Graphical Data Displays and Database Queries: Helping Users Select the Right Display for the Task

Beate Grawemeyer and Richard Cox

Representation & Cognition Group,
Department of Informatics,
University of Sussex, Falmer, Brighton BN1 9QH, UK
{beateg, richc}@sussex.ac.uk

Abstract. This paper describes the process by which we have constructed an adaptive system for external representation (ER) selection support, designed to enhance users' ER reasoning performance. We describe how our user model has been constructed - it is a Bayesian network with values seeded from data derived from experimental studies. The studies examined the effects of users' background knowledge-of-external representations (KER) upon performance and their preferences for particular information display forms across a range of database query types.

1 Introduction

Successful use of external representations (ERs) depends upon skillful matching of a particular representation with the demands of the task. Numerous studies (*eg.* [6] and [13]) have shown how a good fit between a task's demands and particular representations can facilitate search and read-off of information. For example, [18] provides a review of studies that show that tasks involving perceiving relationships in data or making associations are best supported by graphs, whereas 'point value' read-off is better facilitated by tabular representations.

Numerous factors are associated with ER-task matching skill. Firstly, it is known that individuals differ widely in terms of their preferences for particular forms of ERs [3]. Better reasoners organise their knowledge of ERs on a 'deeper' semantic basis than poorer reasoners, and are better at correctly naming various ER forms [4].

Secondly, some types of tasks require a particular, specialised type of representation whereas for other types of tasks, several different ER forms may be useful. The extent to which a problem is representationally-specific is determined by characteristics such as its degree of determinacy (extent to which it is possible to build a single model of the information in the problem).

ER selection skill requires, *inter alia*, knowledge of a range of ERs in terms of a) their semantic properties (*eg.* expressiveness, b) their functional roles together with information about the 'applicability conditions' under which a representation is suitable for use ([17][2][14]).

Considerable advances in intelligent automatic matching of information to visual representations have been made, beginning with APT [11], which included a composition algebra and primitives to generate a wide range of information displays. Later, SAGE [16] extended APT's graphic design capabilities. Another system, BOZ, [1] utilised a task-analytic approach. However, the APT, SAGE and BOZ systems do not accommodate differences between users in terms of their background knowledge of ERs or ER preferences. Individuals differ widely in terms of their ER knowledge and selection predilections. Whilst most people can use some ER forms effectively (*eg.* common examples like bar or pie charts), some ER types require specialised knowledge and training. Euler's circles are examples of the latter kind - set diagram semantics have to be to be learned specifically [3]. Individuals also differ in terms of their preferences for representing information visually (*eg.* via graphics or diagrams) or verbally (lists, notes, memoranda) [10].

The aim of this paper is to describe the process by which we constructed an adaptive system that recommends ERs taking into account the users' background knowledge-of-external representations (KER) and his/her preferences for particular types of information display.

This paper extends our earlier work (*eg.* [7]) by further researching the relationships between individuals' background knowledge of external representations and their ability to select appropriate information displays. The domain studied was that of diagrammatic displays of database information. The work was conducted using an unintelligent database system (AIVE). The results have informed the design of an adaptive system (I-AIVE) capable of supporting users in their choice of external representations (ERs). I-AIVE's user model is being developed on the basis of empirical data gathered from a series of experimental studies ([7]). This approach is similar to that of [9], who used empirical data in the validation of the READY system. That system models users' performance capacity under various cognitive load conditions.

The structure of this paper as follows: Section 2 covers the experimental procedure used to investigate the effect of users' background knowledge of ERs upon information display selection on different representation-specific database query tasks. The experimental results and implications for the ER recommender system are discussed in chapter 3. Section 4 outlines the user model implementation in the form of a Bayesian network which is 'seeded' with and derived from the empirical data. Part 5 describes the adaptation process, where the user model permits the system to engage in overt adaptive behaviors such as suggesting or providing ER selection hints or covert adaptive behaviours, such as restricting the choice of representations for particular participants for particular tasks in order to encourage good ER-to-task matching behavior. Conclusions are presented in section 6.

2   Experiment

In our experiment a prototype automatic information visualization engine (AIVE) was used to present a series of questions about the information in a database.

Knowledge of External Representations (KER) Tasks. Twenty participants first completed 4 tasks designed to assess their knowledge of external representations (KER). These consisted of a series of cognitive tasks designed to assess ER knowledge representation at the perceptual, semantic and output levels of the cognitive system [5]. A large corpus of external representations (ERs) was used as stimuli. The corpus contained a varied mix of 112 ER examples including many kinds of chart, graph, diagram, tables, notations, text examples, *etc.*

The first task was a decision task requiring decisions, for each ER in the corpus, about whether it was 'real' or 'fake'[1]. This was followed by a categorisation task designed to assess semantic knowledge. Participants categorised each representation as 'graph or chart', 'icon/logo', or 'map', *etc.* In the third (functional knowledge) task, participants were asked *'What is this ER's function'?* An example of one of the (12) multiple-choice response options for these items is *'Shows patterns and/or relationships of data at a point in time'.* In the final task, participants chose, for each ER in the corpus, a specific name from a list. Examples include 'venn diagram', 'timetable', 'scatterplot', 'Gantt chart', 'entity relation (ER) diagram'. The 4 tasks were designed to assess ER knowledge representation using an approach informed by picture and object recognition and naming research [8]. The cognitive levels ranged from the perceptual level (real/fake decision task) to through production (ER naming) to deeper semantic knowledge (ER functional knowledge task).

AIVE Database Query Task. Following the KER tasks, participants underwent a session of 30 trials with the AIVE system. The AIVE database contains information about 10 types of car: manufacturer, model, purchase price, CO^2 emission, engine size, horsepower, *etc.*

On each trial, AIVE (figures 1, 2 and 3) presented a database query (*eg.* 'Which two cars are most similar with respect to their Co2 emission and cost per month?'). Participants could then choose between various types of ERs, *eg.* set diagram, scatter plot, bar chart, sector graph, pie chart and table (all representations were offered by the system for any query). These options were presented as an array of buttons each with an icon depicting, in stylised form, an ER type[2] (table, scatterplot, pie chart, ...)(see figure 1). Participants were told that they were free to choose any ER, but that they should select a form of display they thought was most likely to be helpful for answering the question. Following the participant's selection, AIVE displayed a 'full' (data instantiated) version of that representation using data from the database of car information - examples are shown in Figures 2 and 3.

Across the 30 trials, participants experienced six types of database query: 'identify'; 'correlate' ; 'quantifier-set'; 'locate'; 'cluster' and 'compare negative'. For example, a typical 'correlate' task was: 'Which of the following statements is

---

[1] Some items in the corpus are invented or chimeric ERs.

[2] The spatial layout of the representation selection buttons was randomized across the 30 query tasks in order to prevent participants from developing a set pattern of selection.

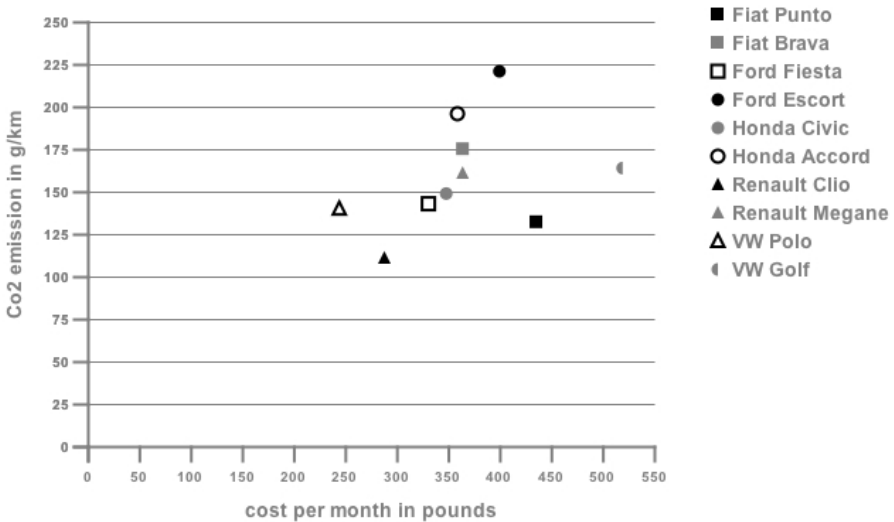**Fig. 1.** AIVE representation selection interface

Fig. 2. Example of AIVE plot representation

true? A: Insurance group and engine size increase together. B: Insurance group increases and engine size decreases. C: Neither A nor B?'; or a typical locate task: 'Where would you place a Fiat Panda with an engine size of 1200 cc inside the display?'.

Based on the literature (*eg.* [6]), a single 'optimal' ER for each database query form was identified[3]. However, each AIVE query type could *potentially* be answered with any of the representations offered by the system (except for some 'cluster' tasks for which a set diagrams was the only really usable ER).

---

[3] Display selection accuracy (DSA) scores were based on this (see results section).

Is it true that all cars which have an insurance group of 6
have an engine size between 1300 and 1400 cc?

A
cars with insurance group 6

B
cars with an engine size
between 1300 and 1400 cc
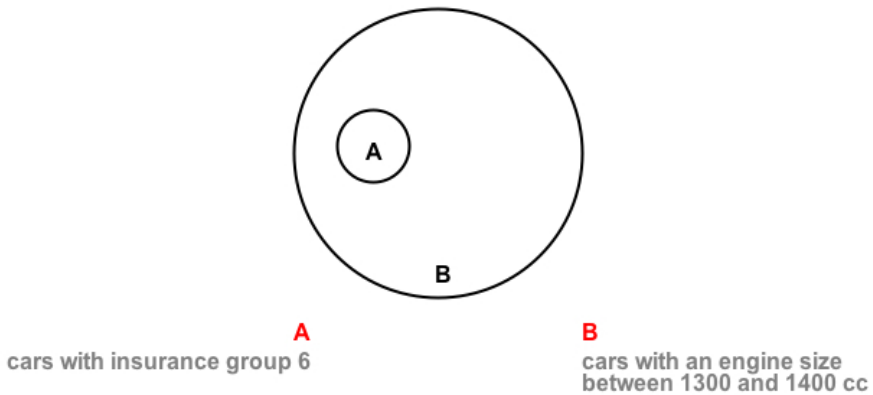
Fig. 3. Example of AIVE Euler's circles representation

Participants were not permitted to select a different 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, participants were presented with the next task and the sequence was repeated. The following data were recorded by the AIVE system: (1) the randomized position of each representation icon from trial to trial; (2) the users' representation choices; (3) time to read question and select representation (selection); (4) time to answer the question using chosen representation (answer); and (5) participants' responses to questions.

3   Results and Discussion

To recapitulate there were 20 participants, each of whom was presented with 30 AIVE tasks (600 data points in total). The independent and dependent variables are shown in Table1. Statistical analyses indicate that the KER tasks are significant predictors of display selection accuracy (DSA) and database query answer (DBQA) performance.

DSA was significantly predicted by one of the KER tasks (ER classification knowledge). For DBQA, the best KER predictor was ER functional knowledge. Hence a degree of conceptual (classificatory) knowledge of ERs predicts success at appropriate information display selection on the AIVE tasks but deeper semantic (functional) knowledge of ERs is associated with success at *using* the selected ER *ie.* reading-off information and using it to respond correctly to the database query. Additionally, appropriate representation selection results in better query performances. This suggests that, for predicting query response accu-

racy, a participant's KER can be as powerful a predictor of question answering accuracy as display selection accuracy.

The selection latency results show that a speedy selection of a display type in AIVE is associated with a good display-type choice. This implies that users either recognise the 'right' representation and proceed with the task or they procrastinate and hesitate because of uncertainty about which display form to choose. Less time spent responding to the database query question is associated with a good display-type choice and correct query response. This suggests that the selection and database query latencies may be used in the system's user model as predictors of users' ER expertise.

The results reported so far were based on all the AIVE query types combined. However, they differed extensively in terms of their representational specificity. Two different query types were contrasted in order to examine the effects of the tasks' 'representational specificity'.

3.1   Comparing a Highly ER-Specific AIVE Task and a Less ER-Specific AIVE Task

Participants' selection behavior and database query task performance for the correlate and locate task are shown in figure 4 and 5. The correlate task is highly representation specific and the locate task much less so.

The AIVE Correlate Task - High ER-Specificity. As shown in figure 4, 77% of AIVE correlate type queries were answered correctly by participants. Moreover, in 77% of the cases they chose the most appropriate ER display (scatter plot) from the array of display types (ERs) offered by AIVE.

Statistical analysis shows that performance on two of the KER tasks (ER classification and functional knowledge) predicts good display selection performance (see Figure 6). Especially ER classification knowledge of set diagrams, and functional knowledge of graphs and charts (as might be expected). Longer display selection latency is associated with longer time spent responding to the database query question.

The AIVE Locate Task - Low ER-Specificity. Figure 5 shows that 4 different data displays are effective for this task. Overall, participants locate task queries were answered with a high degree of accuracy (94%). However, in only 51% cases did participants choose the 'right' representation (table or matrix ER). A range of other AIVE display forms were also effective (bar and pie charts, scatterplots).

KER and database query answer performance were not significantly correlated (Figure 7), which implies that less 'graphical literacy' on the part of participants is required for this task compared to the correlate task. Database query answer performance and display selection accuracy were not significantly correlated - as would be expected on a task in which accurate responding to database queries can be achieved by using any one of 4 different information displays.
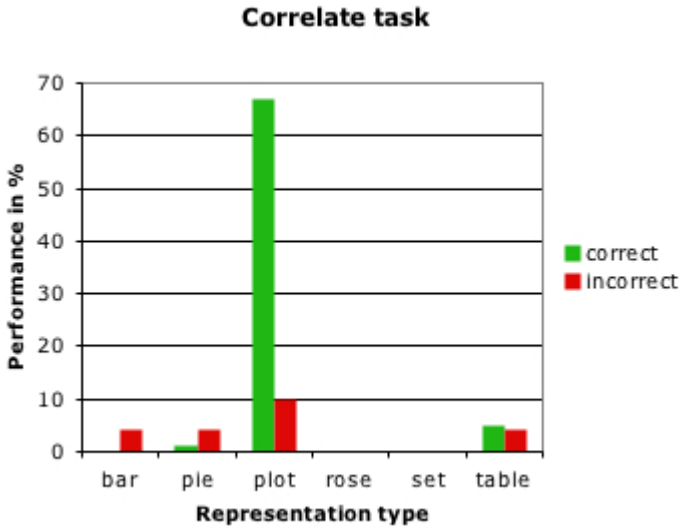
Correlate task

Fig. 4. The highly representation-specific 'correlate' task. DBQA performance in % as a function of chosen representation type.
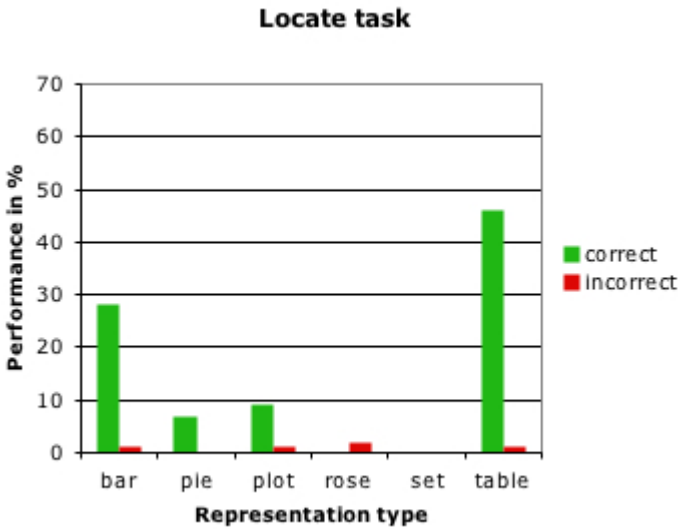
Locate task

Fig. 5. The less representation-specific 'locate' task. DBQA performance in % as a function of chosen representation type.

4   User Model Implementation

The experimental results show that particular types of data are crucial for modeling. Machine learning techniques vary in terms of their advantages and disadvantages for particular applications and domains, as well as for the underlying information or user data needed for the adaptation process ([12]). Our user model needs to reflect the relationship of KER and the varied degrees of representational specificity of the database query tasks. It also needs to track and predict participants' selection accuracy and database query answering performance for various display and response accuracy relationships within and across the various database query task types. The system should be capable of being more stringent in its recommendations to users on highly representationally-specific task types such as correlate tasks but could be able to be more lenient on more display-heterogeneous tasks.

A Bayesian network approach (*eg.* [15]) was chosen as a basis for I-AIVE's user model, because such networks are suitable, *inter alia*, for recognizing and responding to individual users, and they can adapt to temporal changes.

Table 1 shows the independent and dependent variables, that were empirically observed.

Table 1. Independent and dependent variables used for the specification of the network

Independent	
Name	*Description*
KER	This variable represents users' background knowledge of ERs, gathered in the ER tasks.
Task type	Identify, correlate, quantifier-set, locate, cluster or compare negative.
Dependent	
Name	*Description*
DSA	This variables covers users' display selection accuracy score. The value of the variable increases, if the user's selected ER is an appropriate ER for the given task. The variables decreases otherwise.
DBQA	Represents the total score of users' response to the database query question. It increases, if the response with the chosen ER was correct. If an incorrect response is given the variable decreases.
DSL	Time to select a representation in milliseconds.
DBQL	Time to answer the question on each trial in milliseconds.

The structure of a simple Bayesian network based on the experimental data can be seen in figures 6 and 7. The correlations between the independent and dependent variables are represented in this network. Figure 6 presents a graph for the highly representation-specific 'correlate' task, and figure 7 shows a graph for the less representation-specific 'locate' task.

The structure of the network represents the relationships between the independent/dependent variables. For example the arc between DSA and DBQL
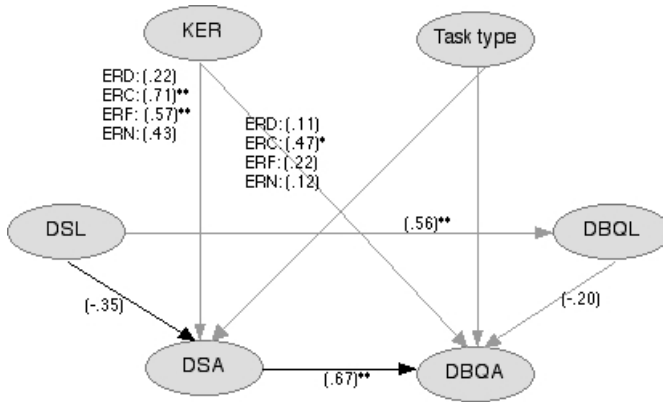
Fig. 6. Graph of a Bayesian network for the highly representation-specific 'correlate' task. Correlations between the KER tasks (ERN: naming, ERD: decision, ERC: categorisation and ERF: functional knowledge) and with the AIVE *correlate* task are shown in brackets. * = correlation is significant at the 0.05 level. ** = correlation is significant at the 0.01 level.

represents the association that good display selection results in better query performance, or the link between DSL and DSA represents that a speedy selection of a display type in AIVE is associated with a good display-type choice. The Bayesian network in I-AIVE's user model has been 'seeded' with the empirical data so that it can monitor and predict users' ER selection preference patterns within and across query types, relate query response accuracy and latencies to particular display selections and contrive query/display option combinations to 'probe' an individual users' degree of 'graphical literacy'. The empirical data is to instantiate values in the relevant conditional probability tables (CPTs) at each node of the model. The network will then dynamically adjust the CPT values and evolve individualised models for each of its users as they interact with the system. For example, for each ER selection and resulting database query performance score the corresponding CPT values will be updated and used from the system for an individual adaptation. The learned network is able to make the following inferences:

– **Predicting ER preferences and performance with uncertainty about background knowledge**
  If there is uncertainty about users' background knowledge of ERs, the system is able to make predictions about the dependent variables, through a probability distribution of each these variables.
– **Learning about users' ER preferences and performance**
  Users' ER preferences and performance can be learned incrementally, through users' interaction with the system. The network can be updated with the individual characteristics and used to predict future actions and system decisions.
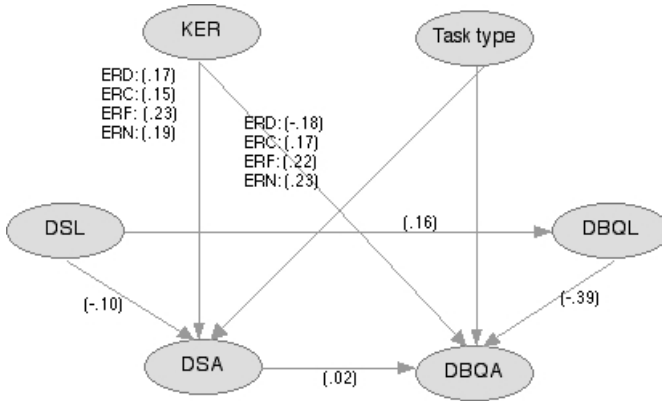
Fig. 7. Graph of a Bayesian network for the less representation-specific 'locate' task. Correlations between the KER tasks (ERN: naming, ERD: decision, ERC: categorisation and ERF: functional knowledge) and with the AIVE *locate* task are shown in brackets.

These inferences are used as a basis for our system to recommend ERs based on background knowledge, task type and ER preferences.

5   The Adaptation Process

I-AIVE's interventions consist of overt hints or advice to users and also covert adaptations such as not offering less-appropriate display forms in order to prevent users from selecting them. The system is able to adapt to the individual user in the following ways:

– **'Hiding' inappropriate display forms**
  The system varies the range of 'permitted' displays as a function of each task's ER-specificity and the user's ER selection skill.
– **Recommending ERs**
  The system will interrupt and highlight the most appropriate ER (based on the user model) if too much time is spent on selecting a representation, after learning an individual's selection display selection latency patterns.

Based on users' interactions the system will adapt the range of displays and/or recommend ERs. For example, if a user manifests a particularly high error rate for particular task/ER combinations, then the system will limit the ER selection choice and exclude the ER with which the user has problems to answer the particular database query task in the past. After users' selection display latency for particular task types has been detected, the system is able to recommend ERs if the user is unclear what kind of ER to choose and spends too much time in selection a representation.

6   Conclusion and Future Work

In this paper we described our process of how we constructed an adaptive system for external representation selection support, based on experimental data. The aim of the system is to enhance users' ER reasoning performance across a range of different types of database query tasks.

At early stages of user-system interaction, the system only offers display options that it believes lie within the users' 'representational repertoire'. After more extensive user-system interactions the user model will be updated and the system will be able to make firmer recommendations to its user.

The next step in our research will be the evaluation of I-AIVE by comparing two versions in a controlled experiment - one version with the adaptive system turned on and the other version with the user modeling subsystem turned off. The results will be used to inform the development and refinement of the user model.

References

1. Casner, A.M.: A task-analytic approach to the automated design of information graphics. PhD thesis, University of Pittsburgh (1990)
2. Cheng, P.C.-H.: Functional roles for the cognitive analysis of diagrams in problem solving. In: Cottrell, G.W. (eds.): Proceedings of the 18th Annual Conference of the Cognitive Science Society. Mahweh NJ, Lawrence Erlbaum Associates (1996) 207-212
3. Cox, R.: Representation construction, externalised cognition and individual differences. Learning and Instruction **9** (1999) 343-363
4. Cox, R., Grawemeyer, B.: The mental organisation of external representations. European Cognitive Science Conference (EuroCogSci), Osnabrück (2003)
5. Cox, R., Romero, P., du Boulay, B., Lutz, R.: A cognitive processing perspective on student programmers' 'graphicacy'. In: Blackwell, A., Marriott, K., Shimojima, A. (eds.): Diagrammatic Representation & Inference. Lecture Notes in Artificial Intelligence, Vol. 2980. Springer-Verlag, Berlin Heidelberg (2004) 344-346
6. Day, R.: Alternative representations. In: Bower, G. (eds.): The Psychology of Learning and Motivation **22** (1988) 261-305
7. Grawemeyer, B., Cox, R.: A Bayesian approach to modelling user's information display preferences. In: Ardissono, L., Brna, P., Mitrovic, T. (eds.): UM 2005: The Proceeding of the Tenth International Conference on User Modeling. Lecture Notes in Artificial Intelligence, Vol. 3538. Springer-Verlag, Berlin Heidelberg (2005) 233-238
8. Humphreys, G.W., Riddoch, M.J.: Visual object processing: A cognitive neuropsychological approach. Lawrence Erlbaum Associates, Hillsdale NJ (1987)
9. Jameson, A., Gromann-Hutter, B., March L., Rummer, R.: Creating an empirical basis for adaptation decisions. In: Lieberman, H. (eds.): IUI 2000: International Conference on Intelligent User Interfaces. (2000)
10. Kirby, J.R., Moore, P.J., Schofield, N.J.: Verbal and visual learning styles. Contemporary educational psychology **13** (1988) 169-184
11. Mackinlay, J.D.: Automating the design of graphical representations of relational information. ACM Transactions on Graphics **5(2)** (1986) 110 - 141

12. Mitchell, T.M. (eds.): Machine learning. McGraw Hill, New York (1997)
13. Norman, D.A. (eds.): Things that make us smart. Addison-Wesley, MA (1993)
14. Novick, L.R., Hurley, S.M., Francis, M.: Evidence for abstract, schematic knowledge of three spatial diagram representations. Memory & Cognition **27(2)** (1999) 288-308
15. Pearl, J. (eds.): Probabilistic reasoning in intelligent systems: Networks of Plausible Inference. Morgan Kaufmann (1988)
16. Roth, S., Mattis, J.: Interactive graphic design using automatic presentation knowledge. Human-Factors in Computing Systems (1994) 112-117
17. Stenning, K. , Cox, R., Oberlander, J.: Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. Language and Cognitive Processes **10(3/4)** (1995), 333 - 354
18. Vessey, I.: Cognitive fit: A theory-based analysis of the graphs versus tables literature. Decision Sciences **22** (1991) 219-241