

Simple animations for organizing diagrams

Mireille Bétrancourt
INRIA Grenoble
655 Avenue de l'Europe
38330 Monbonnot
FRANCE

Barbara Tversky
Department of Psychology, Bldg 420
Stanford University
Stanford, California 94305-2130
USA

Please address communications to:

Dr Mireille Bétrancourt
INRIA Grenoble
655 Avenue de l'Europe
38330 Monbonnot
FRANCE
Tel: (33) 476 61 53 88
Fax: (33) 476 61 52 00
Mireille.Betrancourt@inria.fr

Paper submitted in March 1999 to International Journal of Human-Computer Studies. Revised version sent in April 2000

Simple animations for organizing diagrams

Abstract

Despite a general belief among multimedia designers, there is little evidence that animation improves users' cognitive processing. One reason why animation has not met designers' expectations is that animations are usually complex and difficult to process. In this research we used a simple animation device to guide users in their processing of complex diagrams. The device, called sequential display, consisted of successively displaying portions of the diagrams where the portions were either meaningfully clustered or random. In two experiments, learning from sequentially organized displays was faster from sequentially random displays. However, learning sequential displays was not significantly better than the static display. In all three experiments, order in recall in the sequential conditions followed the presentation order. In the third experiment, processing was faster and more accurate using the organization of the sequential displays. Thus, the way in which spatial information is temporally clustered and arrayed affects the organization of elements in the mental representation. The simple animation device of meaningful sequential presentation of components of a complex graphic display can guide construction of a mental model of the graphic as well as facilitate memory for it.

1. Introduction

Under the assumption that "a diagram is worth thousand words", computer educational materials make extensive use of graphics and diagrams. Use of graphics has increased with the growing development of computerized and multimedia documents. Graphics, animated or static, are widely used in interfaces and in software applications, including web browsers, helps programs and instructional materials. In addition to their attractiveness, graphics are often believed to facilitate comprehension, memory and inference. Designers frequently use diagrams to embellish, reinforce, elaborate, summarize and compare (Hunter, Crismore, & Pearson, 1987). But are graphics as effective as designers believe? Graphics can be complex and confusing. In contrast to text, that is linear, there is no standard way of processing diagrams.

Research comparing learning from illustrated text with learning from text alone provides encouragement for the use of graphics. Overall, graphics facilitate comprehension and memory (Levie & Lentz, 1982; Denis, 1984). Graphics are thought to serve a number of functions. They do so for at least four reasons: attentional, affective, cognitive, and compensatory (Levie & Lentz, 1982; Mayer, 1989; Winn, 1989; Tversky, to appear). First, graphics attract and direct attention to certain points in the text. They make learning more enjoyable. On the cognitive side, they improve the processing of content, and provide additional information. Finally, they offer a non verbal alternative representation.

Nevertheless, there are a non-negligible number of cases where graphics are not only unhelpful, but sometimes detrimental. For instance, irrelevant illustrations have a negative effect (Levin, Anglin, & Carney, 1987; Gyselink & Tardieu, 1994). Moreover, graphics that facilitate one task may be ineffective for another (Mayer & Gallini, 1990). More important, many graphics may fail to benefit, either because readers do not attend to them (Peek, 1993) or are unable to process them (Filippatou & Pumfrey, 1996). The use of animated illustration is even more questionable, since the cognitive load required to process them is greater (Lowe, 1996). Up to now, previous research has not converged on consistent results regarding the actual efficiency of animated displays

(Morrison, Tverky, & Betrancourt, 2000; Rieber, 1990; Scaife & Rogers, 1996). For example, animated displays have not facilitated learning software applications (Palmiter & Elkerton, 1993).

In contrast to graphics, language is sequenced in time, presenting one idea after another. The linear structure of language encourages speakers and writers to structure information and to select an effective order to convey this structure (e.g. Levelt, 1989). On the comprehension side, the order in which information is mentioned has affects on the mental representation (e.g. Denis, 1996). For example, Denis and Denhière (1990) compared the processing of two versions of a text describing a spatial configuration. One version followed a linear and coherent order, whereas the other followed an incoherent order, rarely observed in spontaneous descriptions. Reading times were shorter and recall better for the coherent order. These findings show that mentioning information in a relevant order improves cognitive processing. One explanation for this effect is that a coherent order enables the immediate integration of information into a mental model (Mani & Johnson-Laird, 1982). Order of mention also serves as a guide for later retrieval of information. For example, Taylor and Tversky (1992) found that participants who had read a description of an environment drew the elements of the environment in the order they had been described.

It is often presumed that a diagram is processed immediately, as a whole. However, most diagrams require cognitively demanding processing of the details and relationships between elements in order to reveal the entire meaning. Eye movements reveal that processing is sequential. For diagrams, users have to discover the best order to process the information. This may partly account for the finding that users do not always take advantage of graphics. In some cases, users are provided with auxiliary text that guides them in processing diagrams (Hegarty & Just, 1993). We propose a way that diagrams can do that themselves.

Because animations, like language, change over time, simple animations may be used to guide users to an effective order of processing graphic information. Here we investigate the efficiency of a simple animated device, called sequential display, in facilitating comprehension of graphics. In a sequential display, subgroups of items are displayed one after another. The subgroups are either meaningfully grouped or not. A similar display was used by Wright, Hull, & Black (1990). Participants read a text illustrated with a complex diagram that appeared either gradually or all at once. When the diagram appeared gradually, participants took less time to study the document, for equivalent performance. However, this advantage disappeared if inspection of the diagram was up to the users.

Evidence from the literature on expertise (e.g. Chase & Simon, 1976; Egan & Schwartz, 1992) has shown that experts' memory for graphic displays is better than novices because of their ability to group items in meaningful configurations. Spatial clustering by semantic categories also improves memory for word lists (e.g. Bower & Clark-Meyers, 1980). Similarly, we predict that displaying a configuration using a meaningful temporal clustering of items will improve memory for diagrams compared to random temporal clustering. Moreover, on the basis of Taylor and Tversky's (1992) findings, we also predict that display order will not only organize input but also act as a recall guide. In other words, participants who experience a spatial configuration sequentially will retrieve items of the configuration in the order in which they appeared.

A preliminary experiment using a map of a library (Bétrancourt, Bisseret, & Faure, to appear) compared sequential displays based on meaningful or spatial criteria to a static display. Although participants in the sequential conditions did not learn the diagram faster, they did recall the diagram using the clustering displayed.

Moreover, participants were faster recognizing parts of the diagram that were consistent with the learned clustering. In that experiment, clustering was conveyed through a masking device, that exposed a single cluster of 6 items at a time. In the three experiments we report here, we use a purely temporal display. The first two experiments aim to assess whether a meaningful sequential display of a spatial configuration improves learning and memory compared to a random sequential display and a static display. The third experiment further investigated whether the temporal ordering of elements of a diagram affects the organization of the mental representation.

2. Experiment 1: Sequential display of a small Library

To study the effects of sequential display of meaningful or random parts of a graphic, we chose a map of library for several reasons. A library is familiar to our participants, and it consists of equal parts of undifferentiated elements except by content. Thus it represents a broad class of diagrams, including computer files, web sites and potential customers. In the sequential display conditions, meaningful or random clusters of parts, were sequentially highlighted. In the static condition, the diagram was presented as a whole. We expect that meaningful temporal clustering should be superior to random clustering in speed of learning and amount recalled and that the order of cluster presentation should affect mental organization as revealed in order and clustering of recall.

2.1. METHODS

Participants

Fifty-three undergraduates, 24 male and 29 female, from an introductory psychology course at Stanford University participated individually in this experiment for course credits. They were randomly assigned to conditions.

Material

The library map consisted of 20 rectangle-shaped bookcases labeled by academic discipline (see Figure 1).

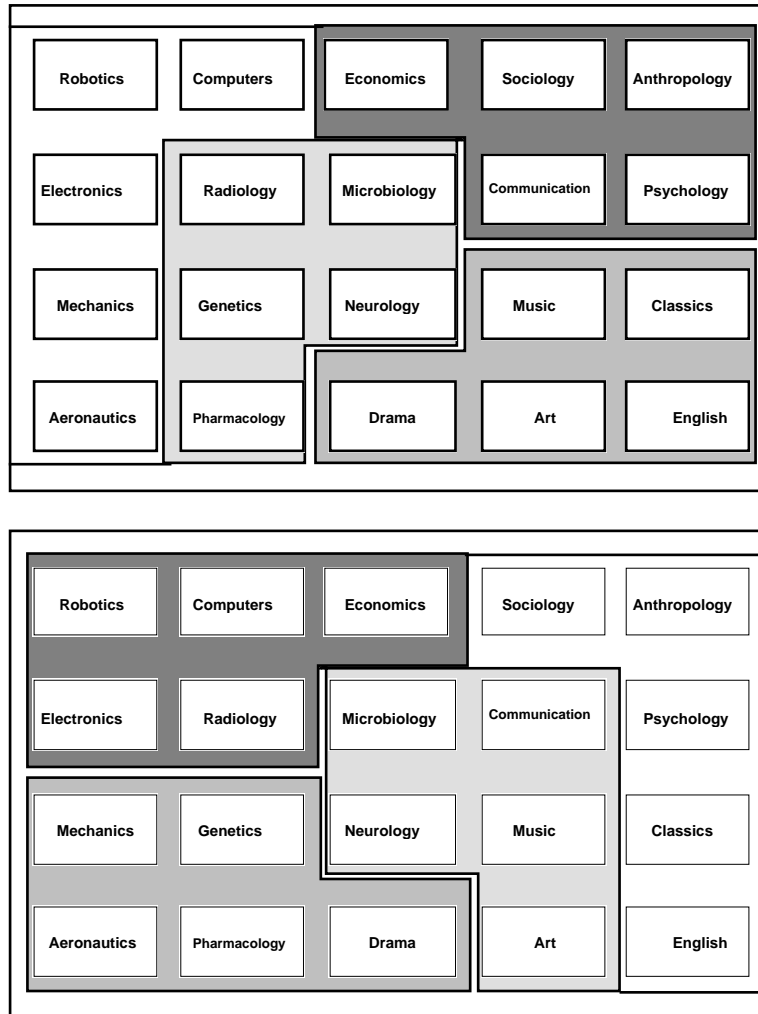


Figure 1. Map of the 20-bookcase library. Spatial clusters indicate the organization conveyed by sequential display: the meaningful organization appears in the upper diagram and the random organization appears in the lower diagram.

There were three display conditions. In the static display, the bookcases appeared all at once, as in traditional diagrams. In the two sequential displays, elements were displayed gradually, one cluster at a time, for 10 seconds. Once displayed, the clusters remained on the screen. Clusters were displayed roughly from left to right and top to bottom in both sequential displays. In the meaningful display, elements were clustered by academic disciplines (humanities, social sciences, medicine and engineering). In the random display, the clustering was obtained by flipping the meaningful clustering along the vertical axis. The items of each random cluster belonged to at least 2 different meaningful clusters.

Two spatial layouts of the same library were designed by flipping along the vertical axis, to insure that the effect of clustering was not due to the spatial display of the clusters. Each physical clustering was used either for the meaningful or the random conditions, depending on the layout. In other words, the spatial clustering of the meaningful condition in layout 1 is the spatial clustering of the random condition in layout 2 and vice versa

Design

The design was a 3 x 2 between-subjects factorial design, with the two factors of display condition (static, meaningful and random) and layout (layout 1 and 2)

Procedure

Participants were stationed in front of a computer and were asked to fill out a simple form (age, gender and major). All participants were acquainted with computers.

Study phase. The participants were given written instructions stating that they would have to memorize the map of a library that would be displayed for about 40 seconds in order to recall the names on the bookcases on a blank sheet. In the sequential conditions, they were told that the map would be gradually displayed. They knew that they would have no more than four trials to recall as many items as possible. The total display time for a session was 42 seconds (2 s for the blank frame plus 40 seconds) equal in all conditions.

Test phase. After the map disappeared, participants were told to write as many names as possible. The first recall was a free recall, where participants had to list the names without locating them in space. Immediately after, the experimenter took away their free recall and participants performed a located recall task. A blank grid 4x5 was provided and they had to recall the names in the right location. Participants went through four identical sessions (study, free recall and located recall) or until their recall was complete. The experimenter recorded the order in which the participants drew the items.

2.2. RESULTS

Memory performance

Table 1 displays the mean number of trials needed in each condition to recall the library (all twenty items accurately located). A score of 5 was assigned to the 9 subjects whose recall was incomplete at the final fourth trial.

Table 1- Mean number of trials needed in each condition to recall the library (SD are indicated into parentheses).

| | Meaningful | Random | Static |
|-----------|------------|------------|------------|
| Mean (SD) | 3.39 (.85) | 3.94 (.80) | 3.71 (.69) |

An analysis of variance was computed on the number of trials with the type of display and layout as between-subjects variables. The effect of the layout was not significant ($F(1,49) = 2.75$, NS), nor was the interaction between the two factors ($F(5,47) = 1.43$, NS). The effect of the display condition was not significant overall ($F(2,50) = 2.27$, NS). However, planned comparison between the sequential conditions showed an advantage of the meaningful groups ($F(1,35) = 4.52$, $p < .05$). The advantage of meaningful clustering over random one was corroborated in detailed analyses of trial-by-trial performance and location errors.

Recall order

To determine whether order of presentation affected order of recall, Kendall's rank correlation coefficients (*taus*) were computed between the order participants recalled the items and the display orders (random and meaningful) for free and located recall. These appear in Table 2. For 10 or more items, the significance of tau can be tested using Fisher and Yates' tableⁱ.

ⁱ reproduced in Johnson, N.L and Leone, FC (1964). Statistics and experimental design in Engineering and the physical sciences, John Wiley & Sons: New York, London, Sydney.

Table 2. Kendall's coefficient correlation (tau) between recall orders and display orders

| | | Meaningful | Random |
|----------------|------------------|------------|--------|
| Free recall | Meaningful order | .23 | .02 |
| | Random order | .20 | .03 |
| Located recall | Meaningful order | .48 | .23 |
| | Random order | .37 | .23 |

Note: Since the analysis required values for all items, we use the order data from the last trial for each subject. We eliminated from the analysis data of participants with more than three missing items (Free recall Located recall: 1 subject from the Static condition, and two participants from the Random condition). When there were three missing values or less, they were allocated the last rank.

None of the correlation coefficients was significant for free recall. For located recall, there was significant agreement between participants' recall order and the function display order only for the meaningful condition as expected.

Clustering in recall

A hierarchical cluster analysis of recall orders reveal details on the recall's clustering. This analysis examined whether participants drew the items of a same cluster successively, regardless of the order of items within and between clusters. The resulting trees are displayed in Figure 2 and 3. Because there were no effects of layout, data were collapsed across layout types for each display condition.

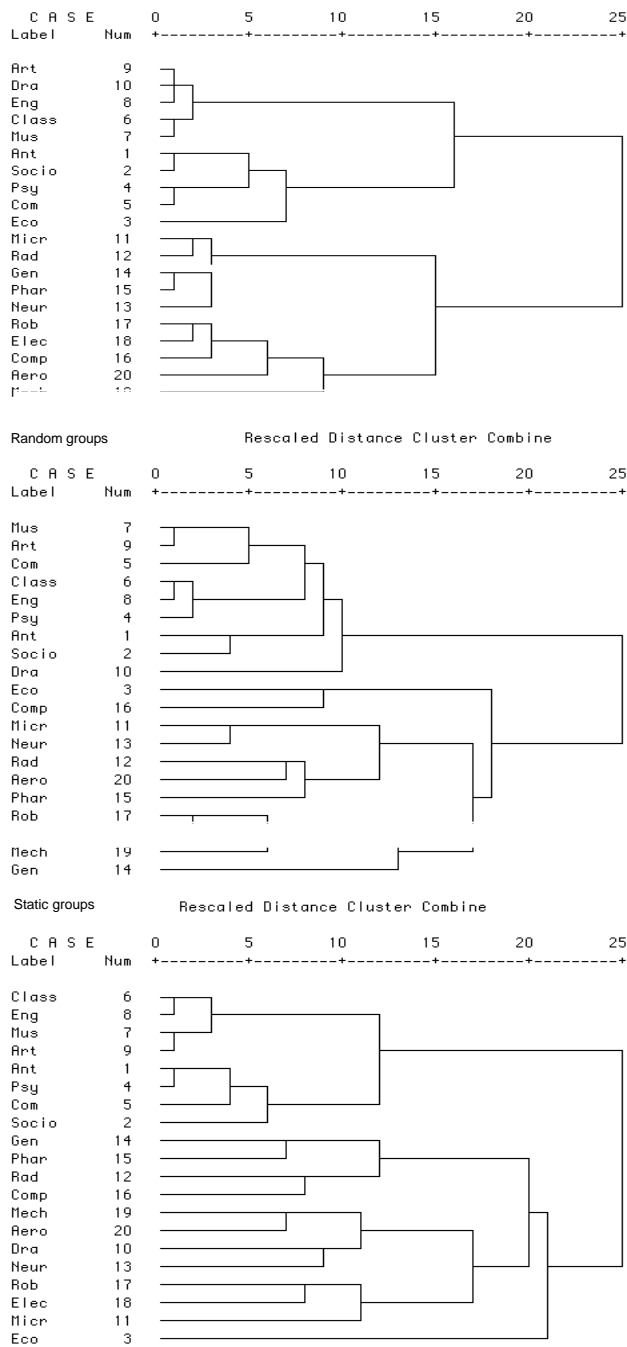


Figure 2. Trees resulting from the hierarchical analysis of drawing order in free recall for, from top to bottom: meaningful condition; random condition; static condition.

Free recall. As expected, clustering for participants in the meaningful condition corresponded to order of sequential display (Figure 2). In contrast, clustering for participants in the random condition did not resemble any of the defined types of clustering. Instead, clustering seemed to follow spatial criteria, but the distinction between clusters was not as clear as in the meaningful condition. Similarly, for the static conditions, the clustering seemed to follow spatial criteria more strictly than in the random conditions, but was not clear either.

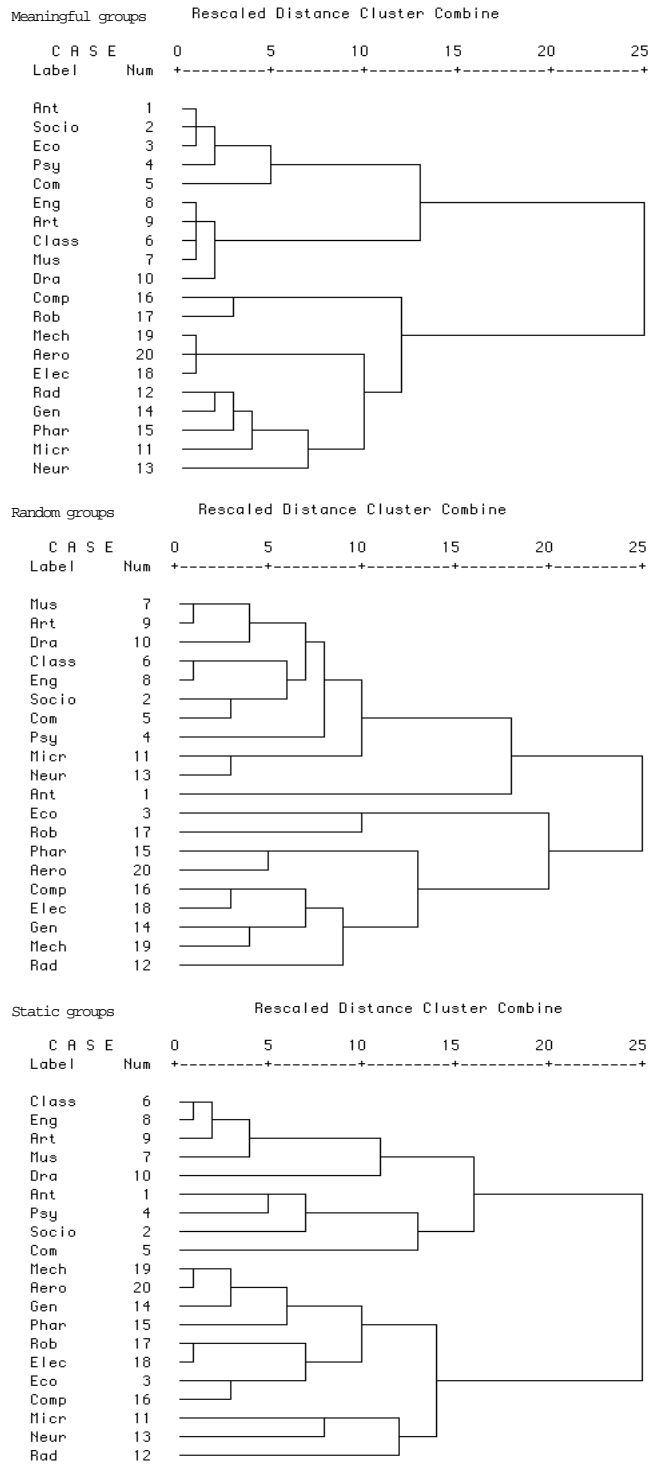


Figure 3. Trees resulting from the hierarchical analysis of drawing order in located recall for a) meaningful condition; b) random condition; c) static condition.

Located recall. In the meaningful condition, participants' recall follow the learning order, except that one of the clusters (Engineering) was split in two, presumably because of spatial constraints (Figure 3). In the random conditions, two spatial clusters were distinguishable (the first cluster corresponded to the three first columns, and the remaining cluster corresponded to the fourth and fifth column). But the aggregation level was very high, suggesting that the clustering was noisy. The clustering in the static conditions seemed to result from a tradeoff

between meaningful and spatial criteria. For example, the Social Sciences items were clustered together, except that “economics” was clustered with its right-hand side neighbor, “Computers”.

2.3. DISCUSSION

The utility of sequentially highlighting meaningful portions of graphic was for the most part supported by the data. Participants in the sequential meaningful condition required fewer trials to learn the diagram than participants in the static and random conditions, but only the difference between the meaningful and random conditions was significant. This advantage was corroborated by detailed trial by trial analysis. Thus, a meaningful sequential display improves memory compared with a random sequential display, but an organized sequential display was not unequivocally superior to static display.

A meaningful sequential display also affected the order and clustering of items in located recall, in contrast to the random display which did not. Thus, a meaningful sequential display acts as a recall guide, but a random organization does not. In the static condition, the clustering seemed to be a compromise between meaningful and spatial properties.

3. Experiment 2: Sequential display of a large library

In the first experiment, the diagram was simple and familiar. The advantage of a sequential display may be more apparent for a larger memory load. The second experiment examined a more complex configuration, using thirty-six less familiar disciplines instead of twenty. The second experiment also added a spatial memory test to equate participants’ spatial memory across conditions. Finally, as the layout of the items made no difference in the first study, only one layout was used in the second experiment.

3.1. METHODS

Population

48 Stanford University undergraduate and graduate students from various majors, 19 of which were female, participated in the experiment for course credits.

Material

Figure 4 displays the enlarged library map. The 36 bookshelves were grouped as: general works (monographs, encyclopedia, dictionaries, periodicals, indexes, almanacs), History (civilization, genealogy, heraldry, archeology, chronology, diplomatic), Geography and Geology (ecology, atlases, cartography, paleontology, oceanography, petrology), Technology (aeronautics, buildings, optics, electronics, hydraulics, manufactures), Philosophy and religion (metaphysics, cosmology, ontology, mythology, ethics, aesthetics) and Mathematics (geometry, logic, arithmetic, statistics, algebra, probabilities). As before, there were three display conditions: static, random sequential and meaningful sequential.

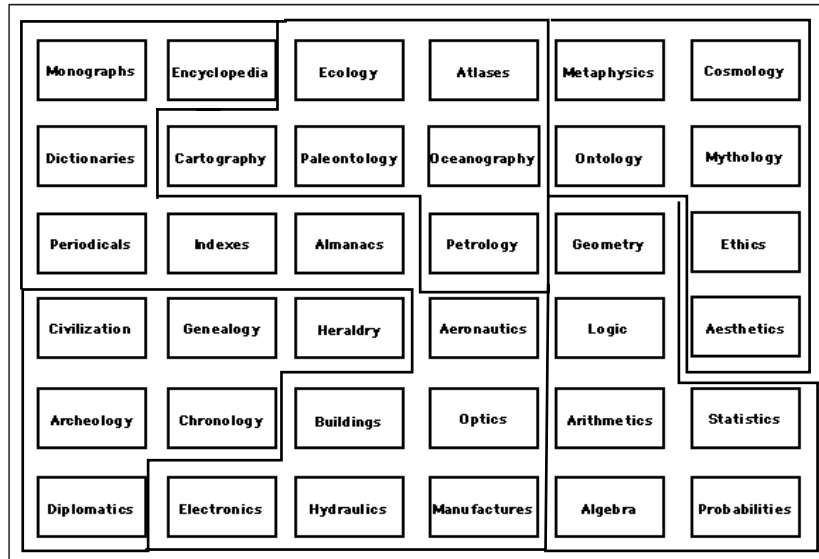


Figure 4 - Library map used as material in Experiment 2

Procedure

The procedure was the same as the first experiment except for the following changes.

First, participants performed a spatial memory test before the experiment began. This test was adapted from the Kelley's Map Memory test I (Kelley, 1964). Participants studied the map of a town containing 25 landmarks for 2 minutes. Then they positioned as many landmarks as possible on a sheet where the roads and the railroad were provided. Participants performed a jigsaw puzzle on the computer while the experimenter scored their performance. Three score groups were determined using a unit normal distribution on the first 15 participants' scores. The test was scored according to Kelley's instructions (2 points for each item correctly recalled and located, 1 point for each item recalled in another location). Participants were then assigned to one of the three display conditions in such a way that the number of participants in each score group was equal across conditions. Study time was increased to 62 seconds (2 seconds for the blank frame and then 60 seconds), equal across conditions. The instructions were changed accordingly.

3.2. RESULTS

Spatial memory test

In order to assess the validity of the spatial test, a correlation coefficient was computed between the score on the spatial memory test and the last trial performance for the recall of the library . The correlation between the spatial memory test and free recall performance was .50 ($p < .0005$) and the correlation between the spatial test and located recall was .54 ($p < .0001$). These high correlations show that the spatial memory test was a good indicator of spatial memory ability for this kind of memory task. Thus, the three groups (function, random, static) can be considered as equivalent with respect to spatial memory ability.

Recall performance

A repeated measures analysis of variance using the type of display as a between participants factor, and the type of recall (free and located) and number of trials (1 to 5) as within-subject factors yielded a main effect of the type of recall. Unsurprisingly, participants were better at free recall than recalling at located recall ($F(1,45) = 69.14$, $p < .0001$). Recall also increased with number of trials ($F(3,45) = 436.08$, $p < .0001$). There was no effect of display on the number of trials to learn. Only few participants (5.6%) recalled all items accurately within the four trials, of which none from the random condition. Figure 5 displays the mean number of items accurately recalled through the four trials.

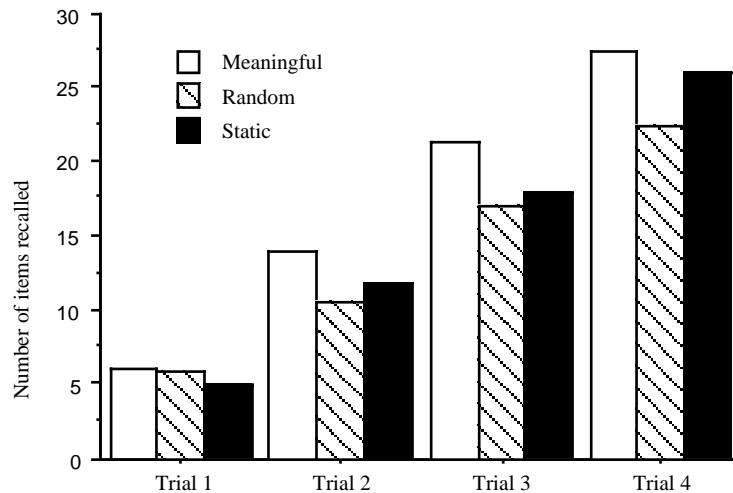


Figure 5. Number of correct items in the located recall test through the four trials.

A repeated measures analysis of variance using the type of display as the between-subjects factor, and the number of trials (2 to 4) as the within-subjects factor was computed on the number of items correctly recalled and located (Figure 5). The first trial was because of a floor effect. The effect of the display factor was not significant overall ($F(2,45) = 2.35$, NS). However, planned comparison yielded a significant difference between the meaningful and the random conditions ($F(1, 30) = 4.67$, $p < .05$).

Because of the increased memory load, recall was lower, too low to allow analyses of recall order.

3.3. DISCUSSION

This experiment replicated many of the effects of the first experiment. In this experiment, using a more complex graphic and equating groups by a spatial memory test. There was a significant correlation between performance at the spatial memory test and performance on free and located recall.

In free recall, there were no differences between conditions. However, in located recall, participants in the meaningful sequential display condition did significantly better than participants in the random sequential condition. Since most participants did not recall all the names, the ordering and clustering of names in recall could not be analyzed in this experiment.

Despite a larger, more complex display, performance in sequential meaningful conditions was not superior to performance in the static condition. However, the meaningful sequential condition was superior to random sequential condition.

4. Experiment 3: Sequential display of a geographic map

The recall order in the first experiment strongly suggested that meaningful sequential presentation affected the way the graphic material was mentally represented. On the present experiment, we evaluate (it?) more directly using a richer graphic, a map, and two uncorrelated meaningful organizations of the towns, historic or size, across participants.

4.1. METHODS

Participants

59 undergraduates participated in the experiment for course credits. The data of one subject were eliminated because of poor recall of the map (7 elements out of 27). There were 15 participants in each sequential condition and 14 in each static condition, randomly assigned.

Material

The material was the map of a fictitious island with 9 towns (see Figure 6) characterized by size (Small, Medium and Large) and the era in which it had been established (Ancient World, Pioneer Age and Modern World). The two dimensions were ordered and crossed in a matrix organization, so that each town realized one of the nine factorial combinations.

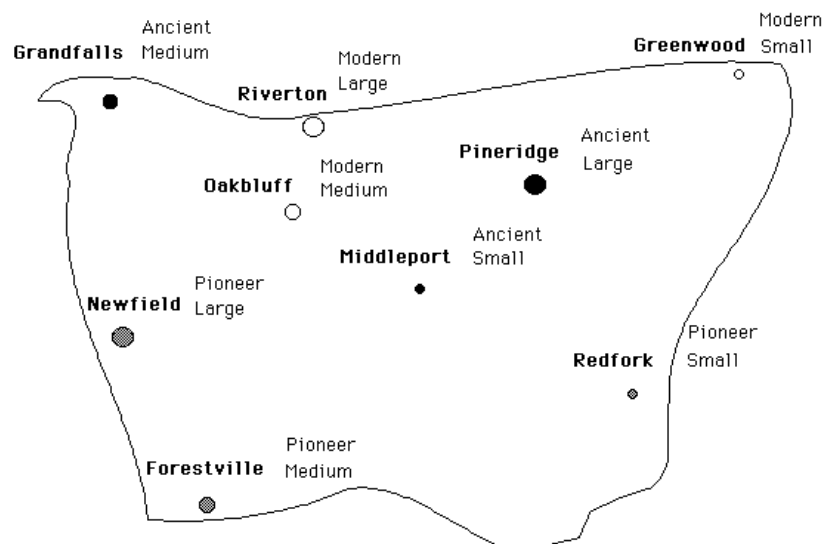


Figure 6. Geographic map used as material in Experiment 3

The towns were located on the map so that the total distance between the three towns of a cluster was equal across clusters for both dimensions. Thus, the spatial organization was not correlated with either of the matrix features. Each town was indicated by a dot that conveyed the size and era dimension in a graphic way (dot size for size and gray shades for the era).

Four types of display were designed. In the *Sequential Era display*, the towns were clustered according to era, and appeared in steps from ancient to modern. In the *Sequential Size display*, the towns were clustered according to size, and appeared in steps from small to large. Finally, in the *Static Era* and *Static Size* displays, all information was available at once.

In both sequential displays, the clusters appeared in steps with the name and related information of the towns within a cluster appearing simultaneously. Once displayed, the information remained on the screen until all clusters were displayed. The three steps of the sequential display ran twice a session. The spatial configuration was the same across displays except for the relative spatial location of the era and size information. In both sequential displays, the information in the highlighted dimension was spatially displayed above the other information for every town. Thus, two static displays were designed as controls, according to the relative spatial location of size and era information—size above era (Static Size) or era above size (Static Era).

Procedure

Study phase. Participants were presented with the material and their tasks. They were told to study the map carefully “in order to be able to answer questions related to each town—name, location, size or era”. Study time was four minutes in all conditions.

Recall of the map information. After the map disappeared, participants were provided with a blank map with dots indicating the location of towns (the dots were all of the same size and color). They were asked to indicate in the appropriate places the names of the town, their size and era. The order in which participants recalled each piece of information was recorded by the experimenter.

Verification test. In the verification test, two towns were displayed and the participants indicated if the towns were either of the same size or of the same era. The dimension according to which towns were compared (size or era) was called *focus* of the test. They were asked to answer as quickly and as accurately as possible. The towns were indicated either verbally or visually. In the verbal test, the names of the towns appeared vertically aligned and horizontally centered on the screen. In the visual test visual, dots indicating the location of the towns were displayed on the outline of the map, but the names of the towns were not provided. Participants performed both the verbal and visual tests successively in counterbalanced order across participants.

Twenty-seven pairs of towns were chosen as stimuli: 9 era pairs (towns of the same era), 9 size pairs (towns of the same size) and 9 both-features pairs (different sizes and different eras). The 9 both-features pairs were chosen among the 18 possible pairs, in such a way that the total distance between the towns in these 9 pairs was equal to the distance between towns of the size pairs and of the era pairs (distance equal across size and era pairs by construction of the material). The 27 pairs appeared twice, once for each focus. For the verbal test, the two names in each pair were presented in reverse location for each focus (e.g. “Pineridge Redfork” for the size focus and “Redfork Pineridge” for the era focus). The 54 resulting trials were blocked in series of same focus (size or era) with 9 trials each, in randomized order across participants. The same 54 trials were used for the verbal and visual test.

A verbal cue was displayed on the screen to indicate the focus (era or size), followed by the names of two towns for the verbal test and by the map with two dots for the visual test. The cue appeared for 5 seconds before each block of trials and for 2 seconds between each trial. The test items remained on the screen until the participants

responded. Participants were told to press the “same” key if they thought the towns belong to the same category, and “different” otherwise. There was no time limit.

Description of the map. The map of the island with the names of the town was provided to the participants. The era and size information were not provided, neither verbally, nor graphically, since the dots were all of the same size and color. Participants were asked to write a description of this map like one they could find in a handbook of geography.

4.2. RESULTS

Recall of the map

Three types of information are distinguished: era, size and name (of the towns). The accuracy of a piece of information is related to its location, irrespective of the accuracy of the two other pieces of information corresponding to each town. The mean number of elements correctly recalled is displayed in Figure 7.

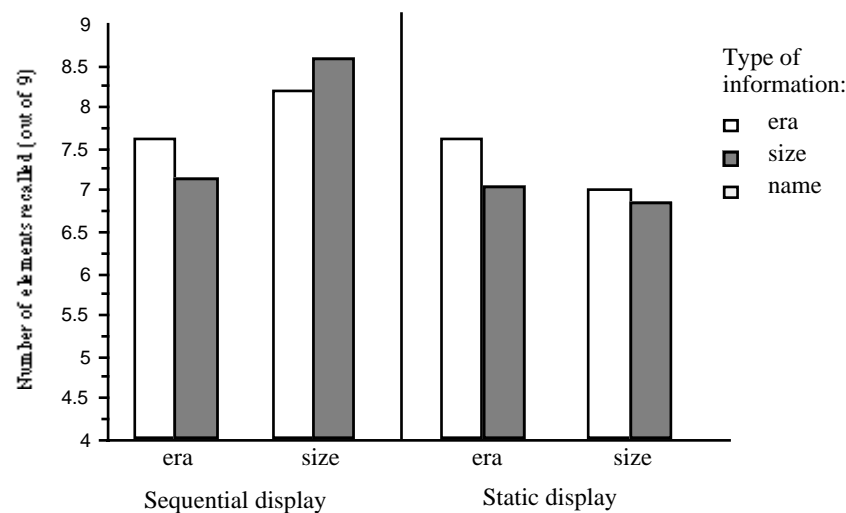


Figure 7. Mean recall of the three types of information as a function of display focus

A repeated measures analysis of variance using the type of display as a between-subjects factor and the type of information recalled as the within-subject factor was computed on the number of elements recalled. Overall, there was a significant main effect of the information type ($F(2,54) = 20.9$, $p < .0001$). Participants recalled more size and era information (respectively 7.3 and 7.5 in average, out of 9) than names of the town (5.8). The correlation between the recall performance for size and era information was significant ($r = .615$, $p < .0001$), whereas the correlation between the name information and other types of information was not (name vs. era: $r = .252$, NS; name vs. size: $r = .076$, NS). As recall of the names did not correlate with recall of other information, we computed a second analysis of variance using the same factors but without the name information.

The effect of the type of display was not significant ($F(3,54) = 2.83$, NS), nor was the effect of the type of information ($F(1,54) = 2.75$, NS), nor was the interaction between type of display and type of information ($F(3,54) = 2.35$, NS).

The main question of interest was whether the organization of the display affected the mental organization of the material. To determine this, a repeated measures analysis of variance was computed with the two sequential

conditions as the between-subjects factor and the size and era information recalled as the within-subject factor. Supporting the hypothesis, this yielded a significant interaction between the condition and the type of information ($F(1,28)=9.03, p <.01$).

Ordering of information in recall

We expected that participants in the sequential conditions would recall successively (or cluster) towns information according to the dimension that was highlighted by the sequential display. The analysis includes only the size and era information that is accurately located.

Kendall’s ranks concordance coefficient was computed for each condition to evaluate the similarity across recall orders within each condition (Table 4).

Table 4 - Kendall’s ranks concordance coefficient ω – Recall order

| | Conditions | | | |
|--------------------|-----------------|-------------|-----------------|-------------------|
| | Static era | Static size | Sequential era | Sequential size |
| Kendall’s ω | .39 | .069 | .40 | .50 |
| Significance | p<.05 | p=.99 | p<.01 | p<.0001 |

As shown in Table 4, the similarity between ordering was significant for all conditions except the static size condition. In other words, participants in both static et sequential era and in sequential size conditions recalled items in similar orders. In order to characterize recall order, we computed a rank correlation coefficient (Kendall’s tau) between the mean recall order in each condition and the two display orders. Only the three conditions where the similarity between subjects’ orderings was significant were considered, since the mean ordering was meaningless otherwise.

There was a significant correlation between the mean recall order in the Sequential Era condition and the era display order ($t = .61, p <.0005$). Similarly, there was a significant correlation between the mean recall order in the Sequential Size condition and the size display order ($t = .82, p <.0001$). In contrast, there was no correlation between the mean recall order in the static era conditions and either of the display orders.

Clustering of information in recall

As for Experiment 1, we carried out a hierarchical analysis on the recall order data across all participants within each condition. The trees resulting from the hierarchical analysis of recall orders in sequential conditions are provided in Figure 8. The same level of cluster distinction (aggregation level) was chosen for all condition.

In the Sequential Era condition, the information related to towns of the same era were clustered together. In the Sequential Size condition, the information was clustered according to the size dimension, except that the information related to a large town (Pineridge) was clustered together with information related to medium-size towns. For the static conditions, clusters were barely distinguishable at this aggregation level and did not match the organization by size or era. Thus, for both sequential conditions, recall order reflected sequential organizations.

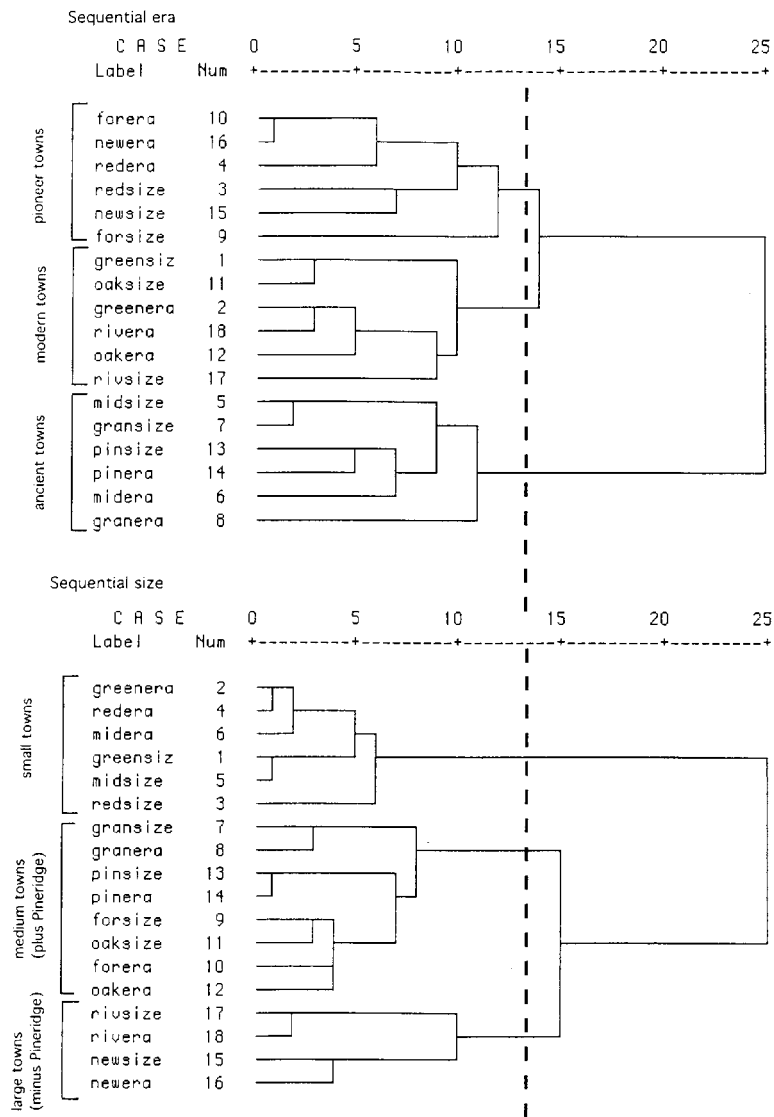


Figure 8. Tree resulting from the hierarchical classification analysis

Description organization

Further evidence that memory organization reflected input organization for only the sequential conditions comes from participants' descriptions of the map. We expected that participants in the sequential conditions would be more likely to use the organization induced by the sequential display (*i.e.* to cluster the towns in the description according to the highlighted dimension) over any other organization.

Four organizations were distinguished: *era*, *size*, *spatial* (clustering by spatial properties, for example, North to South and West to East, etc.) and *indeterminate*. Table 5 displays the proportion of descriptions in each condition falling into each of the four categories.

Table 5: Proportion of descriptions falling into each organization type

| | Static era | Static size | Sequential era | Sequential size | Total |
|---------|------------|-------------|----------------|-----------------|-------|
| By era | 28.6 | 64.3 | 73.3 | 33.3 | 50% |
| By size | 7.1 | 7.1 | - | 53.3 | 17.2% |

| | | | | | |
|----------------|------|------|------|------|-------|
| Spatial | 35.7 | 14.3 | 13.3 | - | 15.5% |
| Indeterminate. | 28.6 | 14.3 | 13.3 | 13.3 | 17.2% |

As seen in Table 5, organization by era was the most frequently used, enhanced by the Sequential Era display. Conversely, organization by size appeared rarely, except in the Sequential Size condition. Even more than participants in the static era condition, participants in the static size condition organized by era. This last result was corroborated by the hierarchical analysis of drawing order, which showed that participants in the Static Size condition clustered items according to the era dimension.

A Chi square analysis on the number of participants using each organization showed a significant difference overall ($\chi^2(9)= 27.35, p <.005$). Whereas the two sequential conditions differed ($\chi^2(3)= 12.25, p <.01$), the static conditions did not ($\chi^2(3)= 3.88, NS$). Participants in the Sequential Size condition organized by size significantly more often than participants in the Sequential Era condition ($\chi^2_{corr}= 3.84, p <.05$), but the two sequential conditions did not differ in organization by era ($\chi^2_{corr}= 3.35, NS$).

Visual and verbal verification tests

In the verification test participants compared the towns (same or different) on the basis of size or era, the focus of the test. Participants in the sequential conditions should be faster and more accurate when the test focus matched the dimension highlighted by the display (size or era).

Accuracy. A repeated measures analysis of variance was conducted on errors using the between-subjects factor of type of display and three within-subjects factors: the type of test (visual or verbal), the type of focus (era or size) and the type of pairs (both-features, era or size pairs). There was no main effect of conditions, nor focus type, nor pairs type. There was a main effect of type of test ($F(1,54)=7.36, p <.01$), with more errors in the verbal than in the visual test (31% errors for the verbal test and 26% for the visual test). This factor did not interact with any other factors, so the data were collapsed across tests. Figure 9 displays the mean number of errors in each condition by test focus (era or size).

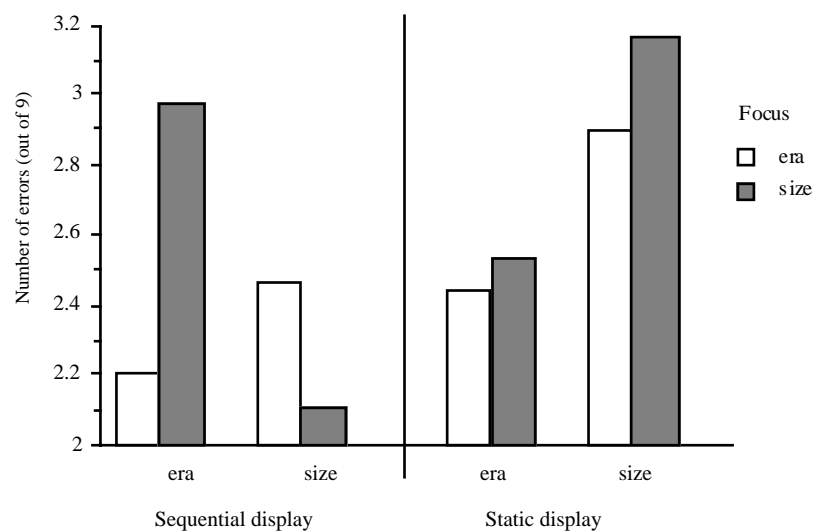


Figure 9. Mean number of errors to the verbal and visual verification tests as a function of the test focus (era or size).

As shown in Figure 9, error rates in the sequential conditions depended on the focus. The interaction between the conditions and the focus was significant ($F(3,54)=3.26, p<.05$). Further analysis showed that this interaction was significant for the sequential conditions ($F(1,28)=11.01, p<.01$) but not for the static conditions ($F(1,28)=.19, NS$). Participants in the Sequential Era condition made fewer errors with an era focus than with a size focus ($F(1,14)=13.48, p<.005$). Conversely, the Sequential Size condition had fewer errors with a size focus than with an era focus, but the difference was not statistically significant ($F(1,14)=1.84, NS$).

Were participants in the sequential conditions more accurate than the participants in the static conditions? Overall, the contrast between sequential and static conditions was not significant ($F(1,56)=1.01, NS$). However, as stated above, the interaction between the conditions and the focus was significant, thus the analysis was conducted separately on each focus. Participants in the Sequential Size condition performed significantly better than participants in Static Size condition when the focus was size ($F(1, 27)=4.47, p<.05$) but not when the focus was era ($F(1, 27)=.81, NS$). However, the difference between the Sequential Era condition and the Static Era condition were not significant either for era ($F(1, 27)=.26, NS$), or for size ($F(1, 27)=.78, NS$).

Participants in the Static Size condition made more errors than did the Static Era condition, but the difference was not significant ($F(1,27)=1.41, NS$). Similarly, there was no difference overall between the Sequential Era and the Sequential Size conditions ($F(1,28)=.46, NS$).

Verification times. As for accuracy, a repeated measures analysis of variance was conducted on the times using the between-subjects factor of type of display and the three within-subject factors: type of test (visual or verbal), type of focus (era or size) and type of pairs. Only correct responses were included in the analysis, but the complete analysis gave the same results.

There was a main effect of the type of the test ($F(1,54)=38.1, p<.0001$). The verbal test took longer than the visual test (mean verbal = 4.7 s, and mean visual = 3.5 s). There were no main effects of display condition, focus type, or pairs type. Data were collapsed across the two tests (visual and verbal), since this factor did not interact with any other factors. Figure 10 displays the mean response times as a function of the focus of the test.

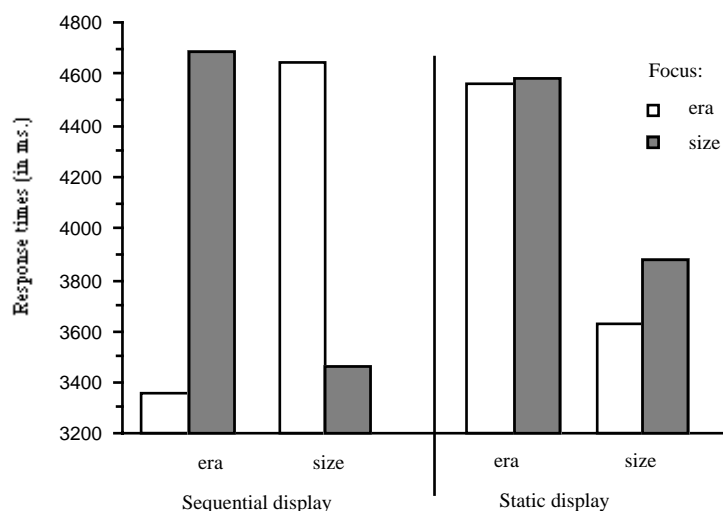


Figure 10. Response times to the verbal and visual verification test as a function of the test focus (era or size).

Participants' response times in sequential conditions varied as a function of the focus, whereas participants' response times in static conditions did not. The interaction between focus and condition was significant for the sequential conditions ($F(1,28)=30.23, p < .0001$) but not for the static conditions ($F(1,26)=.33, NS$). Participants in the Sequential Era condition were significantly faster for the era focus than for the size focus ($F(1,14)=22.59, p < .0005$). Conversely, participants in the Sequential Size condition were significantly faster for the size focus than for the era focus ($F(1,14)=10.73, p < .01$).

Overall, participants in sequential conditions were faster than participants in static conditions. The contrast analysis between sequential and static conditions failed to reach the statistical significance ($F(1,56)=1.28, NS$), but the interaction between condition and focus was significant ($F(1,56)=13.7, p < .001$). Therefore, the analysis was computed for each focus separately. As evidenced from Figure 10, participants in the Sequential Era condition were faster than the participants in the Static Era condition when the test focus was era ($F(1, 27)=4.71, p < .05$), but not when the focus was size ($F(1, 27)=.042, NS$). However, the difference between the Sequential Size and the Static Size conditions was not significant, either for era ($F(1, 27)=2.91, NS$), or for size ($F(1, 27)=.76, NS$).

4.3. DISCUSSION

The results of this experiment support the hypothesis that sequential display of diagram guides the processing of diagrams and affects the organization of them.

Sequential display and organization of the mental representation

First, participants in the sequential conditions recalled the information from the map using the ordering and clustering in which it had been displayed. Furthermore, descriptions of the map matched the input organization for both sequential conditions. In the judgment test, in which participants compared pairs of towns, either on size, or on era. The criterion according to which towns were compared was called focus of the test, participants in the sequential condition were faster and more accurate (significant only for Sequential Era) when the judgement was the dimension highlighted by the sequential display. None of these effects was obtained for the static conditions.

Static vs. sequential display for memory performance

Results provide partial support for an overall advantage of sequential display to static display. For presentation by size, the organization less preferred spontaneously, sequential display led to better performance than the static display. This suggests that sequential display may be especially important when there is no obvious way of organizing a diagram. Similar findings were obtained by Bétrancourt and Bisseret (1995). When an organization schema is obvious, presentation by parts may interfere with participants' own organization.

5. General discussion

Three experiments investigated the effects of a simple animation on memory and information retrieval from a graphic display. The device, sequential display, entails successively displaying parts of a spatial configuration

where the parts are either meaningful or random clusters. Like language, sequential display systematically directs users' attention to parts of a complex graphic, thereby organizing the graphic for the users.

In the first two experiments, for the sequential display, memory for the diagram was better when the organization was meaningful than when the organization was random. In the third experiment, where there were two meaningful organizations, participants in the sequential conditions were better at remembering information on the dimension of organization. In all experiments, participants in the sequential conditions recalled the elements of the configuration in the organization in which the configuration was displayed for a meaningful organization but not for a random organization. Finally, in the third experiment, participants in both sequential conditions were faster and more accurate making judgments on the basis of sequential presentation than on the other basis. Together these findings suggest that displaying parts of a configuration successively on a computer screen has a substantial effect on the organization of users' mental representation of the graphic display.

Despite the unequivocal effects of sequential display on memory organization, sequential display did not lead to overall improvement in memory relative to static display. Recall that for the static condition, all of the graphic was available to users for the entire presentation time, allowing participants to impose their own organization on the graphic. Previous research has sometimes shown that presentation by parts has a deleterious effect on recall, presumably because it disrupts people's natural organization (e.g. Tulving, 1967). Some support for this account comes from the third experiment where sequential display was in fact superior to static display for the subtle organization scheme. However the diagrams used here were relatively simple. That suggests that sequential display should be especially effective for complex graphics whose structure is not immediately apparent. Graphic displays on computers, in manuals and in text books are becoming more and more complex. The present studies have shown that computers can facilitate users' processing of them using a simple animation device borrowed from language: Organizing and integrating relevant parts of the graphic by highlighting the parts that belong together.

Acknowledgements

This research was supported by a grant from the Institut National de Recherche en Informatique et en Automatique to the first author. We are grateful to André Bisseret and Anne Pellegrin, as well as the two anonymous reviewers, for their helpful comments and suggestions on this paper.

References

- Bétrancourt, M., Bisseret, A. & Faure, A. (in press). Sequential display of pictures and its effect on mental representations. in J.-F. Rouet, J. J. Levonen and A. Biarreau (Eds.), *Multimedia learning: cognitive and instructional issues*. EARLI Series "Advances in Learning and Instruction".
- Bétrancourt, M., & Bisseret A. (1995). Influence de la présentation séquentielle d'une figure sur sa mémorisation. [Influence of sequential display of pictures on memory] *Proceedings of the seventh conference on human computer interaction engineering, IHM'95* (pp 135-141), Toulouse : Cepadues.
- Bower, G.H., & Clark-Meyers, G. (1980). Memory for scripts with organized vs. randomized presentations. *British Journal of Psychology*, 71, 369-377.
- Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Denis, M. (1984). Imagery and prose : a critical review of research on adults and children. *Text*, 4 (4), 381-401.

- Denis, M. (1996). Imagery and the description of spatial configurations. In M. De Vega, M.J. Intons-Peterson, P.N. Johnson-Laird, M. Denis & M. Marschark, *Models of visuo-spatial cognition* (pp 128 - 197). New York: Oxford University press.
- Denis, M., et Denhière, G. (1990), Comprehension and recall of spacial descriptions, in *Cahier de Psychologie Cognitive, European Bulletin of Cognitive Psychology*, 10 (2), 115-143.
- Egan, D.E., & Schwartz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, 7, 149-158.
- Filippatou, D., & Pumfrey, P.D. (1996). Pictures, Titles, Reading accuracy and reading comprehension: a research review (1973-1995). *Educational Research*, Vol. 38 (3), 259-291.
- Gonzales C. (1996). Does animation in user interfaces improve decision making ? in *Proceedings of CHI'96* (Vancouver, April 1996), ACM Press, 27-34.
- Gyselinck V., & Tardieu, H. (1994) The role of text illustration in the construction of non-spatial mental models. in F. De Jong et B. van Hout-Wolters (Eds) *Process oriented Instruction, Verbal and Pictorial aids & Comprehension Strategies*, Amsterdam : VU University Press.
- Hegarty, M. & Just, M.A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717-742.
- Hunter, B., Crismore, A., & Pearson, P.D. (1987). Visual displays in basal readers and social studies textbooks. in D.M. Willow and H.A. Houghton (Eds.) *The Psychology of Illustration : Basic research*, Vol. 1, Springer : New-york, Berlin, Heidelberg, London, Paris, Tockyo, pp. 85-95.
- Kelley, H.P. (1964). *Memory abilities: a factor analysis*. New York: Psychometric Society, 1964.
- Levelt, J.M.W. (1989). *Speaking : from intention to articulation*. Cambridge, M.A. : the MIT Press.
- Levie, W.H., & Lentz, R. (1982). Effects of text illustration : a review of research. *Educational Communication and Technology Journal*, 30 (4), 195-232.
- Levin, J. R., Anglin G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose, in D.M. Willows & H.A.Houghton (Eds), *The Psychology of Illustration : I. Basic Research*. New York : Springer, pp 116-135.
- Lowe, R. (1996). Interactive animated diagrams: what information is extracted? in the *Proceedings of the First International Conference on Using Complex Information Systems*, UCIS'96, September 4-6, Poitiers, France.
- Mani, K., & Johnson-Laird, P.N. (1982) The mental representation of spatial descriptions. *Memory & Cognition.*, 10 (2), 181-187.
- Mayer, R. E. (1989). Models for understanding. *Review of Educational Research*, 59 (1) , 43-64.
- Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words ? *Journal of Educational Psychology*, Vol. 82, N°4, pp 715-726.
- Morrison, J. B., Tversky, B., & Bétrancourt, M. (2000) Animation: Does it facilitate learning? *Proceedings of the AAAI 2000 Spring Symposium « Smart Graphics »* (pp 53-60), March 20th-22nd 2000, Stanford, CA, USA.
- Palmiter, S., & Elkerton, J. (1993). Animated demonstrations for learning procedural computer-based tasks, *Human-Computer Interaction*, 8, 193-216.
- Peek, J. (1993). Increasing picture effects in learning from illustrated text, *Learning and Instruction*, 3 , 227-238.
- Rieber, L.P. (1990). Animation in computer-based instruction. *Education Technology, Research and Development*, 38 (1), 18 (3),77-86
- Scaife, M., & Rogers, Y. (1996) - *External cognition : how do graphical representations work ?* *International Journal of Human Computer Studies* (1996) 45, 185-213.
- Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions, *Journal of Memory and Language*, 31, 261-292, 92a.

- Tulving, E., & Osler, S. (1967). Transfer effects in whole/part free-recall learning. *Canadian Journal Of Psychology*, 21(3) 253-262.
- Tversky, B. (in press). Spatial schemas in depictions. in M. Gattis (Ed.) *Spatial schemas and abstract thought*.. Cambridge: MIT Press.
- Winn, W. (1989). The role of graphics in Training documents : toward an explanatory theory of how they communicate. *IEEE Transactions on Professional Communication*, 32 (4), December 1989.
- Wright, P., Hull, A., & Black, D. (1990). Integrating diagrams and text. *The Technical Writing Teacher*, XVII, 244-254.