The Mental Organisation of External Representations

Richard Cox (richc@cogs.susx.ac.uk)
School of Cognitive & Computing Sciences; University of Sussex
Brighton, BN1 9QH, UK

Beate Grawemeyer (beateg@cogs.susx.ac.uk)
School of Cognitive & Computing Sciences; University of Sussex
Brighton, BN1 9QH, UK

Abstract

This paper examines the ways in which people organise their knowledge of external representations (ERs). These include diagrams, graphs, charts, textual forms such as lists and notations, maps, tables, pictures *etc*. We focus upon the kinds of ERs that are used in reasoning and problem solving. Twenty-eight subjects were given an untimed card-sorting task. Cluster analysis of the card-sort data revealed that the cards were classified into 9 major categories of ER - maps, set diagrams, logic/math notations, tables & graphs, lists, music, pictures of objects and scientific diagrams. The card sorts of two sub-groups of subjects were compared. The first group was selected on the basis of skilled (upper quartile) performance on ER reasoning tasks and the second was a group of subjects who performed in the lower quartile on ER reasoning tasks. The card-sort behaviour of the better performing subjects differed from that of their poorer performing peers in that they had fewer but more distinct categories. Their ER naming accuracy was also significantly better. The upper-quartile subjects tended to use structural characteristics and ER semantics as a basis for classifying representations to a greater extent than their lower-quartile peers. The results are discussed in relation to theories of category formation and the graphic-linguistic distinction.

Introduction

This paper examines the mental organisation of external representations (ERs). Here, ERs are defined as representations that are useful for problem solving and reasoning and include diagrams, graphs, charts, textual forms such as lists and notations, maps, tables, *etc*.

Several questions are addressed. First, what categories do people typically use in structuring their ER knowledge? How does the structure compare with studies that have used pictures of objects and written names of objects and categories as stimuli (*e.g.* Rosch, 1975, 1978)? Categories vary in terms of distinctness, the degree to which they feature typicality gradients that are anchored by stereotypes and in terms of their granularity. Granularity refers to the degree of subordinacy versus superordinacy of the category (*ie.* its degree of specificity/generality). Rosch and her associates identified a 'basic level' - *e.g.* 'chair' rather than the more general category 'furniture' or a highly specific one such as 'armchair'. Basic-level categories were found to be the ones that adults spontaneously use to name objects and the ones that are recognised quicker. They are also associated with early ages of acquisition in children (Eysenck & Keane, 2001; Pinker, 1998). Here we seek to find out whether ER knowledge has an analogous 'basic level'.

A second question concerns the consistency of ER knowledge structures. How stable are ER categories? Do they reflect expertise levels (skill at using ERs to problem solve and reason)? How does the ER knowledge organisation of skilled and less skilled subjects differ? These are sub-topics of our main aim which, to recapitulate, is to explore the nature of the mental organisation of knowledge of ERs.

Method

Overview of General Approach This study involved subjects pooled from two separate research studies. In one of the studies (study 1), 16 subjects solved analytical reasoning problems using a computer-based ER support environment. In a second study, a different group of 12 subjects chose computer-generated data visualizations to use in answering database questions. Further details of these studies are provided below. Subjects performing in the lower (LQ) and upper quartiles (UQ) were identified separately in each study. This was done on the basis of analytical reasoning performance scores (in the case of study 1 subjects) or on the basis of the database question performance (study 2 subjects). The LQ and UQ subjects from each study were pooled to form single LQ and UQ groups, each with 7 subjects (4 LQ and 4 UQ subjects from study 1, and 3 LQ and 3 UQ subjects from study 2). The gender balance in the LQ and UQ subject groups was similar.

ER Card-Sort Task Subjects' semantic knowledge of a wide range of ERs was assessed by means of a card-sort task. Card sorts are used in neuropsychological assessment (e.g. Wisconsin Card Sorting Test) and as a technique for eliciting and structuring experts' knowledge in the knowledge engineering field (*e.g.* Schreiber et al., 1999). Eighty-seven ER stimuli were sourced from a wide variety of texts in the diagrammatic reasoning and related literatures, such as ones on graphics, physics textbooks, fragments of computer programs, formulae, instructions, charts, plans, schematic sketches, maps, tables, music, childrens' drawings, set diagrams (Euler & Venn), illustrations, a cartoon picture sequence, X-Y graphs, logic, directed graphs, entity-relation graphs, network diagrams, maps, tables, tree diagrams (*e.g.* hierarchies, decision trees), column graphs, bar charts and circuit diagrams. A sample of 12 of the 87 items is shown in Figure 1[1].

Each representation was mounted on 8 × 5" white index cards which were numbered on the reverse. Each card showed an example of one type of representation. The 87 numbered cards were shuffled for each subject and given to the subject with a pen and a pad of 'post-it' notes.

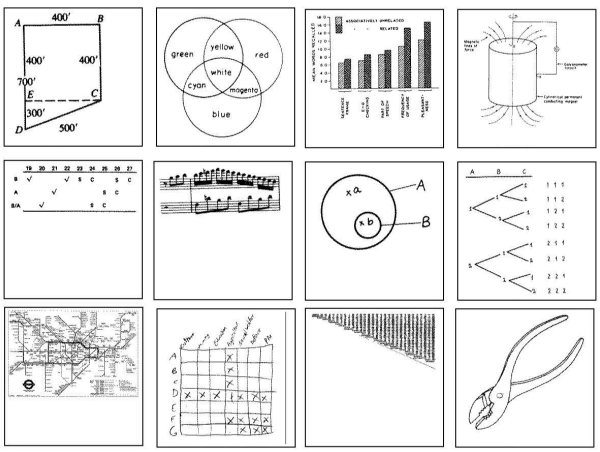

Figure 1: Twelve of the 87 card sort stimuli

Subjects were instructed *"Here is a stack of representations that are used in a variety of problem solving tasks. I would like you to sort them into heaps. You may decide on what kinds of categories to use and how many categories to use. I would like you to label your categories when you have finished."*

The task was untimed. The number of categories, and the numbers of the cards in each category were recorded for each subject.

ER Category Naming Task Subjects named each pile of cards in their card sort with a label written on a 'post-it' note. The card stack was shuffled thoroughly between subjects.

ER Reasoning Study 1: Analytical Reasoning The subjects were undergraduate and postgraduate students at the University of Edinburgh and Edinburgh College of Art. There were 9 female participants and 7 male participants. They were administered the ER card-sort as a pre-test. Participants then used a computer environment (*switchERI*) to solve a series of analytical reasoning problems. The *switchERI* system provided a range of simple computer-based support tools for the selection and construction of diagrams, logical representations, textual epresentations and tables (Cox & Brna, 1995; Cox, 1999).

Analytical reasoning problems generally involve constraint satisfaction solution strategies based on an understanding of the relationships between fictitious things, events, places or persons described in a narrative passage or problem 'stem'. An example of an analytical reasoning puzzle is provided in Figure 2. Subjects were given a practice question which the experimenter used to illustrate *switchERI's* features. They then solved the practice problem, taking as much time as they wished. When they were ready, the subjects attempted three experimental problems. No time limit was imposed.

An office manager must assign offices to six staff members. The available offices are numbered 1–6 and are arranged in a row, separated by six foot high dividers. Therefore sounds and smoke readily pass from one to others on either side. Ms Braun's work requires her to speak on the phone throughout the day. Mr White and Mr Black often talk to one another in their work and prefer to be adjacent. Ms Green, the senior employee, is entitled to Office 5, which has the largest window. Mr Parker needs silence in the adjacent offices. Mr Allen, Mr White, and Mr Parker all smoke. Ms Green is allergic to tobacco smoke and must have non-smokers adjacent. All employees maintain silence in their offices unless stated otherwise.

1. The best office for Mr White is in 1, 2, 3, 4, or 6?

Figure 2: Example of an analytical reasoning problem.

ER Reasoning Study 2: Information Display Selection Task AIVE (Grawemeyer & Cox, submitted) is a prototype intelligent information visualisation system which contains a structured database of information about types of car (fuel efficiency, CO^2 emissions, number of doors, *etc*) and which generates a range of alternative displays (bar charts, scatter plots, 'rose' (sector) graphs, tables, and text (lists) in response to user queries. The subjects consisted of 5 female and 7 male students at the University of Sussex. They were administered the ER card-sort as a pre-test. Each subject responded to 25 multiple choice questions generated from the database. There were 8 types of question. For example, some questions asked the subject to identify a particular entity ('Which car has a standard driver and passenger airbag and an optional side airbag?'), whereas others required subjects to rank items according to some criterion (*e.g.* CO^2 emission). Subjects were allowed to choose the information display they wished to use for each question. Each problem could be answered with any of the representation types offered by the system, but some representations were better suited to particular tasks than others.

[1] From left to right, top to bottom, these correspond to cards numbered 9, 10, 22, 23, 14, 72, 27, 28, 81, 55, 79 & 20 at the leaves of the dendrograms in Figs 3 & 4

Results

ER Card-Sort Results

ER Categories The subjects of studies 1 and 2 did not differ in terms of the number of categories that they identified in the ER card-sort task, or in terms of their category naming accuracy scores (see next section). The number of piles of cards that subjects produced ranged from 4 to 20 (study 1) with a mean of 12.06 and standard deviation of 4.33 and, for study 2, the mean was 15.25, with a standard deviation of 5.64. A t-test of the difference between the means was not significant. The subjects in study 1 and study 2 were considered to be well-matched samples.

ER Category Naming Accuracy The labels applied by subjects to each of their card-sort piles were assigned to each item within that pile. Thus, names for each of the 87 card-sort stimuli were obtained for each subject. ER names were rated against reference definitions for each of the card-sort stimulus items. Reference definitions were based on the name used for the ER in the source of the item, or on the basis of authoratitive sourcebooks (*e.g.* Harris, 1999) or on ER classification frameworks (*e.g.* Twyman, 1979; Lohse et al., 1994). Two independent raters assessed each subject's category labels against the reference definitions and *naming accuracy scores* of 0 or 1 were assigned to each of the 87 items. Each subject provided names for 87 ER card-sort items, hence 2436 (28x87) scoring judgments were made per rater. The raters disagreed on only 15 naming judgments (99% agreement). Those disagreements were resolved by means of a post-scoring discussion between the raters. ER naming accuracy scores out of 87 were computed for each subject by summing individual item scores. The mean score for study 1 subjects was 27.62 (s.d.=13.13) and for study 2 subjects it was 32.33 (s.d.=15.18). These means were not significantly different. For LQ subjects the mean was 17.86 (s.d.=12.09) and for UQ subjects the mean was 35.14 (s.d.=13.52). This difference was significant ($t=-2.52$, $p<.03$).

Table 1 shows the naming accuracy scores computed for all subjects, LQ subjects and UQ subjects for each major ER category in Figure 3. The ER naming accuracy of UQ subjects was superior for every category, with the exception of music. Note, though, that UQ subjects tended to (correctly) subsume music within a cluster containing logic and maths notations, computer code and various (natural language) text forms.

Cluster Analysis - All Subjects' Data

Each subject's card sort was represented in an 87 by 87 matrix. Each cell of the matrix coded the relationship between one distinct pair of items. Considering any two items, the subject either placed them in the same pile or not. For example, if the subject sorted card 3 and card 46 into the same pile, then a one was coded at the cell corresponding to the intersection of row 3 with column 46. Individual subject matrixes were summed to produce a combined matrix for all 28 subjects. Summed matrixes were also computed for LQ and UQ subjects.

Table 1: Mean percentage ER naming accuracy scores, for each ER category (1-9) of Figure 3

ER category	*All n=28*	*LQ n=7*	*UQ n=7*
1. Maps	78.2	54.0	85.7
2. Set diagrams	25.4	0.0	35.7
3. Logic/math notation	41.7	16.7	47.6
4. Informal groupings	11.5	5.0	16.0
5. Tables, graphs	35.7	20.0	42.8
6. Lists	34.8	14.3	41.1
7. Music	61.9	71.4	28.6
8. Pictures of objects	45.2	34.9	69.8
9. Scientific diagrams	22.7	16.0	25.2

The summed matrices were input to SPSS PROXIMITIES procedure to produce a similarity matrix. The similarity matrix formed the input to the SPSS CLUSTER procedure which was used to compute a multilevel, agglomerative, hierarchical cluster analysis (*e.g.* Everitt, 1993). The item clusters are arranged hierarchically with individual items at the leaves and single cluster at the root. Dendrograms provide a graphical display of cluster analysis output. The dendrogram for all subjects' data (Figure 3) shows which card-sort clusters are joined or fused at increasing levels of dissimilarity from the root node on the right towards the leaves on the left. A scaled index of dissimilarity ('Rescaled Distance Cluster Combine') based on the squared Euclidean distance measures is computed by the SPSS PROXIMITIES procedure (see top of Figure 3).

Inspection of Figure 3 suggested that adopting the clusters at scaled distance 15 represented a good compromise between detail and interpretability. At this level there were 9 major clusters (labelled 1 to 9 in Figure 3). These major clusters consist of: **1. Maps:** The 9 items in this cluster consisted of street maps, topographic projections, roadmaps and travel-oriented network diagrams such as the London Tube 'map', plus a 3D perspective drawing of New York streets and buildings. **2. Set diagrams:** This cluster contained set diagrams of both the Euler and Venn kinds but also, anomalously, a geometry figure consisting of 2 concentric circles with labelled radii. **3. Logic notation, math formulae:** A six item cluster containing handwritten (workscratching) examples of first-order logic, semi-formal notations, and also a mathematical equation and a computer program code fragment. **4. Informal text groupings:** This cluster is best described by the term 'informal groupings' *i.e.* where information was circled, placed in close proximity, connected by a hyphen, *etc*. Most of the handwritten workscratchings were subsumed under this cluster, together with hierarchies, tree diagrams, and conceptual graph ('mindmap') type representations. **5. Tables, line**

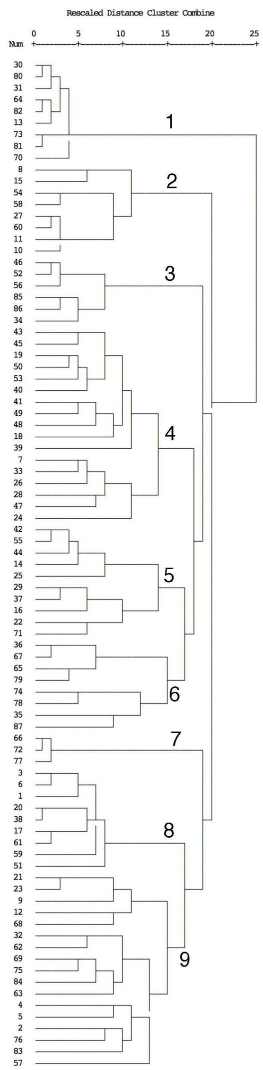

Figure 3: Dendrogram for all subjects (n=28). Labels 1 to 9 represent the major clusters at squared Euclidean distance 15.

graphs: A very distinct cluster containing all of the tabular, matrix representations including both sparse matrices (*e.g.* feature tables with empty cells and cells containing symbols such as ticks, or X's), multiway tables containing numerical or textual information. Line (X-Y) graphs and column graphs were also sorted into this category. **6. Lists:** The items in this cluster consisted of a telephone directory excerpt, timetables, various ordered and formatted text examples, and the computer code fragment. **7. Music:** The three examples of musical notation were sorted into this small discrete category. **8. Pictures of objects:** This cluster consisted exclusively of line drawings (2D & perspective) of concrete objects. **9. Scientific diagrams:** This category differed from the others in that it was characterised by a functional classification rather than being based on structural features or ER semantics. It included most of the ERs that were designed to be instructional in terms of their communicative rôle. These included a scientific illustration of planetary motion, an 'exploded' diagram of a machine, sequential arrays of pictures showing aircraft marshalling signals, dance choreography illustrations and a strip-cartoon.

Table 2: Legend to cluster numbers in Figure 4

category	*LQ group*
1	maps
2	tables, lists, semantic networks
3	tables (handwritten), line/column graphs
4	trees, nodes & arcs/lines, E-R diagrams, textual forms
5	plans, scientific diagrams & illustrations
6	pictures of objects
7	music
8	instructional representations
9	logic notation, math formulae
10	X-Y, line graphs, set diagrams, scientific diagrams

category	*UQ group*
1	maps
2	set diagrams
3	pictures of objects
4	plans, instructions, illustrations, scientific diagrams
5	line, column, X-Y graphs
6	logic notation, math formulae, music, computer code, text
7	trees,node & arc, E-R diagrams, semantic networks
8	textual forms, lists, tables

Cluster Analysis of LQ and UQ Groups' Data Inspection of the LQ and UQ dendrograms (Figure 4) and the legend to the major clusters (Table 2) shows that subjects in both of those groups made clear top-level distinctions between 'map' and 'non-map' representations. Similarly, pictures of objects emerged as clear categories for both groups. LQ subjects classified 'music' examples as a tighter, more distinct category than UQ subjects did. UQ subjects subsumed music into a broader cluster containing logical and math formulae, computer code and various (natural language) text forms. LQ subjects also demonstrated a tendency to classify ERs according to functional criteria. For example, the LQ dendrogram reveals a cluster that might be called 'instructional representations'. LQ subjects also tended to be influenced by visual factors such as whether the representation was produced informally (handwritten or hand drawn) rather than on the basis of semantic similarity - this was particularly striking in the case of table representations (Table 2).

LQ subjects classified set diagrams with (non-set diagram) circular ERs such as a scientific illustration of planetary motion and a geometry figure showing the radii of two concentric circles. This was another example of LQ subjects clustering on the basis of visual similarity. Unlike UQ subjects, the LQ group tended to associate

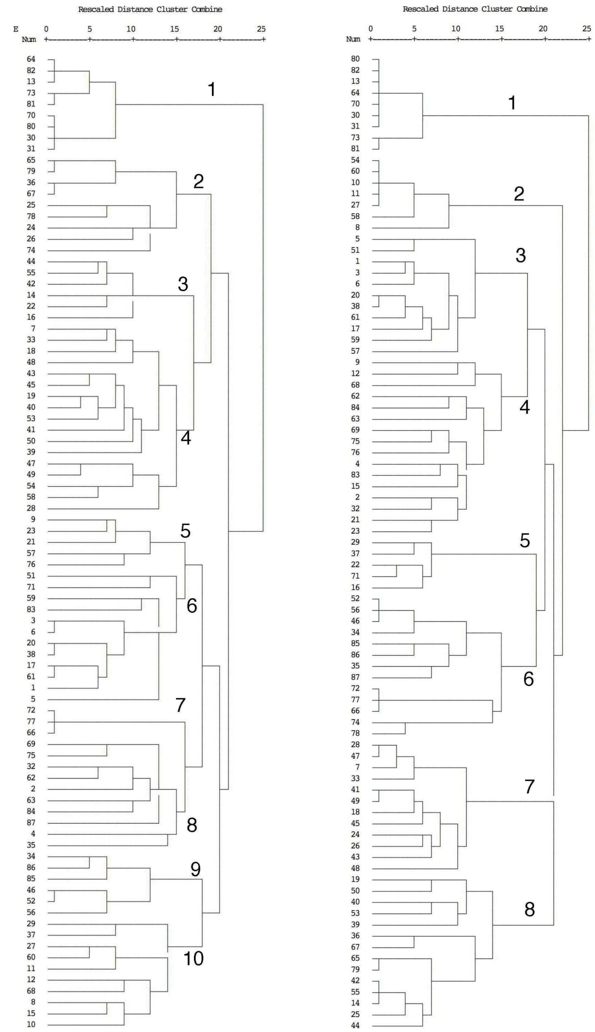

Figure 4: Dendrograms - LQ subjects (left) and UQ subjects (right). Table 2 provides a legend to numbered clusters.

concept graphs (‘mind maps’ or ‘spider maps’) with tables and lists rather than to see them as members of the same class of ERs as node and arc or tree graphs. Moreover, UQ subjects perceived column graphs, line and X-Y graphs as a distinct category of ERs, whereas LQ subjects tended to confound them with set diagrams and scientific illustrations.

Discussion

In general, subjects in this study identified 9 major categories of ERs. Eight were discriminated structurally, on the basis of the semantics of the ER. The clearest top-level distinction for all subjects was between map and non-map representations. Maps are ‘symbolic representations of physical geography’ (Lohse et al., 1994) and are ‘pervasive across time and cultures’ (Liben, 2001). Maps are perhaps the most spatial type of ER. Also, they are isomorphic with the real world and are not metaphorical in the way that, say, set diagrams are (where spatial inclusion acts as a metaphor for set membership).

Comprehension of map semantics seems to be natural and is acquired early in life. Liben (2001) describes research showing that children as young as 3 years old can understand that one space can stand in representational relation to another in mapping tasks. The cluster analysis allowed each emergent clusters to be assessed along a crispness/fuzziness dimension. Maps are crispest (high level branching factor and tight clustering of items at the ‘leaves’). With an early age of acquisition and conceptual crispness, therefore, maps seem to qualify as a ‘basic level’ in Rosch’s (1978) terms.

Comparing the card-sort data for LQ and UQ subjects revealed that expert ability to *use* ERs in problem solving and reasoning is associated with more accurate naming of ERs and a tendency to create categories on the basis of ER semantic distinctions to a greater extent than LQ subjects. As an illustration of this, UQ subjects (correctly) placed music notation examples with other notations such as those of logic, maths and computer programming. This is one of several examples in which UQ subjects seem to be more semantically-based in their conceptualisation of ERs. The semantic framework that guides their categorisation is based on a proper understanding of ER semantics. Their superior performance on tasks that involve reasoning with ERs (as opposed to merely sorting ERs) supports this view though an ability to match tasks to the demands of a task is also crucial. UQ subjects are also able to name ERs more accurately. They produce fewer categories in their ER card sorts because their knowledge of ERs allows them to perceive semantic commonality between visually different ERs. This is because they have greater familiarity with the domain of ERs (*i.e.* they are more graphically literate), and they have better comprehension of ERs. Their card sort categories are internally consistent and they demonstrate superiority in ER naming accuracy. These factors suggest that UQ subjects possess well-organised mental representations of ER knowledge.

In contrast, LQ subjects have less familiarity and ER knowledge and they are less accurate in naming. They respond more to superficial aspects of ERs, particularly to spurious visual features (such as whether the ER is hand-produced or printed, or circular (in the case of the set-diagram/geometry diagram confound). They are also inclined to make distinctions along what might be called functional dimensions, as illustrated by their creation of an ‘instructional ERs’ category. These spurious distinctions betray their lack of well-formed semantic knowledge of ERs to guide their classification. LQ subjects’ card sorts were driven more by more feature-based processes which may be based on stored instances of ER forms that they have encountered (perhaps in a somewhat more *ad hoc* fashion than their UQ counterparts) during the course of their education.

Precisely characterising a ‘graphical-linguistic distinction' remains a work-in-progress (Shimojima, 1999), subjects' mental organisation of ERs reflects that dichotomy to some extent. The emergence, in the whole group cluster analysis, of four graphical categories (maps, set diagrams, tables/graphs, object pictures & scientific diagrams) and three linguistic categories (math/logic notations, informal text groupings & lists) is consistent with such a distinction. However, LQ subjects tended to co-sort textual lists with tables and semantic networks (*e.g.* 'spider diagrams' or 'mind maps' used in note taking) to create a more hybrid or heterogeneous category containing list, linear branching and matrix arrangements of text. A name for such a category might be 'arrangements of text' or 'ordered text'. UQ subjects, on the other hand, seem to have a more formal conception. They tend to see semantic network ERs as members of a class of representations that consists of graphs of types directed/undirected and cyclic/acyclic.

An educational implication of these findings is a need, perhaps, for a domain-independent 'ER curriculum' in schools. This would serve to standardise each student's representational experience. As argued by Cox (1996), the range of ERs that a particular student is familiar with reflects the way he or she was taught. A particular student may *happen* to be familiar with semantic networks because their biology teacher taught food webs using that diagrammatic form. A student may *happen* to understand set diagrams because they studied set theory in mathematics. In current curricula, students are often introduced to ER formalisms in highly domain-specific contexts. This means that ER forms that are potentially useful in a wide range of situations might remain, in some students minds, too narrowly 'locked' into one subject domain. Teachers differ, too, in their enthusiasm for graphical teaching methods. Thus a particular student's mental organisation of ERs may reflect his/her choice of school subjects and the teaching styles of his/her instructors.

In this study, a card sort paradigm proved to be a useful means by which to gain insight into the mental organisation of ERs in skilled and less-skilled reasoners. The method allows normally implicit conceptions of the ER domain to be made explicit. Many studies of diagrammatic reasoning employ tasks which can be solved with a narrow range of ERs. This can afford detailed insights into a participant's representational behaviour with one or two representational systems. In contrast, card-sort tasks using large and varied corpora offer a means to gain a more comprehensive 'birds eye' view of the representational landscape.

Acknowledgements The first author gratefully acknowledges the support of the British Academy and the Leverhulme Foundation (Leverhulme Trust Research Fellowship G/2/RFG/2001/0117).

References

Cox, R. & Brna, P. (1995) Supporting the use of external representations in problem solving: The need for flexible learning environments. *Journal of Artificial Intelligence in Education*, **6(2)**, 239–302.

Cox, R. *Analytical reasoning with multiple external representations.* (1996) PhD thesis. Department of Artificial Intelligence, University of Edinburgh, Scotland.

Cox, R. (1999) Representation construction, externalised cognition and individual differences. *Learning and Instruction*, **9**, 343–363.

Everitt, B.S. (1993) *Cluster analysis*. (3rd edition). London: Edward Arnold.

Eysenck, M.W. & Keane, M.T. (2001) *Cognitive psychology: A student's handbook*. Hove, Sussex: Psychology Press.

Grawemeyer, B. & Cox, R. (submitted) Modeling individual differences in representation selection behavior. Submitted to the *Fifth International Conference on Multimodal Interfaces (ICMI-PUI'03)*, Vancouver, November, 2003.

Harris, R.L. (1999) *Information graphics: A comprehensive illustrated reference*. Oxford: OUP.

Lohse, G.L., Biolsi, K., Walker, N. & Rueter, H. (1994) A classification of visual representations. *Communications of the ACM*, **37(12)**, 36–49.

Liben, L.S. (2001) Thinking through maps. In Gattis, M. (Ed.) *Spatial schemas and abstract thought*. Cambridge, MA: MIT Press.

Pinker, S. (1998) *How the mind works*. London : Allen Lane.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104(3)*, 192–233.

Rosch, E.(1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale: H.J.: Erlbaum.

Shimojima, A. (1999) The graphic-linguistic disctinction: Exploring alternatives. *Artificial Intelligence Review*, **13(4)**, 313–335.

Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W., & Wielinga, R. (1999) *Knowledge engineering and management: The commonKADS methodology*. Cambridge, MA: MIT Press.

Twyman, M. (1979) A schema for the study of graphical language. In Kolers, P.A., Wrolstad, M.E. & Bouma, H. (Eds.) *Processing of Visible Language: Volume 1*, 117–150. New York: Plenum Press.