked whether they felt confident enough to complete a task i.e. to recall the images in correct order. If they did not feel confident enough, the experimenter would reset the environment from start point until the participant completed the task successfully. The participant had to recall each painting by saying the name of the artists, the title of the painting or by describing the themes of the images.

For the first test, all nine paintings including the artists' name and the title were presented randomly on a desk. The participants were asked to place them in historical chronological order, the order in which they were displayed during training. The participant was then asked to place the name and the title of the corresponding music in chronological order on a desk. For the final part of the experiment, the experimenter instructed the participants to match the music details along with the name and the title of the images. The experimenter marked the order of the music as well as the paintings. There was no time limit to perform this task. The whole procedure would typically take about 4–5 min to complete.

The dependent measures of the present study were: the number of correct images placed in chronological order; the total number of error made; the number of passes until the learning criterion was met; the amount of information remembered in three testing conditions (correct chronological order in music, placing paintings and matching music and paintings together); ability to place items in an orderly chronological sequence, assessed using a REM score.

Analysis of the variables showed that the REM picture measurement was significant, $F(2,22) = 3.98$, $p < .05$. The REM music variable showed no significant difference between conditions, $p > .05$. REM music and pictures also showed that there was no significant difference between control and experimental groups. However, the Tries variable revealed that condition differed significantly, $F(2,22) = 7.087$, $p < .05$. The REM picture variable showed that there was a significant difference obtained between VE and PowerPoint groups suggesting that VE trained participants made fewer errors when they were tested on placing pictures in order, $p < .05$.

Contrary to hypothesis, when the additional variable was added—music—participants' performances varied but were not universally enhanced. Not all information was equally well remembered. Clearly, the addition of music might have distracted and detracted from the learning of the art materials. While placing pictures in order benefitted, other variables failed to yield any significant differences. A very surprising aspect of the study was that participants who were exposed to the VE condition required more passes compared to the PowerPoint condition to reach criterion.

6.2 Experiment Twelve: Can Undergraduate Students Acquire Knowledge Effectively in Three Domains Simultaneously Using a VE with Three Parallel Timelines?

The final experiment dealt with a two dimensional time structure situated in time-space, rather than a mere line with attached objects. Spatial memory systems are distinct but interacting [37]. Multiple cues and landmarks can be used as navigational aids that allow the formation of organizational relationships with other points in space [38]. Thus people acquire knowledge about a route by seeing objects sequentially [37], that can be encoded in relation to other locations rather than from a particular stand point [29, 39]. The spatial relationship between objects is durable and can remain stable over a long period of time; it can encompass large complex, vivid and detailed spaces [29, 40–42]. We wanted to know whether presenting events in a triple timeline structure would take better advantage of spatial memory than did a single line.

A previous study using a nine item fly-through showed that undergraduate participants benefited significantly from learning about the history of an imaginary planet by using VEs, when exposed (without challenge) to just one timeline (Experiment Two, above). A further series of studies working with primary school children also showed the benefit of using VEs, especially when children had adequate time to explore the environment and when challenged by using a game format. In the present study a different form of environment was used, incorporating 12 items in three different timelines, history of psychology, general history, and art. Participants were given more time to explore the VE (over a 2 week period) after which they were asked to return and participate in a series of tests. From previous research, and experiments above, it was evident that longer exposure to the environment improves participants' performance in the short term; despite some authors having exposed participants to virtual environments for only a few minutes [43] the acquisition of spatial information from very large scale virtual environments has been said by others to require a considerable period of time [44].

Twenty-seven participants (21F, 6M) took part, fourteen in the VE group (4M, 10F) and thirteen in the Booklet group (2M, 11F). They were randomly selected from a Year 1 university student population. It was ascertained that they did not have specialist knowledge in advance of any area covered by the timelines beyond a Year 1 knowledge of Psychology. Their average age was 24 years.

A triple timeline VE visualisation was used. The same materials (images and information) were used to produce three booklets (in A4 format with coloured images) were produced. Events in the three domains—psychology, general history, and art—were matched according to the year in which they occurred (Fig. 9).

Participants were asked to read a brief introduction to the study which specified what they needed to do. Participants were randomly divided into two groups, one (experimental group) that was exposed to the VE and another (control group) who worked with a paper version of the environment designed in a booklet format. The VE group received training that consisted of passing through the environment

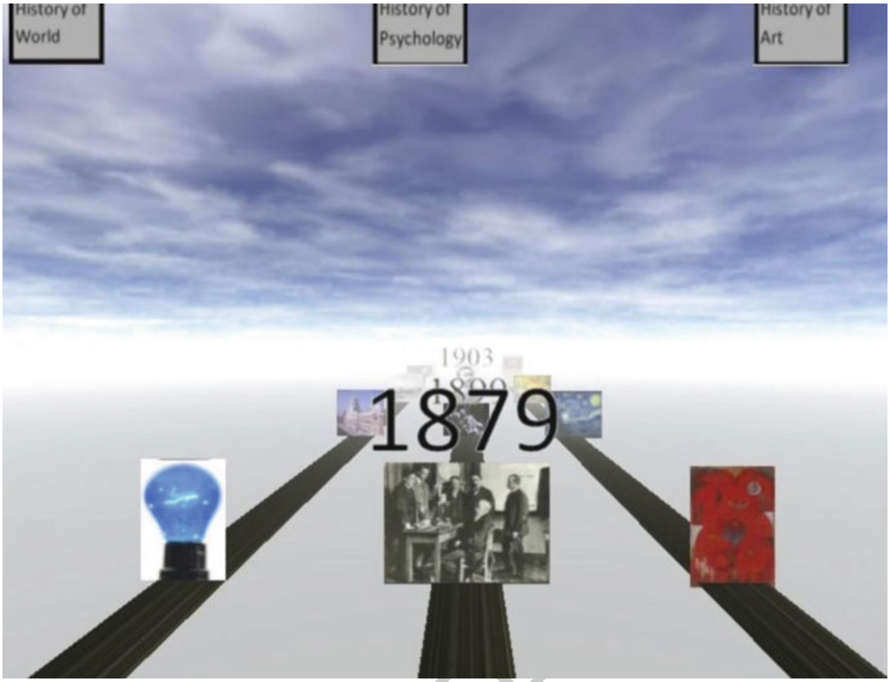

Fig. 9 The three-timeline environment representing History of Psychology, History of Art and General History (Experiment Twelve)

together with the researcher, who ensured that the participant knew how to operate (load, run and fly through) the environment. After the training procedure was complete, the participant was asked to take the environment home (or they were sent it as an e-mail attachment) where he/she could explore it in greater detail at their leisure. The latter was strongly emphasized by the researcher. Also, the researcher pointed out that all information presented in the environment should be considered, as if the participant was being asked to revise for an examination. The control (booklet) group was effectively given the same task, but asked to learn the materials in the three timelines by using three separate booklets depicting the same historical events as presented in the VE. A similar amount of time was spent with controls, explaining the booklets and required procedure, as was spent with the VE group explaining the fly-through. All participants were provided with a chart, on which they had to log the number of hours they had spent working with the materials (VE or booklet). The participants were asked to return after 2 weeks for a testing stage, although the objectives of the test were not disclosed in advance. The testing stage, for both groups, consisted of four parts. In Test 1, the participants had to recall the items learnt in their condition, but not in any particular order. In Test 2, they had to place events presented in a selected timeline in the correct chronological order. The same procedure was repeated for each component timeline. In Test 3,

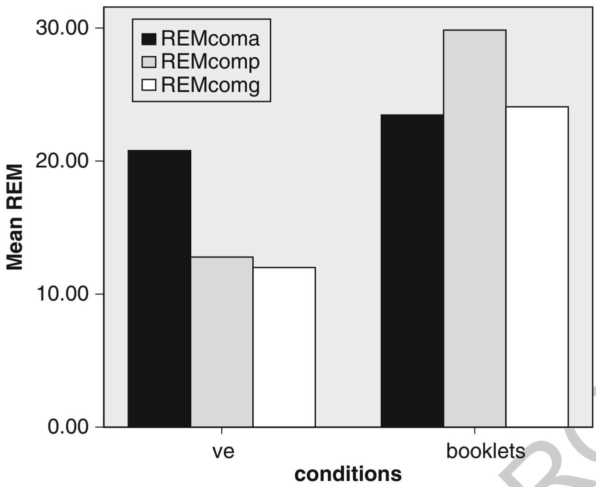

Fig. 10 REMcom: Mean REM scores for each domain/timeline when the three were tested together (*a* art, *p* psychology, *g* general history) (Experiment Twelve)

participants had to place together the events that took place in the three domains, i.e., History of Art, History of Psychology and General History, simultaneously. Finally, a questionnaire was designed to investigate whether participants could relate one timeline to another, and whether simultaneity could be identified between the events in the timelines. For instance, one of the questions asked: "What happened in the History of Psychology when event X occurred in the History of Art?"

The independent variables in the present study were the domain (art, psychology, general history), condition (Virtual Environment versus Booklets), and the gender of participants.

Fourteen dependent variables were measured: six revealed a significant statistical difference between the two groups. The VE group performed better in terms of correct recall of list positions for all three timelines (Figs. 10, 11, and 12). Participants from the Virtual group could also answer more questions correctly than controls. Total items correctly remembered, across all three timelines, approached significance. The VE group performed much better overall than controls in terms of their ability to relate together the events occurring simultaneously in the three timelines.

There was no difference between the groups in terms of the amount of time they spent in studying the materials, either reading the booklets or learning the materials from the VE. On average the two groups spent 3 h on the activities prescribed by the researcher.

This study differed from its predecessors in that a VE group was compared only with a group learning from a booklet, though using a booklet to learn historical materials is a suitable control since it resembles the materials often used in teaching. Participants were given much longer familiarization periods, to encourage the use of spatial encoding and the memorizing of materials rather like learning the layout of a small town when making daily trips through its streets.

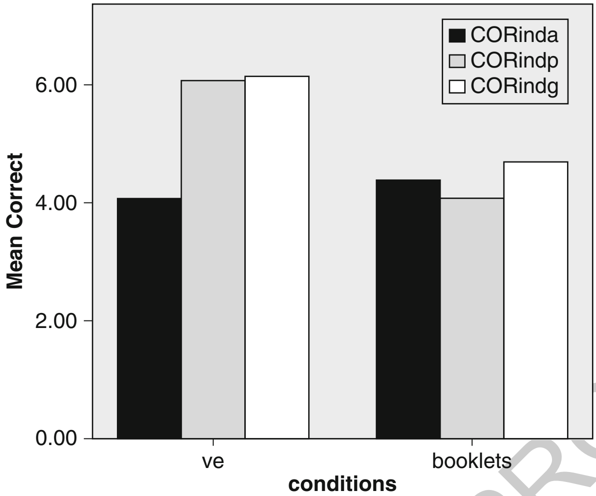

Fig. 11 Mean number of correctly placed items for each domain/timeline when tested independently (*a* art, *p* psychology, *g* general history) (Experiment Twelve)

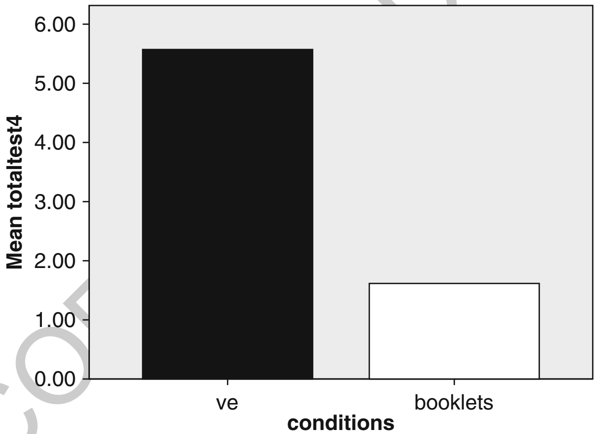

Fig. 12 The total number of items recalled across all three timelines in Test 4 by two groups (Experiment Twelve)

The results suggest that this protocol was successful. They consistently showed that using a VE gave significantly better performance than learning from a booklet. Six variables showed a statistically significant difference between the two groups, demonstrating the effectiveness of a VE, at least for this undergraduate age group. The amount of information to be remembered was substantial, in total 36 items to remember, with dates and textual information, yet still VE participants remembered more than their counterparts using booklets. According to the verbal reports of VE participants, the most important factor that helped them was their ability to connect events with each other—to see a structure and a point of reference, being able to look across the three timelines, suggesting in turn that they were genuinely using a "survey" form of cognitive representation (c.f. [11, 38]). This may be explained in terms of the fact that initially VE participants, unlike booklet participants, were exposed to the three timelines simultaneously, movement in virtual space giving them a better experience of time-space, allowing them to change their position fluently in relation to landmarks, historical images, and facts. As for the control group, they were limited in performing these activities in the sense that they could not easily visualise which historical facts happened simultaneously. For them the information that they were asked to memorize was presented with items isolated from each other, so lacking a sense of historical coherence, structure and organization. The result of the present study strongly suggests that VE visualisations of time have potential for further investigation, because although only three timelines were used in the present study, there is no reason why a more elaborate spatial environment could not encompass many timelines and a quantity of information similar to that remembered (as buildings, streets, squares) in a familiar town.

7 Concluding Discussion

This research has generated interesting but challenging results. It started from a naïve hypothesis, that just moving past events presented as pictures and other markers placed spatially in a virtual environment would instil these as places in users' spatial-temporal memory and make them memorable in correct order. The hypothesis, although naïve, still proved to be a good starting point, being plausible from previous work in which VEs have successfully conveyed spatial information. The results overall have suggested that the technology could be developed in such a way as to be valuable for specific purposes if used in the correct ways, but there are still questions over the effect of the users' age. Indeed the most striking finding, one which may turn out to be applicable to other forms of visualisation, concerns the effect of the age of the users. As we have seen, this is not a story of increase or decrease with age: the youngest and oldest participants performed better than those in between.

As we move around a real world environment, even at first visit to a new location, it seems as though we more or less effortlessly store some model of the place/space, and can remember other information with the assistance of that model. From our experiments, it seems that simply encountering events in time modelled as a spatialised environment does not have the same 'automatic' benefits. We have shown that we often had to cajole our participants using in-built game-like challenges (though these were also introduced across the other conditions) in order to produce significant gain.

An unresolved question is whether the spatial visualisation we were using carried all the potential benefits of experienced spatiality. This might be part of the explanation of our failure to get the effects associated with learning the layout of a real place. Although earlier work had shown that VE models of physical places seem to be learned in similar ways to real places, important cues may be missing when the user merely sits at a monitor navigating a virtual space. Further work is needed to discover what benefits for a similar domain might arise from (1) using immersive VE technologies in place of the screen and (2) harnessing the physicality of movement and proprioception as discussed for example by Price and Rogers [45].

It might be expected that an environment having a variety of engaging and 'realistic' features would promote the greatest learning, especially where the motivation of young children is concerned. However, when a more complex environment was used that included many animations and sounds, primary school children appeared to be distracted and consequently did not gain as much historical information as expected and, when tested, they showed no improvement in retaining information compared to other conditions. There is thus no evidence of benefit from 'decorative' motivational objects and experiences in the environment. Indeed, deciding for any given visualisation which aspects are 'necessary' or functional in itself requires investigation. In the domain of charts and similar visualisations there is unresolved controversy over the question of 'chart-junk' [46, 47], a concept analogous to some of the features we introduced into timelines here.

Some prominent questions raised therefore include:

1. To what extent are the findings of the research reported here, in particular relating to age-related difference, generalizable to other domains, users and formats?
2. What more can be discovered about the benefits and drawbacks of non-functional elements in diagrammatic visualisations?
3. More fundamentally, how can we define the borderline between those aspects of a visualisation which are strictly functional and those which are 'decorative'?
4. Would the results be different if the users' encounter with the VEs was immersive, using, for example, head-mounted displays with stereoscopic viewing?

Further work is also needed on the dynamic relation of the user to the visualisation. We explained that we constrained the movement of the user following the pilot studies. Work is needed to explore the most favourable kinds of allowable movement and constraints.

Acknowledgements The authors would like to thank David Newson for creating the VEs; head-teachers, teachers and pupils at Alexandra Park School, North London, Hoddesdon School in Hertfordshire, Northumberland Park School in Tottenham, North London (especially Peter Molife), Worcesters and George Spicer Schools in Enfield, London, and School No. 23 in Sumy, Ukraine for their participation; and the Leverhulme Trust for financial support (grant number F/00 765/D).

References

1. W. James. *The Principles of Psychology*. (2vols.) New York: Henry Holt, p. 86, 1890.
2. S. Boyd Davis, E. Bevan and A. Kudikov. Just In Time: defining historical chronographics. *Proc. Electronic Visualisation and the Arts: EVA London 2010*. 5th-7th July 2010, British Computer Society, London. 355–362. [Online]. Available: http://www.bcs.org/content/conWebDoc/36111 (Accessed 30 July 2012), 2010.

3. J. Barbeu-Dubourg. *Chronographie, ou Description des Tems* (Chronography, or Depiction of Time). Pamphlet accompanying Chronographic Chart. Paris. (With a translation by S. Boyd Davis, 2009). [Online]. Available at http://goo.gl/vNhN (Accessed 30 July 2012), p.8, 1753.
4. S. Ferguson. The 1753 Carte Chronographique of Jacques Barbeu-Dubourg. *Princeton University Library Chronicle*. (Winter 1991). [Online]. Available at http://www.princeton.edu/~ferguson/ (Accessed 30 July 2012), 1991.
5. A. Le Sage. [Emmanuel, comte de Las Cases]. *Genealogical, Chronological, Historical and Geographical Atlas by Mr Le Sage*. London: J. Barfield. 1801.
6. L. Korallo. *Use of virtual reality environments to improve the learning of historical chronology.* Unpublished Ph.D. thesis, Middlesex University, p. 40, 2010.
7. A. Hodkinson. Historical time and the national curriculum. *Teaching History*, 79 (April 1995), pp. 18–20. 1995.
8. J. West. Young children's awareness of the past. *Trends in Education*, 1, pp. 8–15, 1978.
9. S. Wood. Developing an understanding of time-sequencing issues. *Teaching History*, 79 (April 1995), pp. 11–14, 1995.
10. R.L. Kullberg. Dynamic Timelines: visualizing historical information in three dimensions. MSc Dissertation, Massachusetts Institute of Technology, September 1995. [Online]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.5278&rep=rep1&type=pdf (Accessed: 1 August 2012). 1995.
11. J. O'Keefe and L. Nadel. *The hippocampus as a cognitive map*. Oxford: Clarendon Press. 1978.
12. N. Foreman, S. Boyd Davis, M. Moar, L. Korallo and E. Chappell. 2007. Can Virtual Environments Enhance the Learning of Historical Chronology? *Instructional Science* 36(2). ISSN 0020-4277, pp. 155–173, 2007.
13. N. Foreman, P. Wilson, D. Stanton, H. Duffy and R. Parnell. Transfer of spatial knowledge to a two-level shopping mall in older people, following virtual exploration. *Environment and Behaviour*, 37, pp. 275–292, 2005.
14. M. Tlauka, P. Donaldson and D. Wilson. Forgetting in spatial memories acquired in a virtual environment. *Applied Cognitive Psychology*, 22, pp. 69–84, 2008.
15. N. Foreman, D. Stanton, P.N. Wilson and H. Duffy. Spatial knowledge of a real school environment acquired from virtual or physical models by able-bodied children and children with learning disabilities. *Journal of Experimental Psychology: Applied*, 9, pp. 67–74, 2003.
16. G. Sandamas and N. Foreman. Spatial reconstruction following virtual exploration in children aged 5–9 years: effects of age, gender and activity-passivity. *Journal of Environmental Psychology*, 27(2), pp. 126–134, 2007.
17. F.A. Yates. *The Art of Memory*. Chicago: University of Chicago Press, 1966.
18. L. Korallo, N. Foreman, S. Boyd Davis, M. Moar and M. Coulson. 'Can Multiple 'Spatial' Virtual Timelines Convey the relatedness of Chronological Knowledge across Parallel Domains?' *Computers & Education* 58(2), February 2012, pp. 856–862, 2012.
19. L. Korallo, N. Foreman, S. Boyd Davis, M. Moar and M. Coulson. Do challenge, task experience or computer familiarity influence the learning of historical chronology from virtual environments in 8–9 year old children? *Computers and Education* 58(4). May 2012, pp. 1106–1116, 2012.
20. L. Korallo. Using VEs to enhance chronological understanding of history. *Research Studies, Literature Reviews and Perspectives on Psychological Sciences.* Edited by Thanos Patelis. Athens Institute Of Education and Research, pp. 113–143. In press.
21. S. Boyd Davis. History on the Line: time as dimension. *Design Issues* 28(4). September 2012, pp. 4–17, 2012.
22. J. Pedley, L. Camfield and N. Foreman. Navigating memories. in B. Ahrends and D. Thackara (Eds.), *Experiment: Conversations in Art and Science*. London: Wellcome Trust, pp.173–235, 2003.
23. A. Cockburn and B. McKenzie,. Evaluating spatial memory in two and three dimensions. *International Journal of Human-Computer Studies*, 61, pp. 359–373, 2004.

24. R.A. Ruddle, S.J. Payne and D.M. Jones. Navigating buildings in "desk-top" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, 3, pp. 143–159, 1997.
25. I. Dawson. Time for Chronology? Ideas for developing chronological understanding. *Teaching History* 117, pp. 14–22, 2004.
26. J. Howson. Is it the Tuarts and then the Studors or the other way round? The importance of developing a usable big picture of the past. *Teaching History*, 127, pp. 40–7, 2007.
27. L. Smart. *Using IT in primary school history*. Cassell, London. 1996.
28. J.D.M. Underwood and G. Underwood. *Computers and Learning*. Basil Blackwell Ltd., Oxford. 1990.
29. T. Hartley, I. Trinkler and N. Burgess. Geometric determinants of human spatial memory. *Cognition*, 94, pp. 39–75, 2004.
30. G.A. Miller. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, pp. 81–97, 1956.
31. T. Haydn. Multimedia, Interactivity and Learning: some lessons from the United Kingdom. Current Developments in Technology Assisted Education (2006), *Proceedings of m-ICTE2006*, vol. 1, 110–114. University of Seville, Spain, 22–25 September 2006, pp. 110–114, 2006.
32. P.N. Wilson. Use of virtual reality computing in spatial learning research. In N. Foreman & R. Gillett (Eds.), *Handbook of spatial research paradigms and methodologies: Volume 1, Spatial cognition in the child and adult* Hove: Psychology Press, pp. 181–206. 1997.
33. G. Wallet, H. Sauzeon, A. Prashant and B. N'kaoua. Benefit from an active exploration on the transfer of spatial knowledge: Impact of graphic richness. *13th International Conference on Human-Computer Interaction*. 19–24 July, San Diego, 2009.
34. A. Flickinger and K.J. Rehage. Building time and space concepts. *Twentieth yearbook*. National Council for the Social Studies. Menasha, WI: George Banta Publishing, pp. 107–116, 1949.
35. G. Jahoda. Children's concepts of Time and History. *Education Review,* 15 (February 1963), pp. 87–104, 1963.
36. W.J. Friedman. Conventional time concepts and children's structuring of time. In W.J. Friedman (Ed.) *The Developmental Psychology of Time*. London: Academic Press, pp. 171–206, 1982.
37. G. Jansen. Memory for object location and route direction in virtual large-scale space. *The Quarterly Journal of Experimental Psychology* 59(3), pp. 493–508, 2006.
38. P. Jansen-Osmann and G. Wiedenbauer. The representation of landmarks and routes in children and adult: a studying a virtual environment. *Journal of Environmental Psychology* 24, pp. 347–357, 2004.
39. S. Moffat, A. Zonderman and S. Resnick. Age differences in spatial memory in a virtual environment navigation task. *Neurobiology of Aging* 22, pp. 787–796, 2001.
40. E. Maquire, R. Nannery and H. Spiers. Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain* 128, pp. 2894–2907. 2006.
41. M. Moscovitch, L. Nadel, G. Winocur, A. Gilboa and R.S. Rosenbaum. The cognitive neuroscience of remote memory: a focus on functional neuroimaging. *Current Opinion: Neurobiology* 16, pp. 179–190, 2006.
42. R.S. Rosenbaum, S. Priselac, S. Kohler, S.E. Black, F.Q. Gao, L. Nadel and M. Moscovitch. Studies of remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nature Science*, 3, pp. 1044–1048, 2000.
43. D. Waller. Individual differences in spatial learning from computer-generated environments. *Journal of Experimental Psychology: Applied*. 6, pp. 307–321. 2000.
44. R. Darken and J. Silbert. Navigating large virtual spaces. *International Journal of Human-Computer Interaction* 8, pp. 49–71, 1996.
45. S. Price and Y. Rogers. Let's get physical: the learning benefits of interacting in digitally augmented physical spaces. *Computers & Education* 43(1–2). August 2004, pp. 137–151, 2004.

46. E. Tufte. *Envisioning Information*. Cheshire, Conn.: Graphics Press, pp. 34–35, 1990.
47. S. Bateman, R.L. Mandryk, C. Gutwin, A. Genest, D. McDine and C. Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. *Proceedings of the 28th international conference on Human factors in computing systems (CHI '10)*. ACM, New York, NY, USA, pp. 2573–2582, 2010.

AUTHOR QUERIES

AQ1. Please check if inserted citations for Figs. 1 and 4 are okay.
AQ2. Please specify significant of Fig. 15.