

Supporting the Use of External Representations in Problem Solving: the Need for Flexible Learning Environments

Richard Cox
Department of Artificial Intelligence,
The University of Edinburgh
80 South Bridge
Edinburgh EH1 1HN
Scotland, UK

tel: +44 131 650 4449
fax: +44 131 650 4587
email: rcox@cogsci.ed.ac.uk

and

Paul Brna
Computing Department
Lancaster University
Lancaster LA1 4YR
England, UK

**Note to printer: figures printed in large size for your
convenience. Please photoreduce according to need.
The figures are intended to be in appendices as shown.**

Supporting the Use of External Representations in Problem Solving: the Need for Flexible Learning Environments

Richard Cox
Department of Artificial Intelligence
University of Edinburgh
Edinburgh EH1 1HN, Scotland

and

Paul Brna
Computing Department
Lancaster University
Lancaster LA1 4YU, England

Abstract

External representations (ERs) are effective in reasoning due to their cognitive and semantic properties. We investigated subjects' use of ERs in their solutions to analytical reasoning problems. Two sources of data were analysed. The first consisted of a large corpus of ERs ('workscratchings') used by students in their solutions to problems administered via paper and pencil tests. The second source of data was collected using *switchER*, a computer-based system that administered the problems, provided a range of ER construction environments for the subject to choose between and which dynamically logged user-system interactions. *SwitchER* was developed in order to study the *process* and time-course of ER use and to investigate the mechanisms (such as ER switching) by which subjects resolve impasses in reasoning.

The results showed great diversity of ER use across subjects, allowing the utility of various ERs under differing task conditions to be studied. The range of ERs used by subjects included plan diagrams, various tabular representations, directed graphs, set diagrams, logic, lists, and natural language. Subjects' prior knowledge of ER formalisms was shown to be an important determiner of effective ER use. A review of research on individual differences in problem solving with ERs is also provided.

The conclusions indicate the importance of a number of issues relevant to support both for analytical reasoning and for problem solving in general. These results are currently being used to inform the design of the coaching component of *switchER II* — an interactive learning environment that offers limited learner support in the use of ERs and which is intended to facilitate learning via ER switching.

Introduction

In this paper we examine the argument for supporting students in their use of multiple external representations (ERs). We do this via detailed analyses of two separate corpora. The analyses are of the use of ERs by subjects solving constraint satisfaction puzzles of the kind commonly found on the analytical reasoning section of the U.S. Graduate Record Exam (GRE).

We also elucidate the conditions under which multiple ERs enhance performance and we outline the factors that are important in subjects' decisions to switch ERs in the course of problem solving. We also

determine what kind of support might usefully be provided to subjects at each stage of problem solving.

We test the prediction, derived from specificity theory (to be introduced below), that effective solutions to indeterminate problems require the use of ERs capable of expressing abstraction. Finally, we examine the effectiveness of specific ERs which allows us to establish a principled basis for providing an improvement in the level of support over that found in current GRE coaching texts¹.

Experiment 1 examines a data corpus consisting of workscratchings from a large sample of undergraduate students who completed paper and pencil GRE-like tests. The workscratchings show great diversity of ER use across subjects, and permitted an examination of the utility of various ERs under differing task conditions.

The second corpus of data was collected in two studies that examined the *time-course* of reasoning with ERs. We sought to identify the major events at each of four stages of problem solving with ERs including how subjects resolved impasses and the role of prior knowledge about ER formalisms. The stages of reasoning with ERs parallel the four phase plan of Polya (1957) — understanding the problem, devising a plan, carrying out the plan and examining the solution (reflection). In the case of solving analytical reasoning problems with ERs, these phases correspond to: problem comprehension, ER selection, ER construction and ER use.

A computer-based system (*switchER*) was built in order to dynamically record subjects' behaviour as they reasoned with ERs. The *switchER* environment permitted the subjects to select and construct, from scratch, a wide range of ERs. We used *switchER* to determine the circumstances under which multiple representations enhance performance, to identify the factors that are important in deciding whether to switch representation and to learn about the kind of support needed by learners at various stages of ER use. Again, the results showed wide between-subject variation in the types of ER selected. Subjects often had difficulties with representation construction and with reading off results from their representations.

We also emphasise the importance of ER *construction*, since self-constructed representations have been shown to be more effective than prefabricated ones (Grossen & Carnine, 1990). There is a major issue as to the nature and extent of the support offered to subjects during the ER construction phase of problem solving. This is particularly germane in the context of solving analytical reasoning problems where a range of ERs might plausibly be used and where subjects are permitted to choose their own ER. Additionally, very little work has been done on the kinds of support necessary for ER construction in relation to the cognitive and semantic properties of ERs and subject's prior knowledge. Consequently, the investigation of the kinds of support that might be valuable is an important task, and forms a major aspect of the work reported here.

¹ This is necessary since we suspect that this advice is insufficiently detailed, and insensitive to the range of task demands that analytical reasoning problems typically present.

What this paper is not about

Mental representations may be external or internal. The nature of internal representations is hotly debated within cognitive science — advocates can be found for the position that internal imagery is causally implicated in reasoning and for the position that internal imagery is merely epiphenomenal. The focus of this paper, however, is on the use of *external* representations.

This paper is not about viewpoints (Moyse, 1989; Cheng, 1993; diSessa, 1979) — research on viewpoints is concerned with assessing the educational utility of *providing* multiple representations of domain knowledge which subjects can choose between. In contrast, the work described in this paper addresses the issue of *subject-constructed* representations of domain knowledge. Nor is this paper about mental model developments or mental model transitions (*e.g.* Bibby, 1992; White & Frederiksen, 1990). In some respects it is related to the work of Kieras & Bovair (1984) but whereas those authors were concerned with the facilitating effect of an appropriate internal representation (mental model) upon learning to operate a device, we are concerned with the role of external representations in problem solving.

Finally, this paper is not about graphic communication — it is about the spontaneous construction and use of a variety of external representational forms² in a self-communicative manner as subjects develop and examine their own ideas.

External Representations

The term ‘external representation’ (ER) refers to a wide variety of representations in both the linguistic and graphical modalities. A taxonomy of visual representations is provided by Lohse, Biolsi, Walker & Rueter (1994). They identify 11 basic categories: graphs, tables, graphical tables, time charts, networks, structure diagrams, process diagrams, maps, cartograms, icons and pictures.

The highest level modality distinction is between propositional/sentential and graphical/diagrammatic (*e.g.* Larkin & Simon 1987, Stenning & Oberlander, in press). There are also intermediate representations — tables — which exemplify matrix graphics and which combine elements of both lists and graphics.

The term ER as it is used in this paper, includes sentences of natural language, sentences of formal languages (*e.g.* first-order logic), tables, lists, graphs, maps, plans, and set diagrams.

Graphical ERs such as freehand idea sketches are an invaluable aid to creativity in design disciplines such as architecture (Goldschmidt, 1991). ERs are also an everyday phenomenon. When we buy new floor covering, we take along an annotated plan to the carpet store. If we need to communicate the directions to a party to our friends we draw a map. We take shopping lists to the supermarket. All these are examples of the use of ERs in problem solving or related activities.

In addition to ‘everyday’ examples, it is well known that ERs are effective aids to problem solving for

²Including but not limited to graphical ones.

a range of more formal problem types. These include analogical reasoning (Beveridge & Parkins, 1987), classification of hierarchical information (Greene, 1989), vector arithmetic (Katz & Anzai, 1990), algebra word problems (Singley, Anderson, Gevins & Hoffman, 1989), programming (Merrill, Reiser, Beekelaar & Hamid, 1992), logical and analytical reasoning (Gardner, 1983; Barwise & Etchemendy, 1995; Cox, Stenning & Oberlander, 1994; Stenning & Oberlander, in press), physics (Anzai, 1991) and, more generally, in scientific and mathematical discovery (Davis & Hersh, 1981).

We will now outline the properties of ERs that are responsible for their effectiveness in reasoning. These properties can be classified as ‘cognitive’ and ‘semantic’. In the sections that follow, we will raise a number of issues — a selected number of them will be returned to in later discussions.

Cognitive properties Graphical ERs such as diagrams have received the most attention in the cognition and instruction literatures. The cognitive effects of graphical ERs are to reduce search and working memory load by organising information by location. Semi-graphical ERs such as tables make information explicit and can direct attention to unsolved parts of a problem (*e.g.* empty cells of a tabular representation).

Graphical representations can also permit problem solving by facilitating perceptual judgements of a kind which are almost effortless for humans, and can act as aids to retrieval (Larkin & Simon, 1987). Generally, linguistic representations require more active search, comprehension and inference than graphical representations, though for some tasks such as the comprehension of computer programmes, textual representations have been shown to be more effective than visual ones (Green & Petre, 1992).

Graphical representations probably make use of the visual-spatial scratch-pad component of working memory (Baddeley, 1990). Exploiting this modality of working memory does not consume resources from phonological encoding via auditory channels although there is an attentional overhead. The dimensions of visual and spatial information seem to be orthogonal (Eysenck & Keane, 1990; Bryant, 1992). The spatial component of visually processed information is probably encoded automatically and independently of attention (Mandler, Seegmiller & Day, 1977). In contrast, the encoding of visual information is thought to involve attentional switching by the central executive (Baddeley, 1990). The encoding of spatial information does not depend upon the sensory channel of input. Blind people can use embossed diagrams by touch just as sighted people can use diagrams visually (Goldstein & Moore, personal communication to Stenning and Oberlander, in press). Spatial representations can also be constructed in memory from linguistic inputs such as verbal descriptions (Bryant, 1992) and are equivalent to those constructed as a result of direct visual observation.

The construction of internal representations (and probably external ones also) has a major effect upon recall. Kintsch (1989) points out that when internal models (such as mental maps) are constructed from problem descriptions, reconstruction of the text during recall is based on the model rather than the original stimulus. The same applies to representations constructed for arithmetic word problems (Kintsch, 1989).

Graphics serve to illustrate structure (Larkin, 1989). The visual form displays (the components of

objects and tools in the environment) constrain the ways in which they can be fitted together and used. Errors often result from the invisibility of function and the non-salience of form (Larkin, 1989). External graphical representations must be both well-constructed and capable of representing the information in a problem. If both criteria are met, then the high-bandwidth, rapid processing capabilities of the human visual system are exploited and very easy perceptual judgements are substituted for more difficult logical ones (Paige & Simon, 1966). Thus the computational efficiency of graphics derives from the excellent match between the structure of the data and the program(s) that operates on it. The human perceptual system is limited in the judgements it is capable of making. It is easy to judge which of two circles is larger but in the case of complicated shapes the same task can be almost impossible without measurements and calculation.

Koedinger & Anderson (1990) suggest that, when used in the development of a geometry proof (ostensibly a deductive reasoning task), accurately drawn and skillfully used diagrammatic representations are employed to aid in the generation, by induction, of possible statements that may be provable and which may also lie on the path to the problem goal. In other words, diagrammatic models, compared to syntactic representations, act to constrain the set of possibly provable statements and shift the mode of reasoning from deduction to induction.

The effectiveness of ERs may also be mediated in a way parallel to the 'self-explanation' effect (Chi, Bassok, Lewis, Reimann & Glaser, 1989). Good students tend to explain and justify actions to themselves to a greater extent than poorer students. Good students also tend to monitor their comprehension performance more accurately than poor students. Good students refer back to an example for a specific piece of information, whereas poor students refer back in order to search for a solution (Chi et al., 1989).

The self-explanation effect may also operate during translation *across* modalities (*e.g.* from verbal to diagrammatic or vice versa). In other words, the modality in which an ER is constructed (*i.e.* linguistic or graphical) may affect the operation of the processes underlying the self-explanation effect.

In the case of drawing a diagram for example, the semantic properties of graphics (discussed below) may confront the learner with his or her poor problem comprehension since, unlike language, graphics force a determinate representation that is severely limited in terms of the amount of abstraction that can be expressed (Stenning & Oberlander, in press). As Hall, Kibler, Wenger & Truxaw (1989) have observed, much of a problem solver's activity is devoted to reaching an understanding of the problem. With language, learners may re-write or translate a problem in somewhat abstract terms and may even conceal from themselves their less than complete comprehension. In other words, the modality in which an ER is constructed (*i.e.* linguistic or graphical) may affect the operation of the self-explanation effect.

Stenning & Oberlander (in press) suggest that graphical representations compel certain classes of information to be represented and are less expressive of abstraction than sentential representations. For example, a diagram typically represents a single state of affairs, and represents at least some aspects of it completely. One may say that the spoon is above the plate, and the knife is beside the plate,

but a diagram of this situation cannot be drawn without showing whether the knife is to the right or the left of the plate. The weak expressiveness of graphics makes inference more tractable. It seems likely that graphical ERs, by their limited ability to express abstraction, may provide more salient and vivid feedback to a comprehension-monitoring, self-explaining student than ‘self-talk’ in the linguistic modality. Specificity theory (to be introduced below) provides good grounds for the expectation that *cross-modal* self-explanation may be as effective as unimodal self-explanation of the kind investigated to date. Limited support for an account of this kind is provided by Katz & Anzai (1991) who report that using diagrams to represent vectors enabled the student to recognise useful calculations that she may not have otherwise discovered. Also Lewis (1989) reports that ‘translation training’ (in which students were taught about the types of statements found in arithmetic word problems) was ineffective and even counterproductive. However, when translation training was combined with training in problem-diagramming strategies, performance gains were significantly greater than from either type of training alone. The role of diagramming in the Lewis (1989) study may have been to focus attention on the task, facilitate learning-by-doing and to provide a channel through which the self-explanation effect could operate.

Semantic properties Stenning & Oberlander (in press) have proposed a theory of specificity of graphical information in which they argue that graphical representations compel the representation of certain information whereas non-graphical representations (*e.g.* sentences of natural or logical language) permit the expression of abstraction or indeterminacy. As an illustration, consider the linguistic proposition that ‘All As are Bs’. Two diagrammatic models of this premiss can be constructed using Euler’s circles. The first is the identity diagram (circles representing A and B overlap completely). A second valid representation consists of a small circle A contained within a larger circle B. A comprehensive diagrammatic representation therefore requires two models for the representation of a single (linguistic) premiss. Expressing the abstraction therefore requires multiple diagrams.

Graphical representations are said to possess the property of specificity in that where several models are possible, a single graphical representation can usually represent only one of them unless special conventions such as shading, annotation or animation are employed (Stenning & Oberlander, in press). It is the specificity of graphical representations that makes diagrams so cognitively tractable as Stenning & Oberlander argue that they share the property of specificity with the internal representations used by humans in their reasoning. They also argue that the weak expressiveness of graphical representations is more apparent to users — compared to sentential representations they ‘wear their constraints on their sleeve’.

Specificity theory has important implications for reasoning with ERs. For example, with indeterminate problems (for which more than one model can be constructed), an ER in the linguistic modality such as natural language or first-order logic may be more efficient than the construction of multiple diagrams. In

the case of natural language, Stenning & Oberlander argue that discourse conventions limit the range of interpretations of natural language and hence natural language is closer to graphical representations, in terms of specificity, than formal languages such as first order logic. Both determinate and indeterminate problems were included in the current study in order to examine some of the implications of specificity theory.

ERs and Intelligent Learning Environments

Several intelligent learning environment (ILEs) have employed graphics or graphical interfaces. Some systems, such as BRIDGE (Bonar & Cunningham, 1988), ANGLE (Koedinger & Anderson, 1993) and GIL (Merrill et al., 1992), have exploited proof-tree type graphics in the design of their interfaces, often as a means of making planning processes more salient. Other systems utilise diagrammatic representations of —*e.g.* geometry and optical problems as found in GEOMETRY and REFRACT (Reimann, 1991). Hall (1989) studied ERs as intermediate steps en route to representing algebra word problems in terms of equations. Only three systems, though — HYPERPROOF (Barwise & Etchemendy, 1995); ALGEBRA (WORD PROBLEM TUTOR — Singley et al., 1989) and GIL (Merrill et al., 1992) — are centrally concerned with graphics and reasoning.

No system to date, of which we are aware, has attempted to offer learner support in the selection, construction and use of a range of graphical (and non-graphical) representations during reasoning. Those processes are the central concerns of this paper. Many of the studies listed above have been concerned with how subjects use prefabricated ERs such as textbook illustrations or partially prefabricated diagrams. Other studies have examined subjects as they construct ERs as aids to reasoning, but none (except Katz & Anzai, 1991) have done so in the context of computer-based systems or intelligent learning environments. This paper is concerned with the spontaneous representations that individuals construct ‘from scratch’ during problem solving. Our focus is upon providing an account of the *processes* of problem comprehension, ER selection, construction and use with a view to determining the degree of computer-based support actually needed.

Previous Research

A range of studies have attempted to classify types of representation and elucidate factors associated with their effective use.

Constructivism Grossen & Carnine (1990) have demonstrated the importance of active ER construction in the domain of graphical reasoning about categorical syllogisms. They provided instruction in the use of Euler’s circles and compared a group of students who self-constructed their own ERs with a group who used only prefabricated ERs. As far as the authors are aware, Grossen & Carnine (1990) are the

only researchers in the literature to have compared the effectiveness of self-constructed diagrams with that of prefabricated diagrams. They taught 25 high school students to use a method based on Euler's circles to reason about the relationships between plant species. A computer-based tutoring system was employed. One group of students were required to construct diagrams before progressing through the resource material whereas the other group used only pre-drawn diagrams. Instruction plus self-constructed diagrams was more effective than instruction plus diagram selection. Students in the diagram-construction condition scored more highly on difficult problem types³ and demonstrated fewer trials to mastery within the course. Gains were retained for at least the duration of a two week follow up. Grossen & Carnine (1990) conclude that active drawing produces deeper processing than more passive diagram selection.

Characteristics of analytical reasoning problems A few studies have examined the kinds of analytical reasoning problems used in the studies reported here. For example, Schwartz (1971); Schwartz & Fattaleh (1972) and Polich & Schwartz (1974) have shown that the number of dimensions and the number of values along each dimension determine the difficulty of this kind of deductive reasoning problem. They have also shown that problems containing negative or implicit information (i.e. requiring inference) are more difficult than problems in which all the information is positively expressed.

Chalifour & Powers (1989) analysed the content characteristics of analytical reasoning items and found that item difficulty is predicted by a number of factors. Factors that are positively correlated with difficulty include: the usefulness of drawing diagrams (the greater the usefulness, the more difficult), the number of words in the stimulus, the number of rules and the amount of information from the rules or conditions needed for a solution. The number of unvarying assignments of entities to position⁴ was negatively correlated with item difficulty; that is, the more explicitly given determinate information, the easier the problem.

ER types and recommended representations Students sitting the analytical reasoning sections of the GRE exam are instructed "In answering some of the questions, it may be useful to draw a rough diagram" (Educational Testing Service, 1992). However, to date, there has been much folk wisdom and speculation but little empirical work on the issue of representation selection. Marzano, Brandt, Hughes, Jones, Presseisen, Rankin & Suhor (1988) suggest that categorical information is best represented using a hierarchy, and that event sequences are best represented by links in a chain or a series of boxes. They recommend a web, or 'spider map' for a major idea or concept. Similarly, the authors of a popular 'crammer' for the GRE (Brownstein et al., 1990) advise the use of lists, tables, maps and diagrams in the solution of GRE analytical reasoning problems but do not provide many guidelines for which to select

³ Syllogisms without valid conclusions.

⁴ For example, in the office allocation example, statements of the kind "Ms Green, the senior employee, is entitled to Office 5, which has the largest window."

other than to state that maps or diagrams are ‘particularly helpful’ for problems involving the physical or temporal order of things.

Schwartz (1971) and Schwartz & Fattaleh (1972) noted considerable diversity in the types of ERs that their subjects produced in the course of solving analytical reasoning problems. They classified the ERs into 5 types: matrix graphics (*e.g.* tables), informal groupings, graphics, sentence re-write and miscellaneous. Schwartz (1971) and Schwartz & Fattaleh (1972) showed that subjects who chose tabular representations in their solutions achieved significantly greater success rates than subjects who chose other kinds of ER. Tabular representations were not the most frequently *chosen* type of representation, however (Schwartz, 1971).

The problems used in the studies reported here were selected from a GRE exam ‘crammer’ (Brownstein, Weiner & Green, 1990). Brownstein et al. recommend a ‘summary chart’, a kind of pseudo- set diagram that they refer to as a ‘circle diagram’ and a ‘four-by-four grid’ (*i.e.* tabular representation) for problems 1 to 3 (Appendix A), respectively.

The Domain

The domain of the present investigation is that of analytical reasoning. Analytical reasoning problems are often best solved by constructing ERs. They 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’. Typically, the stem consists of a set of about three to seven related statements about entity relationships followed by three or more questions that test understanding of their structure and any implications. Relationships can be orderings, set membership or cause and effect. The problems are designed to not require domain knowledge for their solution and are used in a subscale of the Graduate Record Exam (GRE — see, for example, Educational Testing Service, 1992). It is claimed that the test does not require specialised domain knowledge and is relatively resistant to the effects of coaching. This claim however has to be considered in conjunction with results such as those obtained by Swinton & Powers (1983) who showed that certain types of item in the GRE analytical scale were susceptible to the effects of a ‘brief curriculum of special preparation’ (p. 104).

The three problems utilised in the studies can be found in Appendix A. All three problems contain implicit information which the solver must infer from the information given in the problem stem. The problems were ‘model’ problems. Model problems are defined here as those for which external representations are useful in finding solutions, though a stricter definition of model-based reasoning does not necessarily require the use of external representations — analogies are a rich source of models for example (Fischbein, 1987).

Two of the problems were determinate in that a single, unique model of the information given may be constructed. ERs such as a diagrams, tables, maps etc therefore facilitate finding the solution. One

problem (problem two) was indeterminate and requires the modelling of quantifier information. The three items used were analytical reasoning problems taken from a GRE test ‘crammer’ (Brownstein, Weiner & Green, 1990). Problems 1, 2 and 3 required responses to sets of 4, 4 and 5 associated questions, respectively (see Appendix A).

Individual questions within the sets associated with each problem vary quite widely in terms of task requirement. For example, in problem 3 subjects can answer the first two questions via relatively straightforward read-off from their external representations. Question 3 is more complex and requires an ordering of three entities. If the subject’s ER has not efficiently represented ‘prize order’ information then reading-off the answer to this question will become much more difficult. Question 4 is similar to questions 1 and 2. Question 5 is substantially different in that it asks the subject to re-solve the problem on the assumption that 3 of the 7 original statements in the problem are no longer available. This renders the subject’s original ER obsolete — the results of the studies reported here indicate that it is more effective to construct a new representation than to attempt to modify the original ER.

The three problems selected were chosen in order to examine the effects of two factors upon problem solving with ERs. The first factor was problem difficulty level. Problems 1 and 3 differ in terms of their level of complexity. Problem 1 is relatively easy and involves the assignment of 6 individuals to offices under various constraining conditions. Problem 3 has more dimensions than problem 1 and some information is presented negatively (*e.g.* ‘Mr Grossmans dog wins neither first nor second prize’). The four dimensions of problem 3 (prize won, dog name, dog breed, owner) each have 4 values. Problem 1, in contrast, is essentially a one-dimensional array of 6 values.

The second factor was level of determinacy. One aim of the study was to investigate the effects of problem determinacy level upon reasoning with ERs in terms of ER selection, construction and use. We wished to examine the prediction, derived from specificity theory (Stenning & Oberlander, *in press*) that effective reasoning on indeterminate problems requires the subject to use ERs capable of expressing abstraction.

Problem 2 was therefore included in order to examine the kinds of ERs that subjects used to represent quantifier information in a problem solving context. The range of representations that are capable of representing quantifier information consists of first-order logic, natural language and set diagrams. Stenning & Oberlander argue that natural language is closer to graphical representations in terms of specificity since in ordinary expository discourse, constraints on interpretation are usually present in order to reduce the range of interpretations that can be placed on utterances — these constraints render natural language less useful in some problem solving contexts than more formal languages such as logic. Set diagrams and logical formalisms require specialised knowledge if they are to be used effectively and so we hypothesised that prior experience with these formalisms would be an important prerequisite for their effective use. We predicted that the indeterminate problem would pose considerable difficulties for subjects who were unfamiliar with appropriate representational formalisms.

Sources of Empirical Data

The first of the two sources of data consisted of a large corpus of residual ‘work scratchings’ from subjects who attempted paper and pencil GRE-like tests under exam conditions.

In a second experiment (to be described below), we also examined the circumstances under which multiple representations enhance performance and to examine ER switching as a mechanism by which problem solvers resolve impasses in reasoning. The incidence of ER switching was difficult to discern from the workscratching data. The second source of data was therefore derived from studies using a prototype of a system (*switchER*) which logged subjects’ interactions dynamically and which therefore permitted analyses of phenomena that emerged over the time-course of reasoning.

Experiment 1 — ‘Workscratching’ Corpus

Subjects

Subjects consisted of first-year Philosophy undergraduates. The number of subjects responding to the three problems were 77, 91 and 51 for problems 1, 2 and 3, respectively.

Procedure

GRE-like tests of analytical and verbal reasoning were administered to four classes of students as part of an investigation into logic teaching (Cox, Stenning & Oberlander, 1994; Stenning, Cox & Oberlander, in press). The tests contained *inter alia*, the same problems as those used in the study that provided the second source of data (i.e. experiment 2 — to be described below). When answering the analytical reasoning problems, subjects were instructed “In answering some of the questions, it may be useful to draw a rough diagram”.

Results and Discussion — Workscratching Corpus

The workscratching data permitted an examination of a range of issues such as assessing the extent of between-subject variation in ER use and examining the utility of various ERs under differing task conditions.

Tables 1,2 & 3 summarise the ER behaviour on the three problems. They show the proportion of subjects using a particular ER who responded correctly to each question.

Tables 1 to 3 show that the majority of subjects used some form of ER — 91% on problem 1, 71% on problem 2 and 92% on problem 3. What is also striking is the *variety* of ER forms used — examples

Table 1: Problem 1 ‘OFFICE allocation’ (determinate)

Question	NONE	OT	PLAN
1	0.86	0.82	0.94
2	1.0	0.82	0.94
3	1.0	0.85	0.94
4	0.86	0.50	0.55
$\bar{x}$	0.93	0.74	0.84
n	7	34	36

NONE= no ER used

OT = ordered text/lists/proto-tables

PLAN= graphical plan

Table 2: Problem 2 ‘POETS’ (indeterminate)

Question	DG	LOG	NONE	SET	TABL	TEXT
1	0.75	0.61	0.73	0.57	0.50	0.55
2	1.0	0.96	0.92	1.0	1.0	0.90
3	0.25	0.58	0.54	0.57	0.0	0.40
4	0.62	0.81	0.73	0.71	0.50	0.65
$\bar{x}$	0.65	0.74	0.73	0.71	0.50	0.62
n	8	26	26	7	2	20

DG = directed graphs (lines/arrows connecting ER elements)

LOG = logic

SET = set diagrams

TABL= tables

TEXT= textual (natural language based)

Table 3: Problem 3 ‘DOG show prizes’ (determinate)

Question	CTAB	LETL	NONE	NTAB	TTAB
1	0.29	0.50	0.25	0.83	0.93
2	0.14	0.50	1.0	1.0	1.0
3	0.14	0.50	0.75	0.83	0.93
4	0.14	0.50	0.50	0.83	0.93
5	0.14	0.25	0.25	0.50	0.43
$\bar{x}$	0.17	0.45	0.55	0.80	0.84
n	7	4	4	6	30

CTAB= contingency tables

LETL= lines connecting letters

NTAB= non target (ie prize) ordered tables

TTAB= target ordered tables

are provided in Appendix B. Subjects used a very wide range of ERs, not just the ones recommended by GRE crammers. Furthermore, subjects often used ‘wrong’ representations quite successfully.

When subjects are given free choice in ER selection as they were here, the range of different and effective ERs is quite broad, especially on the more complex problems (2 and 3).

Problem 1

Problem 1 is relatively easy to solve and most subjects selected similar ERs. Thirty-four subjects constructed ERs of spatially arranged text (ordered text). Twenty-four of these were arrayed horizontally with office 1 to the left. Ten subjects arrayed the text vertically with office 1 at the top in all but one case. Thirty-six subjects produced homomorphic graphical representations (plans) in which rectangles or lines represented offices — see Figures 2, 3, 4 and 5, Appendix B. These graphical ERs were more varied than the ordered-text ERs such as Figure 6, Appendix B. Most subjects produced either a single large rectangle inside which 5 shared walls divided the offices or, alternatively, drew six discrete rectangles, one per office. Five subjects produced vertically arrayed plans (*e.g.* Figure 5, Appendix B). Three subjects drew ‘minimal’ plans consisting of horizontally arrayed vertical lines that represented office partitions (Figure 4, Appendix B). Three subjects drew ‘cubicles’ — *i.e.* 3-sided offices with an open wall — an example is shown in Figure 2, Appendix B.

Most subjects chose to explicitly number the offices in their ERs — this was evident in 72% of the ordered-text ERs and 61% of the plans.

In terms of correct responses, plans were marginally superior to ordered text (Table 1) arguably because they provided clearer read-off of information during question answering.

It is interesting to consider why ‘no ER’ was seemingly more effective for answering question 4 than tables or plans (Table 1). Question 4 requires the subject to assess the impact of hypothetical changes to the originally given information (which of the following events, occurring one month after the assignment of offices, would be most likely to lead to a request for a change in office assignment by one or more employees? 1. Ms Braun deciding that she needs silence in the office(s) next to her own, 2. etc...). The task requirement is not one that is facilitated by straightforward read-off from an ER — there is a need to reason internally about the chain of events that follow and the number of individuals affected.

It is possible that the process of comparing the relative effects of the various scenarios is best done internally. Constructing a graphical external representation may constrain the evaluation of the scenarios due to limitations on the ability of graphics to express abstraction (*i.e.* their specificity — Stenning & Oberlander, *in press*). This is an issue that warrants further research.

Of the 77 subjects, 53 (69%) constructed correct ERs, 18 (23%) constructed erroneous ERs and 6 subjects (8%) showed no trace of having used an ER at all in their solutions. Of the eighteen incorrect ERs, one error pattern accounted for a third. That error pattern involved a reversal of the office positions of two of the 3 office workers described as smokers (*i.e.* ‘Allen, Parker, White’ in offices 1,2 & 3, instead

of ‘Parker, Allen, White’). An example of this error is provided in Figure 4, Appendix B. Because of the particular questions posed, however, only one of the four questions was affected by that particular ER error. In other words, the questions vary in the extent to which they ‘depend’ upon fully correct ER construction. This is illustrated by the fact that, of the 18 incorrect ERs detected in the sample, only three were associated with scores of zero i.e. incorrect answers to all four questions. In general, most responses to questions (whether correct or incorrect) tended to be consistent with the ERs that the respondents had constructed. However, considering the 18 subjects who produced erroneous ERs, only 7 subjects gave question responses that were fully consistent with their (wrong) ERs. The remaining 11 subjects seemed to be selective about which parts of their ER they ‘believed’ in. In other words, one interpretation is that those subjects showed some awareness of their ER’s inadequacy. An interesting and seemingly paradoxical finding emerged from the analysis in relation to this point. Subjects who answered the questions in a manner that was consistent with their (wrong) ERs tended to score better than those whose answers were inconsistent with their (wrong) ERs. This may suggest that even if a learner suspects that his or her ER is incorrect, an intelligent interactive learning environment should encourage ER-congruent responding rather than the use of less systematic strategies such as trial and error or guessing.

Problem 2

Problem 2 (the indeterminate, ‘poets’ problem) involves reasoning with several quantifiers and produced, across subjects, the greatest variety in the types of ERs selected. On the other hand, a higher proportion of subjects chose not to use an ER on problem 2 than the other two problems — it is arguable that those subjects were not familiar with representations capable of expressing the necessary abstractions (i.e. quantifier relations).

The *actual* behaviour of subjects on problem 2 may be contrasted with the recommended solution strategy for that item. Brownstein, Weiner & Green (1990) recommend a kind of set diagram referred to as ‘circle diagram’. It is interesting to note that set diagrams were not the most frequently selected ER in the samples that we studied. Set diagrams were reasonably successful for the subjects that used them, however. Natural language (text) and tabular representations were less successful than other ERs on problem 2 (Table 2).

In terms of general performance across the four questions, logic was the most generally effective ER for the 26 subjects that chose to use it. An example is shown in Figure 8, Appendix B. Specificity theory (Stenning & Oberlander, in press) predicts the utility of logic as an ER modality for expressing indeterminate information. The semantics of first-order logic readily represent quantifier information and permit useful inferences.

All of the set-diagram using subjects in the current study built single models of the premisses —

characteristic⁵ diagrams (e.g. Newstead, 1989; Stenning & Oberlander, in press) were more popular single models (e.g. use of A within B diagram to represent the premiss ‘All A’s are B’s’ rather than the identity diagram — see Figures 10, 12 in Appendix B and Figure 18 in Appendix C). Three of the 7 subjects who used set-diagrams responded incorrectly to question 2 possibly because they constructed only one of many possible diagrammatic models of the information (see Figures 10, 12 in Appendix B and Figure 18, Appendix C).

The superiority of set diagrams and logic on problem 2 is not surprising since these representations are conventionally used for representing set membership and quantifier information. The result provides some empirical support, though, for the prediction (by specificity theory) that ERs capable of expressing indeterminacy either weakly (in the case of set diagrams) or strongly (as in the case of first-order logic) will be most effective in solving indeterminate problems.

As on problem 1, there was also evidence on problem 2 for the superiority of reading off answers from an incorrect ER (ER congruent responding) over complete abandonment of the representation. Consider set diagram notation — one subject (Figure 11, Appendix B) consistently drew a small circle ‘E’ within a larger circle ‘B’ to represent the first premiss of problem 2 (‘All those who enjoy...Browning also enjoy...Eliot’). This misconception resulted in that subject erroneously representing 3 of the 7 premisses. Despite this, the subject answered two of four questions correctly. It is difficult to know whether this was an error of interpretation or one of representation but the point is that flawed ERs are not totally useless, neither is reasoning with external representations necessarily totally external. Subjects who suspect that their representation is incorrect may shift their cognitive resources and rely more on internal strategies.

As in Problem 1, the second problem’s questions vary widely in terms of task requirement. Hence a single ER type is not going to be uniformly useful across all the questions in a set. Again, the data support this view. A higher proportion of subjects using directed graphs or no ER responded correctly to question 1 of problem 2 than their counterparts who chose other ER forms. On the other hand, if overall performance on this problem is considered, logic should be the best choice if one had to choose a *single* representation for use in answering all 4 questions.

Eighteen of the 20 subjects who used text to re-represent the problem information in problem 2 answered the second question of that problem correctly. In fact question 2 was relatively easy. The answer can be inferred from only 2 of the 7 premisses in the problem. Given ‘All those who like Browning also like Eliot’ (premiss 1) and ‘those who enjoy Eliot despise Coleridge’ (premiss 2) it is straightforward to conclude that Browning-likers also despise Coleridge (by transitivity). Here an external representation of any kind is not really required and so the rate of correct response is more or less ER independent.

⁵The characteristic diagram is the diagram that represents the maximum number of types of individual consistent with the premiss. The characteristic diagram for ‘All A are B’, for example, is the diagram in which a smaller circle ‘A’ is inside a larger circle ‘B’ and is not the identity relation diagram in which a circle ‘A’ totally overlaps circle ‘B’.

However, the data suggest that the differing task requirements of each question within problem 2 interact with the expressive properties of different representations. It is likely (on the basis of studies of individual differences to be discussed below) that the subjects in the current study differed in their ER modality preferences — hence it is unclear whether the effects of constraining them to use particular representations would be beneficial.

Problem 3

The third problem was a determinate problem like problem 1. Problem 3 contained 4 dimensions (prize order, dog owner, dog name, breed) with 4 values along each. As reviewed earlier, studies by Schwartz and others (1971;1972) have shown that tabular representations are associated with greater success rates than other kinds of ER. Examples of tabular representations from Experiment 1 are shown in Figures 13,14, and 15 in Appendix B. In contrast to the study by Schwartz (1971), in this study tables were the most frequently chosen ER (Table 3).

There were, however, important differences between the *types* of tables used. Contingency tables were associated with poor performance. These were matrix representations in which a separate two dimensional table was constructed for each possible pairing of variables in the problem. An example can be seen in Figure 16, Appendix B. Two of sixteen subjects in the *switchER* study (to be described below) also produced contingency tables.

The most effective tabular ER form was a target-ordered table with either the first column or row of the table representing prize order. Target-ordering means that the ER is constructed with a view to its utility at the read-off stage. In contrast, discourse ordered ER construction means that elements of the ER are produced in the order that they are listed in the problem. ‘Prize’ ordering of the information during ER construction requires more effort during construction but facilitates read-off compared with less systematically tabulated information. Non-prize ordered tables are usually constructed according to the order that the information is listed in the problem (*i.e.* discourse ordered) — see Figure 14, Appendix B. Having the columns or rows disordered with respect to prizes appears to make the representation more prone to search errors at the read-off stage.

Table 3 shows that contingency tables were pathological in that they require as much effort to construct as ‘unified’ target or discourse ordered tables but are extremely difficult to read information from when answering questions (Appendix B 2 shows examples of contingency (Figure 16) and unified (Figures 13 and 15) tables).

On the whole, there were very few incorrect ERs on problem 3. Six subjects produced incomplete ERs — most common omissions being of ‘Lad’ (name of dog winning 4th prize — 6 instances), ‘boxer’ (breed of dog winning 2nd prize — 5 instances) and ‘Edwards’ (name of owner of dog winning 2nd prize — 5 instances). Information about those three entities are merely implied in the problem and are not explicitly stated. Deductive reasoning is thus required in order to infer the relationships between them.

Only 4 of the 51 subjects actually built ERs in which the relationships between the entities of the problem were actually wrong (as opposed to incomplete). The scores for those subjects were either zero (2 cases) or 1 out of 5 (2 cases). Both of the latter subjects responded correctly to question 2. That question was relatively easy and can be answered without an ER because all but two response options can be eliminated directly from information explicitly stated in the problem stem, resulting in a 50% possibility of guessing the correct answer.

Multiple representations In a significant proportion of solutions (17% averaged across three problems), more than one representation was evident, suggesting the use of multiple ERs. Multiple ER use was associated with good performance and suggested that those subjects were skilled at matching ER properties to task requirements. Heterogeneous reasoning was associated more often with correct responding⁶ to items than incorrect responding⁷. The ratios were 13:1 (problem 1), 1:1 (problem 2) and 3.25:1 (problem 3). For subjects who used single ERs, the comparable ratios were 4.12:1, 1.83:1 and 2.3:1 for the three questions, respectively. Thus for problems one and three, the use of heterogeneous representations was associated with higher proportions of correct item responses. On problem two, selecting an ER capable of representing quantifier information was the important factor. Two examples of heterogeneous ER use are provided in Appendix E — the first shows a subject using textual notes and a tabular representation on problem 1 and the second shows the use of a restricted logic notation together with a plan representation.

Summary and Conclusions — Workscratching Data

In brief, the findings of experiment 1 might be summarised as:

- Subjects use a wide range of ERs
- Subjects sometimes use more than one representation
- Multiple ERs are effective
- No single ER is universally best for answering all questions associated with a problem
- Subjects use 'wrong' ERs successfully
- It is often better to use a partially incorrect ER than to abandon it completely
- Subjects sometimes use a formalism they don't understand
- Consistency of answers with ERs is associated with correct responding

⁶ Defined as correct answers to at least three-quarters of a problem's questions.

⁷ Correct responses to less than three-quarters of a problem's questions.

- Contingency tables yield uniformly poor performance
- The data provide support for specificity theory

The empirical support for these general conclusions will now be discussed.

As shown by the examples in Appendix B, for a given problem there is considerable inter-subject heterogeneity in ER selection. This may reflect, *inter alia*, preferences in cognitive reasoning modality (*e.g.* graphical versus sentential).

Support for specificity theory (Stenning & Oberlander, in press) was provided by the results for problem 2 (indeterminate problem). Set diagrams and logic were the most effective ERs for that problem. Both ERs are capable of expressing the indeterminacy introduced by that problem's quantifier information. Those two ER formalisms differ in their modality, however since set diagrams are graphical and logic is sentential.

Another general conclusion seems to be that subjects sometimes attempt to use an ER formalism that they clearly do not fully understand. Often this results in errors of interpretation such as that shown in Figure 11, Appendix B in which a single set diagram model for each premiss ER was constructed. The first premiss ("All those who enjoy Browning also enjoy Eliot") is represented by a small circle "EL" inside large circle "BR"). Figure 19 in Appendix D (taken from the *switchER* data to be described below) shows that when the subject does not fully understand the representational formalism, unconventional annotations — various types of arrow — are invented 'on-the-fly'. This issue was further explored in Experiment 2, in which subjects' prior knowledge about a range of ER formalisms was assessed.

A further conclusion is that, often, a single ER type is not uniformly useful across a series of questions or tasks relating to a problem. The requirements of the task are an important factor. For an ER to be effective, it must (at least): 1. express the information in the problem (expressivity/specificity — Stenning & Oberlander, in press) and 2. be useful computationally (Larkin & Simon, 1987; Stenning & Oberlander, in press) for answering a particular question (task requirement).

A single ER type may be best suited to requirement 1 but less useful for requirement 2. In some circumstances the subject could construct and use different ERs for each question. In the case of problem 2 for example, the data in Table 2 suggest that, on average, a directed graph (or no ER) is useful for question 1; any or no representation is adequate for question 2; logic, no ER or set diagrams are most appropriate for questions 3 and 4. In fact there was frequently evidence in the workscratchings of subjects having constructed and used more than one ER — *i.e.* heterogeneous reasoning in which complementary but different ERs were used in problem solving. Fourteen subjects used heterogeneous reasoning on problem 1, 2 subjects did so on problem 2 and 17 did so on problem 3.

Use of multiple representations The use of multiple representations was associated with correct responding more often than was the case for single representations.

This finding is consistent with the observations of Lesh and his colleagues in the context of mathematical problem solving. They studied primary and secondary students' use of written symbols, diagrams, manipulative models and language. For example, Lesh, Post and Behr (1987, p37) write:

... the act of representation tends to be *plural, unstable, and evolving* ... we have found that students seldom work through solutions in a single representational mode ... Instead, students frequently use several representational systems, in series and/or in parallel, with each depicting only a portion of the given situation.

Furthermore, diSessa (1979, p250), in a paper on learnable representations of knowledge, has written that:

The fundamental assumption behind ... (the) ... idea of multiple representations is that rich, overlapping collection of different views and considerations is much more a characteristic of preciseness in human knowledge than a small, tight system. In terms of problem solving the claim is that the parity of restatement or translation is as or more important to problem solving itself than the hierarchy of deduction.

More recently, the use of multiple ERs has been described as 'heterogeneous reasoning'. Barwise & Etchemendy (1995) have recently recommended the use of such reasoning with both sentential and graphical representation in their approach to teaching logic — it is embodied in their programme for teaching first-order predicate calculus 'Hyperproof'⁸.

However, empirical evidence for the effectiveness of multiple representations is thin on the ground and the circumstances under which reasoning with multiple representations improves performance are unclear. For example, should subjects shift or switch from one representation to another during the course of reasoning?

If learning material is *presented* via multiple representations, how well can students integrate the information into a coherent whole? Recent work by Schwarz & Dreyfus (1993) has identified some of the conditions necessary — they stress, however, that it is crucial to monitor and measure the degree to which students are able to integrate information since there are few guarantees that students can do this despite the provision of appropriate learning environments and tasks.

In solving analytical reasoning problems, the processes of problem comprehension, representation selection and multiple representation *construction from scratch by the student* is bound to place very high cognitive loads upon the learner.

⁸Barwise & Etchemendy's (1995) system, Hyperproof, presents information graphically and sententially. Subjects develop logical proofs by entering sentences of first-order logic into a text window and making small modifications to a pre-fabricated diagram - this can be contrasted with reasoning in situations where subjects select and construct ERs 'from scratch'.

Experiment 2 — *switchER* Studies

Experiment 2 was conducted in order to examine the time-course of reasoning with ERs, to identify the major stages and study the dynamic relationships between them and to investigate the role of prior knowledge of ER formalisms upon problem solving performance. The second experiment therefore utilised a computer-based system which provided a problem-solving environment and which dynamically logged user/system interactions.

Subjects

Two groups of subjects solved the three problems using the *switchER* system. One group consisted of subjects with strong formal backgrounds in numerate disciplines such as computer science and mathematics (subjects S1 to S8). The second group (subjects S9 to S16) consisted of visual communication (art) students. There were 5 female and 3 male subjects in group 1 and 4 female and 4 male subjects in group 2. The subjects' ages ranged from 20 to 29 years.

Procedure

The *switchER* system consists of a Macintosh Hypercard program that provides a range of simple computer-based support tools for the selection and construction of representations. The subject was able to select environments that supported the construction of diagrams, logical representations, textual representations or tabular representations. The system logged and time-stamped all subject/system interactions.

The system was designed to be both easy to use and to provide a sufficient number of ER tools with which the user may construct a wide range of representations. The user was permitted to change ERs during the course of reasoning (ER switching). The study was also expected to yield findings that would inform the design of an intelligent ER reasoning environment capable of offering support to the user at each stage of reasoning.

Subjects were given a practice question and then allowed as much time as they needed to solve the three experimental problems. All the subjects used in the study were familiar with the Apple Macintosh graphical user interface. Figure 1 shows a *switchER* screen display as subject 1 used the *switchER* diagram tool to construct a set diagram ER in the course of solving problem 2. The results of the first group and a description of the system can be found in (Cox & Brna, 1993a,b).

RepSupportProb2S1

Professor Kittredge's literature seminar includes students with varied tastes in poetry.

All those in the seminar who enjoy the poetry of Browning also enjoy the poetry of Eliot.

Those who enjoy the poetry of Eliot despise the poetry of Coleridge.

Some of those who enjoy the poetry of Eliot also enjoy the poetry of Auden.

All those who enjoy the poetry of Coleridge also enjoy the poetry of Donne.

Some of those who enjoy the poetry of Donne also enjoy the poetry of Eliot.

Some of those who enjoy the poetry of Auden despise the poetry of Coleridge.

All those who enjoy the poetry of Donne also enjoy the poetry of Frost.

Miss Garfield enjoys the poetry of Donne.

Which of the following must be true?

A. She may or may not enjoy the poetry of Coleridge.

B. She does not enjoy the poetry of Browning.

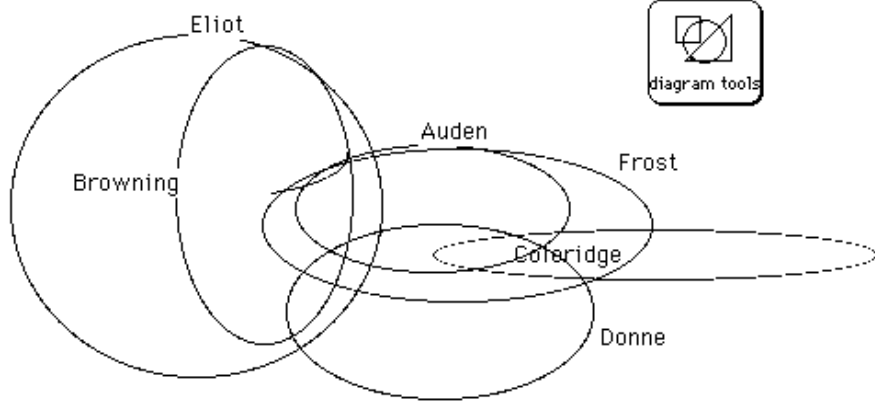
C. She enjoys the poetry of Auden.

D. She does not enjoy the poetry of Eliot.

E. She enjoys the poetry of Coleridge.

Show questions

 diagram tools



To recover the "hand" pointer, hold down the Apple key and press any letter

 Answer

Figure 1: Subject 1 uses the *switchER* diagram environment to construct a set diagram on problem 2.

Method

Taxonomy pre-test

Subjects' prior knowledge about a wide range of ERs was assessed by means of a ER-taxonomising (card sort) task administered before they attempted the reasoning problems. Subjects were given 87 numbered cards together with a pen and a pad of 'post-it' notes. Each card showed an example of one type of representation. The 87 images depicting a wide range of ER forms (diagrams, maps, tables, formulae, text etc) taken from a wide range of sources. Further details are provided in Cox & Brna (1993b).

The following instructions were read:

"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 what kind and how many categories to use. I would like you to label your categories when you have finished."

Subjects completed the task in their own time. Each card was numbered and the subjects category names and the numbers of the cards placed in each category were noted. The card stack was shuffled thoroughly between subjects.

Then, the 16 subjects each attempted the 3 GRE-style problems described above. Their interactions with the system were dynamically recorded using Farallon Inc.'s 'ScreenRecorder' utility⁹ for later protocol analysis. The screen recordings were replayed to the subjects at the end of the session and they were encouraged to verbally describe their actions and decisions at each stage of reasoning on the three problems. Also, the subjects were either videotaped (study 1) or audiotaped (study 2) during problem solving and the replay of the screen recordings.

Using this methodological approach, we were able to chart the time-course of the stages of problem comprehension, ER selection, ER construction and ER use (read-off). A range of significant ER events and issues were identified following systematic analyses of the replayed recordings.

Results and Discussion — *switchER* Studies

Statistical tests revealed that the two subject groups differed significantly from each other in terms of total time-to-solve on problem 2 ($t = -1.93, df = 14, p < .05$) and total score on problem 3 (Mann-Whitney 'U', $z = -1.99, p < .05$ — see Table 4). No other significant differences between the groups were found either in terms of score or times¹⁰. Since the group differences were not extensive, it was decided to pool the data into a single group of sixteen subjects for the purposes of this paper. Detailed analyses of the taxonomy pre-task data can be found in Cox & Brna (1993b). Here, the results will be discussed below in relation to ER selection and construction.

⁹Part of Farallon Inc.'s 'MediaTracks' package.

¹⁰Total time spent on solution, time spent on problem comprehension, time to construct the ER, use of ER (read-off) and question answering.

The *switchER* system logs permitted analyses of the *process* and time-course of problem solving. The presentation of the results will therefore be organised on the basis of problem solving stages rather than in the problem by problem format used in reporting the ‘workscratching’ results of the first experiment.

Problem Reading and Comprehension

The time-course analysis of the current study reveals that students typically spend only 10% of the total time on problem reading and ER selection. Also, subjects often read only the problem ‘stem’ but not the questions. Hence they rush into ER construction, often demonstrating what Green (1989) has termed ‘premature commitment’ to an unsuitable ER form or modality.

Previous research has highlighted the problem comprehension stage as being critical in the problem solving process (Proudfit, 1981; Hall et al., 1989; Reed & Ettinger, 1987; Schwartz, 1971; Polich & Schwartz, 1974). Proudfit (1981) compared the effect of two treatments (Polya’s problem solving model vs simple practice) upon the mathematical problem solving performance of 24 5th-grade children. Polya’s 4-phase method consists of understanding the problem, devising a plan, carrying out the plan, and reflecting upon the solution. In the ‘Polya’ condition of Proudfit’s study, children were questioned about the appropriateness of their solutions and were encouraged to discuss their strategies. The Polya method produced significant improvements at two of the 4 problem solving phases (devising plan & reflection). Among nine behaviours associated with successful problem solving was “drawing a diagram”. Proudfit (1981) reports that most errors were due to mistakes made at the comprehension phase of problem solving.

Reed & Ettinger (1987) also found that problem comprehension difficulties were a major source of error. They studied algebra word problems in a sample of 53 college students. The problems were of 2 kinds — ‘mixture’ problems and ‘work’ problems. Their research question concerned the usefulness of tables (matrix graphics) for solving those kinds of problems. Subjects were not required to construct their ER from scratch, instead they were provided with table templates which they could fill in with information from the problem as an aid to deriving an algebraic expression. Results showed that asking students to fill in a table had little effect upon their ability to construct equations. Students often failed to enter the correct values due to problem comprehension difficulties. However, subjects provided with completed tables improved in their performance but the effect did not transfer to isomorphic problems where the completed tables weren’t provided.

Polich & Schwartz (1974) discovered that the representation of implicit information (ie inferred from problem statement) was a substantial cause of error. This source of error was minimised by the use of matrix (tabular) representations however, compared to other representations. Errors of omission exceeded errors of commission by three to four times.

ER selection

ERs marked with an asterix in Table 4 indicate that the subject used an ER formalism that they did not identify accurately in the taxonomy task. This tended to result in poorer scores¹¹. The effect is particularly pronounced for the indeterminate problem where the range of useful, abstraction-expressing ERs is more restricted than for determinate problems. The mean scores on problem 2 (excluding the 2 subjects who used no ER) were 3.11 for familiar-ER users and 1.8 for unfamiliar-ER users. This difference is statistically significant ($t = 3.41, df = 13, p < .005$). There was no significant difference in terms of time to solve the problem, however. As in Experiment 1, the highest scores on the indeterminate problem tend to be associated with abstraction-expressive ERs such as those based on natural language, formal languages (logic) or graphical ERs that can represent set conjunctions and disjunctions (set diagrams). Again, this result provides support for specificity theory (Stenning & Oberlander, in press).

During the ER selection phase, subjects are confronted with the task of deciding which subset of the information given in the problem stem actually contributes to the conclusions that they wish to draw. Only some of the given information may need to be represented in the ER in order for the questions associated with the problem to be answered.

There are 2 major selection issues: dimensionality and abstraction.

Errors of dimensionality are those in which subjects choose ERs that are unsuited for the task. Good ER selection requires comprehension of the deep structure of a problem and knowledge of the range of ERs available in both the sentential modality (logical and natural languages) and in the graphical modality (set diagrams, semantic networks, conceptual graphs, directed graphs, tables, plans, maps etc) as well as comprehending the nature of the relationship between the problem's entities — *i.e.* one of group membership or causal, temporal, hierarchical, spatial etc.

For example, expressing quantifier information such as 'Some of those who enjoy the poetry of Donne also enjoy the poetry of Eliot' in a graphical ER is often difficult because of the specificity of the graphical modality (Stenning & Oberlander, in press). Subjects have to be familiar with relatively domain specific ER forms such as Euler's circles or Venn diagrams and have suitable strategies for their use in order to do this successfully. Even then, they often readily build one model of the information but rarely build alternative, but equally valid, models. Set diagrams and logic are both capable of expressing quantifier relations but require specialised knowledge for their effective use. Easier-to-use and more ubiquitous ERs such as tabular representations are much less useful. Many subjects seem to suspect that set diagrams, for example, are a useful formalism but do not fully understand the metaphor or conventions of the representation. This is not surprising in the case of set diagrams, since the strategies needed are rather complicated. The *switchER* data shows that the use of idiosyncratic annotations such as lines connecting circles of set diagrams, boundaries of overlapping circles erased and other idiosyncratic annotations

¹¹ This finding is commensurate with that of Schwartz (1993) who found that students who had prior experience of directed graphs (in the context of learning about food webs) were able to use them effectively on new problems.

(Appendix D, Subject S11) is associated with lower efficiency in terms of score per unit time (Table 4, Problem 2, subjects 11 and 14).

ER Construction

Several issues were observed at the ER construction stage. On problem 1, all of the subjects (except S2 who didn't build an ER) began by representing the 'anchor' premiss that explicitly gives the location of 'Ms Green' in the 'office 5', the 'room with the large window'.

As in experiment 1, discourse ordering of ER construction was associated with lower scores than target ordering — we consider that this was a significant contributory factor towards poor performance.

The user/system interaction logs revealed that *switching* occurred quite commonly during the ER construction phase of problem solving. Two of the sixteen subjects (S7,S14) switched on problem 1, three subjects (S14,S15,S16) switched on problem 2 and five subjects (S2,S3,S4,S12,S16) switched on problem 3. Switching represents a strategic decision by the subject to abandon the current ER and construct a new one. Switching is more costly in terms of time than of score. On problem 1 switching subjects took an average of 1556'' to solve the problem whereas non-switchers took a mean time of only 740''. In terms of answer correctness, switchers scored an average of 3/4 answers correct whereas non-switchers scored 3.4/4 correct. For problem 2 these figures were 1635'' (switchers) and 826'' (non-switchers) and 1.7/4 correct (switchers) versus 2.7/4 (non-switchers). On problem 3 switchers took an average of 1391'' and scored a mean of 3.8/5 whereas non-switchers took 831'' and scored 3.9/5.

The frequency with which subjects switched increased as a function of problem difficulty. Problems 1 to 3 increase in difficulty from a fully determinate, one dimensional array (problem 1), the need to reason with several quantifiers (problem 2) and multi-dimensionality (4 dimensions) plus implicit information (problem 3). Schwartz (1971), Schwartz & Fattaleh (1972) and Polich & Schwartz (1974) have shown that more dimensions and the presence of implicit information increases the difficulty of this type of problem. There is therefore a suggestion that subjects switch as part of their strategy for resolving impasses in reasoning. In this view, ER switching indicates an adaptive response since switchers show an awareness of poor initial ER selection and attempt to remediate the situation via rebuilding either the same ER or a different one. The re-representation of previous information is time-consuming — for each of the 3 problems the problem solving times for switchers are approximately twice those of non-switchers which indicates that the re-representation of information in the new ER is as time consuming 'second time around' as it was in the first ER.

Like the workscratching data, the *switchER* results provide support for those of Schwartz (1971) who found that tabular (matrix) representations were the most successful for determinate problems such as problems 1 and 3 in this study (which we characterise as 'determinate'). On problems 1 and 3, all of the switching subjects (except S16) switched, to tabular (or closely related 'ordered-text'/proto-tabular)

Table 4: Results

Problem 1 'Office allocation', determinate, recommended ER = map or plan

Problem 2 'Some/All Poets', indeterminate, recommended ER = circle/set diagram

Problem 3 'Dogs, owners & prizes', determinate, recommended ER = table

S	Problem 1			Problem 2			Problem 3		
	Rep. Used	#correct out of 4	Time	Rep. Used	#correct out of 4	Time	Rep. Used	#correct out of 5	Time
1	table* (1x6)	4	631''	set diag.*	4	378''	table* (3x4)	4	766''
2	none	3	370''	none	4	828''	table* (4x4)	5	945''
3	table* (6x5)	4	858''	set diag.*	4	1263''	ordered text*	5	1499''
4	table (1x6)	3	1048''	table (4x3)	2	926''	ordered text	5	1624''
5	table* (1x6)	3	342''	table* (3x2)	2	402''	table* (4x4)	4	1058''
6	table* (1x6)	3	572''	restrict.logic*	3	586''	table* (4x3)	5	613''
7	table* (6x4)	2	1438''	set diag.	2	747''	table* (4x4)	4	734''
8	table (6x1)	4	877''	list*	4	976''	ord. text*	4	1206''
9	table* (6x3)	4	598''	set diag.*	3	709''	table* (4x3)	4	435''
10	table* (6x1)	4	561''	text*	3	884''	table* (4x4)	0	572''
11	plan diag.	4	721''	pseudo set diag.	2	1096''	table (6x4)	5	1170''
12	table (6x1)	0	1203''	none	0	953''	table (4x4)	2	1003''
13	plan*	4	1077''	tree diag.*	3	986''	ord. text	1	1193''
14	table (6x2)	4	1675''	pseudo set diag.*	2	2439''	-	-	-
15	table (6x1)	4	741''	list	1	981''	table (4x4)	4	566''
16	table* (6x1)	3	761''	table (6x3)	2	1486''	text	2	1886''

NB Numbers in brackets alongside tables indicate number of cells (X by Y). An asterix indicates that the subject accurately and precisely identified that ER type on the taxonomy pre-task. 'Set' diagram refers to a representation of circles or rectangles where spatial inclusion is used as an analogy for set membership. Ord.text refers to ordered text. Note that subject 14 did not attempt problem 3.

ERs from some other kind of representation. This suggests a self-mediated improvement in ER selection strategy on the part of switching subjects, at least on the two determinate problems.

This finding is consistent with results reported by Schwartz & Fattaleh (1972). Those authors manipulated the modality in which the deductive reasoning problems were presented. A third of the subjects were presented with the problems in matrix format, a third received them in sentence format and a third received the problems in the form of a network diagram. No effect for the mode of presentation was found. When constructing ERs in their solutions, almost half of the subjects actually switched from the modality that the problem was presented in. The presentation modality most frequently changed-from was the network diagram (74% of subjects switched). The least commonly changed-from presentation modality was the matrix format (17%) with sentence format in between (57%). The most commonly switched-to ER was the matrix. Of the subjects presented with sentence problems, 59% of the switchers chose the matrix representation. For network formatted problems, 68% of subjects switched to a matrix representation in their solutions. Schwartz & Fattaleh conclude that subjects recognise the appropriateness of the matrix representation for these problems by not switching from it when problems are presented in that form and by often switching to it when the problems are not presented in matrix form.

On the indeterminate problem (problem 2), switching did not seem to be as effective in resolving reasoning impasses. The switched-to ERs for subjects S14, S15 and S16 respectively were a pseudo set-diagram, list structure and a tabular representation. None of those ERs were easily able to represent the level of abstraction necessary to solve the problem. Although S15 used a kind of set diagram, he seemed to have only a hazy notion of how set diagrams represent conjunctive and disjunctive information — his ‘set diagram’ used lines to interconnect circle segments and thereby represent relations such as ‘likes’ and ‘dislikes’ in a manner that betrayed a poor understanding of the usual spatial-inclusion-by-overlapping-circles metaphor (Appendix D).

Subjects S2 and S12 both switched ERs on problem 3 — S2 attempted problems 1 and 2 without using an ER and S12 did not use an ER on problem 2. Hence subjects who are somewhat ambivalent about using ERs may be more prone to switch and prevaricate over which type of ER to use. The computer trace revealed that S2 actually *did* use a subtle kind of ER on problem 2 — he used the multiple choice ‘check’ boxes as an elimination array — placing crosses in all of them and then unchecking them systematically as he mentally eliminated response options. This strategy was also occasionally used as a supplement to the constructed ER — some subjects check off boxes as they reject response options, other subjects first check all the response boxes and uncheck them as they eliminate response options. The check-box technique does not represent the domain information in the same way as, for example, a table or set diagram. Rather, it can be seen as a minimal *aide memoire* on progress through the problem.

Read-off from the ER and question answering stages

Some subjects attempted to begin answering questions while still engaged in ER construction — in the *switchER* data, ‘mixing’ was observed in two subjects on problem 1, in four subjects on problem 2 and in three subjects on problem 3. In seven of those nine cases, subjects constructed target ordered ERs. While reading the questions and constructing target-ordered ERs is a good strategy, the cognitive load of answering questions in addition to constructing an ER is very high and performance may be compromised.

Summary and Conclusions — *switchER* Studies

The findings of experiment 2 can be summarised as:

- Subjects often allocate too few resources to problem comprehension
- ER switching is relatively common during ER construction
- Switching is positively related to problem difficulty
- Switching extends solution time
- Switching occurs at an impasse
- Tabular (matrix) representations are best for determinate problems
- Target oriented ER construction is superior to discourse-ordered construction
- Idiosyncratic representations are associated with poor performance
- Students do better if they fully comprehend the semantics of the ER formalism that they attempt to use in their solutions

To elucidate these points, the discussion will first address the interaction of subject and task variables and will then discuss the ILE design implications of our findings, focussing particularly upon (a) ER switching as an adaptive mechanism for resolving impasses in reasoning and (b) the provision of system generated ERs (ER co-construction).

Subject Variables

Individual Differences

It is clear from the results of both the workscratching and *switchER* studies that, for any given analytical reasoning problem, there is large variation between subjects in the types and modalities of ER that they use in their solutions. There is also large variation in the kinds of ER that individual subjects use

on different problems. One source of the variation is likely to be individual differences in cognitive style. There is compelling evidence that a ‘visualiser–verbaliser’ (VV) dimension of cognitive style has a profound effect on reasoning with ERs. Several studies report that subjects who differ along the VV cognitive style dimension use different strategies in the performance of tasks such as syllogistic reasoning, sentence-picture verification and hypertext navigation (Matsuno, 1987; Macleod et al, 1978; Frandsen & Holder, 1969; Cox, Stenning & Oberlander, 1994; Stenning, Cox & Oberlander, in press; Campagnoni & Ehrlich, 1989; Riding & Douglas, 1993).

Riding & Douglas (1993) found that subjects who were classified as ‘visualisers’ in terms of their cognitive style used more diagrams in their answers to questions about the workings of a car braking system than subjects classified as ‘verbalisers’ when the stimulus information was presented in the form of text and pictures. In a condition where the material was presented in the form of text only, there was no difference between the subject groups in the use of drawings.

MacLeod, Hunt and Mathews (1978) have shown that subjects who differ in spatial ability (but not in verbal ability) differ in their strategies on a sentence-picture verification task. Of a sample of 70 subjects, 43 subjects used a linguistic strategy and 16 subjects used a pictorial-spatial strategy. Independent psychometric measures confirmed a difference between the groups in that the subjects who used the pictorial-spatial strategy performed significantly higher on a test of spatial ability. Macleod et al. (1978) argue that their results severely limit the generalisability of purely linguistic theories of performance.

Frandsen & Holder (1969) selected subjects who were matched in terms of verbal reasoning ability but who differed in terms of their spatial visualization abilities. They provided instruction in the diagrammatic representation of verbal problems¹² to half of the high spatial visualization subjects and half of the low spatial visualization subjects. The instruction consisted of explanations, step by step demonstrations and guided practice. The diagrammatic techniques taught included Venn diagrams, time lines and ‘symbolic maps’. High and low spatial visualization control groups received no instruction. Pre and post–instruction tests revealed that, compared to controls, low spatial visualization subjects benefited from diagrammatic instruction. The scores of subjects high in spatial visualization ability were not improved by the instructional intervention (they scored at close to ceiling levels on both pre and post tests). Those results suggest that this dimension of cognitive style is responsive to educational intervention and is not an immutable character trait. A further implication is that students of lower spatial–visualization ability should undergo a ‘bridging’ programme to encourage them in learning to make use of graphical reasoning techniques before they are exposed to teaching methods that exploit graphical representations.

Matsuno (1987) asked subjects to subjectively report their internal representations during a syllogism task in which subjects reasoned about the relationships between patterned geometric objects. Subjects reported three kinds of internal representation — imagined diagrams, imagined concrete figures and intuitions based on reasoning with verbal expressions. Subjects who reported imagined internal diagrams

¹² Syllogisms, time-rate-distance problems and logical-deduction problems.

performed significantly better on ‘no valid conclusion’ (NVC) syllogisms than subjects whose reported internal representations were either concrete figures or sentential. The results suggest that *internal* graphical representations have a facilitatory effect similar to that of external graphical representations. However, the use of introspective reports from the subjects is problematic and therefore the findings must be interpreted with caution. Further work is required in order to elucidate the mechanisms by which the facilitation occurs. The use of the single term ‘visualiser’ to describe subjects who habitually use internal graphical imagery may be too simplistic since Matsuno (1987) found that some ‘visualiser’ subjects reported using graphical internal imagery that was pictorial in nature but others reported using graphical internal imagery of diagrammatic representations. Thus, psychometricians may need to consider finer distinctions when characterising individual differences in mental representation modality preference.

Cox, Stenning & Oberlander (1994) and Stenning, Cox & Oberlander (in press) report that subjects who are skilled at reasoning with ERs demonstrated faster acquisition of first-order logic from Hyperproof, a computer-based learning environment that employs both graphical and sentential modalities (Barwise & Etchemendy, 1995). Hence the effectiveness of ERs in reasoning cannot be expected to be uniform across individuals and attempts to model these aspects of student behaviour pose an interesting and important challenge for cognitive science.

Superficially, it seems reasonable to assume that the VV dimension should predict the kinds of ERs that subjects use when reasoning. However, the relationship between internal (mental) representation and external representations is not well understood. More research is required on the relationship between internal and external representational behaviour. To the authors knowledge only one study (Riding & Douglas, 1993) has shown that subjects classified as ‘visualisers’ tend to use diagrammatic ERs more than subjects classified as ‘verbalisers’.

Prior knowledge

Another important source of between subject variation is prior knowledge. The *switchER* studies highlight the importance of this factor particularly when indeterminate information has to be represented. Many of the less ubiquitous ER forms require specialised knowledge for effective use. They include the types of ER that are useful for solving indeterminate problems (*i.e.* set diagrams, logic). Most subjects are capable of using tabular representations or plans effectively on less complex problems such as problem 1. With more complex, multi-dimensional problems (such as problem 3), however, not all tabular representations are equally effective. The skill of matching ER formalisms to the semantics of a problem is not often the subject of direct instruction and a subject’s repertoire may have been acquired in a relatively ad hoc fashion. For example, students may encounter semantic network diagrams only if they happen to study food webs in a biology course. Perhaps a domain-independent ‘graphics curriculum’ should be devised and generally taught? Subjects who habitually reason either internally or externally in the sentential modality may be amenable to training in the use of graphical representations. The results of the intervention study by

Frandsen et al. (1969) (reviewed in section 2) suggest that even one hour of instruction in diagramming techniques produces significant score gains in subjects classified as low in spatial visualization¹³ and it may be the case that diagrammatic modellers respond to efforts to broaden their ER repertoire by the addition of sentential representations, though this remains to be demonstrated. On the other hand, individual differences in cognitive modality preferences may militate against prescriptive advice and the development of a general ‘ER curriculum’ as the basis for instructional interventions. This issue warrants further investigation.

Task Variables

The analytical reasoning tasks posed by the constraint satisfaction puzzles vary between problems and within problems. The three problems differ in terms of level of determinacy and degree of complexity. Some, such as problem 2 in the current study, are concerned with categorical reasoning and require an ability to represent and reason with several quantifiers. The variety of ERs used on the indeterminate problem was greater than the range used on the determinate problems. This may suggest that subjects were uncertain about which type of ER to select. The results indicated that logic and set diagrams were slightly more effective, across the range of questions, than directed graphs, tabular representations or text. This finding is commensurate with specificity theory. Subjects sometimes choose ERs that cannot express crucial aspects of the problem. This is exemplified by attempts by several subjects to use a tabular representation for problem 2. Subjects also select ERs that represent the information adequately but which are computationally intractable. This is exemplified by attempts to use contingency tables on problem 3. Other kinds of problems are determinate, constraint satisfaction puzzles such as the office allocation problem (problem 1) and the more complex dog-show prizes problem (problem 3). Further common types of analytical reasoning item include family relationship (genealogical) problems that require hierarchical representations such as tree diagrams and verbal reasoning problems that require the analysis of arguments. Analytical reasoning problems thus vary in terms of their semantics, complexity, determinacy and in the extent to which particular ERs are useful in finding solutions. Skilled reasoners therefore require a large repertoire of representations, an ability to discern the ‘deep’ features of a particular problem and the skill to choose an appropriately expressive ER.

Within each problem the set of questions are heterogeneous with respect to their task demands — some require straightforward read-off from ERs, some don’t require the use of an ER and some require the re-construction of an ER. The computational efficiency of an ER varies with the task requirement (*e.g.* Larkin & Simon, 1987; Vessey, 1991; Day, 1988; Green, Petre & Bellamy, 1991). For example, the data in a spreadsheet contains precise values but is difficult to search. A bar chart generated from the spreadsheet is much more useful for comparing sets of data because our perceptual subsystems make

¹³It should be noted, though, that subjects who score poorly on psychometric tests of spatial visualization may not necessarily use the linguistic modality either internally or externally when reasoning.

visual comparison seem effortless. One representation is suited to one task requirement (read-off of precise values) whereas another is suited to a different task (comparison). There is evidence in the workscratching data that some subjects used two ERs in their solutions (heterogeneous representations), presumably capitalising on the expressive strengths of each. In the analytical reasoning domain, a system that assists in providing alternative representations should facilitate problem solving by allowing the user to select the most appropriate ER for a particular task on a question by question basis.

Questions also differ in terms of their ‘ER dependency’ — data from problem 1, for example, indicate that common ER errors have a greater effect on some questions than others. Incorrect ERs, therefore, are not necessarily useless and we present suggestive evidence that (if switching ERs or ER reconstruction is not an option) subjects in some circumstances may be better off using their incorrect ER than guessing. Ideally, subjects need to be able to switch from one ER to another as their attention turns from question to question within the problem. Switching is also to be encouraged as an adaptive response to impasses in reasoning.

Implications for the Design of an Intelligent Learning Environment (ILE)

The empirical results from both the ‘workscratching’ corpus data and the *switchER* studies strongly suggest that flexible environments are needed for supporting analytical reasoning with ERs. The *switchER* and ‘workscratchings’ data suggest a range of ways in which support could be given to someone reasoning with ERs. However, subject and task variables interact in subtle ways to make prescriptive interventions impossible. Moreover, the type of support needed varies with the stage of reasoning.

Using ERs is a multi-staged process that involves problem comprehension, an ER selection decision, ER construction and then problem solving via reading-off solutions from the ER. The distribution of cognitive effort over the stages is far from even. Compared to the stages of ER selection and using the ER, the stage of ER construction requires high levels of cognitive effort. There is a trade off between cognitive load expended upon ER construction and ease of subsequent use of the ER for reading-off problem solutions. Therefore support at the construction stage is particularly valuable. This phase of reasoning could be supported in an ILE via two mechanisms: ER switching and ER co-construction.

We identified a range of ER ‘issues’ that an intelligent computer-based support environment might detect and use in coaching interventions, though our focus here is not upon the criteria under which interventions should occur. Other studies (e.g. Burton & Brown, 1982; Breuker, 1988) have elucidated principles and criteria for intervention. Rather, our data yield a range of suggestions for the kinds of knowledge that a coaching system might draw upon. Some of the findings are currently being incorporated into the *switchER* system. The second version of the system will be capable of delivering limited user-

support.

Support at the Problem Comprehension Stage

One method for facilitating greater attention and diligence at the comprehension stage is to provide a ‘problem summary’ window into which the subject has to post information such as the number and labels of problem dimensions (*e.g.* ‘dog names’, ‘breeds’, ‘owner names’, ‘prizes’) and values along those dimensions (*e.g.* ‘Lad’, ‘labrador’, ‘Smith’, ‘2nd’). This serves to inform the system about the user’s comprehension, encourage the user to spend more time on problem comprehension and also, arguably, facilitate self-explanation on the part of the user if comprehension is incomplete. For the learner, abstracting elements of the problem into the summary window might function to increase the extent to which s/he reflects upon the problem resulting in improved comprehension and preventing premature commitment to inappropriate ERs.

A facility to support the re-ordering of problem premisses may aid comprehension by permitting related information to be juxtaposed. As an example, in the second problem, premiss re-organisation would make transitivities much more salient. For example premiss 4 “All those who enjoy the poetry of Coleridge also enjoy the poetry of Donne” could be placed immediately adjacent to premiss 7 “All those who enjoy the poetry of Donne also enjoy the poetry of Frost” to make more salient the transitive conclusion that ‘All Coleridge lovers enjoy Frost’.

Support at the Representation Selection Stage

A major finding was that subjects need to look ahead at task demands of the problem’s questions before and during ER construction and not simply represent the information in the problem stem. In other words, ER planning is required.

Vessey (1991) coined the term ‘cognitive fit’ to describe congruence between the problem solving task, the external representation and the mental representation. However, cognitive fit is difficult to achieve — it requires, *inter alia*, that the subject accurately discerns the semantics of the problem, that the subject has acquired the appropriate ER formalism and that the modality of the ER is commensurate with the subject’s cognitive modality preference. Vessey (1991), however, does not include individual differences in mental representation in her model. For a subject with a marked mental representation modality preference, there may be a good semantic fit between the external and mental representations of the problem but poor fit between the mental representation and the subject’s preferred mode of internal representation.

Subjects permitted to choose their own ERs often select poorly in that the ERs they choose cannot express important aspects of the problem. This can be due either to the semantic properties of the ER (*e.g.* graphical specificities) or due to ER selection errors of ‘dimensionality’. As mentioned earlier, common

dimensions are set membership, causality, hierarchy, temporal relations and spatial relations. It would be very difficult to develop an algorithm that could determine the ER best suited to a particular problem, task and individual. Individuals differ in terms of their prior experience and cognitive modality preferences.

Subjects should be permitted to make poor ER choices but not be ensnared by them indefinitely — a limited amount of ‘productive thrashing’ (Foss, 1987) can result in rich learning outcomes. There is a fine line, however, between intervening immediately prior to the moment of self-discovery (robbing the learner of the experience) and allowing the learner to flounder and become frustrated by repeated failure. In our view the learning outcomes from self-directed learning are much more valuable in the long term than efficient acquisition, lengthy retention and other performance based indices of learning.

The effective guidance of ER selection hinges upon what constitutes a ‘good’ representation. The results of the studies reported here suggest that the answer depends heavily on the attributes of the problem. For determinate problems, a range of ER forms ‘work’ in the sense that subjects achieve correct answers with them. Ordered texts (proto-tables) and tables were associated with respectable scores on the determinate problems. On the indeterminate problem, tables clearly were less useful — the highest scores on that item were associated with set diagrams, logic or, in one case, no representation. Indeterminate information requires ERs that are capable of expressing abstraction — usually through the use of special conventions for overcoming the specificities inherent in the case of graphical ERs. Non-graphical ERs such as language (natural or formal *e.g.* logics) are more often used to express disjunction and other abstractions (Stenning & Oberlander, in press). Unfortunately, there are few such principled sources of instruction available to students who wish to improve their skills at matching appropriate ERs to problem characteristics.

For example, students sitting the analytical reasoning sections of the GRE exam are instructed “In answering some of the questions, it may be useful to draw a rough diagram” (Educational Testing Service, 1992). However, to date, there has been much folk wisdom and speculation but little empirical work on the issue of representation selection. To assist with ER selection, an intelligent and flexible educational environment could maintain a database of those representational forms that have been empirically associated with successful solutions to each problem. This domain model can then be used to determine whether there might be a need for intervention if an unwise choice is made. An alternative strategy would be to adapt the ER selection suggestions of the ‘crammer’ and GRE practice texts and incorporate them into the system, though the basis upon which these are chosen seems somewhat arbitrary. Further work is required on this important question.

In the current study the worst scores were associated with ER selection errors of *dimensionality*. Appropriate ER selection can be considered a two-stage process. First, an accurate assessment of the relationships (dimensions) between entities must be made. Accuracy at this stage depends heavily on question comprehension. The most common dimensions for word problems are **set membership** (ER’s = natural language, logic, set diagrams); **causality** (natural language, logic, directed graphs); **hierarchy**

(natural language, tree diagrams); **temporal relations** (natural language, table, 1 or 2 dimensional diagrams, graphs) and **spatial representations** (natural language, ordered texts/tables, plans/maps).

Secondly, an optimal ER from the range available within a dimension must be selected. Failures of the first kind (*e.g.* when S4 and S5 chose tables to solve problem 2) are costly and more so if the problem contains uncertainty (indeterminacy).

Supporting ER Construction

A facility for highlighting the problem premiss being represented during a particular phase of ER construction may help to reduce the user's cognitive load. Mechanisms for re-casting negatively phrased statements, and increasing the salience of 'signal' words¹⁴ would also assist ER construction. The system should ensure that explicitly given information¹⁵ is represented early in ER construction and before inferences are made on the basis of implicit information.

Where a subject is using a specialised ER such as set diagrams or fragments of first-order logic, it is important to establish that s/he understands the semantics of the representation, especially where problems involve quantifier reasoning. This may be difficult to establish indirectly, so a straightforward query to the user may be the most practical way to implement this feature.

If the subject attempts to build a table for problems that involve quantifier reasoning (such as problem 2 in the current study), then this should be discouraged and a switch to a more appropriate ER recommended. The use of contingency tables should always be discouraged. The results suggest that target-ordered ER construction is associated with efficient problem solving. An intelligent system should be capable of detecting ER construction sequences that merely reflect the order in which information is presented in the problem (*i.e.* discourse ordering). The detection of discourse ordering should trigger a guiding suggestion to the user that the *questions* as well as the problem information should be read closely before proceeding. One option for facilitating target ordered ER construction would be to require the subject to actively make visible each question via a mouse click, thus informing the system that the problem questions have been consulted.

If the subject's ER is incomplete, and the missing information is implicitly given in the problem, then the subject should be encouraged to engage in deductive reasoning in order to complete the representation. If the un-represented information is explicitly given in the problem, draw the user's attention to it overtly by highlighting relevant regions of the interface.

Events such as repeated re-readings of the problem, multiple deletions of elements of the ER, long periods of inactivity, erroneous responses to questions should signal to the system that the user has reached an impasse. At this point a representation switch could be suggested as one of a small number of sensible

¹⁴ Words such as 'always', 'some', 'all', 'not necessarily' that indicate constraints or which flag temporal, causal or spatial relations between entities.

¹⁵ *e.g.* 'Ms Green, the senior employee, is entitled to Office 5, which has the largest window.'

options.

ER Switching In impasse situations¹⁶, a frequent response was for subjects to switch ERs. There is much scope for debate about the merits of switching. We are of the view that switching should be encouraged if it is principled and exploits the complementary expressive strengths of ERs from both modalities (*i.e.* heterogeneous reasoning as, for example, in the case of graphical and sentential reasoning). However, it can be argued that encouraging a subject to switch might lead to unproductive ‘thrashing’. There may be no reason to suspect that the subject’s performance will improve through using a second ER if they failed to use the original ER effectively. On the other hand, we observed that subjects *do* switch. Moreover, from a constructionist standpoint (*e.g.* Foss, 1987) a limited form of thrashing can be productive. Therefore we do not propose to encourage ER switching without providing support to help the student avoid the more damaging aspects of thrashing.

Two kinds of switching were apparent from the data — task requirement driven vs impasse driven. The former takes place when the subject switches ER because the task requirement changes and the previous ER is now computationally inefficient compared to some alternative. Subjects also seem to switch when they reach a reasoning impasse. In our view, switching is an adaptive response to the resolution of an impasse in problem solving and should be encouraged and supported.

Supported switching (*i.e.* switching with ER co-construction), we feel, will offer support to the subject during the extremely demanding activity of ER construction. We have demonstrated that unsupported switching is time consuming — reconstruction doubles the time to respond to the problem questions.

ER co-construction The data show that switching was costly in terms of time and it would be desirable to reduce the overhead. One method of providing support would be to implement ER co-construction. An intelligent version of the *switchER* system that was capable of co-constructing the user’s ER in a different modality would eliminate the severe time cost of ER switching. For example, if the user was attempting to represent (on problem 2) ‘All those who like Donne also like Coleridge’ by constructing a diagram in which a small circle labelled ‘Donne’ is contained within a larger circle labelled ‘Coleridge’, the system could list (in another, unseen, ER construction environment) sentences of either natural language or first-order logic corresponding to ‘Some of those who like Donne also like Coleridge’ and ‘All those who like Donne also like Coleridge’.

An intelligent learning environment for ERs could enable ER construction effort to be re-cycled by intelligently incorporating aspects of the original ER into subsequent ones (with intelligent error-spotting). Whether this is desirable is a contentious issue. Perhaps automated construction would render the subject’s role too passive — the constructionist view would favour active construction on the part of

¹⁶Identified in the *switchER* recordings as episodes where the subject failed to progress with ER construction, re-read question and erased all or part of the current ER.

the subject. However, the cognitive load of constructing even a single ER is very high. Certainly, it would be very reasonable to offer this possibility as a selectable option for the user.

It could also be argued that making partially constructed ERs available for switching-to by the subject might cause them to lose track of the relationship between the elements in the original ER and lead to confusion about the mappings from one ER to the other. It is also possible though that establishing the mappings is itself an activity that benefits the subject by encouraging reflection. This issue warrants further empirical work and will be the focus of a study utilising a new version of *switchER* (*switchER II*).

Supporting the Use of the ER in Question Answering

It is desirable and feasible that an ILE perform some checks on the consistency of the subject's responses (answers) with read-off from his/her ER. If the answer is incorrect *and* inconsistent with the subject's ER then a switch of representation could be suggested. If the subject chooses not to switch, then s/he should be warned that the ER is incorrect.

If the subject uses the response check boxes as a secondary, 'elimination array' ER then the system should ensure that the subject's ER actually provides a basis for the valid rejection of response options. Sometimes subjects use the response boxes in this manner by first checking all of them and then unchecking options as they are rejected or, conversely, by checking each eliminated option until only one blank remains and then reversing the check marks.

When the subject attempts a 'what if' question (*e.g.* question 4 of problem 1 and question 5 of problem 3) s/he should be encouraged to construct a new ER from scratch (reconstruct) rather than attempting to modify the existing ER.

Another potentially useful support feature would be to provide the subject with the opportunity to verify the accuracy of his or her ER by permitting read-off conclusions to be validated against the system's representation of the problem. This could take the form of a query to a simple Prolog 'expert' system¹⁷.

Ideally, the system should be able to check whether the user's answers to the problem questions are consistent with his or her ER. The results show that even in the case of a wrong ER, subjects should be encouraged to use it in their answers rather than abandon it completely in favour of guessing. To implement such a capability would require accurate parsing of the subject's ER — this is a difficult AI problem but one which it should be possible to address, at least to a limited extent.

Conclusion

We argue, on the basis of empirical evidence derived from two sources of data, that an adaptive ILE must be flexible if it is to provide useful support to users.

¹⁷ Not so expert: the system would be able to solve the problem but not explain how it went about choosing its own representation!

The system must be sensitive to user differences. Subjects differ greatly in the size and sophistication of their ER repertoires, often as the result of ad hoc educational experiences. They also differ in terms of their position along the ‘visualiser–verbaliser’ dimension of cognitive style and therefore in their predilection for graphical or sentential ER formalisms. Hence subjects, when reasoning, must be free to choose their preferred representational formalism and must not be constrained to use particular ERs that may be incompatible with their prior knowledge and cognitive style.

ER selection emerges as a crucial phase of reasoning — selecting an appropriate ER is often very difficult because the requirements of tasks vary considerably between and within problems. The expressive properties of the chosen ER must be capable of representing the semantics of the problem. Thus the subject must accurately discern problem characteristics such as dimensionality and level of determinacy and then select an appropriate representational formalism from his or her ER repertoire.

A significant proportion of the information in analytical reasoning problems is given implicitly and therefore must be inferred before it can be represented. An important function of ERs is to guide the search for implicit information. However, the combination of inference and ER construction places a heavy cognitive burden on the subject which intelligent support can help to alleviate.

We believe that our findings also have implications for the design of ILEs in related domains such as, for example, scientific data visualisation systems and visual programming environments.

References

- Anzai, Y. & Simon, H. [1979]** The theory of learning by doing. *Psychological Review*, **86**(2), 124–140.
- Anzai, Y. [1991]** Learning and use of representations for physics expertise. In K.Anders-Ericsson & J. Smith (Eds.) *Towards a general theory of expertise: Prospects and limits*. Cambridge University Press.
- Baddeley, A.D. [1990]** *Human Memory: Theory and practice*. Hillsdale, NJ: Lawrence Erlbaum & Associates.
- Barwise, J. & Etchemendy, J. [1995]** *Hyperproof*. CSLI Publications. Cambridge University Press.
- Beveridge, M. & Parkins, E. [1987]** Visual representation in analogical problem solving. *Memory and Cognition*, **15**(3), 230–237.
- Bibby, P.A. [1992]** Mental models, instructions and internalization. In Y. Rogers, A. Rutherford, A. and P.A. Bibby (Eds.) *Models in the mind: Theory, perspective and application*. London: Academic Press.
- Bonar, J. & Cunningham, R. [1988]** BRIDGE: Tutoring the programming process. In J. Psotka, L. Massey, and S. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned*. Hillsdale, NJ: Lawrence Erlbaum Associates, 409–434.
- Breuker, J. [1988]** Coaching in help systems. In J. Self (Ed.) *Artificial intelligence and human learning*. London: Chapman & Hall.
- Brownstein, S.C., Weiner, M., & Green, S.W. [1990]** *How to prepare for the GRE*. New York: Barron’s Educational Series.
- Bryant, D.J. [1992]** A spatial representation system in humans. *Psychology*. (electronic journal sponsored by American Psychological Association), Internet, May.
- Burton, R.R. & Brown, J.S. [1982]** An investigation of computer coaching for informal learning activities. In D.H.Sleeman & J.S. Brown (Eds.), *Intelligent tutoring systems*. London: Academic Press.
- Campagnoni, F.R. & Ehrlich, K. [1989]** Information retrieval using a hypertext-based help system. *ACM Transactions on Information Systems*, **7**, (3), 271–291.

- Chalifour, C.L. & Powers, D.E. [1989]** The relationship of content characteristics of GRE analytical reasoning items to their difficulties and discriminations. *Journal of Educational Measurement*, **26**, (2), 120–132.
- Cheng, P.C.-H. [1993]** Multiple interactive structural representation systems for education. In Cox, R., Petre, M., Brna, P. & Lee, J. (Eds.) *Proceedings of the AI-ED93 Workshop on Graphical Representations, Reasoning and Communication*, August, Edinburgh.
- Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. & Glaser, R. [1989]** Self explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, **13**, 145–182.
- Cox, R. & Brna, P. [1993a]** Reasoning with external representations: Supporting the stages of selection, construction and use. *Proceedings of the World Conference of Artificial Intelligence in Education (AI-ED93)*, Charlottesville, VA: Association for the Advancement of Computing in Education.
- Cox, R. & Brna, P. [1993b]** Analytical reasoning with external representations: The relationship between prior knowledge and performance. *Research Paper 646, Department of Artificial Intelligence, University of Edinburgh*.
- Cox, R., Stenning, K. & Oberlander, J. [1994]** Graphical effects in learning logic: Reasoning, representation and individual differences. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 237–242.
- Davis, P.J. and Hersh, R. [1981]** *The mathematical experience*, Harvester Press.
- Day, R.S. [1988]** Alternative representations. In G.H. Bower (Ed.) *The psychology of learning and motivation*, Vol. 22, New York: Academic Press, 261–305.
- diSessa, A.A. [1979]** On ‘learnable’ representations of knowledge: A meaning for the computational metaphor. In J. Lochhead & J. Clement (Eds.) *Cognitive Process Instruction*, Philadelphia, PA: The Franklin Institute Press.
- Educational Testing Service GRE Board, [1992]** *Practising to take the GRE General Test Number 9*. NJ: Warner.
- Eysenck, M.W. & Keane, M.T. [1990]** *Cognitive psychology: A student’s handbook*. Hillsdale, NJ: Lawrence Erlbaum & Associates.
- Fischbein, E. [1987]** *Intuition in science and mathematics: An educational approach*. Dordrecht, Holland: Kluwer.
- Foss, C.L. [1987]** Learning from errors in ALGEBRALAND. *Technical Report IRL-87-0003*, Palo Alto, CA: Institute for Research on Learning.
- Frandsen, A.N. & Holder, J.R. [1969]** Spatial visualization in solving complex verbal problems. *The Journal of Psychology*, **73**, 229–233.
- Gardner, M. [1983]** *Logic machines and diagrams*. Brighton, UK: Harvester Press.
- Gick, M.L. & Holyoak, K.J. [1983]** Schema induction and analogical transfer. *Cognitive Psychology*, **15**, 1–28.
- Goldschmidt, G. [1991]** The dialectics of sketching. *Creativity Research Journal*, **4**, 123–143.
- Green, T.R.G., Petre, M. & Bellamy, R.K.E. [1991]** Comprehensibility of visual and textual programs: A test of superlativism against the ‘match-mismatch’ conjecture. In J. Koermann-Bellivega, T.G. Moher & S.P. Robertson (Eds.) *Empirical studies of programmers*, Ablex.
- Green, T.R.G. [1989]** Cognitive dimensions of notations. In A. Sutcliffe & L. Macaulay (Eds.), *People and Computers V*, Cambridge University Press.
- Green, T.R.G. & Petre, M. [1992]** When visual programs are harder to read than textual programs. *Proceedings of the 6th European Conference on Cognitive Ergonomics (ECCE-6)*, Hungary.
- Greene, T.R. [1989]** Children’s understanding of class inclusion hierarchies: The relation between external representation and task performance. *Journal of Experimental Child Psychology*, **48**, 62–69.
- Grossen, G. & Carnine, D. [1990]** Diagramming a logic strategy: Effects on difficult problem types and transfer. *Learning Disability Quarterly*, **13**, 168–182.
- Hall, R. [1989]** Qualitative diagrams: supporting the construction of algebraic representations in applied problem solving. In D. Bierman, J. Breuker and J. Sandberg (Eds) *Artificial intelligence and education*. Proceedings of the 4th International Conference on AI and Education, May, 1989, Amsterdam, IOS.

- Hall, R., Kibler, D., Wenger, E. & Truxaw, C. [1989] Exploring the episodic structure of algebra story problem solving. *Cognition and Instruction*, 6, (3), 223–283.
- Jonassen, D.H. & Grabowski, B.L. [1993] *Handbook of individual differences, learning and instruction*. Lawrence Erlbaum & Associates, Hillsdale:New Jersey.
- Katz, I.R. & Anzai, Y. [1991] The construction and use of diagrams for problem solving. *Proceedings of the International Conference on Advanced Research on Computers in Education*, July, Tokyo: International Federation for Information Processing and the Information Processing Society of Japan, 191–195.
- Kieras, D.E. & Bovair, S. [1984] The role of a mental model in learning to operate a device. *Cognitive Science*, 8, 255–273.
- Koedinger, K.R. & Anderson, J.R. [1990] Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14, 511–550.
- Koedinger, K.R. & Anderson, J.R. [1993] Effective use of intelligent software in High-school math classrooms. In P. Brna, S. Ohlsson & H. Pain (Eds.), *Proceedings of AI-ED93 World Conference on Artificial Intelligence in Education*, Charlottesville, VA: Association for the Advancement of Computing in Education, 241–248.
- Larkin, J.H. & Simon, H.A. [1987] Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–100.
- Lesh, R., Post, T. & Behr, M. [1987] Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics*, Hillsdale, NJ: Lawrence Erlbaum Associates, 27–40.
- Lohse, G.L., Biolsi, K., Walker, N. & Rueter, H. [1994] A classification of visual representations. *Communications of the ACM*, 37, (12), 36–49.
- Lewis, A.B. [1989] Training students to represent arithmetic word problems. *Journal of Educational Psychology*, 81, 521–531.
- MacLeod, C.M., Hunt, E.B. & Mathews, N.N. [1978] Individual differences in the verification of sentence–picture relationships. *Journal of Verbal Learning and Verbal Behavior*, 17, 493–507.
- Mandler, J.M., Seegmiller, O. & Day, J. [1977] On the coding of spatial information. *Memory and Cognition*, 5, 10–16.
- Matsuno, T. [1987] Cognitive style and representational strategies in categorical syllogistic reasoning. *Tohoku Psychologica Folia*, 46(1-4), 97–102.
- Marzano, R.J., Brandt, R.S., Hughes, C.S., Jones, B.F., et al. [1988] Dimensions of thinking: A framework for curriculum and instruction. *Association for supervision and curriculum development*. Alexandria, VA.
- Merrill, D.C., Reiser, B.J., Beekelaar, R. & Hamid, A. [1992] Making processing visible: Scaffolding learning with reasoning-congruent representations. In C. Frasson, G. Gauthier and G.I. McCalla (Eds.) *Intelligent tutoring systems, Proceedings of the 2nd International Conference, ITS'92*, Montreal, June. Lecture notes in Computer Science 608, Berlin: Springer Verlag.
- Moyse, R. [1989] Knowledge negotiation implies multiple viewpoints. *CITE Report 63*, Open University, UK.
- Newstead, S.E. [1989] Interpretational errors in syllogistic reasoning. *Journal of Memory & Language*, 28, 78–91.
- Polich, J.M. & Schwartz, S.H. [1974] The effect of problem size on representation in deductive problem solving. *Memory and Cognition*, 2, (4), 683–686.
- Polya, G. [1957] *How to solve it: A new aspect of mathematical method*, Princeton, NJ: Princeton University Press.
- Proudfit, L.A. [1981] The examination of problem-solving processes by fifth-grade children and its effect on problem solving performance. PhD Thesis, Indiana University. *Dissertation Abstracts International*, DAI-A 41/09, March, p.3932.
- Reed, S.K. & Ettinger, M. [1987] Usefulness of tables for solving word problems. *Cognition and Instruction*, 4, 43–59.

- Reimann, P. [1991]** Detecting functional relations in a computerized discovery environment. *Learning and Instruction*, **1**, (1), 45–65.
- Riding, R. & Douglas, G. [1993]** The effect of cognitive style and mode of presentation on learning performance. *British Journal of Educational Psychology*, **63**, 297–307.
- Singley, M.K., Anderson, J.R., Gevins, J.S. & Hoffman, D. [1989]** The algebra word problem tutor. In D. Bierman, J. Breuker and J. Sandberg (Eds) *Artificial intelligence and education. Proceedings of the 4th International Conference on AI and Education*, May, Amsterdam, IOS. 267–275.
- Schwartz, D.L. [1993]** The construction and analogical transfer of symbolic visualizations. *Journal of Research in Science Teaching*, **30**, 1309–1325.
- Schwartz, S.H. [1971]** Modes of representation and problem solving: Well evolved is half solved. *Journal of Experimental Psychology*, **91**, (2), 347–350.
- Schwartz, S.H. & Fattaleh, D.L. [1972]** Representation in deductive problem solving: The matrix. *Journal of Experimental Psychology*, **95**, (2), 343–348.
- Schwarz, B. & Dreyfus, T. [1972]** Measuring integration of information in multirepresentational software. *Interactive Learning Environments*, **3**(3), 177–198.
- Stenning, K. & Oberlander, J. [in press]** A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*.
- Stenning, K., Cox, R. & Oberlander, J. [in press]** Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. *Language and Cognitive Processes*.
- Swinton, S.S. & Powers, D.E. [1983]** A study of special preparation on GRE analytical scores and item types. *Journal of Educational Psychology*, **75**, 104–115.
- Vessey, I. [1991]** Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, **22**, 219–240.
- White, B.Y. & Frederiksen, J.R. [1990]** Causal model progressions for intelligent learning environments. *Artificial Intelligence*, **42**, 99–157.

Acknowledgements

The authors thank members of the SIGNAL project for their helpful comments and advice. We especially thank Keith Stenning, Doug Merrill and the referees for their comments on the draft versions of this paper. The support of the Economic and Social Research Council UK (ESRC) is gratefully acknowledged. The work was partly supported by the ESRC funded Human Communication Research Centre and the following grants: NATO Collaborative Research grant 910954 & Joint Council Initiative UK (JCI) grant G9018050 (SIGNAL project).

A The Three Problems

Problem 1

An office manager must assign offices to six staff members. The available offices, numbered 1-6 consecutively, are arranged in a row, and are separated only by six-foot high dividers. Therefore, voices, sounds and cigarette smoke readily pass from each office to those on either side.

Ms Braun's work requires her to speak on the telephone frequently throughout the day.

Mr White and Mr Black often talk to one another in their work, and prefer to have adjacent offices.

Ms Green, the senior employee, is entitled to Office 5, which has the largest window.

Mr. Parker needs silence in the office(s) adjacent to his own.

Mr. Allen, Mr. White and Mr. Parker all smoke.

Ms Green is allergic to tobacco smoke and must have non-smokers in the office(s) adjacent to her own.

Unless otherwise specified, all employees maintain silence in their offices.

Questions:

1. The best location for Mr White is in Office 1,2,3,4,5 or 6 ?
2. The best employee to occupy the office furthest from Mr Black would be Mr Allen, Ms Braun, Ms Green, Mr Parker, Mr White ?

3. The 3 employees who smoke should be placed in Offices 1,2 & 3; 1,2 & 4; 1,2 & 6; 2,3 & 4; 2,3 & 6 ?
4. Which of the following events, occurring one month after the assignment of offices, would be most likely to lead to a request for a change in office assignment by one or more employees?

Ms Braun's deciding that she needs silence in the office(s) adjacent to her own
 Mr Black's contracting laryngitis
 Mr Parker's giving up smoking
 Mr Allen's taking over the duties formerly assigned to Ms Braun
 Ms Green's installing a noisy teletype machine in her office.

Problem 2

Professor Kittredge's literature seminar includes students with varied tastes in poetry.
 All those in the seminar who enjoy the poetry of Browning also enjoy the poetry of Eliot.
 Those who enjoy the poetry of Eliot despise the poetry of Coleridge.
 Some of those who enjoy the poetry of Eliot also enjoy the poetry of Auden.
 All those who enjoy the poetry of Coleridge also enjoy the poetry of Donne.
 Some of those who enjoy the poetry of Donne also enjoy the poetry of Eliot.
 Some of those who enjoy the poetry of Auden despise the poetry of Coleridge.
 All those who enjoy the poetry of Donne also enjoy the poetry of Frost.

1. Miss Garfield enjoys the poetry of Donne. Which of the following must be true?

She may or may not enjoy the poetry of Coleridge.
 She does not enjoy the poetry of Browning.
 She enjoys the poetry of Auden.
 She does not enjoy the poetry of Eliot.
 She enjoys the poetry of Coleridge.

2. Mr Huxtable enjoys the poetry of Browning. He may also enjoy any of the following Poets, except:

Auden	Coleridge	Donne
Eliot	Frost	

3. Ms Inaguchi enjoys the poetry of Coleridge. Which of the following must be false?

She does not enjoy the poetry of Auden
 She enjoys the poetry of Donne
 She enjoys the poetry of Frost
 She does not enjoy the poetry of Browning
 She may enjoy the poetry of Eliot

4. Based on the information provided, which of the following statements concerning the members of the seminar must be true?

All those who enjoy the poetry of Eliot also enjoy the poetry of Browning
 None of those who despise the poetry of Frost enjoy the poetry of Auden
 Some of those who enjoy the poetry of Auden despise the poetry of Coleridge
 None of those who enjoy the poetry of Browning despise the poetry of Donne
 Some of those who enjoy the poetry of Frost despise the poetry of Donne

Problem 3

In this year's Kennel Show

1. an Airedale, a boxer, a collie and a Doberman win the top four prizes in the show. Their owners are Mr. Edwards, Mr. Foster, Mr. Grossman and Ms. Huntley, not necessarily in that order. Their dogs' names are Jack, Kelly, Lad and Max, not necessarily in that order.
2. Mr Grossman's dog wins neither first nor second prize.
3. The collie wins first prize.
4. Max wins second prize.
5. The Airedale is Jack.
6. Mr. Foster's dog, the Doberman, wins fourth prize.
7. Ms. Huntley's dog is Kelly.

Questions:

1. First prize is won by: Mr Edward's dog, Ms Huntleys dog, Max, Jack, Lad ?
2. Mr Grossmans dog: is the collie, is the boxer, is the Airedale, wins 2nd prize is Kelly ?
3. Which statement correctly lists the dogs in descending order of their prizes?
 - I. Kelly; the Airedale; Mr. Edwards dog
 - II. The boxer; Mr. Grossman's dog; Jack

III. Mr. Edward's dog; the Airedale; Lad

- I only?
- III only?
- II and III only?
- II only?
- I and III only?

4. Lad:

- is owned by Mr. Foster ?
- is owned by Mr. Edwards ?
- is the boxer ?
- is the collie ?
- wins third prize ?

5. On the basis of statements 1,3,4,5 and 6 only, which of the following may be deduced?

- I. Max is the boxer
- II. The Doberman is Kelly or Lad
- III. Jack wins third prize

- I and II only ?
- I and III only ?
- II and III only ?
- I, II and III ?
- Neither I, II nor III ?

B Example ERs

1	2	3	4	5	6
Park	Alen	White	Black	Green	Braun
			7A		7A
			7W		7W
			7P		7P
			7G		7G

Figure 2: Plan, Problem 1



Figure 3: Vertical plan, Problem 1

site is in Office

at	Parkes	white	Black		
	white	Black		green	Braun
—	—	—	—	—	—
1	2	3	4	5	6

Figure 4: Minimal plan, Problem 1

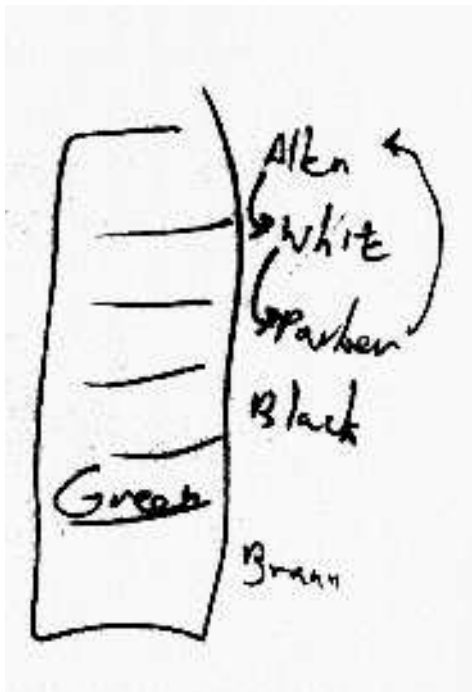


Figure 5: Vertical plan, Problem 1

A, W, P

1	2	3	4	5	6
B	W	A	P	G	Br
W	B	P	A		
P	A	W	B	G	Br
s	s	s	T		T

Figure 6: Ordered text, problem 1



Figure 7: Directed graph, Problem 2

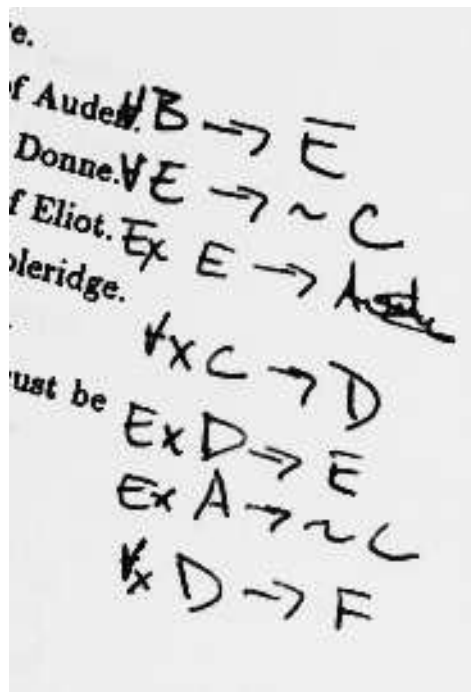


Figure 8: Logic, Problem 2

Brown	B
El	E
Col	C
Aud	A
Donne	D
Frost	F

$C = D$
 $B = E$
 $D = F$
 $E \neq C$

$SE = SA$
 $SD = SE$
 $SA \neq C$

Figure 9: Text, Problem 2

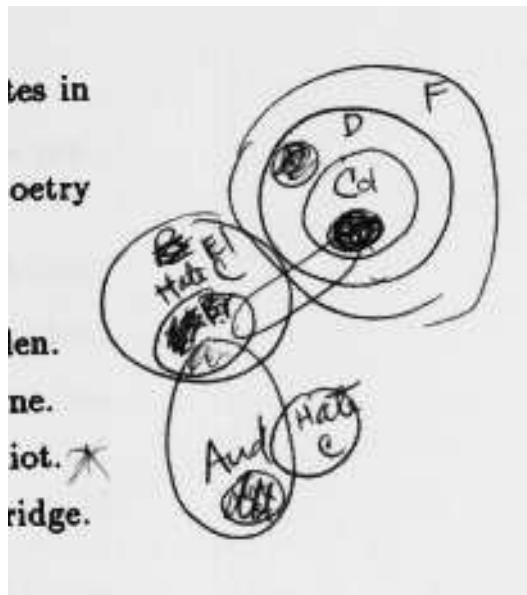


Figure 10: Set diagram, Problem 2

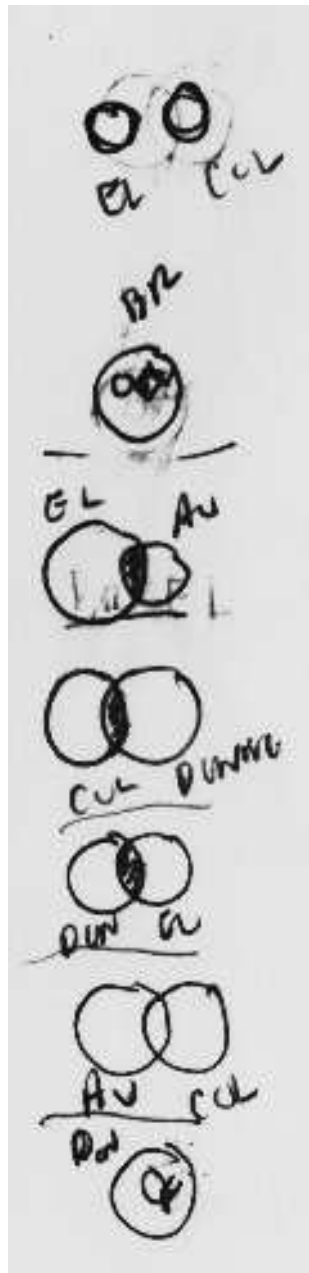


Figure 11: Set diagram (non unified), Problem 2

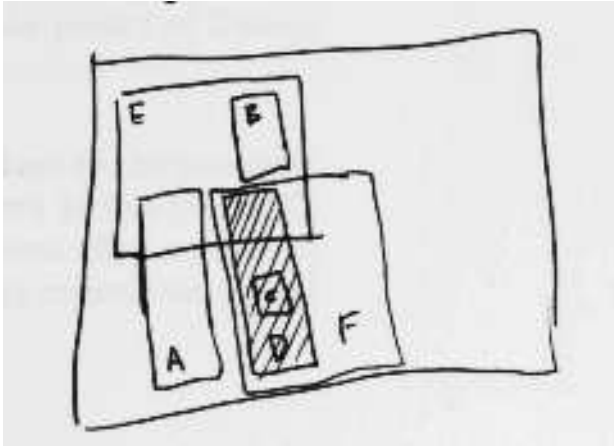


Figure 12: Set diagram using rectangles, Problem 2

an, wins fourth prize.

		1	2	3	4
Aird	D Name	Kelly	Max	Jack	Lad
	O Name	MS. Hunt	Mr. Eds	Mr. G	Mr. Foster
Jack	type	Cellie	boxer	Aird	Oldman

Figure 13: Tabular representation, Problem 3

Colbie Kelly Hurley
Baker Max Edwards
Airedale Jack Grossman
Doberman Lad Foster

A	B	a	b	a, b, c, d
H	E	G	F	E, F, G, H
K	M	J	L	J, K, L, m

Figure 15: Tabular representation, Problem 3

	E	F	G	H	I	2	3	4	J	K	L	M
A		X		X	X				⊙	X	X	X
B		X			X				X			
C		X			⊙	X	X	X	X			
D	X	⊙	X	X	X				X			
1		X	X									
2		X	X									
3	X	X	⊙	X								
4	X	⊙	X	X								
J				X		X						
K	X	X	X	⊙		X						
L				X		X						
M				X	X	⊙	X	X				

Figure 16: Contingency table, Problem 3

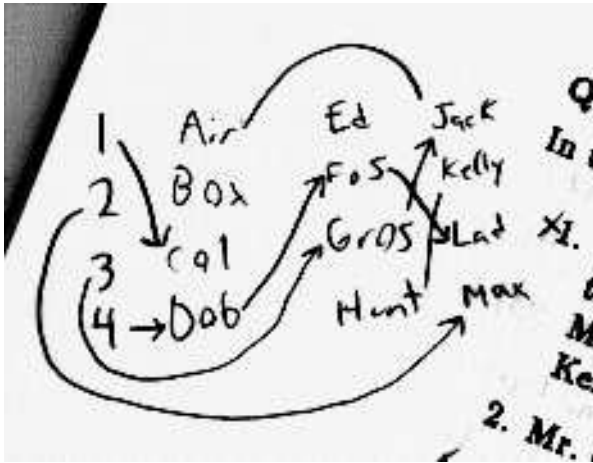


Figure 17: 'Letters and lines' representation, Problem 3

C Single Models — Ignoring Alternatives

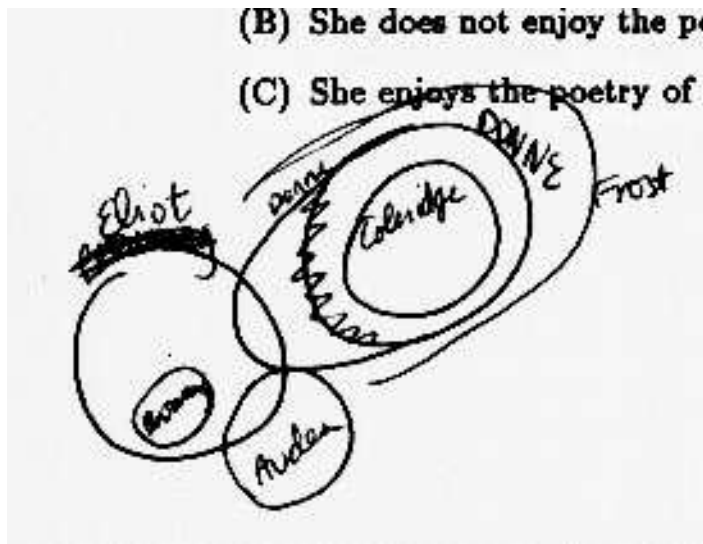


Figure 18: Set diagram, Problem 2 — only one of many possible models of problem

D 'Invented' Annotations

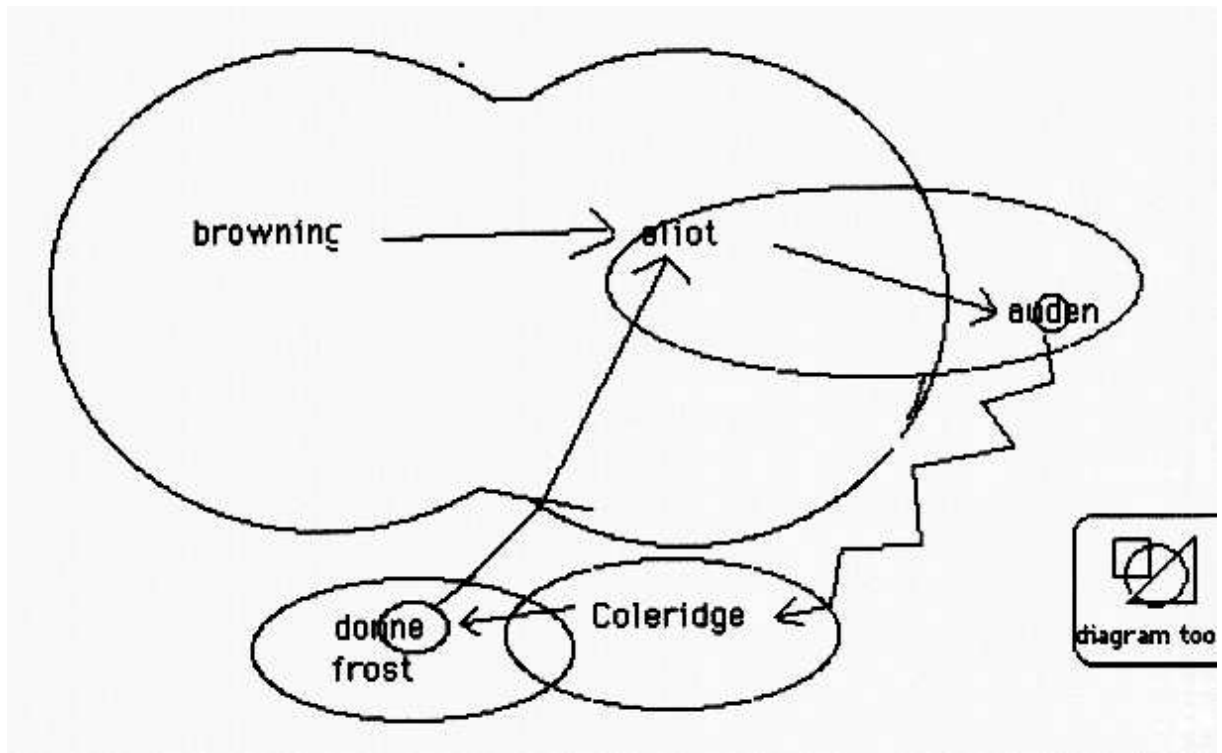


Figure 19: Subject using *switchER* diagram tool to construct 'set diagram' — annotated with various arrowed lines

E The Use of Multiple Representations

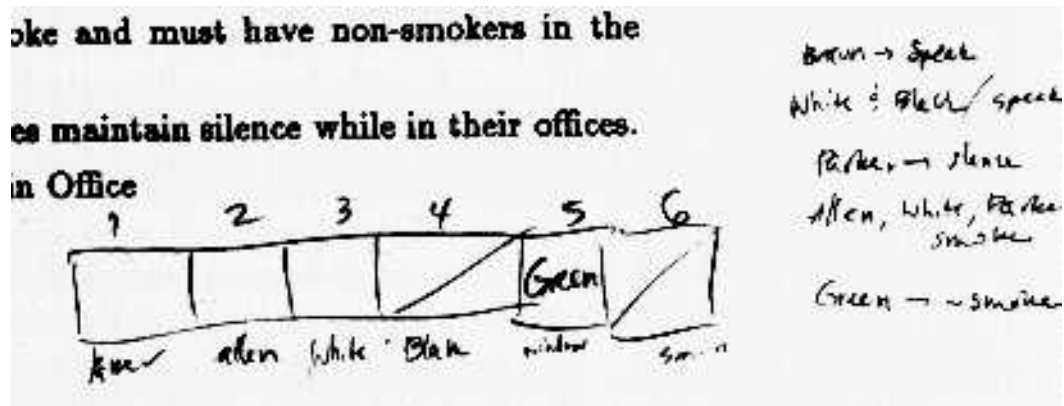


Figure 20: Use of multiple representations on Problem 1 — restricted logic plus plan