The effects of knowledge of external representations and display selection upon database query performance

Beate Grawemeyer and Richard Cox

Department of Informatics, University of Sussex, Falmer, Brighton BN1 9QH, UK
{beateg,richc}@cogs.susx.ac.uk

Abstract. This study investigated the representation selection and reasoning behaviour of participants who were offered a choice of informationally equivalent data representations (tables, bar charts, scatter plots, sector graphs, pie charts, lists - *i.e.* matrix graphics). A prototype automatic information visualisation system (AIVE) was used to present a series of questions about the information in a database. The database contained car information (manufacturer, model, purchase price, CO2 emission, engine size, horsepower, *etc*). The database queries posed to participants varied in terms of task demand (*e.g. identify* a car with a particular attribute, *rank* cars in terms of some feature, *compare* cars on dimension(s)). Participants read a question then chose, via an array of icons, the representation that they thought would best assist them with answering the query. AIVE generated the chosen representation and instantiated it with the appropriate data. Response accuracy and timing data were recorded. Results indicated that some tasks are more 'representation specific' than others. The ratio of correct to incorrect responses also varied widely across representation/task combinations. Participants were observed to make surprisingly quick representation selection decisions. Participants were divided into two groups on the basis of their prior ER knowledge level (assessed via a card-sort pre-task). The groups differed most in their use of scatterplots, bar charts and sector graphs. On the whole participants in the high ER knowledge group tended to be more graphically heterogeneous in their representational behaviour than those in the low ER knowledge group. Results are discussed in relation to work on representation schemas ([13]) - the suggestion is that participants in the high ER knowledge group possessed more developed representation schemas, including better knowledge of representation applicability conditions than participants with less well-structured ER knowledge.

1 Introduction

External representations (ERs) such as diagrams (graphs), text (notes, lists) and hybrid forms (tables, concept maps) are powerful aids to reasoning and problem solving. This has been established in various domains including analogical reasoning, vector arithmetic, algebra word problems, and logical and analytical reasoning *e.g.* [1, 4, 3, 8, 10, 19].

In this study we focussed upon a subset of ERs, namely lists, tables, scatter plots, sector ('rose' or 'star') graphs, pie charts and bar charts/column graphs. With the exception of lists, this class of representations has been termed 'matrix graphics' [20]. They are defined as a representation system 'family' in which graphical representations of populations of data cases are multi-dimensional, with all cases specified on all dimensions. Matrix graphics are also weakly expressive 'Minimal Abstraction Representation Systems' (MARS) in which one graphic represents one model [21].

It has been proposed [2] that at least three factors interact to determine whether a particular representation will assist reasoning on a particular task: 1. the subject's representational repertoire (the range of representations with which s/he is familiar) and cognitive style factors; 2. the semantic and cognitive characteristics of the representation, and 3. a sensitivity to the cognitive demands posed by the task (*e.g.* is the aim to spot trends, compare values or read-off of a single value?)[1].

Successful use of matrix graphics depends upon skillful matching of a particular representation with the demands of the task. [6] and [12] provide numerous examples of how a good fit between a task's demands and particular representations can facilitate search and read-off of information. [6] provides an illustration. Information about medication is best laid out, from a pharmacist's perspective, in list form. For the patient, however, compliance with the therapeutic regime is better assured if information is laid out in matrix form (drug A,B,C against breakfast, lunch, dinner and bedtime). A contingency table such as this allows the patient to easily decide which pills he should take (say) at lunchtime, and which pills should be taken together. [22] provides a review of studies that show that tasks involve perceiving relationships in data or making associations are best supported by graphs (*e.g.* line graphs, bar charts) whereas 'point value' read-off is better facilitated by tabular representations.

The process of choosing the 'right' representation has been the subject of research by [14]. She studied college-aged students' use of tabular, network and hierarchical representations. It was found that participants have rudimentary domain-independent abstract schemas for tabular, network and hierarchical representations which include 'applicability' conditions. [13] has suggested that these (implicit) schemas are defined in terms of structural properties of representations. For example, selecting a tabular representation seems to be 'driven' mainly by three factors: a) the information requires values to be represented along two distinct variables, b) values on the same dimension (row or column) cannot be linked, and c) 'movement' (*eg.* path traversal as, say, in a network) is not required. However, it seems that participants vary in the extent to which they possess and/or exploit representation schemas and applicability conditions. Puzzle-solving studies by [17], [18] and [15], for example, have found that participants tend to remain with the same representation despite changes in problem characteristics.

[1] [22] uses the term 'cognitive fit' to describe a good match of factors 2 and 3.

Skill at assigning representations to information might be expected to depend, at least in part, upon the breadth of a person's ER knowledge and upon how it is organised in semantic memory. In order to select an appropriate representation one must be familiar with a range of alternative possibilities. In this study we included measures of ER knowledge in order to explore the relationship between ER knowledge and ER selection skill.

We investigated participants' ability to select representations that are well-suited for answering a range of query types, using a database with a matrix-graphic production facility. Participants attempted a varied range of database query tasks. For each query, participants were permitted to choose (from an array of 5), a representation to assist them in responding (Figure 1). Computer-generated arrays of alternative, informationally equivalent representations were separately generated for each of a range of questions about data in a database. The questions varied in terms of their cognitive task demands (identify single value, compare, spot trends) and were based on the task taxonomies of [23] and [24].

A prototype automatic information visualization engine (AIVE) contained a database of information about cars and generated a range of alternative displays (bar charts, scatter plots, 'rose' (sector) graphs, tables, and text (lists)). For any given question, any of the representations was *potentially* usable but one (or some) supported efficient search to a greater extent than others. In other words they were informationally equivalent but not computationally equivalent [10]. On the basis of the research cited above (*e.g.* [17] [10] [6]), it was predicted that tables would provide better support for reading-off particular values, bar charts would provide better support for qualitative comparisons, and scatter plots would best support tasks requiring the identification of data clusters. However, representation-to-task matching ability was expected to correlate positively with participants' prior knowledge of ERs. As a measure of ER knowledge, we employed the card-sort and ER naming task used by [2] and [5]. The prediction was that participants who were familiar with a wide range of ERs would demonstrate more careful selection and (perhaps) more flexibility and virtuosity in matching representations to tasks. In contrast, low background knowledge

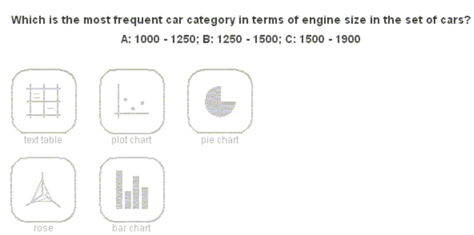

Fig. 1. AIVE representation selection interface

would be expected to be associated with less-principled selection, a tendency either to experiment with a range of ER types or to repeatedly use the same (familiar) representation despite variations in task demands.

To recapitulate - we studied the relationship between query type and selection behaviour in a group of participants, and we were also interested in within-group individual differences in selection patterns and preferences. Ultimately, the aim is to inform the design of a user-model for an adaptive automated data visualisation system (I-AIVE) in the tradition of Mackinlay's [11] APT system.

2 The study

The study utilised a computer-administered task in which participants made judgments about, and compared cars in terms of their features. Car information was contained in a database and participants selected the type of information display that they wished to use for each task. The database query session was preceded by a card-sort task [2] in which participants sorted and labelled a large corpus of ER examples.

Participants were 8 postgraduate students, 2 research fellows, 1 engineer and 1 computer scientist (5 females and 7 males). Ten of the twelve participants had a strong maths or computer science background, whereas two had a language/ linguistics background. The experimental procedure comprised a single session in which the card-sort task was administered. This was followed by the AIVE database query problem solving task. The session took between 50-65 minutes to complete.

2.1 ER card-sort task

As a measure of general knowledge about ERs, a card sort task developed by [2] was used. This task assesses participants' semantic knowledge of a wide range of ER forms. Card sorts have been 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.* [16]). 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, tables, music, set diagrams (Euler & Venn), illustrations, maps, tree diagrams, etc. Detailed descriptions of the task and stimuli can be found in [2]. Participants were instructed to sort the cards into groups and were given paper slips and pencil to subsequently label their 'piles' of cards. No constraints were placed upon the number of card groups and they were permitted to produce hierarchies of groups and subgroups.

2.2 AIVE database query task

In the AIVE database query tasks, participants were asked to make judgments and comparisons between cars and car features based on database information.

The database contained information about 10 cars: manufacturer, model, purchase price, CO2 emission, engine size, horsepower, *etc.* The question types were based on those used in studies described in [23] and [24], which made use of a 'visual task taxonomy'. This means that for a given task, there is an associated visual perceptual act. For example selecting similar items involves perceptually clustering visual data. Other tasks (*e.g.* locate, rank, compare) entailed different visual processes, such as detecting proximal values, or grouping or sequencing them. Table 1 gives an overview of the different representation types and example queries used in this study. Each subject responded to 25 database questions,

Task type	**Example query**
Associate	Which is the most frequent insurance group in the set of cars?
Cluster	Which cars are most similar with respect to their monthly average running costs?
Compare	Which car has the most environmentally friendly engine, in terms of the CO2 emission rating?
Correlate	Is it true, that at least one car has a CO2 emission below 120 g/km?
Distinguish	One of the cars is unlike any of the others. Can you identify it?
Identify	Which is the fourth most powerful car?
Locate	Where would you place a Rover 800 with an engine size of 1994cc inside the display?
Rank	Rank Honda Civic, Renault Megane and VW Polo in respect of their safety equipment (most to least).

Table 1. Example queries of the different task types used in the AIVE experiment

which were of 8 types: associate (3 questions); cluster (3); compare (3 positively,3 negatively phrased); correlate (3); distinguish (1); identify (3); locate (3) and rank (3). All response formats were multiple-choice (*e.g.* Figure 1). Participants were informed that to help them answer the questions, the system (AIVE) would supply the appropriate data from the database. AIVE also offered participants a choice of representations of the data. Participants could choose between various types of graph, diagram and tables of information. The representations which were offered from the system varied acording to the type of data (nominal, ordinal) to be displayed. On 5 of the 25 questions (one each of type rank, correlate, compare, associate and identify), the representation array offered bar chart, plot chart, rose, table and text for nominal data. For ordinal data items (20 questions) the representation selection array did not include a text option but, instead, included a pie chart option. Participants were told that the choice was theirs, but that they should select a form of display they thought was most likely to be helpful for answering the question. Participants then proceeded to the first question, read it and selected a representation (Figure 1). The spatial layout of the representation selection buttons was randomized across the 25 query tasks in order to prevent participants from developing a set pattern of selection. Each problem could potentially be answered with any of the representations offered by the system. After the subject made his/her representation choice, AIVE generated and displayed the representation instantiated with the data required for answering the question (*e.g.* Figure 2). The subject then answered the question using the chosen visualization. Participants were not permitted to change

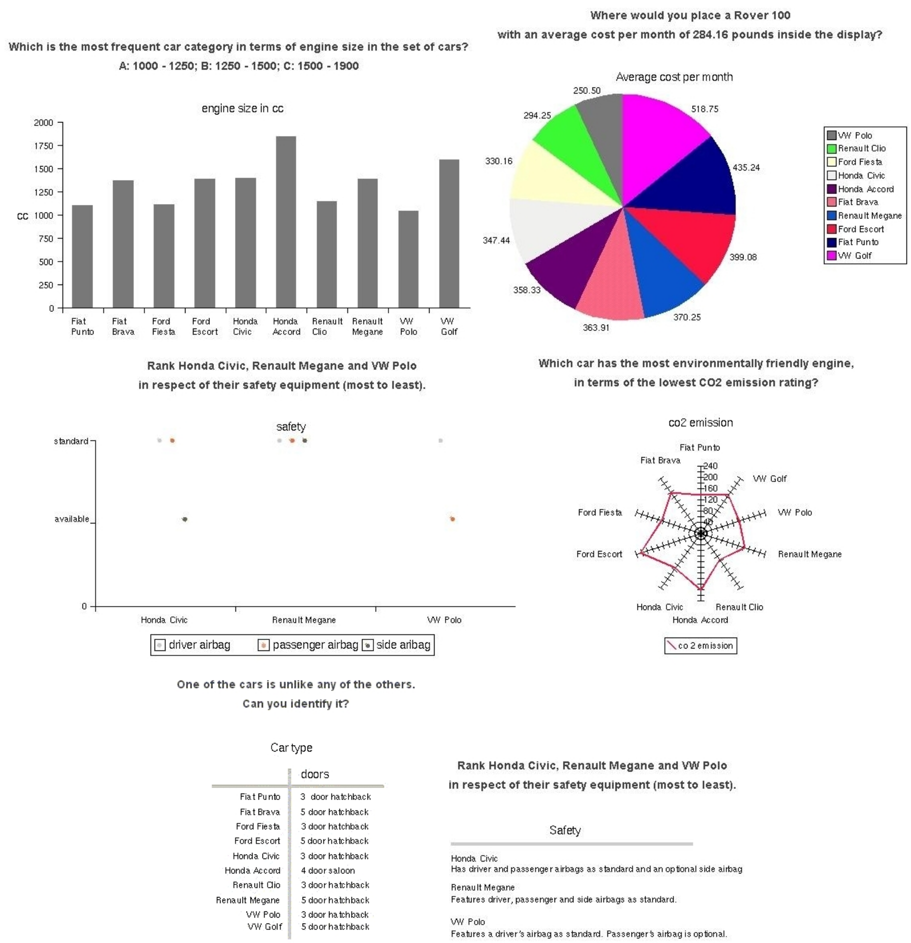

Fig. 2. Examples of AIVE bar, pie, plot, rose (sector) graph, table and text (list) representations.

representation following their initial selection. This constraint was imposed in order to encourage participants to carefully consider which representation was best matched to the task. Following a completed response, the subject was presented with the next task and the sequence was repeated. AIVE recorded (1) the randomized position of each representation icon from trial to trial; (2) the user's representation choices; (3) time to read question and select representation (selection); (4) time to answer question using chosen representation (answer); and (5) participants' responses to questions.

3 Results

3.1 ER card-sort task

subject	*dbQueryTask*	*cluster*	*label*	*subject*	*dbQueryTask*	*cluster*	*label*
S1	81.8	11267	65.79	S7	64.0	4035	36.71
S2	72.0	5295	69.94	S8	84.0	5941	66.27
S3	79.2	6093	49.62	S9	80.0	6225	74.7
S4	72.0	5439	45.7	S10	84.0	6935	75.99
S5	52.2	4720	40.96	S11	84.0	5957	33
S6	68.0	5509	44.62	S12	80.0	8905	58.01

Table 2. Subject performance on the database query task (%), card sort cluster ('typicality' - higher = more typical) and cluster labelling scores.

Each subject's card sort data was represented in matrix form. The individual subject matrices were added to form a group matrix. The summed matrix was input to the SPSS PROXIMITIES procedure which outputs 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. [7]). The item clusters were arranged hierarchically with individual items at the leaves and a single cluster at the root. Dendrograms provide a graphical display of cluster analysis output. The dendrogram for all the participants' data (Figure 3) shows which card-sort clusters were joined or fused at increasing levels of dissimilarity from the root node on the top towards the leaves on the bottom. The dendrogram shows that, as a group, participants distinguished 9 main categories of representation (Figure 3), These were: logic / mathematical formulae, music, text, x/y graphs, set diagrams, tree diagrams, list/tables, illustrations and maps. These categories are similar to those found by [2] in a different subject sample. 'Maps' versus 'non-maps' was the most striking distinction made by participants (evidenced by a major branch very close to root node) and that most map examples in the corpus were seen by participants as equally good examples of maps (evidenced by the flat branching factor at the leaves of the map sub-cluster).

For each subject, two scores were derived from the card-sort data. One score was for card sort 'typicality' and the other was for cluster labelling (naming) accuracy. To derive the first score, each individual's matrix was multiplied by the group matrix to produce 'distance' scores for each subject. This resulted in

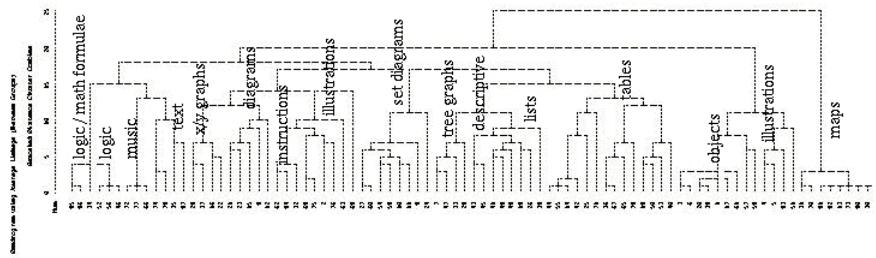

Fig. 3. Dendrogram for participants card sort clusters

a card-sort 'typicality' score - the higher the score, the more typical of the group as a whole that subject's card-sort.

The second score represents the cluster naming accuracy of each subject. Each label that participants applied to their cluster was analyzed as another method of assessing their comprehension of the semantics of the representation in particular clusters. Higher cluster label scores were given for precision and accuracy. For example, if a subject discriminated 'set diagrams' from other circular ERs (*e.g.* geometry diagrams featuring circles), and labelled them precisely, a higher score was assigned. A lower score was given for more vague labels such as 'things with circles' (a real example from the data)[2]. Table 2 shows the clustering and labelling score results.

3.2 AIVE database query task

Score totals on the query task were out of 25 except for three participants for whom system failures occurred on some trials - S1 (2 trials), S3 (1) and S5 (2). Since these three participants underwent slightly fewer trials than others, raw scores were converted to percentages (Table 2). Subject 5 selected 'bar chart' for every question and also showed idiosyncratic timing data and was therefore dropped from the analyses. To recapitulate, there were 8 types of tasks: associate, cluster, compare, correlate, distinguish, identify, locate and rank. Six representations were available: bar, pie, rose, plot, text and table. The frequency of representation by task type for correct/incorrect response can be seen in Figure 4. It shows that the most used representations (irrespective of accuracy) were tables, bar and pie charts, and scatter plots. If accuracy is considered, the most effective representations were text, plot and table. It can be seen that the most popular representations, such as bar charts, were not necessarily the most effective ones in the case of the cluster, identify and locate tasks. The 'top 3' representation/task pairings, in terms of optimal performance, were bar chart for compare tasks, tables for rank tasks and scatter plots for correlation tasks.

[2] The naming data analysis is in progress and will not be discussed here.

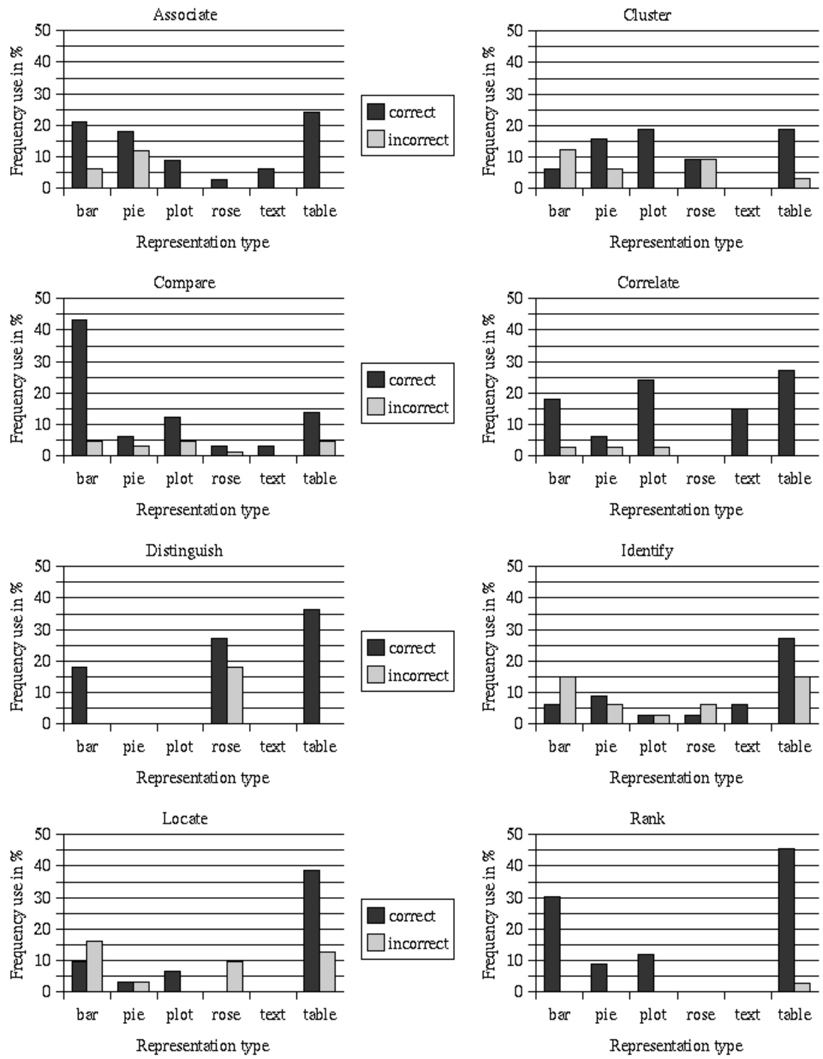

Fig. 4. Frequency with which each type of representation was used for each type of database query task type in % (all participants' data, n=11). Note that text was not an option for cluster, distinguish, and locate questions. Text was substituted for pie charts on a third of the rank, correlate, associate and identify questions and a sixth of the compare questions.

The most difficult task types were cluster, identify, compare and locate, which had a high proportion of errors. Identify, compare and locate in particular seemed to be more 'representation critical' in that only one or two representations were associated with correct responses. This can be contrasted with the associate, correlate, distinguish and rank tasks where a broader range of representations were associated with correct responses (Figure 4). The text representation was not offered as an option for cluster, distinguish, and locate questions and was only substituted for pie charts on a third of the rank, correlate, associate and identify questions and a sixth of the compare questions. Nevertheless, when it was available, it was chosen and used constructively on all tasks except rank.

The time taken to select representations was comparable for both correct and incorrect responses, with a means of 12.96 and 13.10 seconds for correct and incorrect responses, respectively (SDs = 11.68, 9.0 seconds). This could imply that participants were not anticipating difficulties when they first encountered problems that they subsequently responded to incorrectly, since they did not deliberate longer over their choice of representation. More time was spent on problem solving and answering the questions. The means were 36.06 and 42.24 seconds for correct and incorrect responses, respectively (SDs 25.52, 28.37 seconds). Representation selection time for identify questions tended to be lower (8.42 seconds) than for other task types. Identify questions (*e.g.* 'Which car has the second biggest engine size') were shorter and less complex than other task types and therefore are quicker to read and comprehend. The visual complexity of the rose representation (see Figure 2) was reflected in increased read-off time during answering (mean = 50.0 seconds) compared to other forms of representation (range of means 29.6 to 44.45 seconds). Particular combinations of representation and task were associated with distinctive timing patterns. For example, bar charts were quickly chosen for correlate tasks (8.76 seconds) and identify (6.86 seconds) and solution times also tended to be quite efficient in these combinations (11.14 and 22.14 seconds respectively). However, the accuracy data in Figure 4 shows that bar charts had high rates of inaccuracy on identify questions. Tables also tended to be quickly selected on identify tasks (7.07 sec) and, like bar charts, they tended to be reasonably efficient in the read-off and answering phase (25.86 sec). Figure 4 shows that, unlike bar charts, they were associated with accurate performance across all tasks except identify.

3.3 Relationship between ER knowledge and performance

Participants were divided into two groups on the basis of a median-split on ER card-sort cluster scores. This yielded two groups - 'typical' card-sorters (high ER knowledge) and 'more idiosyncratic' card-sorters (low ER knowledge). Idiosyncratic card sorts tended to indicate that the subject has a mental organisation of ER knowledge that was not based on the semantic properties of representations to the same extent as it was in skilled reasoners [5]. A between-groups t-test with database query task performance as the dependent variable revealed that lower ER knowledge was associated with poorer query answering performance (means = 72.00, 81.50, $t=-2.99$, $df(9)$, $p<.05$). The two groups also differed in

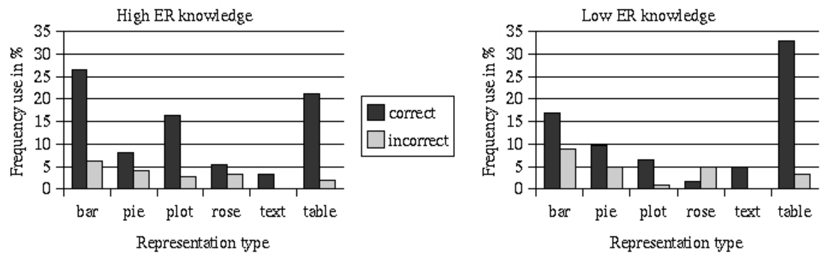

Fig. 5. Mean frequency with which high/low ER knowledge groups used each type of representation in % (all tasks combined). Average use per subject expressed as the number of times it was chosen as a proportion of number of occasions it was offered by AIVE.

terms of representation use across tasks (Figure 5). In general, participants with higher ER knowledge used a wider range of ER types, and used them more successfully than low ER knowledge participants. In particular, they tended to use scatter plots more frequently and effectively. Both groups of participants used bar chart representations very frequently across tasks but the success rate was greater for higher ER knowledge participants. Higher ER knowledge participants also tended to use the unusual sector graph ('rose') representation more often, and more effectively, than lower ER knowledge participants.They also showed a tendency to use tables less frequently than lower-ER-knowledge participants - an effect that was evenly spread across task types. Again, the success rate was greater for higher ER participants.

Separate versions of Figure 4 were generated for low and high ER knowledge groups. Inspection of these figures revealed that the tendency of low ER knowledge participants to use tables more frequently was true across all tasks but most marked in the case of correlate questions. On these tasks, high ER participants markedly preferred bar charts compared to low ER participants (mean within-task % use on correct responses 6.7 versus 27.8 - *cf* figure 4). For scatterplot use the corresponding means were 6.7 and 38.9). High ER knowledge participants therefore demonstrated more appropriate representation selection for correlate tasks particularly, compared to low ER participants, presumably due to greater knowledge of applicability conditions. Another large between-group contrast was observed on the distinguish task - a large difference in the tendency of low ER participants to select tables (corresponding Figure 4 means for correct responses were 60.0 versus 16.7). Finally, higher ER knowledge participants chose bar charts for rank questions to a much greater extent than low ER participants (44.4 versus 13.3 on correct responses).

4 Discussion and Conclusions

The prediction that tables would provide better support for reading-off particular values was supported, on the whole. On each of the identify, distinguish and locate tasks tables were used frequently and associated with relatively high success rates. They were also the most frequent choice for ranking type questions (bar charts were the second most frequent choice). This supported earlier findings by [17] and [18] who have suggested that the superiority of tables results from the way that they support cognitive operations (suggesting fruitful orders of operation, facilitating consistency checking, *etc*). A second prediction from the literature was that bar charts would provide better support for qualitative comparisons. This too was supported by our results. For compare tasks, bar charts were the most frequently chosen representation and were associated with a high degree of response accuracy.

The prediction that scatter plots would best support tasks requiring the identification of data clusters was also borne out to some extent. Scatter plots were selected most frequently for cluster tasks with tables a very close second. Rose, pie and bar charts were also selected but, unlike scatter plots, they were associated with high proportions of erroneous responses. The comparison of low and high ER knowledge groups' selection patterns suggests extensive differences in expertise. This was most apparent in the case of scatterplots. High ER knowledge participants appear to use scatterplots more frequently, and more effectively than their low ER knowledge counterparts. This was also true for the use of bar charts and, to some extent, rose (sector) graphs. Participants in the high ER knowledge group tended to be more 'graphical' and heterogeneous in their selections than low ER participants. Low ER participants favoured tabular and textual representations to a greater extent. The analysis of between-group representation selection patterns for each task type showed that, on the whole, wherever differences were marked, it was in the direction of high ER knowledge demonstrating more appropriate representation selection compared to low ER participants. This was presumably due to greater knowledge of applicability conditions and more developed representation schemas in those participants.

Overall, the results supported the contention of ([13]) that most participants have at least rudimentary representation schemas which include 'applicability conditions' for the appropriate assignment of representations to information. Participants' mental organisation of representational knowledge, as assessed by the card-sort task, was predictive of performance differences between participants in terms of correct responses to database queries and representation selection and use. Findings from another study ([5]) suggested that the mental organisation of ER knowledge in skilled reasoners differs from that of less skilled reasoners in that it consists of fewer (and more semantically coherent) categories. For example, skilled reasoners demonstrate in their card-sorts that they understand the 'spatial inclusion for set membership' metaphor of set diagrams by the fact that they co-sort Euler and Venn diagrams and differentiate them from geometry diagrams of circles and pictures of circular objects. The findings of this study suggest that card-sort tasks offer a useful non-verbal methodology for assessing implicit representational knowledge schemas which are not accessible verbal articulation.

Another findings was that participants did not spend very long on representation selection - the average time taken for an initial reading of a question *and* representation selection was only 12-13 seconds. This tendency to 'hastily' (prematurely?) commit to a representation has been observed in other studies. In a study of analytical reasoning with external representations, [4] concluded that students allocate too few resources to problem comprehension - students spend relatively little time on question reading and representation selection. Also [9]), in a study of information translation between representations, noted that "one of the most surprising aspects of our results is that our ...(participants)... made their choice of representations extremely early, perhaps even before they understood the details of a given problem." (p.221). If, as suggested by [13] (p.201), representational knowledge (and applicability condition knowledge) 'might be to some extent implicit' then representation selection decisions would be expected to be automatic and fast. This is an issue that warrants further research.

As mentioned in the introduction, one of the motivations for the work presented here is to eventually inform the design of a user-model for an adaptive version of the AIVE system. Perhaps an adaptive system version of AIVE might, as one strategy of intervention in cases of repeatedly poor representation selection, attempt to move its user from an implicit to a more reflective, considered explicit mode of cognition. The system may adapt to the users with low ER knowledge by constraining the diversity of representations it offers. In the immediate future, however, the aim is to run more participants and continue the analysis of the naming accuracy data with a view to gaining further insights into how the representational schemas or mental organisation of ER knowledge varies with expertise.

4.1 Acknowledgements

We thank the three anonymous reviewers for their thoughtful suggestions. The support for Richard Cox of the Leverhulme Foundation (Leverhulme Trust Fellowship G/2/RFG/2001/0117) and the British Academy is also gratefully acknowledged.

References

1. A. Blackwell. *Thinking with diagrams.* Kluwer, Dordrecht, 2001.
2. R. Cox. *Analytical reasoning with external representations.* PhD thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.
3. R. Cox. Representation interpretation versus representation construction: A controlled study using switchERII. In B. du Boulay & R. Mizoguchi, editor, *Artificial Intelligence in Education: Knowledge and media in learning systems (Proceedings of the 8th World Conference of Artificial Intelligence in Education Society)*, pages 434–441, Amsterdam, 1997.
4. R. Cox and P. Brna. Supporting the use of external representations in problem solving: the need for flexible learning environments. *Journal of Artificial Intelligence in Education*, 6(2-3):239–302, 1995.

5. R. Cox and B. Grawemeyer. The mental organisation of external representations. In *European Cognitive Science Conference (EuroCogSci - joint Cognitive Science Society and German Cognitive Science Society conference)*, Osnabrück, 2003.
6. R. Day. Alternative representations. In G. Bower, editor, *The Psychology of Learning and Motivation*, volume 22, pages 261–305, New York, 1988. Academic Press.
7. B. Everitt. *Cluster analysis.* Edward Arnold, London, 3rd edition, 1993.
8. J. Glasgow, N. Narayanan, and B. Chandrasekaran. *Diagrammatic reasoning.* AAAI Press/MIT Press, Menlo Park, CA, 1995.
9. D. Jones and D. D.A. Schkade. Choosing and translating between problem representations. *Organizational Behavior and Human Decision Processes*, 61(2):214–223, 1995.
10. J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11:65–100, 1987.
11. J. Mackinlay. Automating the design of graphical presentations of relational information. In J. M. S. Card and B. Schneiderman, editors, *Readings in information visualization: Using vision to think*, San Francisco, 1986. Morgan Kaufman.
12. D. A. Norman, editor. *Things that make us smart.* Addison-Wesley, MA, 1993.
13. L. Novick and S. Hurley. To matrix, network, or hierarchy, that is the question. *Cognitive Psychology*, 42:158–216, 2001.
14. L. Novick, S. Hurley, and M. Francis. Evidence for abstract, schematic knowledge of three spatial diagram representations. *Memory & Cognition*, 27:288–308, 1999.
15. J. Polich and S. Schwartz. The effect of problem size on representation in deductive problem solving. *Memory and Cognition*, 2(4):683–686, 1974.
16. G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. V. de Velde, and R. Wielinga. *Knowledge engineering and management: The commonKADS methodology.* MIT Press, Cambridge, MA, 1999.
17. S. Schwartz. Modes of representation and problem solving: Well evolved is half solved. *Experimental Psychology*, 91(2):347–350, 1971.
18. S. Schwartz and D. Fattaleh. Representation in deductive problem solving: The matrix. *Experimental Psychology*, 95(2):343–348, 1972.
19. K. Stenning. *Seeing reason: Image and language in learning to think.* Oxford University Press, Oxfort, 2002.
20. K. Stenning and R. Inder. Applying semantic concepts to the media assigment problem in multi-media communication. In N. N. J. Glasgow and B. Chandrasekaran, editors, *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, Cambridge, Ma., 1995. MIT Press.
21. K. Stenning and J. Oberlander. A cognitive theory of graphical and linguistic reasoning: Logic and implication. *Cognitive Science*, 19(1):97–140, 1995.
22. I. Vessey. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22:219–241, 1991.
23. S. Wehrend and C. Lewes. A problem-oriented classification of visualization techniques. In *Proceedings IEEE Visualization 90*, pages 139–143, 1990.
24. M. Zhou and S. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proceedings of the CHI 98*, pages 392–399, 1998.