Analytical reasoning with multiple external representations

Richard Cox

Ph.D.
University of Edinburgh
1996

Abstract

This thesis presents work on analytical reasoning with external representations (ERs) using problems similar to those used in the US GRE college-entrance examination. The work investigates the factors associated with effective ER use in situations where subjects select, construct and reason with their own ERs. Practically all previous work has tended to focus solely upon performance rather than process. In this thesis the emphasis is upon cognitive processes during the entire time-course of reasoning with ERs, from problem comprehension through to answer selection. A background to the work is provided by 2 comprehensive reviews of: 1.) previous research on ERs and reasoning and 2.) the cognitive and semantic properties of ERs.

Results from three empirical studies are reported. The first study examined a large corpus of 'workscratchings' produced by subjects as they solved paper and pencil-based analytical reasoning problems under test conditions. The workscratching ERs showed great diversity between and within subjects and across a range of problems. They included lists, various kinds of table, set diagrams, node and arc diagrams, first-order and propositional logic, plans and natural language. It is shown that problem-solving performance is related to the type of ER used in the solution.

The second study utilised a computer-based system (*switchERI*). The system administered analytical reasoning problems and provided a range of ER construction environments for the subject to choose and switch between. User-system interactions were recorded dynamically during problem solving. This methodology permitted micro-analyses of the cognitive events at each stage during the time-course of problem solving. A process account of analytical reasoning with ERs is developed in which five major stages are identified - problem comprehension, ER selection, ER construction, read-off from the ER and answer selection/responding. A range of common slips and misconceptions are identified at each stage. The results show, *inter alia*, that subjects whose responses are consistent with their ERs perform better than subjects whose responses are inconsistent with their ERs *even if the ER is partially incorrect*.

The data from the workscratching analysis and *switchERI* study informed the design of *switchERII*, a second system. *SwitchERII* incorporates a representation of the semantics of Euler's Circles, dynamically parses the user's representation and provides feedback and advice. A third study was conducted with the *switchERII* system.

Few, if any, studies to date have attempted to relate subjects' prior knowledge of ER formalisms to their reasoning performance. Subjects' prior knowledge of ER formalisms was assessed in both *switchER* studies. It was observed that subjects' performance on representation *interpretation* tasks does not necessarily predict their performance in conditions where they *select* and *construct* their own representations. The reasons for the decoupling are discussed.

Data from all three studies show that subjects often utilise multiple representations in their solutions, either concurrently or serially via ER switching. Two distinctly different types of switching were observed. One kind ('thrashing') is associated with poorer performance and reflects less comprehensive prior knowledge, inability to select an appropriate ER and hazy problem comprehension. Judicious switching, on the other hand, is associated with high levels of problem comprehension and skilled matching of the ERs' properties to changing task demands.

It is claimed that effective reasoning with ERs involves complex interactions between at least three factors: (a.) within-subject variables such as the subject's representational repertoire (prior knowledge) and representational modality preferences (cognitive style); (b.) skill at overcoming a variety of barriers to comprehension and an ability to discern the salient attributes and characteristics of different problem types and (c.) an understanding of the semantic and cognitive properties of graphical and non-graphical ERs coupled with an ability to match those properties to the problem's task demands. It is suggested that the role of *externalisation* in reasoning with ERs may be to facilitate the swapping of information between cognitive subsystems. A mechanism by which the use of diagrammatic ERs may facilitate self-explanation is also proposed.

The thesis concludes with an argument in favour of a domain-independent 'ER curriculum' . It is suggested that direct instruction in the use of a range of ERs might equip students with wider representational repertoires and hence allow them more scope to indulge their representational preferences. Finally, several directions for future work are proposed. These include extending the representational semantics of *switchERII*, evaluating various types of system feedback and implementing a mechanism for checking for slips during read-off from ERs.

Acknowledgements

This thesis is dedicated to my family.

Many individuals are owed my thanks and gratitude for helping bring this thesis to its conclusion.

First, profound and special thanks are due to my primary supervisor, Dr. Paul Brna, for many hours of patient help and supervision and for being a constant source of advice and encouragement.

Secondly, I would like to thank my work colleagues in the Graphics and Language Working Group of the Human Communication Research Centre - Professor Keith Stenning (my second supervisor), Drs. Corin Gurr, Robert Inder, John Lee, Jon Oberlander, and Richard Tobin for innumerable stimulating discussions, some in formal settings but mostly in our offices and in the tea room! I also acknowledge the invaluable assistance that Robert and Richard gave me with Prolog and LaTeX.

Thanks, for many stimulating discussions in Cellar One (and more formal settings), to my friends and colleagues in the A.I. and Education and COED groups, especially Maziah Aziz, Diana Bental, Susan Bull, David Duncan and Judith Good. I also wish to thank everyone who participated in my experiments.

I must also express gratitude to Dr. Helen Pain and Professor Rosamund Sutherland for agreeing to formally examine the thesis.

I would also like to acknowledge the continuing influence of my Masters degree supervisor, Dr. Tom Matyas (School of Behavioural Health Sciences, La Trobe University, Australia). I remain indebted to him for any skills in research methods and statistics that I may possess. Thanks too to Dr. Geoff Cumming of the Psychology Department for many useful and thought-provoking discussions.

Finally, I am indebted to Dr. Carmel Lum, my dear friend and partner of 19 years, whose love and constant encouragement enabled me to realise my aspirations.

[redacted]

Some of the work in this thesis was published prior to thesis submission. Some of the analyses of 'workscratchings' presented in Chapter 5 and the *switchERI* study described in Chapter 6 appeared as:

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

Some of the workscratching analyses reported in Chapter 5 were also reported in:

Cox, R., Stenning, K. & Oberlander, J. [1995] The effect of graphical and sentential logic teaching on spontaneous external representation. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, **2(4)**, 56–75.

Several conference papers were also based on the work:

Cox, R.J. & Brna, P. [1993] Reasoning with external representations: Supporting the stages of selection, construction and use. *Proceedings of the World Conference on Artificial Intelligence in Education (AI-ED93)*, Charlottesville, VA: Association for the Advancement of Computing in Education, 185-192.

Cox, R.J. & Brna, P. [1993] Analytical reasoning with external representations: The relationship between prior knowledge and performance. In R.Cox, Petre, M., Brna, P. and Lee, J. *Proceedings of the workshop on Graphical representations, reasoning and communication held at World Conference on Artificial Intelligence and Education (AI-ED93)*, August, Edinburgh, 33-36.

Cox, R. & Draper, S. [1995] External representations, choice and task effectiveness. *4th ESRC Seminar on internal and external representations (Representations and Constructivism)*, Institute of Education, London, 21-22 April, 1995.

Cox, R. [1996] The role of externalisation in reasoning with self-constructed representations. Presented at the IEE Colloquium on Thinking with Diagrams, Savoy Place, London, January 18. *IEE Digest No. 96/010*, London: Institute of Electrical Engineers, 1–7.

The following Departmental papers were also based on the work:

Cox, R.J. & Brna, P. [1993] The relationship between prior knowledge of external representations and analytical reasoning performance: Implications for the design of a learning environment. *Research Paper RP-646*, Department of Artificial Intelligence, University of Edinburgh.

Cox, R. & Brna, P. [1994] Analytical reasoning with external representations: Supporting the stages of selection, construction and use. *Research Paper RP-686*, Department of Artificial Intelligence, University of Edinburgh.

Contents

List of Figures

Chapter 1

Introduction

What this thesis is about

This thesis examines the use of graphical and linguistic external representations (ERs) by subjects solving constraint satisfaction puzzles (analytical reasoning problems). The work addresses the role of ERs in problem solving and the focus is upon external representation as an *activity* of the reasoner.

The work adopts a constructivist perspective and examines the processes of ER selection, construction and use by subjects who spontaneously choose and build their own representations. The intention is to develop a *process* account of reasoning that moves beyond those based solely on outcome (performance) data.

Analytical reasoning problems are constraint satisfaction puzzles – they are computationally tractable and programs in languages such as Prolog can be written to solve them. This is an important characteristic since one of the aims of the work reported here was to gather information in order to inform the design of an intelligent learning environment capable of providing support to problem solvers using ERs in their solutions.

Analytical reasoning can be characterised as a five component process in which the subject:

1. reads and comprehends the problem
2. selects an ER

3. constructs the ER
4. uses the ER to read off solutions
5. responds to (answers) the problem questions

It should be emphasised that the components are not sequential or fixed – they are iterative. Problem solving evolves as the process of representation proceeds.

At each stage of ER selection, construction and use, the interplay of three factors should be considered. The first factor is concerned with cognitive processes within the subjects — these include individual differences in cognitive style and prior knowledge. Secondly, the cognitive and semantic properties of various types of ER must be taken into account. The third important factor consists of the task demands of the problem and the linguistic/structural features that are associated with 'puzzle' problems.

Switching between different representations and the use of multiple representations are also examined since they are strategies used by subjects in order to resolve impasses in reasoning and in response to changing task demands. The term 'switching' refers to situations where, during problem solution, the subject constructs a representation in one modality but subsequently builds another, different, ER.

The results of three empirical studies are reported in Chapters 5, 6 & 7. The first study examined a large corpus of 'workscratchings' produced by subjects as they solved paper and pencil-based analytical reasoning problems under test conditions. The results informed the design of *switchERI*, a computer-based system which administered analytical reasoning problems and which provided a range of ER construction environments for the subject to choose and switch between.

Data from the workscratching analysis and *switchERI* study informed the design of *switchERII*, a second system[1]. *SwitchERII* incorporates a representation of the semantics of Euler's Circles, dynamically parses the user's representation and provides feedback and advice.

The use of interactive learning environments in data collection was motivated by a

---

[1] This process might be termed 'iterative design', but this is a secondary theme which will not be developed in this thesis.

number of factors. The first of these was a wish to develop a *process* account of ER use, rather than to simply analyse the residual ER *products* of reasoning. As Vygotsky observed, when signs (language, diagrams *etc.*) are included in an action, they do more than facilitate manoeuvres that are impossible in the absence of the sign system. They fundamentally transform the action (Wertsch & Toma, 1995; Wertsch, 1991). Therefore, a major reason for using interactive learning environments was that the *switchER* systems facilitated the study of the action-transformation aspects of ER use. This was achieved by means of recording detailed, time-stamped, dynamic protocols of the students' interactions with the systems.

Another factor concerned the role of audience and its effect upon the use of ERs. ERs produced for private use differ from those produced for others. For example, a diagram drawn for purely personal use may not be annotated linguistically, whereas diagrams produced for publication almost always are. Subjects sometimes produce and use ERs purely privately (*e.g.* 'workscratchings' produced on scrap paper whilst problem solving under exam conditions). Mostly, however, a person's ERs are also seen by others, and may be commented upon by others, as in classroom settings. ER use often represents socially shared cognition. *SwitchERII* approximates ER use in naturalistic settings since the system parses the student's diagram and provides feedback.

In naturalistic educational settings, human tutors may or may not notice errors in students' representations and they may or may not provide feedback to the student. The inability to control curriculum delivery and feedback to the student poses severe problems for educational researchers who wish to study classroom behaviour. Hence, another advantage of using interactive learning environments for data collection is that the antecedants and consequences of particular representational behaviours can be precisely specified and that the system's responses are consistent. Thus the use of interactive learning environments permits the study of ER use in quasi-naturalistic settings, but with considerable methodological advantages. As Lepper & Gurtner (1989) observe, computers provide a 'particularly propitious vehicle for examining a number of classic issues in education in a more controlled and precise fashion than has been possible in the past' (p. 176).

What this thesis 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 thesis, however, is on the use of *external* representations in reasoning.

This work is not about viewpoints (Moyse, 1989; Cheng, 1993) — 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 thesis addresses the issue of *subject-constructed* representations of domain knowledge.

Nor is this thesis 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, this thesis is concerned with the role of external representations in problem solving.

The concern here is not with rare, highly creative and 'radical' re-representations. Peterson (1994) and Norman (1993) present examples of that kind of problem re-representations such as, for example, the re-representation of nine-card number scrabble[2] as noughts and crosses played over a 3 by 3 magic square (Newell & Simon, 1972). Such innovative and creative re-representation improves performance drastically. In the work reported here, however, the interest is in less radical re-representations of problem information.

Another category that is not addressed by this thesis is that of 'everyday' representational activity *i.e.* drawing, graphing or diagram-making associated with study. Those representations were the subject of a series of studies by Van Sommers (1984), which will briefly reviewed here, since they serve to elucidate some of the ways in which ERs

---

[2] Nine cards, numbered one to nine, are placed face up. Two players draw cards alternately. First player to pick 3 cards that sum to 15 wins.

produced for private use differ from those produced in the knowledge that they will be seen by others (this notion of 'audience' will be discussed further in Chapter 4). Van Sommers (1984) asked 86 adult subjects to recall their recent 'everyday' external representations. For females the most frequently reported five categories of private drawings were, in rank order, doodling, defacing pictures (*e.g.* in magazines), drawing an imaginary person, expressing feelings (*e.g.* drawing a sad or 'smiley' face) and sketches of clothing. For males, the equivalent categories were very similar: doodling, drawing an imaginary person, expressing feelings, defacing pictures and sketches of clothing. In the case of public drawing, females produced, in order of frequency, local district maps, clothing sketches, puzzle or game related drawings, drawings to amuse a child and plans of the house. For males the equivalent five most frequent public drawings were local district maps, doodling, game or puzzle related drawings, drawing of a real person and house plans for maintenance tasks. Hence in the private category, recreational or expressive drawing are most common with a second, less common, group of personal planning related drawings (house plans, clothing , hair styles, time tables, flow charts...). When drawings are produced for an audience, direction-giving maps are most frequent. Child-related public drawings (homework help, amusement) are also common. Interestingly, some items that might be thought of as private are produced relatively often in the presence of an audience - these include doodling, defacing pictures, and drawing imaginary people.

Finally, this thesis is not about graphic communication or the role of sketches in design (*e.g.* Goldschmidt, 1991; Goel, 1995) — it is about the spontaneous construction and use of a variety of external representational forms[3] in a self-communicative manner as subjects develop and examine their ideas.

Defining the term 'external representation'

A definition of representation has been proposed by Davis, Young & McLoughlin (1982)

> A representation may be a combination of something written on paper, something existing in the form of physical objects and a carefully con-

---

[3] Including but not limited to graphical ones.

> structed arrangement of idea in one's mind

This definition includes both internal and external representations, but does not explicitly include graphical external representations. It could be extended to include representations constructed on paper other than those that are written, since external graphical representations can be said to be drawn or constructed rather than 'written'. An attractive feature of the Davis et al. definition, however, is that it highlights the *interaction* between internal and external representations in reasoning.

Another definition is provided by Mason (1987a):

> ...it is not clear that 'representation' is a sensible or consistent way to describe what goes on inside a person, because their inner experiences *are* their world, and not merely a representation of *the* world, whatever that may be....it is more sensible to speak of inner experiences as a person's world, and to speak of their manifestations in terms of pictures, diagrams, words, and symbols as a 'presentation' of their world. Furthermore this perspective emphasizes the importance of getting students to use and become fluent with a variety of modes such as diagrams, symbols, and metaphor to express what they perceive...

Mason would speak, not of reasoning with ERs, but of construing (making sense) with external presentations. In his view, presentations are the record of what someone says about what they see. Using ERs (EPs?) serves to make person aware of their learning.

Anderson & Helstrup (1993) studied the effectiveness of mental imagery with and without drawing support and use the term 'perceptual assistance' to describe the facilitatory effect of externalisation (drawing) upon the synthesis of novel patterns from simple shapes. Anderson & Helstrup (1993) conclude that mental imagery seems to be the initial source of discovery and synthesis but that drawing seems to be useful in production and refinement of patterns.

The term ER, as it is used in the work reported here, includes propositional or linguistic representations such as sentences of natural language and sentences of formal languages (*e.g.* first-order logic), as well as graphical or analogical representations such as graphs, maps, plans, and set diagrams. Various matrix graphics such as tables and lists can be considered as intermediate forms between 'pure' examples of graphical or linguistic representations. In practice, however, there is no such thing as an example of a 'pure' modality — sentences of natural language have graphical elements (*e.g.* punctuation symbols) and graphics almost always require linguistic annotations if they are to be successfully interpreted by persons other than the producer.

Graphical ERs such as freehand idea sketches are an invaluable aid to creativity in design disciplines such as architecture (Goldschmidt, 1991; Goel, 1995). 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 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, 1982; Barwise & Etchemendy, 1994; Cox, Stenning & Oberlander, 1994; Cox & Brna, 1995; Stenning, Cox & Oberlander, 1995; Stenning & Oberlander, 1995; Stenning & Cox, 1995; Cox, Stenning & Oberlander, 1995), physics (Anzai, 1991) and, more generally, in scientific and mathematical discovery (Davis & Hersh, 1981).

To summarise, an adequate definition of the term 'external representation' must distinguish between modalities (graphical versus linguistic) and must acknowledge the interaction between internal and external representations and the role of ERs in helping to disambiguate internal representations. It must also acknowledge a distinction between formal and 'everyday' external representational activity.

ER taxonomies

Several taxonomies of external representations have been developed. For example Twyman (1979) organises representations by means of a matrix with methods of configuration and modes of symbolisation as column and row headings, respectively. Methods of configuration are the 'graphic organisation or structure of a message which possibly determines the 'searching' and 'looking' strategies adopted by the user' (*pp* 119—121). The 7 'methods of configuration' range from pure linear forms to lists, linear branching and matrix representations to non-linear forms. The 4 modes of symbolisation consist of verbal/numerical, pictorial and verbal/numerical, pictorial and schematic. The example given by Twyman for the intersection of the column 'pure linear' with row the 'pictorial & verbal' is the Bayeux tapestry. A 'pure linear' and 'verbal' example is the Phaistos disc (an ancient Minoan artifact consisting of pictograms read in a spiral from the outside to the centre). Twenty-eight types of 'graphic language' can be identified with Twyman's schema, though finding examples for two of the matrix cells is problematic. Twyman can find no example of a list configuration with a schematic mode of symbolisation. Twyman also notes that it is almost impossible to find an example of a non-linear, open pictorial representation (*i.e.* a photograph that does not influence the viewer by its organisation).

Lohse, Biolsi, Walker & Reuter (1994) also present a classification of visual representations. Their scheme is based on cluster analyses of subjects' subjective ratings of graphics along 10 dimensions. They identify eleven basic categories of graphics: graphs, tables, graphical tables, time charts, networks, structure diagrams, process diagrams, maps, cartograms, icons and pictures.

Treu (1992), in a paper on HCI design, attempted to identify a set of structures 'amenable both to the user's mind and to the computer-based application'. On the basis of the cognitive science literature, he developed a taxonomy of representations that a typical computer user can understand. Treu starts with the classification of representational system proposed by Rumelhart & Norman (1988) in which representations can be propositionally based, analogical, procedural or based on distributed knowledge. All representational systems also consist of data structures and processes that operate on them. The basic interface structures Treu (1992) identifies include:

1. knowledge *Objects* (nodes) within patterns of association links (arcs)
2. *Sets* of features, logical clustering
3. *Vertical* layering - ordering, levels of abstraction, sequences, lists....
4. *Horizontal* layering - ordering, levels of depth (perception), lateral sequences
5. *Language* bi-directional human-human language, NLG, word sequencing

The basic structures can be combined as in the case of tree structured representations (*Objects + Vertical*), hierarchically related sets of objects (*Sets + Vertical*), and 2-D arrays (*Horizontal + Vertical*). In Treu's taxonomy, ubiquitous ER forms, such as a matrix or table and trees, are composite representations rather than basic forms.

Stenning & Inder (1995) provide an analysis of three families of representational systems - matrix graphics (tables, histograms, graphs and maps), logic diagrams (Venn diagrams, Euler's circles) and semantic networks (node and link diagrams). In their account, tables are a mixed modality in which both spatial relations and linguistic components are intertwined in interpretation. In the view of Stenning & Oberlander (1995) matrix graphics and tables are Minimal Abstraction Representation Systems (MARS) in which one graphic represents just one model[4]. Such graphics are said to be *weakly expressive* (Stenning & Oberlander, 1995) or *vivid* (Levesque, 1988). ERs in the linguistic modality, such as first-order logic and natural language, are expressive and pallid. They are capable of expressing unlimited amounts of abstraction (UARS). Some graphics can be made to express a limited amount of abstraction by means of 'tricks' such as special notations or multiple diagrams. These then become limited-abstraction representational systems (LARS — Stenning & Oberlander, 1995).

In the artificial intelligence domain, Hayes (1985) describes representational schemes consisting of logical calculi, programming languages, the systematic use of data structures to depict a world (*e.g.* the use of an array as a room map), musical notation, maps

---

[4] 'Generally speaking, a system B represents a model of system A if, on the basis of a certain isomorphism, a description or solution produced in terms of A may be reflected consistently in terms of B and vice versa' (Fischbein, 1987, p. 121).

and circuit diagrams. A particular map, he argues, is one configuration in a scheme. A scheme is a set of configurations. Hayes distinguishes between well-formed configurations (*e.g.* formal representations) and non-schemes (*e.g.* drawings, photographs, conversational English). In formal schemes, knowledge can be stored and used by a computer program whereas 'informal scenes' or 'perceptual situations' require the deployment of large amounts of background and general knowledge for their successful interpretation.

Petre & Green (1993) suggest that graphical representations are distinguished from non-graphical ones because graphics have 'secondary notations'. For example, in diagrammatic representations, lines represent connectedness and adjacency represents relatedness. In electronic circuit schematics, the adjacent placement of electronic components indicates a relatedness of function or purpose over and above the information contained in the wires (lines) that connect the components. In textual representations, however, adjacency indicates *both* connectedness and proximity. They write (p. 57)

> The strength of graphical representations—almost universally—is that they complement perceptually something also expressed symbolically.

In addition to secondary notations, Green (1989) has proposed the following further 'cognitive dimensions' of ER notations:

- viscosity (resistance to editing or modification)
- hidden dependencies (*e.g.* the inability of a spreadsheet to indicate that data in a cell is used by a formula in another cell)
- premature commitment (learner forced to make choices too soon)
- perceptual cueing (indenting subroutines in computer programs)
- role expressiveness (visibility and parsability - this varies with the user's understanding and prior knowledge)

When subjects are free to choose a representation for problem solving, they should, ideally, choose one that is capable of expressing the problem's information adequately and which is computationally efficient to construct and use. The issues of expressivity and efficiency will be addressed in more detail below. In practice, as it will be seen, subjects often have limited repertoires of representational forms on which to draw, often make poor choices of representation, make errors during ER construction and make slips when reading off information from their ERs.

As this brief review illustrates, there is little consensus in the literature about the type or number of dimensions to use in categorising ER formalisms. Twyman (1979) proposes two - methods of configuration crossed with modes of symbolisation. Lohse et al. (1994) identify 11 discreet categories based on subjective judgements. Treu (1992) identifies 8 basic structures of which 3 are 'composites'. Stenning & Oberlander (1995) suggest that the ER family divides into linguistic and graphical representations at top level. Those two modalities differ in the 'strength' with which they are capable of expressing indeterminacy. Hayes (1985) introduces a distinction between schemes and non-schemes – this points up the difference between representations that have underlying semantics that are computationally tractable and representations that require a large amount of background knowledge for their interpretation and so are difficult to program (*e.g.* photograph recognition). Finally, Petre & Green (1993) introduce the secondary notational properties of external representations – these are learned conventions for interpreting, for example, adjacency of graphical elements as representing conceptual relatedness. Amalgamating the studies produces at least seven dimensions along which to characterise ERs, prior knowledge and the nature of the task.

Conceptualisations of representations in related fields

Representations in mathematics and artificial intelligence

In fields such as mathematics, it has been known for long time that individuals differ in use of internal imagery (*e.g.* Poincaré). Poincaré (cited by Fischbein, 1987) classified mathematicians into two groups: 'geometers' (ones who think in images) and 'analysts' (conceptual thinkers). However, it is not clear whether mathematicians of the 'geometer' kind use external graphical representations more than 'analysts'. Conversely, do analysts externalise their reasoning via the use of mainly linguistic forms? Fischbein

(1987) argues that the role of internal images or visualisations in discovery is dynamic and that those kind of representations are not fully developed conceptualisations. Scientists and mathematicians *play* with images (*e.g.* Kekule mentally 'playing' with an image of the benzene ring as a snake with its tail in its mouth). Fischbein states that such mental images, while not representing fully developed ideas, are associated with intrinsic feelings of certainty about their correctness. This intrinsic feeling of certainty, Fischbein points out, is not confined to creative discovery via internal imagery. As Einstein's famous conversation with Wertheimer revealed, the same conviction can derive from non-imagistic mental conceptions:

> These thoughts did not come in any verbal formulation. I very rarely think in words at all. A word comes, and I may try to express it in words afterwards.

Among teachers of logic and mathematics, graphical methods of external representation remain highly controversial. In a recent discussion of beliefs about the nature of mathematics, Eisenberg (1992) points out that, in the mathematics community, the idea that mathematics must be communicated in a non-visual manner is 'deeply rooted'. Sutherland (1995, p.80) also makes the point that:

> In our culture where sentential systems have higher status than visual systems, 'academic' is almost always synonymous with articulate.

Attitudes towards graphical representations are sometimes seemingly inconsistent. Graphical representations can be assigned different status depending upon whether they are used as a conceptual aid or as a medium of communication. Eisenberg (1992) illustrates this by citing Hilbert:

> I have given a simplified proof of part (a) of Jordan's theorem. Of course, my proof is completely arithmetizable (otherwise it would be considered non-existent); but, investigating it, I never ceased thinking of the diagram (only thinking of a very twisted curve), and so do I still when remembering it.

In the logic domain, Barwise (1993) quotes the views of Tennant (1986) as representative of the attitude of traditional logicians to the role of diagrams in logical proofs:

> [The diagram] is only an heuristic to prompt certain trains of inference; ...it is dispensable as a proof-theoretic device; indeed, ...it has no proper place in the proof as such. For the proof is a syntactic object consisting only of sentences arranged in a finite and inspectable array.

There are at least two historical reasons for such negative attitudes towards graphical representations. One is that diagrammatic representations may lack generality - a single diagram of a triangle can portray only one of the whole universe of possible triangles. Mathematics, logic and other formal fields require general proofs which do not suggest particular models but which apply to all possible models. A second reason is that in the history of mathematics there have been instances where diagram-based proofs ('proofs without words') have subsequently been shown to be erroneous or based on accidental properties of a particular diagrammatic model that the reasoner happens to have chosen. However, as Shin (1994) has shown, the misapplication of diagrams is not intrinsically related to the nature of diagrams, at least in the case of Venn diagrams. She cites Barwise & Etchemendy (1992):

> If we threw out every form of reasoning that could be misapplied by the careless, we would have very little left. Mathematical induction, for example, would go.

Another factor in the dispute concerns the issue of whether teaching mathematics or logic is undertaken to improve general reasoning as well as teach the domain material or whether it is taught solely to prepare students for further advanced symbolic logic courses and their application to computer science, for example.

Amarel's (1968) work on representation in reasoning about actions is very important since he suggests that much of problem solving is concerned with selecting the right representational *system* and translating information into that system.

> ...our general thesis...(is that) it is an important function of the problem solver to find the most appropriate representation of his (its) problem.' (p.166)

He addresses the problem of choosing an appropriate (machine) language into which the verbal statement of the problem should be translated. He considers production systems (P systems) and N-state language, extended description language, rules of action and the two N-states that correspond to initial and terminal situations.

The objective of the problem solving system is to find a trajectory between the initial and terminal situations using the 'coarsest possible elements and predicates that are capable of expressing the rules of action in sufficient detail'. In the domain he studied ('missionary and cannibal' (M&C) problems) one size of sets of individuals (coarse grain) is better as a basic reasoning element than the consideration of individuals as elements (*i.e.* fine grain).

In the M&C problem, Amarel noted that improvements in formulation came from a recognition that one of the conditions was redundant. In the M&C problem the rules of action are highly context dependent. Search graphs and array representations of state space facilitate the detection of symmetry.

> ...in order to discover useful properties in the N-state space it is very important to have 'appropriate' representations of that space... in general, the problem of *choosing* a representation of N- state space, and of *discovering useful regularities* of solution trajectories in this representation, require much more study...the choice of appropriate representations is capable of having spectacular effects on problem solving efficiency (pps. 169–170)

Amarel's heuristics for M&C problems can be summarised as:

1. choose the appropriate basic elements and attributes for the representation you use to represent state-transitions
2. choose appropriate representations for the rules of action and for the N-state space

3. use the representation to discover useful properties of the problem that permit a reduction in size of the N-state space (*e.g.* redundant conditions, symmetry, critical points)

4. utilise new knowledge about the problem's properties in formulating better problem solving procedures

Later chapters will demonstrate that elements of those heuristics are utilised by subjects in their solutions to analytical reasoning problems.

Representations and psychology, cognitive science

This section covers quite a large volume of material since most research on external representations can be found within cognitive science. The field includes, *inter alia*, human-computer interaction (HCI) (*e.g.* Green, 1989), analogical reasoning (*e.g.* Gick & Holyoak, 1983), problem solving (*e.g.* Larkin & Simon, 1987; Anzai, 1991), syllogistic reasoning (Stenning & Oberlander, 1995), display-based reasoning (Larkin, 1989) and visual cognition (*e.g.* Humphreys & Bruce, 1989). Some of those studies will be reviewed in Chapter 4 under the heading of 'related work' and some will be reviewed here.

Semantic properties of ERs Stenning & Oberlander (1995) have proposed a theory of specificity of graphical information in which they argue that diagrammatic 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. For example, a diagram typically represents a single state of affairs or single model, 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. Similarly, the word 'triangle' can be used to refer to any kind of triangle, but a single diagram must represent only one triangle. As another 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 coincide exactly). 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, 1995). It is the weak expressiveness (specificity) of graphical representations that makes diagrams so cognitively tractable. Stenning & Oberlander argue that they share the property of specificity with the internal representations that humans use when they reason. They also argue that the weak expressiveness of graphical representations is more apparent to users than is the case for sentential representations because diagrams 'wear their constraints on their sleeve'. Diagrams are 'vivid' representations (Levesque, 1988). ERs capable of expressing indeterminacy[5] can be described as 'pallid' (Holyoak & Spellman, 1993). Stenning & Oberlander (1995) argue that reasoning is facilitated by representations whose degree of pallidity is well-matched to the indeterminacy of the information in the problem. Too much expressivity, however, reduces the cognitive tractability of the formalism. Generally, the best representation is one which has sufficient abstraction-expressing power for the task, *but not any more than is needed.* Empirical support for this contention has been provided by Cox, Stenning & Oberlander (1995).

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.

Studies that included both determinate and indeterminate problems are reported in this

---

[5] Indeterminacy is synonymous with the term abstraction.

thesis — they were studied in order to examine some of the implications of specificity theory.

Cognitive properties of ERs There is a large literature on the cognitive properties of ERs. The kinds of properties that are proposed depends upon the proposer's position in the so-called 'imagery debate' (*e.g.* Paivio, 1986). The debate's central question concerns the extent to which, and at which level, communication occurs between the various perceptual and cognitive subsystems. Is there an interlingua, a kind of internal mentalese, which mediates between linguistic and imagistic internal representations? Pylyshyn (1973, 1979), for example, argues that there must be an internal propositional interlingua in which both verbal and non-verbal processing are carried out. Or is the only route between the subsystems via externalisation? Mason (1987), whose work was discussed earlier, and Reisberg (1987), whose work is introduced later, seem to adopt the latter position.

Both positions tend to downplay the strong possibility that there may be individual differences in cognitive representational style, which can influence the kinds of representations that subjects use internally and externally. Individuals certainly differ in their problem solving strategies and representational habits - but whether this reflects differently arranged cognitive subsystems or the effects of prior experience upon more-or-less invariant hardware is beyond the scope of this thesis. The author has no strong commitment to either position in the debate, since, in discussions about *external* representations, it is not crucial. As Barwise & Etchemendy (1992) point out in a discussion of Hyperproof[6], there are techniques for transferring information from linguistic forms into graphical forms and vice versa that do not need to appeal to cognitive 'Rosetta stones' or internal interlinguas. In analytical reasoning, and reasoning about blocks on chessboards, the two representational schemes (diagrammatic and linguistic) represent the same worlds and cognitive devices exist for moving information between the two modalities. In Hyperproof, invoking the sentential *APPLY* rule permits information to

---

[6] Hyperproof is a computer-based program for teaching proof development in first-order logic. It adopts a heterogeneous reasoning approach — information is presented and manipulated in both diagrammatic and sentential modalities. Students reason about situations in a blocks world. Hyperproof's graphical upper window shows geometric objects arrayed on a chessboard. Sentences of first-order logic are entered in a lower window.

be applied from sentences to the diagram. Conversely, the use of *OBSERVE* permits the opposite movement. However, a theory about the exact cognitive mechanisms by which the analogues of 'apply' and 'observe' function in humans is not crucial for the arguments advanced in this thesis.

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 aid 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, in some respects (Green & Petre, 1992; Petre & Green, 1993). Thus, Larkin & Simon (1987) emphasise the difference between the *informational* and the *computational* equivalence of representations. Stenning & Oberlander (1995) use the terms 'expressiveness' and 'facility of inference' and contrast the *logic* of a task with its *implementation*. Unlike Larkin & Simon (1987), who argue that graphics facilitate perceptual judgements, Stenning & Oberlander (1995) explain the computational efficiency of diagrams in terms of representational system semantics. They argue that graphical representations such as diagrams are less expressive than sentential representations and that they therefore aid processability. They term that property of graphical representational systems 'specificity', *i.e.*

> ...the demand by a system of representation that information in some class be specified in any interpretable representation (Stenning & Oberlander, 1995, p.98)

In other words, graphical representations compel specification of classes of information, in contrast to systems that allow arbitrary abstractions. Specificity theory is further discussed below in the section on the semantic properties of representations.

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. 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.

External graphical representations must be both well-constructed and capable of representing the information in a problem. If both criteria are met, the high-bandwidth, rapid processing capabilities of the human visual system are exploited and very easy perceptual judgements[7] are substituted for more difficult logical ones (Paige & Simon, 1966). They argue that the computational efficiency of graphics is due to the congruence between the structure of the data and the program(s) that operates on it.

Koedinger & Anderson (1990) suggest that the use of ERs can facilitate a shift of reasoning mode. When used in the development of a geometry proof (ostensibly a deductive reasoning task), accurately drawn and skillfully used diagrammatic representations 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

---

[7] Not all perceptual judgements are easy — the visual-spatial systems are limited in the judgements they are 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.

of possibly provable statements and shift the mode of reasoning from deduction to induction.

Constructing an ER is equivalent to building a model of the information in the problem. However, semantically equivalent models may not be equally suited for all types of problem solving task (Day, 1988). Some representations facilitate the rapid comparison of values (*e.g.* histograms) whereas other representations facilitate read-off of precise values (*e.g.* tables). Hence, effective ER use requires both an adequate representation of the information in the problem and a representation suited to the type of task posed by the problem.

Representations also differ in their ability to convey what might be termed 'progress through the problem' information. Tables or spreadsheets, for example, are particularly good at highlighting (via empty cells) missing or not-yet-represented information.

Externalisation and the self-explanation effect The effectiveness of ERs may be mediated by mechanisms that parallel the ways in which self-explanations are associated with successful problem solving — the 'self-explanation' effect (Chi, Bassok, Lewis, Reimann & Glaser, 1989). In the Chi et al. (1989) study, subjects studied a chapter on Newton's laws from a physics textbook and then studied worked-out solutions to problems. The worked-out examples were illustrated by diagrams. Afterwards, they were asked to solve similar problems. Self-explanations (remarks that students made as they studied the text and worked-examples) were recorded (Chi & Bassock, 1989).

Good students tend to generate self-explanations when learning from worked-out solution examples. They actively construct an interpretation of the example that they are studying. Good students produce more self-explanations than poor learners. The mechanism of self-explanation is believed to be one by which students generate tacit knowledge that links pieces of explicitly stated knowledge. In contrast, poor learners tend to re-read and paraphrase parts of the example. Furthermore, good learners are able to articulate their comprehension difficulties more clearly than poor learners who tend to indicate that they understand the statements given in the worked-out examples. 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). VanLehn, Jones & Chi (1992) argue that students invent new knowledge during example studying and problem solving rather than merely engaging in the recollection and operationalisation of knowledge acquired from reading text. The most productive source of new knowledge, according to VanLehn et al. (1992), is explanation-based learning[8] of correctness (EBLC) in which new domain knowledge is created by specialisation of overly general knowledge.

The self-explanation effect has been replicated in the domains of physics and computer science, but the underlying cognitive processes have not been fully explored. ER construction may affect self-explanation but the issue has received little research attention to date. For example, the self-explanation effect may operate during translation *across* modalities (*e.g.* from verbal to diagrammatic or vice versa). 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 constructing a diagram for example, the semantic properties of graphics 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, 1995). 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 incomplete comprehension.

As mentioned earlier, Stenning & Oberlander (1995) suggest that graphical representations compel certain classes of information to be represented and that these representations are less expressive of abstraction than sentential representations.

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 (Stenning & Oberlander, 1995) provides grounds for predicting that the process of translating informa-

---

[8] Explanation-based learning occurs where there is a single training example together with knowledge of the task domain on the part of the learner (Wusteman, 1992).

tion *from* a linguistic representation such as natural language or logic *to* a graphical representation might be more effective than translation from one representation to another within the same modality.

Externalisation — turning one's representations into stimuli Reisberg (1987) sees the process of constructing an ER as a procedure for 'widening the context of understanding' and 'turning ones representations into stimuli'. ER selection and construction consist of dynamic iterations and interactions between external and mental models. Some tasks can be performed with internal representations but are very difficult. Consider the task reported in a study by Hinton (1979) who instructed subjects to imagine picking up a cube and holding it such that one corner is vertically above another. He then asked the subjects about the location of the corners that they weren't 'holding'. Most subjects believe that the corners will form a square along the 'equator' of the cube. The middle edges of the cube, in fact, form a zig zag. Hinton argues that most subjects' mental images are not fully elaborated — we mentally reconstruct the cube on the basis of an incorrect approximation of the transformations involved. We work from some rather poorly elaborated, structural description. Externalising the representation — drawing a diagram of the cube — assists greatly with determining the correct arrangement of corners. Graphical ERs force consistency and help to turn an initial internal representation into an external stimulus which, upon re-processing, assists with finding a solution.

In other words, the process of externalisation helps to disambiguate ambiguous mental images. In a similar way, the process of drawing an ER such as a diagram (*i.e.* externalising a mental model) can facilitate problem solving. In the author's view, externalisation may also facilitate the transfer of information between cognitive subsystems in ways that are not possible internally according to the dual-coding hypothesis (*e.g.* Paivio, 1986).

Plan

In the remainder of this thesis, Chapter 2 outlines the domain (analytical reasoning) and describes the structural and linguistic characteristics of analytical reasoning problems. Chapter 3 presents an overview of the cognitive processes involved when subjects solve analytical reasoning problems. The nature and extent of support that might be offered to subjects during representation construction is also considered.

Chapter 4 reviews related research. The review is organised around the issues of:

- self-selection/construction of external representations versus the use of pre-determined, pre-fabricated external representations
- domain-specific external representational formalisms versus more generic forms
- instructional intervention in the use of external representations versus unguided, spontaneous use

Chapter 5 presents the first study which examined a large corpus of 'workscratchings' produced by subjects as they solved paper and pencil-based analytical reasoning problems under test conditions. The workscratchings show great diversity of representation use across subjects and permitted an examination of the utility of different external representations under differing task conditions.

A second study, presented in Chapter 6, utilised a computer-based system (*switchERI*). The system administered analytical reasoning problems and provided a range of ER construction environments for the subject to choose and switch between. User-system interactions were recorded dynamically during problem solving. This methodology permitted micro-analyses of the cognitive events at each stage during the time-course of problem solving. *switchERI* was used to determine some of the circumstances under which multiple representations enhance performance, to study subjects' behaviour during impasses in reasoning and to identify the important factors surrounding decisions to switch representations. The data also suggested the kinds of support that might be useful to learners at different stages of reasoning.

Chapter 7 presents the results of a study using a second system, *switchERII* – an intelligent learning environment. *SwitchERII* incorporates a representation of the semantics of Euler's Circles, dynamically parses the user's representation and provides feedback and advice. The *switchERII* study examined the relationship between subjects' prior knowledge and reasoning performance. The *switchERII* study also investigated the extent to which subjects' prior misconceptions affected external representation selection, construction and use.

Chapter 8 relates the findings of the 3 studies to the issues raised in earlier chapters. The roles of representation switching in the resolution of impasses in reasoning and the cognitive effects of externalisation are discussed. The thesis concludes with an argument in favour of a domain-independent 'ER curriculum' . It is suggested that direct instruction in the use of a range of ERs might equip students with wider representational repertoires and hence allow them more scope to indulge their representational preferences.

Finally, in Chapter 9, the contributions made by the thesis are summarised and several directions for future work are proposed. These include extending the representational semantics of *switchERII*, evaluating various types of system feedback and implementing a mechanism for checking for slips during read-off from ERs.

Chapter 2

The domain: Analytical reasoning problems

The characteristics of analytical reasoning problems

This chapter describes the linguistic and structural characteristics of analytical reasoning (AR) problems since subsequent chapters assumes some familiarity with the domain.

The GRE exam is taken by US undergraduates, and, together with course grades, plays a major role in determining entry into US Graduate Schools. In its full form, the GRE has sections that test verbal ability (analogies, antonyms, sentence completions, reading comprehension), quantitative ability (arithmetic, algebra, geometry, quantitative comparison, discrete quantitative, data interpretation) and analytical ability (analytical reasoning, logical reasoning). The analytical ability scale of the GRE was introduced into the exam in 1977, in response to the view of graduate school teaching staff that the then current exam lacked a test of abstract reasoning ability. Abstract reasoning was favoured by staff and students as a means of broadening the GRE over alternative scales measuring scientific thinking and study style (Miller & Wild, 1979). However, the early GRE *verbal* and *quantitative* subscales did not add a great deal of predictive utility to the Scholastic Aptitude Test (SAT), which is taken at secondary school four or five years earlier than the GRE. In a sample of 22,923 subjects, the GRE verbal scale was observed to correlate highly with SAT verbal scores ($r = .858$) and GRE quantitative also correlates highly with SAT mathematical scores ($r = .862$)

(Angoff & Johnson, 1990). Hence there was a need for a GRE scale that was designed to include aspects of reasoning additional to those measured by the SAT. The revised GRE is claimed to predict graduate school performance in a wide range of disciplines separately from other measures such as domain specific undergraduate performance.

Swinton & Powers (1983, p. 104) write that ...

> the (analytical portion of the) test is intended to measure analytical reasoning abilities that, like the verbal and quantitative skills measured by the test, are assumed to develop over a relatively long period of time.

The GRE analytical test has two subscales, termed *analytical reasoning* and *logical reasoning*, respectively (Duran, Powers & Swinton, 1987). Logical reasoning items take the form of verbal reasoning or argument analysis problems and were not used in the studies reported here. Rather, the items used in this research were derived from the other subscale — analytical reasoning. Analytical reasoning problems are usually constraint satisfaction puzzles for which diagrams are often useful.

Analytical reasoning problems[1] 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. Some of the information is given explicitly but some is implicit and must be inferred. The given information is followed by a series of questions that require deductive reasoning for their solution. Examples of these *analytical reasoning* puzzles are provided in Appendix A.

It is claimed that the GRE test does not require specialised domain knowledge and is relatively resistant to the effects of coaching. However, Swinton & Powers (1983) 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). Those item types were

---

[1] Sometimes referred to as 'Who-done-it?' or 'Who owns the zebra?' deductive reasoning problems (*e.g.* McGuiness, 1986).

eliminated from the analytic measure in 1981 (Emmerich, Enright, Rock & Tucker, 1991).

An analysis of the content characteristics of analytical reasoning items by Chalifour & Powers (1989) revealed that the difficulty of analytical reasoning items 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 position[2] was negatively correlated with item difficulty; that is, the more explicitly given determinate information, the easier the problem.

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). 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 which 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.

Analytical reasoning problems are often best solved by constructing ERs. GRE analytical reasoning problems can involve deductive reasoning about seating plans for dinner parties, order of speakers at a conference, the assignment of individuals to offices, committee membership etc. This although specialised domain knowledge may not be required, problem solvers may need to be familiar with various 'scripts' of everyday Western cultural experience (Schank & Abelson, 1977). Norman (1988) partitions 'constraint satisfaction' into physical, semantic and cultural constraints. Physical constraints are exemplified, in a common analytical reasoning context, by the fact that only one person at a time can sit on a chair and this is assumed knowledge in the 'seating plan' type of problem. Semantic constraints are illustrated by the tacit assumption in the 'dogs' problem that each dog wins only one prize, has only one owner and only one name. Cultural constraints are often represented as scripts, such as

---

[2] 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.'

Schank & Abelson's restaurant script which constrains the range of culturally acceptable behaviours at each stage of entering, ordering, eating and leaving. It is therefore questionable whether GRE analytical reasoning problems are 'culture fair' for non-US and perhaps students of low socio-economic status.

Attractiveness of the domain

Analytical reasoning problems were an attractive domain for the study of reasoning with ERs because of the *a priori* evidence that indicates diagrams are often useful in finding solutions and that diagram drawing is associated with item difficulty. A range of GRE 'crammers' (*e.g.* Brownstein et al., 1990) are available and these provide a ready source of suitable problems. Also, the crammers frequently recommend particular ER formalisms for solving particular problem types. The utility of the recommended ERs has not been subjected to empirical scrutiny and so one of the aims of the current work was to examine the usefulness of the crammer–recommended ERs.

Analytical reasoning problems have an additional advantage over many types of stimulus used in cognitive research in that subjects often report that they enjoy solving them. This is important from the motivational standpoint. The non-requirement of specialised domain knowledge is also an advantage in that they can be administered to subjects from a wide variety of backgrounds. In the US, skilled performance on analytical reasoning problems is very important because of their inclusion on the GRE. Hence the findings of empirical studies of solution strategies are likely to be received with interest. Another further attraction of the domain is that item validity and other psychometric data is available, usually from the ETS. Much of the ETS research, however, is concerned with comparing the effects of various multiple-choice response formats or the ability of particular problems to discriminate between individuals. The ETS research focusses very little upon the kinds of strategies that subjects use in their solutions.

Finally, another important factor in the choice of domain was that a substantial corpus of paper and pencil tests was available because analytical reasoning items were used in the pre and post course assessments of a graphical approach to teaching logic (Cox, Stenning & Oberlander, 1994, 1995; Stenning, Cox & Oberlander, 1995; Oberlander, Cox & Stenning, 1995a,b).

The linguistic and structural properties of analytical reasoning problems.

The characteristics of analytical reasoning problems stem from both their linguistic and structural properties.

Linguistic characteristics and attributes of AR problems

Many of their linguistic properties have already been discussed and are summarised below in list form

- problems have 'stem' and questions
- information about entities (people, things, places) and relations between them
- some information is given negatively
- a proportion of the information is usually implicit
- information presentation violates Grice's 'cooperative principle' of manner (orderliness, obscurity of expression). This will be discussed further in Chapter 3.
- instructions to subjects 'Each question or group of questions is based on a passage or set of conditions. In answering some of the questions, it may be helpful to draw a rough diagram. For each question, select the best answer choice given.[3]'

Structural characteristics and attributes of AR problems

The structural attributes of analytical reasoning problems are listed below:

- they are constraint satisfaction deductive reasoning puzzles that involve various kinds of relationship between entities, such as:
  - attribute assignments

---

[3] 'Practicing to take the GRE General Test Number 9' ETS/Warner Books, 1992.

  - conditional relationships
  - familial relationships
  - ordering relationships
  - spatial relationships
  - assignment relationships

- the problem questions are heterogeneous in terms of their task requirement
- the questions can consist of 'what if' (hypothetical cases) *e.g.* question 4 of the 'Office' problem; question 5 of the 'Dogs' problem (Appendix A). 'Hypothetical' questions affect the number of versions of representations that may be required in a subject's solutions, since a new model of the problem information is required for each 'what if' scenario.
- AR problems vary in complexity (number of dimensions, values per dimension)
- AR problems vary in their level of determinacy, for example, the 'Office' and 'Dogs' problems in Appendix A are determinate and the 'Poets' problem is indeterminate
- AR test item designers (Educational Testing Service (ETS)) have data on the factors that are associated with item difficulty (*e.g.* need for diagrams is associated with greater difficulty)
- coaching texts ('crammers'), especially 3rd party, non-ETS ones, recommend particular ERs for solutions (*e.g.* Brownstein et al., 1990; Research & Education Association, 1994)

The effects of linguistic factors and structural characteristics upon comprehension are discussed further in Chapter 3.

GRE analytical reasoning items have an additional advantage for the studies reported in this thesis: they are verbal tests posed in English, eliciting selections of verbal answers. They present no diagrams and there is no opportunity to present the results of reasoning diagrammatically. However, there is room on the test questionnaire for the construction of external representations and therefore candidates can engage in diagrammatic and other representational activity in the course of the test.

The structural and linguistic characteristics of analytical reasoning problems have profound effects upon comprehension. But comprehension is only the first stage of analytical reasoning. Subsequent stages involve ER selection decisions, ER construction, reading-off information from ERs, building a second ER (i.e. switching ER), selecting an answer from the multiple choice array, *etc.* These are addressed in the next chapter.

Chapter 3

Stages in analytical reasoning with ERs

Introduction

This section outlines the cognitive processes involved at each stage of solving analytical reasoning problems using ERs. The stages loosely follow Polya's (1957) stages: understand the problem, devise a plan, carry out the plan, examine and reflect upon the solution. The overview will be used to elucidate and highlight important issues that will be discussed in more detail in following chapters.

Five components in reasoning with ERs can be identified:

1. problem comprehension and interpretation
2. ER selection
3. ER construction
4. read-off from the ER or use of ER
5. responding to (answering) problem questions

It should again be emphasised that the components are not linear and fixed, rather they are iterative — problem solving evolves as the process of representation proceeds.

Interpretation and comprehension

> Understanding is a constructive process, in which a representation is developed for the object that is understood. The difference between understanding and not understanding is in the nature of the representation. (Greeno, 1977, p44)

> ... word problems ... involve learning to use ordinary language in a special way ... (Kintsch, 1991, p.241)

Solving AR problems begins with the interpretation and comprehension of the (linguistically) presented information.

As soon as the question of understanding is addressed, difficulties emerge with a stage conception of reasoning with ERs. One perspective argues that understanding is defined by the process of representation construction (which, in turn, pre-supposes representation selection).

But what is comprehension? In problem solving, Greeno (1977) suggests that a problem solution is a cognitive product, generated by the problem solver, which can be evaluated in terms of the degree of understanding it shows.

Understanding requires background knowledge and conceptual knowledge.

The older tradition of Kohler, Duncker, and Wertheimer emphasised insight - a sudden realisation of critical relations between elements of a problem. More recently, Greeno (1977) has offered three criteria for understanding: coherence, correspondence and connectedness. Greeno (1977, p.45) proposes three criteria for a theory of understanding:

- achievement of a coherent representation (*i.e.* unified at high level, via analogy for example, or global thematic content)
- an internal representation that corresponds closely to the object that is understood
- relating the understood object and its components to the understander's other knowledge

A problem solution possesses coherence if its components are related in a compact structure. For some problems, Greeno, maintains, there are alternative solutions which differ in coherence. In discussing coherence, Greeno describes how Duncker (1945) distinguished between organic proofs (which make higher-order relations between steps explicit) and mechanical proofs (which proceed step by step with attention focussed on the justification for deriving each step from preceding statements rather than on overall plan or proof structure).

> Good understanding is often achieved in solutions involving transformations of the problem and sensible constructions that preserve the main structure of the problem. The undesirable cases involve mechanical applications of rules that fail to preserve important relational properties needed for understanding. (Greeno, 1977, p.47)

Greeno's second criterion of comprehension, correspondence, refers to the relationship between the cognitive representation and the object that is understood. Many failures of correspondence are due to slips. Greeno writes: 'persons frequently miscopy some information in a problem, or omit some relevant information. Many such errors are probably best explained as random lapses of attention.' (p.46). The more central type of correspondence concerns whether or not the 'solution is a natural one in the domain of the problem, or whether the problem has been translated in some way that makes the solution artificial.' Greeno cites an example from Wertheimer (1959) and contrasts two methods of finding the area of a parallelogram: a) use of the formula $A = b \times h$ and b) a transformation that shows how a parallelogram is related to a rectangle. In terms of Greeno's correspondence criterion, Newell & Simon's (1972) 'radical' re-representation of nine-card number scrabble (Chapter 1) would also be classified as an artificial solution in Greeno's terms.

The third criterion concerns the integration of a problem's cognitive representation with prior knowledge. Generalisation of a solution pattern to new situations, and the ability to answer interpretive questions, depends upon the extent to which the cognitive representation of problem and its solution are connected with other components of the person's knowledge. The second and third studies to be reported in Chapters 6 and

7 incorporated measures of prior knowledge of ERs in order to assess the extent of its effect upon, *inter alia*, representation selection.

Greeno's constructivist approach to comprehension is rather general. A large number of problem characteristics and presentation factors influence subject's proneness to 'see' solutions to word problems. These factors can be grouped under two headings - linguistic factors and structural factors. The linguistic and structural factors associated with analytical reasoning problems were briefly listed in Chapter 2 — they will now be discussed in more detail.

Effect of linguistic factors upon comprehension

In the mathematical domain, educators have recognized for some time that language can impede the understanding of mathematical concepts (Austin & Howson, 1979; Bell, 1983; Pimm, 1987; Durkin & Shire, 1991). Phenomena which have been discussed are mathematical register (Pimm, 1987), lexical ambiguity (Durkin and Shire, (1991), and metaphor, (Nolder, 1991). These linguistic factors also affect the interpretation of analytical reasoning problems.

In order for a student to accurately comprehend a problem, they:

- need to avoid a variety of word-problem comprehension errors
- need to overcome 'uncooperative' aspects of information presentation (*i.e.* violations of Gricean maxims of co-operative discourse)
- need to give a precise reading to (apparently) informal language
- must not be misled by the surface form of a problem - its 'cover story'

Test item designers exploit these factors in order to manipulate the difficulty of the problems.

Many errors are made at the stage of problem comprehension and this emerges as a crucial phase of problem solving (Proudfit, 1981; Reed & Ettinger, 1987; Schwartz, 1971; Polich & Schwartz, 1974). Schwartz (1971) has shown that the syntactic complexity of analytical reasoning problems affects the error rate — negative wording and the use of disjunctions are associated with high error rates. In contrast positively expressed problems and the use of conjunctions are much easier. Polich & Schwartz (1974) found that the representation of information that has to be inferred (*i.e.* is implicitly given) is particularly error-prone. Problem complexity (in terms of the number of dimensions and the values along them) is another determining factor (Schwartz, 1971; Polich & Schwartz, 1974).

Information extracted from a problem, and how it is interpreted, influence ER selection. There are many sources of errors of interpretation - many errors are due to the language in which the problems are posed in the sense that formal readings of information can be contrasted with natural (cooperative) dialogue readings. Other errors result from misinterpretation of quantifiers, implicatures etc. The influence of such 'Gricean' factors is discussed further below.

The 'cover story' of a problem radically affects problem difficulty as demonstrated by studies that have compared the Tower of Hanoi (TOH) problem with various isomorphs such as various 'Monster - Globe' versions (*e.g.* Lewis & Toth, 1992). Versions in which transitions between problem states are difficult to envisage (imagine) are more difficult than more salient versions. For example, a monster-globe TOH isomorph, in which spheres in the monsters' hands change shape 'magically'[1] was more difficult to solve than one in which the monsters passed globes between each other in a more 'natural' and more easily imagined fashion *i.e.* the monsters handed globes to each other in the same way that humans might hand things around.

Reversal errors are also common correspondence errors (to use Greeno's (1977) term). This is a tendency to match the word order in the problem to entity order in the problem representation — as in the 'students and professors' type of problem. These erroneous representations often 'pop' quickly into the solver's mind and, in experts, are actively suppressed via self monitoring and checking. Much of research on errors of interpretation and comprehension comes from the word arithmetic and word algebra literatures - however, many of the findings are relevant to analytical reasoning. Translation errors are frequently studied in those domains — *e.g.* Wollman (1983) reports

---

[1] An analogue of disk moves between pegs in the Tower of Hanoi.

that one in three college students produce the erroneous algebraic representation 6S=P for the sentence 'There are 6 times as many students as professors'. Clement, Lochhead and Monk (1981) found that 37% of engineering students and 57% of non-science students made this type of error. The reversal error is due to several causes - a tendency to match word order in the sentence to the sequence of algebraic symbols, and/or a 'set match' error where the equals sign is interpreted to represent 'for every'.

Wollman (1983) showed that highly proficient subjects (*e.g.* mathematical physicists) may initially make a reversal error, but self-monitoring processes and checking result in the initial equation being corrected. With more typical subjects, Wollman (1983) has shown that brief (10 mins) training in monitoring, comparing and checking either during or following translation produces dramatic improvements in performance - 16 of 17 students (from a sample of 43) students who made reversal errors initially were able to arrive at a correct equation following the intervention. Wollman (1983) characterises the intervention as an 'active operation' approach.

Kaput (1987) argues that the high error rate in word-problem-to-algebra translations (mostly of the 6S=P variety) is due to natural language overriding the rules of algebraic syntax and rules of reference[2].

Analytical reasoning problems also offer plenty of scope for reversal errors - as in the case of the 'Poets' problem used in the investigations to be reported in Chapters 5–7[3].

Grice (1975) analysed conversational implicatures and has proposed the 'cooperative principle'. Several maxims of the cooperative principle assist with understanding sources of difficulty in puzzles such as analytical reasoning problems. The cooperative principle has maxims arranged under four headings: quantity, quality, relation and manner.

Under **quantity** there is the maxim that contributions (in dialogue) should be as informative as required for the purposes of the exchange, and that they should not be

---

[2] Curiously, when Students-Professors types of problems are presented with a schematic diagram in addition to the word-problem, the error rate increases dramatically (Sims-Knight & Kaput, 1983). This finding might be born in mind by those who advocate heterogeneous reasoning such as Barwise & Etchemendy (1992). Heterogeneous reasoning will be introduced later in this chapter.

[3] 'All those who enjoy the poetry of Browning also enjoy the poetry of Eliot' is sometimes represented in a manner that is commensurate with a reversal error.

more informative than necessary.

The **quality** of a contribution is determined by its truthfulness - do not say what you believe to be false, do not say that for which you lack adequate evidence.

Grice's heading **relation** has only one maxim: 'be relevant'. Writing about relation, Grice states:

> I expect a partner's contribution to be appropriate to my immediate needs at each stage of the transaction.

The final category, **manner**, is perhaps the one of most relevance to the domain of analytical reasoning problems. It requires that contributions should avoid obscurity of expression, should avoid ambiguity, and that they should be brief and orderly.

If the sentences in the problem stem of an analytical reasoning puzzle are examined (Appendix A), it can be seen that they violate several of the Gricean maxims. The maxim of quantity is flouted in the sense that the contribution of the problem poser is to only provide as much information as is required for a solution to the problem and not to be as informative to the extent of assisting the solution process (*i.e.* optimally informative). Analytical reasoning (AR) puzzles are also obscure since they express problem information in natural language (*e.g.* 'Some of those who enjoy the poetry of Eliot also enjoy the poetry of Auden') but require a formal interpretation of the (quantifier information *i.e.* 'some' must be taken to mean 'at least one and possibly all'. In natural language, the subject term denotes shared knowledge between communicants, and predicates convey information that is being transferred. In the interpretation of word problems the subject and predicate terms can stand for sets, for example, and attribution is to be understood as asserting relations between sets.

In analytical reasoning problems, information is presented in an arbitrarily ordered manner, there is no attempt to minimise ambiguity, and information is (deliberately) obscurely stated (*e.g.* 'Mr Grossman's dog wins neither first nor second prize'). The violations fall mainly under Grice's categories of quantity, manner and relation.

Analytical reasoning puzzles are, by design, 'uncooperative' in the sense that information is dispensed in a manner calculated to complicate the solver's task. That is, information is presented in a manner that is inappropriate to the reasoner's immediate needs. One of the reasons why AR problems are puzzles is because they violate Gricean maxims and the cooperative principle.

A possible benefit of ERs in analytical reasoning is that they assist in overcoming violations of Gricean maxims and the tension between natural language and formal interpretations of, for example, quantifiers. ERs assist by re-ordering the information in ways useful for solutions and by laying out the range of possible models of the information, making missing information easy to detect, and by enforcing the explicit representation of implicit information.

In another respect, though, the analytical reasoning problems *do* cooperate with the reasoner - it is common for the entity labels to be abbreviatable to alphabetic sequences. In the poets problem, for example, the poets Auden, Browning, Coleridge etc can be abbreviated to A,B,C,D... Another example is the dogs problem - owners Edwards, Foster, Grossman, Hunt (E,F,G,H) , dog names Jack, Kelly, Lad, Max (J,K,L,M) and dog breeds Airdale, Boxer, Collie, Doberman (A,B,C,D).

Analytical reasoning problems are, then, a curious mixture of cooperatively and uncooperatively presented information which makes them differ markedly from normal narrative text or everyday dialogue.

Lewis (1989) provides another example of how the way in which problems posed in natural language can be expressed in ways that are fixed against the formal operation required for a solution. She argues that the use of terms such as 'more than' agree with the arithmetic operation of addition whereas often problem presentation is made deliberately inconsistent as in:

'At ARCO gas sells for $ 1.3 a gallon. This is 5c less per gallon than gas at Chevron. How much do 5 gallons of gas cost at Chevron?'

In this case *addition* is the formal arithmetic operation required for solution but the natural language phrasing and syntactic form are inconsistent with priming such an operation because of the use of the phrase 'less per gallon'.

Puzzle solving therefore involves a mind set of linguistic wariness and requires a bag of representational tools for recognising and defusing traps laid by non-cooperative modes of expression. This mind-set is akin to Levesque's (1988) 'puzzle mode' of problem solving and the need to adopt what has been termed an 'extensionalist stance' (Stenning & Cox, 1995).

It also represents what Wollman (1983) has characterised as an 'active operation' approach. Hence an important factor in the interpretation of analytical reasoning problems is the need for more-precise-than-usual interpretations. For example, when reasoning in puzzle situations, subjects often recast the term 'if' into 'and'. This error of interpretation, and others, have received a great deal of attention in the syllogistic reasoning literature where they are termed 'belief biases' (*e.g.* Oakhill & Garnham, 1993). Kintsch (1991) reports, too, that in his studies of word arithmetic problem solving:

> ...a fairly small number of linguistic misunderstandings yielded most of the typical errors that children made' (p. 241)

For example, 'Have more than' is interpreted as 'more', 'altogether' as 'and'; 'some' treated as an ordinary modifier instead of as a number. Kintsch (1991) goes so far as to state:

> word problems are not suitable to train or test the development of logico-mathematical skills, but rather involve learning to use ordinary language in a special way' (p. 241).

Kintsch points out that the avoidance of special, abstract, language and the embedding of the problem in a rich, familiar context has dramatic positive effects upon solution frequencies. He states that comprehension failures are central to the difficulty of word algebra problems.

Kintsch (1991) reviews Larkin's (1989) work on display-based reasoning. In display-based reasoning, the problem is structured in a way that permits it to be solved via procedures immediately available in the world. However, he points out that more complex problems require symbolic manipulations in some problem model. Hence there is a differentiation between reasoning in the problem model and reasoning in the world. Decontextualization (loss of connection between problem model and real world) is a major problem in education - for many students, physics, for example, remains as 'book knowledge'. Kintsch (1989; 1991) distinguishes between the *textbase* and *situation model*. The textbase is the mental representation of the text that a reader constructs from propositions in the process of comprehension. The situation model is a mental representation of the situation described by the text. The two mental representations are not independent of each other — each has its own characteristics and supports some types of behaviour but not others (Kintsch, 1989). The formation of a situation model does not necessarily depend upon a full linguistic parse of the presented text. When the situation is familiar (as is the case for many analytical reasoning problems), the situation model can be formed in the absence of a coherent and well-organised textbase. Conversely, subjects sometimes construct a good textbase without being able to form a situation model.

Kintsch (1991) suggests the use of a mediating 'situation' model to map between real objects and events and the abstract symbol level. To illustrate, the situation model for a typical word algebra problem might be:

> 'a slow aeroplane leaves first, then, after a while, the second one follows; at some point it will overtake the first one'

In the case of a typical analytical reasoning problem, (*e.g.* Problem 1, Appendix A), the situation model corresponding to the text of the problem stem might be:

> 'There are 6 workers and 6 offices, everyone must be assigned to an office, one person per office, in such a way that everyone is content'

The use of a situation model prevents students from building formal problem models from text simply by:

> 'plugging numbers into equations, without explicitly considering the relation of the situation to these equations, and as result misinterpret the

problem[4]' Kintsch (1991; p.240).

Lepik (1990) took a different approach to the study of linguistic factors in word algebra problems from that of Kintsch. Lepik investigated the effects of the number of words, mean word length and similar indexes of complexity. Structural variables included the number of given quantities in a problem, the number of wanted quantities, number of equations required for solution, *etc.* Generally, the linguistic variables were not good predictors of performance in terms of correctness of responses. Several, however, were positively correlated with time spent on the problem. These were the number of literals in the problem, the number of letters, words, units and numerals in problem and the number of sentences in the problem. Several were negatively correlated with time spent — number and proportion of words in the problem with more than 6 letters. Structural variables (*e.g.* known quantities, wanted quantities, number of formulae/equations required for solution, number of relations between quantities) predicted both the proportion of correct answers and time spent more extensively than linguistic variables. Several structural variables were negatively correlated with correct solutions and also positively correlated with problem solving time. The most significant of these was a compound variable defined as the total number of formulae, equations, and known/unknown quantities in the problem. Lepik's study shows that for algebraic word problems at least, *structural* variables tend to exert more effect over time and correctness of solutions than the linguistic attributes of problem wording. An analysis of analytical reasoning word-problems using Lepik's approach would be an interesting topic for future research.

Effect of structural factors upon comprehension

The effects of structural factors in analytical reasoning problems have been analysed in a series of studies by Schwartz and colleagues (Schwartz, 1971; Schwartz & Fattaleh, 1972; Polich & Schwartz, 1974). They have shown that difficulty is related to the number of dimensions in the problem, the number of values along each dimension, the amount of information presented negatively, the number of disjunctions and the

---

[4] This tendency to use formulaic solutions is similar to Greeno's (1977) notion of comprehension errors of correspondence.

amount of implicit information contained in the problem stem.

Several more structural factors affect comprehension, though. Problem–solvers do not always recognise the well-structuredness of problems. Schoenfeld (1988) points out that the nonsense problem:

> There are 26 sheep and 10 goats on a ship. How old is the captain?

has achieved 'folklore status in European mathematics circles' because, although it is clearly absurd, students will try to solve it. French and Swiss research has found that three quarters of 4th and 5th grade students attempt solutions — the most frequently given answer being '36'. Like many mathematical phenomena, this one is probably not confined to younger students. The failure to recognise such problems, Schoenfield argues, is often the result of instructional approaches that emphasise computational algorithms at the expense of meaning. Greeno (1977) and Clement (1982) argue that if students were instructed in ways that emphasised meaning, they would be better equipped to identify 'impossible' problems.

In many respects the phrasing of many analytical problems is reminiscent of the 'How old is the captain?' example — they often seem to be deliberately couched and phrased in ways designed to make their degree of well-structuredness difficult to assess. Consider the following 3 sample sentences taken from 11 given in the problem stem of an analytical reasoning problem used by Schwartz (1971):

1. The hyena's owner doesn't live in the white, yellow or green house.
2. Neither the Japanese, the Indian nor the Englishman lives in the green house.
3. Neither the American nor Canadian owns a zebra.

Moreover, the instructions that accompany analytical reasoning problems do not indicate that they are well-stuctured:

> In this part, each question or group of questions is based on a passage or set of conditions. In answering some of the questions, it may be useful to

> draw a rough diagram. For each question, circle the best answer choice given.

The use of the term 'best' rather than 'correct' in the final sentence is probably deliberate.

It is argued that the difficulty of detecting the well-structuredness of a problem has profound implications for ER selection.

A crucial ER selection decision concerns whether or not the representation needs to express indeterminacy (abstraction). Obscuring the problem's degree of structuredness can therefore result in two kinds of ER selection error—assigning an abstraction–expressing (non-weak) representation to a determinate problem or using a weak (determinate, unique and single model) representation on an indeterminate problem. As Cox, Stenning & Oberlander (1995) have shown, both result in poor performance, though much more so in the latter case than the former.

Contextual factors are also very important in analytical reasoning. Recent work by Bernado & Okagaki (1994) in the domain of 'Students and professors' type problems has emphasised the importance of *problem information context* (PIC). Bernado & Okagaki (1994) suggest that in order to arrive at the correct equation subjects must see the equation as a dynamic representation of an operation, rather than as a static representation of verbal information — *i.e.* to use an operative approach. In a series of experiments, Bernado & Okagaki considered two factors they felt were central to an operative approach. These were: (a) knowledge of about the meaning of mathematical symbols and (b) problem information context (PIC) in which this knowledge is accessible. In a first experiment, they provided students with a study sheet containing symbolic knowledge[5]. This had a positive effect on performance. They also point out that the 'students and professors' equation-writing task lacks a specified problem goal of the form 'how many students are there?' It also lacks information that a procedure will be needed to reach the goal (*e.g.* 'There are 94 professors'). Together these produce a lack of problem information context. Bernado & Okagaki predicted that full PIC versions of the problems used in the equation-writing task would enable subjects

---

[5] Sample item: 'Placing numbers next to symbols indicates a mathematical operation'

to access symbolic knowledge. This prediction was supported by data from two experiments. A third experiment manipulated both the amount of symbolic knowledge and the PIC. The results showed that the performance of subjects in 2 groups (knowledge only; knowledge plus PIC) did not differ. Bernado & Okagaki interpret this finding as 'consistent with the hypothesis that having the appropriate problem-information context helps problem solvers by facilitating their access to all relevant problem-related information, including pertinent knowledge about the meaning of mathematical symbols or notations' (p218).

The Berndado & Okagaki PIC results have a bearing on analytical reasoning problems. In analytical reasoning, the information given in the problem stem does not constitute the full problem information context. Subjects need to 'look-ahead' to the questions associated with the information given in the problem stem. They should not just proceeed with their solutions on the basis of information given in the problem stem alone. Empirical results to be presented in Chapters 5,6 and 7 will demonstrate that attempts to select and build ERs that are made without considering the full PIC often end in erroneously constructed representations and poor reasoning performance.

To conclude, there follows a summary of the main points in relation to problem interpretation and comprehension. In order to reason effectively, subjects:

- must negotiate complicated syntactic structures and overcome the uncooperative aspects of the problem's discourse
- need to use the full problem information context (PIC)
- should recognise the number of dimensions (variables) in the problem
- must accurately discern the number of values along each dimension
- need to understand that some information is implicitly stated
- must utilize negatively-stated information
- should accurately gauge the problem's level of determinacy
- should detect redundancy, symmetry and critical points in the problem (Amarel, 1968)

- should adopt an 'extensionalist stance' or 'puzzle mode' mind-set

Following a period of problem interpretation and comprehension, the reasoner next makes an ER selection decision.

ER selection

> ...the choice of appropriate representations is capable of having spectacular effects on problem solving efficiency (Amarel, 1968, p.170)

In solving analytical reasoning problems, the majority of subjects choose to reason with external representations (*e.g.* Schwartz, 1971; Cox & Brna, 1995). ER selection is a crucial phase of a problem solving, but, to date, there has been much folk wisdom and speculation but little empirical work on the issue. 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 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 determinate 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 ER that a subject chooses will depends, *inter alia*, upon:

- interpretation and comprehension of the problem's explicit and implicit information

- interpretation and comprehension of problem solving task posed
- an ability to recognise the salient features and characteristics of the problem
- a matching of the problem's salient features to an appropriate ER formalism
- an appropriate ER formalism being in the subject's repertoire
- the subjects cognitive style and/or tendency to use *external* representations

As suggested in Chapter 1 in the discussion of the semantic properties of ERs, the optimal representation is one which has sufficient expressive power for the task, *but not any more power than is needed.* For determinate problems, diagrams (weakly expressive ERs) are useful because there is a unique model of the problem's information which a diagram can represent. Indeterminate problems require representations whose degree of specificity is well matched to the amount of abstraction contained in the problem. That type of problem requires the use of an expressive representational system such as natural language or logic. Alternatively, graphical abstraction 'tricks' might be employed in order that a single representation stands for multiple models. An example of such a graphical trick might be the use of a dotted circle in a set diagram which stands for an expandable/shrinkable set of entitities. Another example is provided by Hyperproof's blocks-world. In Hyperproof, a graphical, computer-based, logic teaching program, an object shaped like a crumpled paper bag stands for an object of unknown shape[6].

Analytical reasoning problems are posed verbally, in natural language. Natural language is highly expressive and the determinacy level of the problem is one of the goals of comprehension. As mentioned earlier, a subject's ability to assess whether or not a problem can be modelled with a single, unique model or whether adequate representation requires multiple models, is a major goal of comprehension and a crucial precursor to the selection of an effective ER.

---

[6] Possible shapes are cube, tetrahedron and dodecahedron.

The role of prior knowledge in ER selection

A subject's knowledge of a domain can be partitioned into several components — declarative knowledge, procedural knowledge, cultural knowledge *etc.* Another category consists of knowledge about the representational formalisms associated with a domain. The extent of the subject's representational repertoire for a domain will determine what choices he or she has for solving problems within it.

Such knowledge has to be learned. As Guri-Rozenblit (1988, p.221) writes, in the case of graphical representations:

> The once popular view that graphics are self-explanatory is no longer plausible... visual symbols of any greater complexity than those portraying simple objects require knowledge of the symbolic form to be intelligible or explained...literacy is required for pictorial interpretations as much as for textual interpretation...the learner might encounter difficulties in reading, decoding and understanding visual information, not only because he may have a 'low' visual aptitude, but rather because he was not trained to do it.

Petre & Green (1993) also emphasise the roles of training and experience in the interpretation of (graphical) representations. Their work will be introduced below in the section on reading—off solutions from ERs.

The *switchER* studies to be reported later highlight the importance of prior knowledge 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 in Appendix A. With more complex, multi-dimensional problems (such as problem 3), however, not all forms of tabular representation are equally effective, as will be shown later. 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? This is not the first call for such an innovation. In 1965, for example, Balchin & Coleman wrote:

> It is hoped that the concepts of graphicacy and ingraphicacy will be taken up and developed by educationists, to mould the vague idea of visual aids at large into a more integrated goal of education, and to carry it down into the earliest stages to take its rightful role as one of the essential underpinnings. (p. 947)

The educational gains from such interventions might be considerable. The results of the intervention study by Frandsen & Holder (1969) (reviewed in the next section) suggest that even one hour of instruction in diagramming techniques produces significant score gains. In another intervention study, Lewis (1989) studied 'compare' word problems in 96 college students. The subjects were selected from a larger pool of 299 students and were those who had manifested reversal errors when problem solving (*i.e.* who used the inverse of the correct arithmetic operator when solving test problems). The Lewis study is reviewed in detail in the next chapter, but, in brief, Lewis found that, to be effective in her sample of 'buggy' students, training in diagramming techniques needed to be combined with training in translating the problem (*i.e.* developing what Kintsch would term a situation model of the problem). Significantly for this discussion, Lewis noted that students learned and used the necessary ER skills with very little time and effort.

Prior knowledge of representational formalisms can be independent of the problem information context or of a problem solving strategy *using* the formalism. Novick (1990) has shown that transfer of a *representational* strategy (a matrix ER) can occur between two problems in the absence of a shared problem *solution* strategy between the two problems. In the Novick study, 75% of experimental subjects (who had been exposed to matrix ERs in a problem solving context) transferred a matrix ER to the transfer problem compared to 21% of control subjects. The transfer problem was best solved using a matrix representation, but required a different solution strategy from that required on the prior matrix problem. Novick's (1990) results suggest that the transfer effect was more than a mere recency phenomenon because the prior matrix representation problem was the second of three initial problems that experimental subjects received prior to the transfer problem. Thus, the prior matrix problem was embedded in two other problems requiring (each of which required a different ER) and there was a non-matrix problem interposed between the prior matrix problem and the target problem. Lindvall et al. (1982) also report good transfer of representational skill in primary children from less to more complex word problems. They also report that primary children can validly adapt ERs to new problems.

Individual differences in ER selection - effects of cognitive style upon ER selection and the relationship between modes of internal representation and ER behaviour

The results of both the workscratching and *switchER* studies (to be presented in later chapters), show 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.

As mentioned in chapter 1, the idea of cognitive style variation is not new. The mathematician Poincaré believed that mathematicians could be divided into 'geometers' (ones who think in images) and 'analysts' (conceptual thinkers). This dimension of cognitive style has continued to attract a moderate amount of research attention. For example, individual differences along what can very loosely be termed the 'visualiser–verbaliser' dimension have been shown to be important in reasoning with ERs (MacLeod, Hunt and Mathews, 1978; Matsuno, 1987; Riding & Douglas, 1993; Cox, Stenning & Oberlander, 1994, 1995; Stenning, Cox & Oberlander, 1995; Oberlander, Cox & Stenning, 1994, 1995).

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. Individuals vary in their sentence—picture verification strategies: some subjects recode the pictorial information into verbal form and perform an internal verbal-verbal comparison. Others recode the verbal information into pictorial form and perform an internal comparison of images. Of course, as Roberts, Wood & Gilmore (1994) are correct to point out, some subjects may vary their strategy from trial to trial rather than consistently utilize one recoding route. There may be intra-individual variation in strategy as well as inter-individual differences.

More recently, Ford (1995) analysed the paper-based protocols of people trying to solve 3 term syllogistic reasoning problems and who were required to think aloud and to explain to another person how they reached their conclusions. Ford's subjects were presented with the 27 (out of 64) syllogisms for which valid conclusions can be found.

From the protocol data, Ford identified two distinctly different strategies. Of 20 subjects in her study, she identified 8 who she termed 'verbal' reasoners and 8 who she termed 'spatial' reasoners[7]. The strategy preferences of the subjects were strong. Verbal reasoners manipulate the verbal form of the syllogism, creating and following rules and substituting subject terms from one premise into another. Spatial reasoners reason primarily spatially, though they do keep the verbal tag of the premises in mind. The protocols of spatial reasoners show that they used Eulers Circle-like representations. Ford divided the 27 'valid conclusion' syllogisms into two groups — those that are difficult to solve using the verbal strategy and those that are difficult to solve using spatial strategies. She found that for the 6 syllogisms where the set boundaries are constrained (*i.e.* one-model cases), spatial reasoners performed at the level of 80 to 100 percent correctness, whereas for verbal reasoners the results were not so homogeneous.

Cox, Stenning & Oberlander (1994) and Stenning, Cox & Oberlander (1995) report that subjects skilled at reasoning on diagrammatic reasoning problems demonstrated faster acquisition of first-order logic from Hyperproof, a computer-based learning environment

---

[7] Two subjects used mixed strategies and 2 subjects struggled with the task.

that employs both graphical and syntactic modalities (Barwise & Etchemendy, 1994). In the Hyperproof studies, subjects were classified as diagrammatic reasoners or non-diagrammatic reasoners on the basis of their *performance*[8] on analytical reasoning problems of the diagrammatic type.

Hyperproof boosts students previously strong on items which benefit from diagram use, whereas the syntactic course appears to degrade the same group of students' graphical strategies. Cox, Stenning & Oberlander (1995) analysed the students' free choices of representation on the pre- and post-course analytical reasoning tests. Hyperproof improved representation selection accuracy on analytical reasoning post-text items compared to control class subjects. 'Diagrammatic' subjects were less likely than 'non-diagrammatic' subjects to select weakly expressive representations for problems requiring abstraction. That particular error was associated with particularly serious detrimental affects upon performance. Thus it seems that the Hyperproof and traditional logic teaching methods have their differing outcomes at least in part because of their effects on representation selection.

While different teaching has different effects on students' external representation strategies, it has also been demonstrated that the same teaching to students with different pre-course aptitudes results in different proof styles, and that these hinge on the use of Hyperproof's semantic devices for expressing abstraction (Oberlander, Cox & Stenning, 1994; Oberlander, Cox & Stenning, 1995).

Further analyses of subjects' patterns of rule use while using Hyperproof (Oberlander, Cox, Monaghan, Stenning, & Tobin, 1996) suggest that the nature of the individual differences may not be quite as simple as modality preference for graphical reasoning (in the case of 'diagrammatic' subjects) or linguistically based reasoning ('non-diagrammatic' subjects). Rather, non-diagrammatic subjects, who were previously believed to be subjects who prefer the linguistic modality, in fact tend to concretise

---

[8] It should be noted, however, that most learning style studies that have investigated the visualiser–verbaliser distinction have used *psychometric* instruments as the basis for classifying subjects. For example, the paper-folding test has been used by Mayer and Sims (1994) in a recent study of learning from computer-generated animation; and by Campagnoni & Ehrlich (1989) in a study of individual differences in hypertext navigation. It is currently unclear, however, how strongly internal behaviour (as measured by paper-and-pencil psychometric tests) is related to external reasoning performance. For this reason, classifying subjects on the basis of performance measures is far less equivocal than using psychometric tests.

(add determinacy to the graphic) to a greater extent than diagrammatic subjects. What distinguishes diagrammatic subjects is their greater tendency to *translate* between graphical and sentential modalities in *both* directions. Multimodal reasoners *interact* with Hyperproof's graphical situations more than unimodal reasoners — they use current graphical situations as input to later stages of their proofs, they manipulate the graphical situation to mediate between modalities and make much more use of the Hyperproof rules that permit information to be transferred between modalities. In contrast, unimodal reasoners tend to just *output* graphics, without subsequently using them strategically in proof development.

Riding & Douglas (1993) found that subjects who were classified as 'visualisers' on the basis of psychometric tests 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.

Frandsen & Holder (1969) selected subjects who were matched in terms of verbal reasoning ability but who differed in terms of their spatial visualization ability. They used a psychometric test[9] of spatial relations to classify their subjects into high and low spatial visualization groups. They provided instruction in the diagrammatic representation of verbal problems[10] 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 cognitive trait. A fur-

---

[9] Differential Aptitude Test of Spatial Relations.

[10] Syllogisms, time-rate-distance problems and logical-deduction problems.

ther 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 that they imagined internal diagrams 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.

Superficially, it seems reasonable to assume that an individual's location on the V-V dimension should predict the kinds of ERs that s/he uses when reasoning. However, the relationship between internal (mental) representation and external representations is not well understood. Much more research is required. To the author's knowledge only the Riding & Douglas (1993) study has shown that subjects classified as 'visualisers' tend to use diagrammatic ERs more than subjects classified as 'verbalisers'. The Riding & Douglas (1993) study is interesting in that it is the only one, to the author's knowledge, that has demonstrated a correlation between internal representational modality preference (cognitive style) and external representational behaviour.

Even in that study, however, the relationship was far from straightforward since an interaction with presentation format was reported. This accords with the findings of Guri-Rozenblit (1988), who, in a study of the use of abstract diagrams in social science texts, has shown *inter alia* that the mode of presentation of information was a stronger influence on the modality of students' responses than students' initial cognitive style as assessed by verbal and visual aptitude aptitude tests.

An implicit assumption in many studies (*e.g.* Frandsen & Holder, 1969), is that subjects who score poorly on psychometric tests of spatial visualization therefore prefer or tend to use the linguistic modality internally and/or externally when reasoning. This assumption is not justified and the relationship between internal cognitive modality preferences and the use of external representations requires much more research. For example, it would be interesting to examine whether subjects classified as highly spatial reasoners or diagrammatic modellers respond to efforts to broaden their ER repertoire by training in the use of non-graphical external representations. To the author's knowledge, this has never been demonstrated.

Another potentially important individual difference, one that is orthogonal to linguistic/graphical modality preferences, and one that has received no research attention, is the extent to which individuals *externalise* their reasoning. The results of the studies to be reported in this thesis and those of others (*e.g.* Schwartz, 1971) show that most subjects (usually > 80%) use ERs in their solutions to ER problems. But do the ERs serve the same function for all subjects? Subjects differ in the extent of their externalisation; some subjects reason with no ER whatsoever, others use 'minimal' ER strategies and some use 'full blown' diagrammatic models. As we have seen, subjects certainly differ in the modality of the ERs they use - partly due to cognitive style effects, partly due to the representational demands of the problem, and partly due to their prior experience and ER repertoire. But even in the case of subjects who build 'full blown' ERs, the ER may be more 'central' to the reasoning of some subjects than for others. ERs probably serve different functions for different people. Subjects probably vary in the way in which they partition their internalised and externalised cognition. The proportion of reasoning that is externalised by a particular subject may also vary at different stages of reasoning. Some subjects may use an ER to keep track of their progress through the problem, while reasoning internally for the most part. Other subjects may exploit the cognitive and semantic properties of the ER fully in their reasoning, adopting a model-based mode of reasoning.

ER construction

> Solving a problem simply means representing it so as to make the solution transparent. (Simon, 1981)

As Barwise & Shimojima (1995) point out, the process of building an ER necessitates reasoning about the problem, reasoning about the representation and also reasoning about the relationship of the representation to the problem. Compared to reasoning without external aids, they write ...'we seem to replace one problem with three problems' (p. 9). However, ER construction assists problem solving by re-ordering the information in ways useful for solutions and by laying out the range of possible models of the information, making missing information explicit, and representing implicit information explicitly. However, to effectively perform those functions, an ER must be constructed correctly. Slips and more profound errors can occur at any of several points in ER construction.

The process of ER construction also assists reasoning in another way. Attempting to construct a model of the information in a problem helps the problem solver decide whether a single, unique model adequately expresses the information in the problem stem. If not, the user may need to switch to a more expressive representation in the same modality as the first attempt (*e.g.* from a tabular representation to a set diagram) or to a different representation in a different modality (*e.g.* from a table to first-order logic or natural language).

ER construction, then, is a stage of reasoning in which some or all of the following processes may occur:

- reformulation of sententially presented information perhaps by re-ordering sentences given in the problem stem
- translation of given information *e.g.*to graphical from linguistic-sentential

- radical re-representation (*e.g.* Peterson, 1994) - very rare
- heterogeneous reasoning (Barwise, 1993) – concurrent use of ERs from both modalities, also referred to as the use of multiple representations by diSessa (1979)
- heterogeneous reasoning via switching between representations in serial manner (Cox & Brna, 1993a,b;1995)

There is not a great amount of information about ER construction in the literature because most studies examine residual workscratchings in which process information not captured or preserved. A notable exception is work by Katz & Anzai (1991) who used a computer-based logging system to record the ERs produced by a subject. Katz & Anzai (1991) studied an undergraduate as she learned to represent and solve vector arithmetic problems. She studied a physics textbook which contained vector diagrams and attempted a set of 16 problems three times. Verbal protocols were taken and she used a computer-based system which logged her interactions. Katz & Anzai (1991) analysed her written and 'think aloud' protocols. The role played by the vector diagrams changed as she became more proficient. Early on, the diagram was a literal translation of the problem and reasoning was strongly diagram-driven. When more practiced, she was able to abstract more important elements of the problem (*i.e.* vectors) and solve the problem using domain-specific methods she had learned. The acquisition of the domain specific solution strategies seemed to be assisted by her diagrams. As the subject's expertise developed, the vector diagrams assisted the student to recognise useful calculations that she may not have otherwise discovered – this is an interesting finding and was discussed in Chapter 1 in relation to the self-explanation effect.

The use of time-stamped, user—system interaction logs can provide an excellent method of gathering timing information, the order in which elements of an ER are constructed, whether or not the subject uses a representation subsequent to construction and how ERs are used in the resolutions of impasses in reasoning, whether errors in ER construction are made, detected and rectified and so on. Hence the use of the technique in the *switchER* studies (Chapters 6 & 7). Computer logs have several advantages over videotaped protocols in that they can time stamp and capture low level user–system interactions such as keystroke sequences and mouse/cursor positions.

Read-off from ER and responding

When an accurate ER has been constructed, reading off information requires, in comparison to ER construction, far fewer cognitive resources and represents the 'pay off' from the effort expended on construction.

Holyoak & Spellman (1993) point out ...

> Whereas inference rules operate on given premises to yield conclusions, read-out procedures operating on a model completely blur the distinction between 'premises' and 'conclusions.'

But this stage of reasoning with ERs has its share of problems. Later, it will be demonstrated that subjects' responses are not always commensurate with direct read-off from their ER. In the case of an accurately constructed ER, non-commensurate responding may reflect slips made during read-off. These slips may have catastrophic effects for some tasks but may not affect others.

In the case of an incorrectly constructed ER, read-off slips are compounded with the errors already inherent in the model. Additionally, the subject may or may not realise that his or her constructed model is inaccurate. If the subject has insight into the fact that the ER may be incorrect, s/he may use it selectively - in this case the subject may not bother to overtly correct the ER but may overlay 'corrections' via mental representation.

Another source of error at this stage is due not so much to the ER being incorrect but due to the form of its construction. For example, when a table is constructed, decisions have to be made about which dimensions to make salient or visible (see Green, 1989; Gilmore, 1991). Contingency tables are also very difficult to search for information. If the subject has insight into the problem with his or her ER, then s/he may decide at this stage to correct the ER. This may prove quite hard to do if the ER is difficult to edit and modify *i.e.* is viscous (Green, 1989). Alternatively, s/he may reconstruct the ER in a more useful form (*e.g.* recast multiple 2 × 2 contingency matrices into a unified $N \times N$ table. Another option is for the subject to switch to a different type of ER.

Apart from the salience of dimensions in a particular ER type, such as a table, the modality and format of ERs is also important. As Larkin & Simon (1987) point out, a diagram can be described sententially in the form of a list of propositions. However, the list can be extensive if the description is to be comprehensive and search for information is error prone and difficult.

There is practically no research on subject's behaviour at the read-off stage of reasoning with self-selected, self-constructed ERs. As in the case of ER construction, very few studies examine the time course and process of reasoning with ERs using a methodology that permits events at each stage to be monitored.

There is some work on ERs and search, however, but it has all been conducted in studies where subjects are presented with pre-fabricated ERs rather than in situations where subjects select and construct their own representations (O'Donnell, 1992; Bartram, 1980; Guthrie & Mosenthal, 1987; Guthrie & Kirsch, 1987; McGuinness, 1986; Mayer, 1976; Petre & Green, 1993; Day, 1988). For example, Bartram (1980) compared the effectiveness of several representations (conventional and schematic maps, and sequential and alphabetical lists) for the task of working out a bus route between two locations and found that the schematic map was the most effective form of representation for that task in terms of time-to-solution. Lists were affected to a larger extent than maps by changes in task complexity as would be predicted by specificity theory (Stenning & Oberlander, 1995), since the cognitive availability of the information expressed sententially is less than is the case for the equivalently expressive graphical representation.

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 is much more useful for rapid qualitative comparisons between sets of data. One representation is suited to one task requirement (read-off of precise values) whereas another is suited to a different task (comparison).

Day (1988) and Norman (1993) provide numerous examples of how a good fit between the task and the representation can facilitate search and read-off. Day (1988) provides an everyday illustration from the domain of medicine. The information in a medical prescription is best laid out, from the pharmacist's point of view, in the form of a list. From the patient's point of view, however, a better arrangement is a matrix representation in which times of day (breakfast, lunch, dinner, bedtime) form the columns and drug types (lanoxin, inderal...etc) form the rows. The tabular configuration facilitates accurate read-off and better compliance with the therapeutic regime.

Mayer (1976) studied the effects of several modes of information presentation on a large sample of university students. The problem used was one in which various sports teams were matched in a tournament. Test questions asked about the outcomes given certain conditions. The problem was presented in eight formats. Four were verbal formats (propositional), namely: 'jump' and 'short-jump' (itemised sentences arranged in list format with 'go to' statements), 'nested' (in form of computer program listing) and 'example' (matrix table). The remaining formats were graphical and consisted of flow chart analogues of the verbal formats. In answering the test questions, subjects reasoned with the presented representational format and did not construct their own ERs. The results indicated that the graphical flow chart format produced better reasoning performance for the 'jump' and 'nest' conditions. The verbal (propositional) version of the 'example' (matrix) condition was superior to the flowchart version. Mayer (1976) concludes that flowcharts are suited to the representation of complexly structured information whereas verbal formats are optimal for efficiently-structured information. In the Mayer (1976) study, the 'example' (matrix) format of information was more effective in a verbal format than in a flowchart format — the difference between the verbal format of the 'example' stimulus and the other verbal formats studied was that the 'example' verbal format minimised the amount of visual search required to read-off conclusions.

O'Donnell (1992) investigated the influence of three variables upon information search. The three variables were presentation format ('knowledge maps' versus text), subject characteristics (vocabulary level, prior knowledge of domain) and the nature of the information to be located. Her results showed that all of the variables affected search efficacy. Knowledge maps facilitated search for declarative, bottom-up information and subjects with higher prior knowledge outperformed those with lower prior knowledge.

Prior knowledge and expertise has also been shown to be important in studies by Petre & Green (1993). They studied expert electronic hardware designers' use of circuit diagrams and the relative readability of textual and graphical notations in computer program representations. Petre & Green (1993) argue that users of ERs *learn* to read ER formalisms and are active in the process in much the same way as skilled readers use typographic and semantic cues in text. Petre & Green (1993) compared 4 types of program representation (text—'and/or' notation; text—nested conditionals; graphical—nested conditional; graphical—and/or notation) in tasks that required either top-down or bottom-up reasoning. Each subject received top-down and bottom-up questions for each representation type in a within-subjects design. The hypothesis was that nested conditionals support working forwards and that 'and/or' supports working backwards. The results supported the hypothesis but graphics were shown to be slower than text in all conditions for that type of task (program tracing). Expert and novice subjects were similar in terms of response latency. Expert subjects, however, used graphical representations in qualitatively different ways than novices. They used secondary notation cues to greater effect. Petre & Green (1993) conclude that what is salient in a representation largely results from experience — 'what a reader sees is largely a matter of what he or she has learned to look for' (p.69).

McGuinness (1986) investigated a completely different domain. She presented adults with information about family relationships between 2 generations of 4 related families. The information was presented in one of two ER forms - a hierarchy (family tree) or a matrix in which birth order in the 2nd generation formed the rows and birth order in the 1st generation formed the columns. Each subject learned one representation to mastery and then used an internal (mental) version of it to answer two kinds of questions. These were rule based across-family questions (*e.g.* which cousins can go on holiday with other cousins according to birth order and sex rules. The second set of questions were within-family inheritance questions. The results showed that holiday questions were easier in the matrix than in the hierarchy condition and that inheritance questions were equally difficult to answer with both (internalised) representations. McGuinness (1986) explains her findings in terms of the number of steps each representation requires in the search process. There are many more adjacent elements to search in the hierarchical representation than in the tabular representation though McGuinness acknowledges that steps vary in their cognitive demands according to whether they are rule-relevant or other kinds. In a second experiment, in which birth order was made the salient variable in the representations (as opposed to generations as in the first experiment). This reversed mapping produced reversed effects of the representations. Despite the rather unusual feature of having subjects memorise representations, this study, like those of Day (1988), Norman (1993), Gilmore (1991) and Petre & Green (1993), demonstrates the importance of mapping information to representations in ways which make salient the aspects of the information required by the task.

Guthrie & Mosenthal (1987) distinguish between *reading* goals using different kinds of material (prose, forms, tables, schematics, graphs ...) and the *information search* goals that the user may have (knowledge, specific information, evaluation, construction ...). Their distinction is relevant to the issue of reading-off solutions from ERs in analytical reasoning because the task is one of locating information in a text or diagram and not solely one of reading for comprehension. Guthrie & Mosenthal (1987) discuss what they refer to as 'a critically important class of reading tasks termed 'locating information in written documents''. They argue that 'strategic reading' *i.e.* searching for specific information) is a ubiquitous task at home and school, yet most models of reading focus on reading prose for comprehension.

Guthrie & Mosenthal (1987) propose a 5 stage cognitive model of information search:

1. form a clear search goal (What is information being sought?)
2. inspect appropriate categories of information
3. sequence the inspection
4. extracted details from one or more categories
5. recycle to obtain solution

The Guthrie & Mosenthal (1987) model can be applied to reading off solutions to analytical reasoning problems from ERs. The first step is specified by the problem's question. The second step, category selection is more or less straightforward depending upon the representation. For example, in a representation of the information in problem 3 (A), such as that shown in Figure A.12 of Appendix A, searching for the name of the owner of the dog that wins second prize involves identifying the second column from the left and the middle row. Next the row must be searched until the cell representing the intersection with column two is located and the cell content read-off.

Summary and conclusion

This chapter has identified 5 stages in the process of analytical reasoning with external representations.

The first stage, problem comprehension, requires background knowledge and conceptual knowledge. Slips of omission and commission must be avoided and new information must be integrated into existing knowledge through the use of a situation model.

Obstacles to comprehension include linguistic factors such as syntactic complexity, implicitly-given information, and uncooperative discourse structures. The reasoner must be wary and adopt a puzzle mode of thinking. The task involves the use of ordinary language in a special way. Structural factors also result in obfuscation — the reasoner must decide on the well-formedness of the problem, its degree of determinacy, the number of dimension (variables) it possesses and the number of entities along the dimensions.

The development of an adequate problem information context (PIC) cannot result from only comprehending the problem stem information. The reasoner must also read the problem's questions and evaluate the task demands they pose. Results to be presented in Chapters 5, 6 & 7 will show that subjects rarely do this, instead they tend to read the problem questions for the first time following ER construction.

The second stage is ER selection. Good ER selection requires important decisions by the reasoner which will be influenced by at least 3 factors - problem comprehension, prior knowledge (*i.e.* the reasoner's ER repertoire) and the reasoner's cognitive modality preference or cognitive style. An appropriate level of granularity of the representation has to be adopted (Amarel, 1968). There is some folk wisdom but very little empirically-based guidance available that can be given to subjects to aid them in choosing an appropriate ER. One heuristic, however, derived from the work of Stenning & Oberlander (1995) is that the ER should be capable of expressing the indeterminacy of the problem but not any more than is required. The ability to match ERs to problem characteristics requires prior knowledge and experience with a range of ER formalisms. Some representations are ubiquitous and generic (*e.g.* tables and other matrix graphics such as bar charts). Others are relatively domain-specific and have underlying semantics that have to be learned (*e.g.* the spatial inclusion metaphor for set membership that underlies the semantics of set diagrams such as Euler's Circles). Whether or not subjects have representations like set diagrams or semantic networks in their repertoires can often be a matter of chance. The ERs in an individual's toolkit are often acquired in an *ad hoc* manner. Representations such as network diagrams may be familiar to a subject simply because s/he studied food webs in biology, for example. What might loosely be termed 'cognitive modality preference' also plays a role. Individuals differ in their tendencies to use graphical or linguistic representations.

The representation construction phase provides an opportunity for the problem information to be re-ordered, and for it to be re-represented graphically or translated into a different linguistic form such as propositional, first-order logic or an idiosyncratic, 'restricted', logical form. Slips and errors in ER construction must be avoided. The subject may engage in heterogeneous reasoning, using multiple representations either concurrently or sequentially (switching). Effort expended during ER construction is high — the cognitive load is great, but it is well-invested since the payoff comes when the representation is used to speedily read-off solutions. Often, though, the high cognitive load results in construction errors which may or may not be detected by the subject.

Read-off is a relatively easy phase in the cycle but slips are common. Slips and errors in reading-off can be compounded with ER construction errors and result in incorrect responses to problem questions. Read-off is essentially a process of search... a kind of reading that is highly goal-directed. The ease of read-off is greatly facilitated by the choice of ER ... compare reading off solutions from Figure A.12 with read-off from Figure A.15 (Appendix A).

The cognitive demands of analytical reasoning change with each stage and individuals differ in terms of the prior knowledge they bring to problem comprehension and ER selection. Evidence that will be presented later shows that individuals also differ in terms of the modality that they prefer to reason in.

It can be seen from this stage-account that analytical reasoning draws upon a wide range of cognitive processes and prior knowledge.

The next chapter presents a structured review of the literature on reasoning with ERs in a variety of domains. The review attempts to integrate findings from disparate sources. The literature on reasoning with ERs is very distributed - both across subject domains (science, maths, logic *etc.*) and across disciplines (education, pyschology, cognitive science, A.I. *etc*).

Chapter 4

Related work

Introduction

This chapter reviews previous work on reasoning with ERs. It is organised around the issues of:

- self-selection and self-construction of ERs versus using pre-determined, pre-fabricated ERs
- domain–specific ER formalisms versus more generic forms
- instructional intervention in use of ERs or not

Several issues emerge from the review — at the most general level, all of the papers are concerned with external representations and reasoning. However, in some studies subjects construct their own representations and in others pre-drawn (prefabricated) representations are used. Some studies have examined what might be called partial self-construction in that, for example, an empty matrix graphics (table) or diagram template may be filled-in by the reasoner. Problem comprehension emerges as a crucial phase of reasoning and several studies investigated transfer of solution strategies or types of external representation to other problems.

Another issue concerns the question of the domain-specificity of ERs. Some domains (*e.g.* syllogistic reasoning) involve the use of highly domain-specific ERs such as Euler's circles (graphical) or the use of syllogistic premises (sentential). In contrast, in domains such as deductive/analytical reasoning, word arithmetic and word algebra, problems can be solved effectively with the aid of a wide variety of representational forms. Generally, studies that use domain-specific or specialised ERs tend to be the ones in which ER instruction is provided to subjects. The studies vary widely in the extent of instruction given on the use of ERs in reasoning. Some studies offer coaching in the use of ERs whereas other studies allow subjects to spontaneously generate any ER they choose in the absence of instructional intervention.

Two distinctions — self-construction versus pre-fabrication and the extent of domain specificity — will be used to structure the review.

ER construction by the subject: studies utilising domain-specific ER forms

Domain: Mathematics 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.

Lindvall, Tamburino & Robinson (1982) gave 23 primary grade children instruction in the construction and use of diagrams in the domain of arithmetic word problems involving addition and subtraction. The type of diagrams were highly domain specific and emphasised groupings and unit correspondence. The subjects were required to construct the diagram. A 4-stage general procedure was taught which emphasised reading or listening to the 'story', drawing a diagram to represent sets, operations or relations in the story, writing a 'number sentence' and solving the problem. The instruction was intensive — twenty-two 40' sessions over 40 days. Subjects showed significant gains from the instruction on answering, modelling and writing number sentences. There was also evidence of transfer of the representation skills to more complex problems on a post-test. Lindvall et al. (1982) also report that the students validly adapted the taught ER forms to new problems.

Willis & Fuson (1988) studied 24 2nd grade students of high maths ability and 19 students of average maths ability on 'change' and 'compare' arithmetic word problems involving 3-digit numbers. Subjects were taught several categories of domain-specific schematic drawings. One type of diagram consisted of a type of tabular representation with cells for the various subcomponents of put-together (combine) and compare problems. The other ER type consisted of two rectangles for the start and end states of change-more and change-less problems, linked by an arrow. Above the arrow was an ellipse into which the change quantity can be entered. Students were introduced to word problems and taught to identify and label the three important elements in the story. They were also taught to make appropriate schematic drawings for each category of problem and to enter the identified numeric elements from the problem. Students practiced the procedure on worksheets. Results on pre and post tests showed that students in both ability groups mastered the stages of appropriate diagram selection and diagram labelling. There was a strong relation between the correctness of diagram selection/labelling and correct solution strategies. Sometimes, the problem category would be incorrectly identified — for example, average students substituted 'put-together' and 'compare' drawings on about a fifth of each type of problem. Students experienced most difficulty with problems which had subtraction as their underlying semantics ('change-get-less' and 'compare') but for which the best solution strategy is one involving addition. Willis & Fuson (1988) argue that their results support the view that teaching students to use different types of diagram, each tailored to a specific category of problem, is more effective than teaching them to use one kind of diagram for a range of problem types (*e.g.* Resnick, 1983).

A later study (Fuson & Willis, 1989) showed that, for some problem types, the drawings are sometimes used by children to illustrate the solution procedure rather than to represent the problem situational structure. They report that future research might investigate the relative effectiveness of the two types of diagram and also the 'tension' within individual students (and teachers) between the two kinds of classification might also be addressed. They conclude that one important function of the drawings is to provide the teacher and children with a common vocabulary with which to discuss problems, especially on those problems where the solution procedure and problem situation contrast. The facilitation of interaction, rather than the drawing *per se*, might be the more crucial factor, they suggest.

Lewis (1989) studied 'compare' word problems in 96 college students. The subjects were selected from a larger pool of 299 students and were those who had manifested reversal errors when problem solving (ie who used the inverse of the correct arithmetic operator when solving test problems). Three groups of 32 subjects were compared. The diagram group were trained in word problem translation and diagramming using a domain specific diagramming procedure. A second group received translation training only and the third group was a control. Translation training consisted of learning about the types of statements found in arithmetic problems. The diagram plus translation group showed significantly greater gains in terms of problem comprehension and problem representation and also some transfer to more complex problems of the same general kind. Far-transfer to more general types of problem was not found, however. It was concluded that translation training alone encouraged a focus on the surface features of a problem (eg a search for key words) at the expense of deeper semantic understanding. In fact the authors go so far as to say that no training at all is better than translation training alone. Lewis also noted that students learned and used ER skills with very little time and effort.

Domain: Physics As reviewed in Chapter 3, Katz & Anzai (1991) studied an undergraduate as she learned to represent and solve vector arithmetic problems. To recapitulate, the role played by the vector diagrams changed as she became more proficient. Early on, the diagram was a literal translation of the problem and reasoning was strongly diagram-driven. When more practiced, she was able to abstract more important elements of the problem (ie vectors) and solve the problem using domain-specific methods she had learned. The acquisition of the domain specific solution strategies seemed to be assisted by her diagrams.

Domain: Computer programming GIL (Reiser, Ranney, Lovett & Kimberg, 1989; Merrill, Reiser, Beekelaar & Hamid, 1992) is a tool with which student programmers may construct visual representations of LISP programs. As in many problem solving domains (eg logical proofs, geometry) and programming languages (Prolog, C), a difficulty with LISP is that the syntax of a solution does not reflect the reasoning process required to construct it. In GIL, users build a graphical representation of a Lisp program by connecting icons (representing program constructs) in a directed graph. One of the research goals was to overcome the tendency of Lisp syntax to conceal programming solution strategies and program structure.

GIL's program graphs make implicit information visible to the student and vividly illustrate (by means of animation) the propagation of cause and effect when the program is run. Reiser et al. (1989) and Merrill et al. (1992) have shown that GIL is more effective than the traditional textbook and programming environment combination. Using GIL, students are forced to make every step of their reasoning explicit and they can focus, if they wish, on individual solution components. The GIL interface also allows the students to mix top-down and bottom-up reasoning strategies more flexibly than is possible in traditional Lisp programming environments. The aim is to ensure that the structure of the graphically depicted solution mirrors the planning processes used in reaching it. GIL thus assists students with their understanding of the causal structure of programs and how algorithms work. The data computed by the program between initial input and final output are shown explicitly and the path of the program's execution is also displayed. Thus students can see and reason about internal program states that are usually invisible. The graphical approach also allows students to plan in a variety of directions — either from given data towards the goal or backwards from the goal towards the given data. GIL makes progress forward towards a goal and backward reasoning steps salient and provides an example of what Merrill et al. (1992) have termed a 'reasoning-congruent representation'.

Domain: Logic-syllogistic reasoning Grossen & Carnine (1990) 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 author is aware, Grossen & Carnine (1990) are the only researchers in the literature to have conducted a controlled comparison of diagram self-construction with the use 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 computer-based 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 types (without valid conclusions) 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.

ER construction by the subject: studies utilising non domain-specific ER forms

This section reviews studies in which students were permitted a degree of choice in the type of representation that they constructed. This is in contrast to studies reviewed in the preceding section where subjects used one particular type of ER (Euler's circles, vector diagrams, directed graphs etc). Those ERs were 'domain-specific' in the sense that they are most closely associated with, or have their historical origins in, a particular subject domain. A clear example is provided by Euler's circles and syllogistic reasoning. There needs to be a distinction made, however, between domain specific research and domain specific representations. The studies reviewed in this section were all conducted in the context of specific subject domains, but the subjects were not constrained to use 'domain-specific' ERs in their reasoning.

Domain: Class inclusion hierarchies Greene (1989) investigated the ability of children between the ages of 7 and 12 years to construct tree-diagram representations of passages containing hierarchical information. The passages described sets and subsets of mutually exclusive information. In the first experiment, subjects were asked to construct their own spontaneous ERs. The results showed that the subjects used a variety of ERs to represent the information — written representations (re-writing the original information), more structured representations (that showed understanding of the hierarchical nature of the information), drawings and tree diagrams. The results also showed that there was a decreasing tendency for children to draw pictorial representations as age increased, and that older children (4th and 6th grade) tended to use more written and structured representations than 2nd grade children.

An instance of a highly structured ERs was noted in the case of one of the 12 second grade children and one of the 12 fourth grade children. Four out of 12 sixth grade children used structured (ie hierarchy capturing) ERs.

Structured ERs were associated with higher representational quality measured on several scored dimensions such as redundancy, presence or absence of features and proper relations between nodes.

The self-generated ERs were used by the subjects in answering identification and reasoning questions. Performance on the question task showed that poor answers were associated with non-graphical ERs in Grade 2 subjects. Structured ERs produced good answers at all ages.

In a second experiment subjects read a passage and were subsequently presented with a corresponding tree-diagram of the information. They were shown how it related to the information in the passage. They then answered a set of question using the tree diagram only. Next, a second passage was read and they were asked to construct their own tree diagrams.

Fourth and sixth graders constructed perfect tree diagrams and responded to questions at near ceiling level of performance. Second grade subjects' tree diagrams and responses to questions were less adequate than those of the older children. Second graders performed about as well using 'coached' tree diagrams in the second experiment as they did with their own spontaneous ERs in the first experiment. Fourth and sixth graders showed slightly more improvement. Most children spent less time constructing tree diagrams than their freely chosen ERs. They also tended to perform better on the question answering task when they used the tree diagrams they had been coached to produce. Greene's (1989) results show that some children as young as 7 years old can understand tree diagrams and can have a substantial understanding of 4-level class inclusion hierarchies in terms of subset/superset classification, transitivity of information and other measures.

Domain: Mathematics Hall, Kibler, Wenger & Truxaw (1988) collected written protocols from 85 mathematically competent undergraduates as they solved a range of algebra word story problems. The subjects were instructed to show all their working and not to erase after making mistakes. Problem types were motion-opposite direction (MOD), round trip (MRT), work together (WT) and work competitive (WC). Analysis of the protocols revealed that diagrams were used more on motion problems than on work problems. Hall et al. (1988) noted that many subjects construct solutions to problems rather than smoothly execute a highly practiced skill and that the constructions often involve reasoning that is only partly connected with algebraic or arithmetic formalisms. Competent reasoners often use problem solving techniques from "outside" algebraic formalism. Furthermore, they write that "conceptual errors of omission or commission are both more prevalent and more damaging than manipulative errors in algebra or arithmetic"(p.269). They observe that problem comprehension and solution are complementary processes and that integrating dual representations (ie at situational and quantitative levels) of a problem is a key aspect of competence.

In their conclusion they write that "reasoning about the situational context of a problem can serve as a justification for assembling quantitative constraints that may eventually lead to a correct solution. Thus, a substantial portion of a problem solver's activity is devoted to *reaching* an understanding of the problem that is sufficient for applying the routine of formal manipulation." (p.269).

Model-based reasoning techniques such as simulation[1] are often used as a means of recovering from impasses reached while using formal reasoning. In discussing the educational implications of their work, Hall et al. note that instruction based solely on mathematical formalisms may not produce learning outcomes that transfer to non-routine problems. They note that textbooks often instruct students to translate problems from words to algebraic forms, sometimes via the use of graphical ERs such as tables. Hall et al. therefore propose that 'combined interactive illustrations' are likely to be effective representations for problem solving.

As mentioned above, the integration of dual representations of a problem is a central aspect of competence according to Hall et al. As an example they present such a dual representation for a MRT problem in which a 2 dimensional graph is used to represent the time by distance relation. The graph is used interactively with a 'quantitative network' (Shalin & Bee, 1985) in which arithmetic operational relations can be represented by a linked network of cells into which intensive and extensive elements can be entered. The quantitative network provide a spatial abstraction of variables and equivalence relations. The network provides a visually inspectable representation of constraint propagation that is far more salient to students than the traditional algebraic operations on linear equations. The quantitative network concept therefore is similar in principle to the 'reasoning-congruent' representations of Lisp syntax proposed by Reiser et al. (1989). The use of two representational modalities (network representation and 2D graph) also provides an example of heterogeneous reasoning (to be discussed further in a later section).

Van Essen & Hamaker (1990) in two controlled intervention experiments, studied self-generated drawings as heuristic strategy for the solution of arithmetic word problem

---

[1] In which students enter a value on one dimension then 'runs' the model, repeating the process with systematically incremented values.

solving. They studied 1st and 2nd grade children (Experiment 1) and 5th grade children (Experiment 2) as they attempted 'combine', 'change' and 'compare' problems. Subjects were trained for 60 to 90 minutes over 3 sessions during which time the experimenter read aloud word problems and the children were instructed to make drawings depicting what was happening in the story. They were told that drawing is often a useful technique for understanding word problems. The experimenter generated drawings in order to illustrate how a drawing might look but they were not prescriptive and were not presented as models to be emulated by the students. Van Essen & Hamaker (1990) wished to avoid teaching specific drawings for different categories of problem since they felt this may encourage superficial problem analysis strategies such as key-word matching. In the results, the authors report that 'First and second graders normally do not make drawings of word problems in order to facilitate problem solutions' (p.305). This statement, however, contradicts the results of studies by Greene (1989) and Fuson & Willis (1989) who found that the tendency to make drawings *decreases* with age. First and second grade subjects in the experimental group did not produce significantly more drawings than the control group. However, of those that did, the experimental subjects' ERs were of higher quality in terms of accuracy and completeness. Van Essen & Hamaker (1990) classified the solutions into 5 types - category 1 drawings 'adequately mirrored the structure of a word problem' i.e. expressed sets and relations were depicted correctly; category 2 drawings depicted sets but not relations; category 3 drawings depicted the answer but not the solution strategy; category 4 were incorrect drawings e.g. only one of the 2 sets depicted or relation was incorrectly depicted; category 5 was no drawing. Three of the intervention lessons described above were given over 4 weeks. Compared to controls, experimental groups subjects produced more drawings at post test and, importantly, more category 1 drawings.

A second experiment examined the effect of the intervention on 5th graders. A 'difficult-to-visualise' category of problem was added for the 5th graders. Intervention resulted in improved performance (answer correctness) on problems similar to those used during practice and also on near-transfer problems. The performance gains were confined to the easy-to-visualise problems however. The number of drawings on all problem types increased significantly following the intervention, even the difficult-to-visualise ones. Van Essen & Hamaker (1990) analysed the relationship between ER quality and problem-solving performance. With the 5th graders, drawings were used 39% of the time in experimental (intervention) subjects, compared to 5% for control subjects. Of the 39% of ERs, 22% were associated with correct answers and 17% with incorrect answers. Correct answers were usually accompanied by correct drawings and, conversely, incorrect answers were accompanied by incorrect drawings. In 10% of the 5th graders' solutions, the drawn ER helped the child find the correct answer. However, Van Essen & Hamaker (1990) point out that drawing is an heuristic strategy which does not guarantee finding a correct answer — 17% of 5th graders' ERs reflected interpretational errors. The authors conclude that 5th graders are capable of appreciating the usefulness of self-generated drawings in solving word problems and can work out which problems to use the technique with (as evidenced by fewer ERs constructed on the difficult-to-visualise problems. Van Essen & Hamaker (1990) demonstrated that general instructions and modelling in the use of ERs in reasoning can have a significant effect upon problem solving outcomes. It is interesting to note also that in the Van Essen & Hamaker study, training was not given in the use of a domain-specific type of ER.

Koedinger & Tabachnek (1994) took written and "think aloud" protocols from 12 undergraduate students as they solved 2 word algebra problems. Subjects were simply asked to solve the problems and were not directed to use a particular method. Protocol analysis revealed the use of four strategies: algebra (formal), model-based reasoning (informal-guess and test), verbal arithmetic (informal) and diagrammatic (semi-formal). Students who used multiple strategies in the course of their solutions generally seemed to perform better. When a problem solving impasse occurred, approximately half were responded to with a strategy switch. Changes were from schooled to unschooled strategies and also vice versa. Koedinger & Tabachnek (1994) suggest that there is a trade-off between the benefits of schooled and unschooled strategies. Schooled strategies offer efficient calculation at the cost of error-prone comprehension and translation. In contrast, unschooled strategies tend to support comprehension but at the cost of efficient calculation. Evidence from Koedinger & Tabachnek (1994) suggests that multiple strategy users who change strategy during problem solving are generally more successful than single strategy users. Koedinger & Tabachnek (1994) studied four kinds of solution strategy (two of which did not involve ERs). Also, Schwartz &

Fattaleh (1972) showed that a third of the subjects in their experiment switched ERs from that used to present the problem to another kind when solving the problem. In contrast to the studies of Koedinger & Tabechnek (1994) and Schwartz & Fattaleh (1972), here the concern is with ER switching of a different kind — switching during the problem solving process — that is, in situations where, during problem solution, the subject constructs a representation in one modality but subsequently switches and builds a second, different ER.

Domain: Analytical/deductive reasoning Schwartz (1971) examined the written protocols ('workscratchings') of 30 university students who were given analytical reasoning problems. All the problems employed by Schwartz were determinate in that it is possible to build complete, single model representation of the information.

Subjects in the Schwartz (1971) study were free to construct any type of ER and were encouraged to 'show all work'. The problems varied in difficulty. Some problems presented information positively and some problems contained negatively-phrased information. A second dimension of difficulty was the number of variables — some problems had 3 dimensions and some had 4 dimensions.

The ERs used by subjects in their written solutions were classified into 5 modalities: matrix graphics (eg tables) , informal groupings, graphics, sentence re-write and miscellaneous. In terms of solution success, affirmatively (positively) worded problems were easier than negatively worded ones. The most successful ER modality was that of matrix representations (tables). This was superior to all other ER types as an aid to finding solutions. However, tables were not used successfully on negatively worded problems because information was often erroneously represented.

In a second study, Schwartz & Fattaleh (1972) manipulated the modality in which the deductive reasoning problems were presented. As in the first study, subjects were able to self-construct any ER in their solutions. 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. As in the Schwartz (1971) study, affirmatively phrased problems were found to be easier than negatively phrased problems. Also problems involving disjunction (or) were found to be more difficult than those involving conjunction (and). Dealing with disjunction requires a strategy of 'breaking into cases' and the use of multiple diagrams or the use of a representation that is capable of expressing indeterminacy i.e. provides a means of abstracting over a range of cases (Stenning & Oberlander, 1995).

In terms of the ERs constructed by subjects in the course of 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. It was also noted that subjects often changed negatively phrased information into positive phrasing and disjunctive information into conjunctive.

In a third study, Polich & Schwartz (1974) replicated the matrix superiority findings and also discovered that the representation of implicit information (ie inferred from problem statement) was the greatest source of error. This source of error was minimised by the use of the matrix however, compared to other representations. Errors of omission exceeded errors of commission by three to four times.

Domain: Biology Schwartz (1993) analysed the written protocols of Grade 7 to 10 children in standard and advanced biology classes who attempted to solve biology problems about food webs, disease transmission etc. The hypothesis was that advanced students would tend to use path diagrams (directed graphs) to represent the problem information whereas less advanced students would use functional diagrams. Subjects were randomly assigned to work alone or with a partner. They were told that good representations would help them solve the problems and that different questions required different representations. The results confirmed the hypothesis in that more of the advanced biology students used a path diagram representation in their solutions. Students showed 'extreme ingenuity' in their ER construction and used directed graphs, tree diagrams, pictorial representations and various text-based forms. The advanced biology students had more experience with path diagrams as the result of having been taught about food webs and were capable of transferring the ER strategy to a novel problem. Seventh graders showed evidence of being able to modify a previously encountered ER form for use on a new problem.

Studies in which subjects used prefabricated ERs

This section reviews educational studies in which the modality of information presentation was manipulated.

Domain: Analogical reasoning Beveridge & Parkins (1987) studied large samples of 10-11 year olds and university students. They found that overlayed transparent coloured strips provided an effective visual analogue of the solution to Dunckers 'radiation' problem. Demonstrated to the subjects before the target problem was presented, the coloured-strips ER neatly showed the summation of several weak X-ray beams at the central crossing point in a highly salient manner. That ER , they argue, represented the appropriate features of the problem and manifested 'structural correspondence' in terms of the X-rays' direction, intensity and summation. The ER was imageable, and acted a cue for visual recall, 'freeing' pertinent information from the story context in which it was presented. The ER facilitated the solution of a problem analogue in both the adults and children and was superior to a drawn diagrammatic analogue which required for its interpretation graphical knowledge of arrows and shading conventions. The coloured strips analogue was also superior to both a story analogue of the target problem and a less vivid type of diagram such as the type used in an earlier experiment by Gick & Holyoak (1983). Beveridge & Parkins (1987) showed that the failure of Gick & Holyoak (1983) to find that graphical representations were effective was due to the type of diagram that they used. Beveridge & Parkins (1987) study demonstrates the importance of selecting the *right* ER in order to facilitate effective read-off and inference.

Domain: Mathematics Singley, Anderson, Gevins & Hoffman (1989) evaluated an intelligent learning environment (Algebra Word Problem Tutor) using 4 'beginning algebra' students aged 13 years. The system is described as a model-tracing tutor and is designed to facilitate the writing of algebraic expressions. The system, however, does not solve equations. Model tracing consists of learning-by-doing in conjunction with a cognitive process model that allows for fine grained diagnosis and remediation.

The system provides support by assisting with problem representation and supporting 2 strategies — means-end analysis and a diagram strategy.

At the first stage (problem definition), the user is offered a range of unlabelled, qualitative diagrams — only one of which is correct. If the wrong diagram is chosen, the system offers remediation based on the qualitative relationship violated in the erroneous representation. Singley et al.(1989) acknowledge that 'One drawback...is that (diagram selection) is recognition-based and does not require students to build the diagrams for themselves'. Having selected the correct ER, the subject then maps quantitative information from the problem statement onto the diagram. Next, the user generates constraints and is assisted by the tutor in different ways depending upon the strategy being pursued. Constraints are combined in a final equation.

Four evaluation subjects were pre and post tested on a paper and pencil test of the 8 problem types supported by the system. Singley et al report that the tutor had little effect on the correctness of problem answers but had a significant and positive effect on students ability to write a solvable equation.

The Singley et al. (1989) study is problematic, however, in several respects. It was an uncontrolled and small scale evaluation. Assessing the contribution of diagram selection is also impossible since the diagram selection facility was not implemented at the time of the evaluation (presumably the subjects were provided with a single correct diagram to label). The Word Algebra Tutor is, nevertheless, one of the few systems in which reasoning with (graphical) ERs is central — a point which is discussed further below.

Domain: Computer programming As reviewed in Chapter 3, Mayer (1976) studied the effects of several modes of information presentation on a large sample of university students. The problem used was one in which various sports teams were matched in a tournament. Test questions asked about the outcomes given certain conditions. The problem was presented in eight formats, four were verbal and four were graphical. Mayer (1976) concludes that flowcharts are suited to the representation of complexly structured information whereas verbal formats are optimal for efficiently-structured information.

Domain: Logic-syllogistic reasoning Newstead (1989), in two experiments, presented undergraduates with categorical syllogism premises in two modalities — graphical (Eulers circles) and sentential. The design of the 2nd experiment was within-subject, so that all students received both conditions (order was randomised across subjects). In the graphical condition, students were required to indicate which conclusion validly follows from a given premise by selecting the appropriate diagram(s) from five characteristic Eulers circle diagrams. In the sentential condition subjects indicated against each of a range of possible conclusions which were true or false given a single premise. The results showed that subjects make different patterns of errors depending upon whether the task is performed in the graphical or linguistic modes. In general, Euler's circles were associated with more errors of a Gricean nature — i.e. based on a natural, 'everyday' language interpretation of quantifiers such as 'some' excluding the possibility of 'all' and upon the assumption that the information provider is being maximally cooperative. The Gricean effect was observed at a highly significant level in the Euler's circle task despite explicit instruction to the effect that 'some' meant "at least one, and possibly all". In contrast, on the sentential version of the task, subjects were more likely to 'convert' (i.e. assume that All Bs are As, given All As are Bs) than on the graphical version. Newstead (1989) interprets his results as providing support for both the phenomena of Gricean errors and conversion.

Domain: Biology Hesse, Tiberghien, Baker, Picard & Reinhard (1995) compared 'manipulable' graphics with 'static' graphics. The graphics were structure diagrams of a complex dynamic system (a model from the domain of population biology).

Subjects learned about population models from a computer-based instructional hypertext. They subsequently were required to correct computer-based structure diagrams as a learning test. The relations shown in the diagram had to be judged as correct or not in written form, and the changes necessary for correction had to be described and justified. Hesse et al. write ...

> Under the condition "manipulable graphics" in addition to the written description and justification, necessary changes were to be carried out by manipulating the graphics that were shown. Under the condition "static graphics" working on the tasks consisted only in the written description and justification of necessary changes.

Performance was significantly better for subjects in the manipulable graphics condition for both individual problem solvers and for subjects who worked in collaborative pairs, though the effect of manipulation was larger for individual subjects than for collaborating dyads. Interestingly, in the collaborating pairs condition, the effect sizes of the manipulation factor and collaboration factor were equal but did not add together in terms of the interact effect. They write that 'it can be supposed that the complexity of the tasks was not high enough for manipulable graphics to cause any further improvement in addition to the advantage already given by cooperation.'

Review conclusions

Several conclusions emerge from the review. Instruction in the use of ERs is effective in a wide variety of domains — both in those requiring specialised ER forms such as syllogistic reasoning, mathematics and physics and also in domains where a range of more general ER types can effectively be employed such as analytical/deductive reasoning. Results from Lewis (1989) and Lindvall et al. (1982) suggest that subjects can learn to use ERs with comparatively little time and effort. Furthermore, students can often select and use appropriate ERs in the absence of instruction (*e.g.* Schwartz, 1971)[2]. Willis & Fuson (1988) emphasise the need to teach a range of ERs rather

---

[2] But they must be familiar with the ERs semantics prior to their attempt to use it in reasoning (Cox & Brna, 1993a,b).

than a single type. Lewis (1989) expresses concern that instruction in the use of ERs ('translation training') may lead to superficial key-word spotting approaches to problem reading on the part of learners. To avoid a merely superficial analysis of the problem, an interactive learning environment could require the subject to indicate an adequate level of problem comprehension before ER construction begins. This could be implemented by requiring the subject to enter problem information into a 'problem summary' window. This issue will be explored further in the discussion.

Grossen & Carnine (1990) have shown that instruction with self-constructed ERs is more effective than instruction with prefabricated ERs. Hesse et al. (1995) demonstrated improved performance under conditions where subjects can manipulate graphics compared to conditions where subjects merely observe static representations. Van Essen & Hamaker (1990) emphasise the need for instructors to both instruct and model the use of ERs. Evidence from the literature on syllogistic reasoning points to the importance of a. externalising a representation and b. the need to construct that external representation rather than merely use one that has been prefabricated. Matsuno (1987) analysed the responses of subjects who were asked 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 imagined internal diagrams performed significantly worse only on 'no valid conclusion' (NVC) syllogisms than subjects whose internal representations were either concrete figures or sentential. This result is interesting because unpublished data from a study by Stenning & Cox[3] has shown that the use of *external* diagrams (Euler's circles) is associated with better performance on NVC problems than sentential representation. For example, if given 'Some As are not Bs' the correct response to 6 of the 8 possible conclusions[4] is 'can't tell' (CT). 38% of subjects who drew Euler's circle type workscratchings on their test papers produced a valid (3 diagram) model. In contrast, only 6% of subjects who responded sententially gave correct answers. Furthermore, the study reviewed above by

---

[3] A replication of the study by Newstead (1989).

[4] 'All As are Bs', 'No As are Bs', 'Some As are Bs', 'Some As are not Bs', 'All Bs are As', 'No Bs are As', 'Some Bs are As', 'Some Bs are not As'.

Grossen and Carnine (1990) points up the need for active construction of ERs rather than the passive use of prefabricated ones. Taken together, these three lines of evidence tend to suggest that representations need to both externalised and constructed for them to be maximally effective in problem solving. There are, however, degrees of construction — the subject may draw an entire diagram or merely be allowed to modify or tinker with an extensively pre-constructed representation. In the case of matrix graphics such as spreadsheets, the distinction is probably less crucial than in the case of say using Euler's circles to reason about what conclusion may validly follow from syllogisms.

Another interesting issue that emerges from the review is that of switching. Evidence from Koedinger & Tabachnek (1994) suggests that multiple strategy users who change strategy during problem solving are generally more successful than single strategy users. Koedinger & Tabachnek (1994) studied four kinds of solution strategy (two of which did not involve ERs). Also, Schwartz & Fattaleh (1972) showed that a third of the subjects in their experiment switched ERs from that used to present the problem to another kind when solving the problem. In contrast to the studies of Koedinger & Tabechnek (1994) and Schwartz & Fattaleh (1972), in this study we are concerned with ER switching of a different kind — we are concerned with switching during the problem solving process — that is, in situations where, during problem solution, the subject constructs a representation in one modality but subsequently switches and builds a second, different ER.

The review provides good evidence for the transfer of ER skills. Schwartz (1993) showed transfer of the use of path diagrams in the biology domain. Novick (1990) has shown that transfer of a *representational* strategy (a matrix ER) can occur between two problems in the absence of a shared problem *solution* strategy between the two problems. In the Novick study, 75% of experimental subjects (who had been exposed to matrix ERs in a problem solving context) transferred a matrix ER to the transfer problem compared to 21% of control subjects. The transfer problem was best solved using a matrix representation, but required a different solution strategy from that required on the prior matrix problem. Novick's (1990) results suggest that the transfer effect was more than a mere recency phenomenon because the prior matrix representation problem was the second of three initial problems that experimental subjects received prior to the transfer problem. Thus, the prior matrix problem was embedded in two other problems requiring (each of which required a different ER) and there was a non-matrix problem interposed between the prior matrix problem and the target problem.

Lindvall et al. (1982) also report good transfer of representational skill in primary children from less to more complex word problems. They also report that primary children can validly adapt ERs to new problems.

Many errors are made at the stage of problem comprehension and this emerges as a crucial phase of problem solving (Proudfit, 1981; Reed & Ettinger, 1987; Schwartz, 1971; Polich & Schwartz, 1974). Schwartz (1971) has shown that the syntactic complexity of word problems affects the error rate — negative wording and the use of disjunctions are associated with high error rates. In contrast positively expressed problems and the use of conjunctions are much easier. Polich & Schwartz (1974) found that the representation of information that has to be inferred (ie is implicitly given) is error-prone. Problem complexity (in terms of the number of dimensions and the values along them) is another determining factor (Schwartz, 1971; Polich & Schwartz, 1974).

There seems to be contradictory evidence on whether graphical representation (drawing) increase or decreases with age in the primary grades. Greene (1989) and Fuson & Willis (1989) found a decreasing tendency for primary children to draw pictorial representations as age increased. However, in the domain of arithmetic word problems, Van Essen & Hamaker (1990) report that first and second graders produced fewer self-constructed ERs in response to instruction than fifth graders. The explanation of this seeming contradiction almost certainly lies in the nature of the drawings, specifically in the difference between pictorial representations (literal depictions) and ERs specifically generated as aids to reasoning.

One study provides evidence for different error patterns associated with graphical and linguistic modalities (Newstead, 1989). This is an interesting issue and there are important questions that it raises. For example, do subjects who construct their ERs manifest greater or smaller modality effects in terms of error-patterns?

ERs and ILEs

Several interactive learning environments (ILEs) have employed graphics or graphical interfaces. Some systems, such as BRIDGE (Bonar & Cunningham,1988), GEOMETRY (Boyle & Anderson, 1984) 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 geometry and optical problems (GEOMETRY; REFRACT — Reimann, 1991). Hall (1989) studied ERs as intermediate steps en route to representing algebra word problems in terms of equations. Only four systems, though — HYPERPROOF (Barwise & Etchemendy, 1994); ALGEBRA (WORD PROBLEM TUTOR — Singley et al., 1989), GIL (Merrill et al., 1992) and ANIMATE (Kintsch, 1991) - are centrally concerned with graphics and reasoning.

Unlike primary grade level word arithmetic problems, college level word algebra problems require world and language knowledge that is too rich to simulate and hence to incorporate into intelligent tutoring systems. For that domain, Kintsch (1991) advocates the development of unintelligent tutoring environments that can assist with graphical representation of the problem and which can animate the problem model. Such systems assist with the construction of useful, mediating situation models and intervene in Vygotsky's 'zone of proximal development' for many students. Kintsch describes ANIMATE - a system that helps the student to graphically represent the relevant features of the algebraic problem model, resulting in a conceptual model. ANIMATE uses the conceptual model to produce an animation the purpose of which is to make explicit the link between the problem model and the situation model.

The author is not aware of any system, to date, that 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 central concerns of this thesis. Many of the studies listed in the preceding sections 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 thesis is also concerned with the spontaneous representations that individuals construct 'from scratch' during problem solving. The 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. Moreover, in practically all the studies of reasoning with ERs reviewed above, subjects were instructed to 'show their working' – this may well have influenced subjects to produce ERs for an 'audience' rather than as purely private *aides memoire*.

To date, in the vast majority of empirical studies of ERs, subjects have been specifically instructed to 'draw diagrams' or 'show their working'. Where those instructions are not given, subjects are aware that their responses will be examined in detail at a later time. Often, too, the experimenter is present as an onlooker. One or more of those conditions applied in studies by Proudfit (1981), Reed & Ettinger (1987), Lindvall et al.(1982), Lewis, (1989), Katz & Anzai (1991), Schwartz (1971), Polich & Schwartz (1974), Schwartz & Fattaleh (1972), Schwartz (1993), Hall et al. (1989), Van Essen & Hamaker (1990), Novick (1990), Biron & Bednarz (1989), Greene (1989), Carroll et al. (1980) and Koedinger & Tabachnek (1994). Under those circumstances, subjects produce qualitatively different ERs - ones modified for an audience and therefore not produced solely as private, externalised, cognition. Students are likely to omit material that may be embarrassing, as Hall et al. (1989) concede. However, students sitting analytical reasoning problems assume that the only performance measure of interest to the examiner are their multiple-choice responses. They would therefore not be so likely to modify their representations in any way. On the other hand, students may produce more explicit trails of their reasoning and richer protocols if they are specifically instructed to do so - this is an interesting empirical question that perhaps deserves study in its own right. Responses to the analytical reasoning test, like all the GRE subtests, are machine scored via pencil marked multiple choice response cards. Any residual workscratchings on the test booklet have not been produced in response to instructions to 'show your working' or with the assumption that they will perused by an examiner at a later time.

Subjects in the first study of this thesis (Chapter 5) were not instructed to show their working, neither did they have any cause to suspect that anyone would see their private workscratchings at a later time. This increases the validity of the data as records of cognitive processes compared to previous work on reasoning with ERs.

Thesis arguments expounded in detail

ER selection

Most work on ERs and reasoning, to date, has analysed the residual workscratchings of subjects, or has studied how students use prepared diagrams and illustrations (*i.e.* not ERs that they themselves choose and build). There have been very few studies that have analysed the *processes* of problem comprehension, ER selection, ER construction and the use of ERs for reading-off solutions. Hence, a dynamic, constructivist, approach to studying reasoning with ERs is necessary in order to address a range of important issues at each stage.

First, there are good reasons for believing that ER use is not 'all or none'. There is a spectrum of use that ranges from reasoning exclusively internally (no ER), using a'minimal' ER to full, externalised, model-based reasoning with (usually) graphical ERs or, alternatively, reformulation/re-translation via linguistic ERs.

Representation selection follows comprehension. Good comprehension is associated with the development of a situation model (Kintsch, 1991) or coherent representation unified at a high level by analogy (Greeno, 1977). 'Uncooperative' (Grice, 1975) aspects of information presentation must also be overcome. The full problem information context (Bernado & Okagaki, 1994) must be exploited. A 'puzzle mode' (Levesque, 1988) or 'extensionalist stance' (Stenning & Cox, 1995) must be adopted by the reasoner. Natural language interpretation conventions need to be suspended in favour of more formal readings.

The subject needs to recognise the complexity of the problem by discerning number of dimensions, its level of determinacy, and the presence and proportion of negative and implicitly presented information. One of the objectives of the process analysis based approach of the *switchER* studies (chapters 6 & 7) was to determine how well subjects are able to do those things and to find out what aspects of the problem pose most difficulty.

In the analytical reasoning and study skills literatures, it is possible to find several general and domain-specific sources of ER selection advice. An important question concerns the utility of such recommended ('off the shelf') representations. It is argued that they are only useful to the extent that they happen to be commensurate with the ERs that the subject would have chosen had s/he had free choice.

For a given set of analytical reasoning problems, how wide is the range of ERs chosen and used - are the recommended ones the most effective?

What is the role of prior knowledge and a subject's ER 'repertoire'. According to the test developers, analytical reasoning problems do not require domain knowledge. However, knowledge of some representational formalisms and problem solving strategies *is* required. What happens when a subject lacks knowledge of an appropriate representational formalism such as a diagrammatic method for representing indeterminacy, for example?

The questions associated with analytical reasoning problems can be very heterogeneous in their task requirements and in the extent to which solutions to the questions that they pose are 'ER dependent'.

To what extent, therefore, is a *single* ER useful across a set of (typically) 4 or 5 questions? For example, in order to overcome the specificity of a particular diagrammatic ER, do skilled reasoners use more than one ER - *i.e.* reason heterogeneously?

Finally, how extensive are individual differences in ER selection?

Constructivism

> a constructivist theory of knowledge ... is generally shared by all researchers in the domain of representation. (Janvier, 1987)

It is argued, from a constructivist perspective, that individual differences in prior knowledge must be recognised. For that reason, and because of a general paucity of research in the area, it is impossible to issue general guidelines for choosing a 'correct' representation for a given task.

There are strong reasons to believe that subjects reason better with representations that they construct than with provided, 'prefabricated', ERs. However, there is a need for explanations that go beyond merely attributing the superiority of 'a required diagram-drawing response' to 'deeper processing' (Grossen & Carnine, 1990, pp. 179-180). The adoption of a process-based approach to the study of reasoning with ERs can be expected to shed light on the cognitive effects of externalisation.

In the syllogistic reasoning domain, the process of constructing an external representation (Euler's circles) has been shown to result in richer learning outcomes than the use of pre-fabricated diagrams (Grossen & Carnine, 1990). However, the range of representations that are useful for syllogistic reasoning is relatively narrow and, in that study, Euler's circles were the only representations available to the subjects. How do subjects approach such problems when they lack knowledge of Euler's Circles?

The effectiveness of a representation depends on how well suited it is to the task at hand. For example some tasks require qualitative comparisons, others require precise numerical read-off or fast searches. Representations may be internal or external, graphical or linguistic and the form of representation chosen can greatly affect performance. How successful are subjects at discerning the nature of the task(s) posed by a problem and how well do they match their representations to those demands?

Switching

Heterogeneous reasoning either with linguistic/diagrammatic mixtures of ERs or with intra-modal ERs (*e.g.* multiple diagrams) is effective because the expressive properties of several representations can be exploited. Heterogeneous reasoning is encouraged by Hyperproof, a Macintosh program for teaching first- order logic (Barwise & Etchemendy, 1994). There are not many evaluative studies of heterogeneous reasoning. However, the use of Hyperproof in 'real' logic teaching contexts has been shown to be at least as effective as traditional approaches for most students and particularly effective for graphically inclined students (Stenning, Cox & Oberlander, 1995). Also, Molitor, Ballstaedt & Mandl (1989) review a well-controlled study by Stone & Glock

(1981) in which it was shown that the combined presentations of text and pictures produced superior performance in a model assembly task than the presentation of one modality alone.

Selecting and constructing multiple ERs, however, is a challenging task that requires sensible allocation of information to modalities and skilled division of cognitive effort across the stages of problem solving. In some systems (*e.g.* Hyperproof) ERs are prefabricated or partially formed and no ER selection decision is required of the user (the graphics and sentences are a *fait accompli*). Furthermore, the subject does not construct the representation from scratch but modifies and extends the existing ones. Under these circumstances the two modalities are used concurrently, side-by-side. In contrast, where a subject reasons heterogeneously with self-selected and self-constructed ERs, the cognitive load is high and only one representation at a time can be constructed. It is argued, therefore, that switching between representations aids the distribution of cognitive effort. Switching is often a process of translation — initially the problem may be re-cast or translated from the presented (sentential) form to another, perhaps abbreviated sentential form or notation. This might be deemed a direct translation. Subsequently the subject might build a diagrammatic model. This represents two kinds of shift — a shift of modality (sentence to diagram) and a shift of strategy (direct translation to model-based reasoning).

Switching might be triggered when the subject reaches an impasse in reasoning or where an initial representation has failed to assist with the problem solution. Alternatively, it may result from a recognised problem with the existing representation, such as an error in construction or as a result of dissatisfaction with a (sub-optimal) representation.

Switching between some modalities and/or representational formalisms may be easier than switching between others. Lesh, Behr & Post (1987) examined 4th to 8th graders' responses to a paper and pencil test of rational number and proportional reasoning (mathematical) problems. They characterised each type of item in their tests according to the cognitive operations required by each. These consisted of translations from:

- symbols to written language
- written language to symbols

- pictures to pictures
- written language to pictures
- pictures to written language
- symbols to pictures
- pictures to symbols

They concluded that the order of increasing difficulty was that in which the translations are ordered in the list above, with pictures to symbols being most difficult. They write:

> In general, if other factors are held constant: (a) Translations to pictures is easier than translations from pictures; (b) translations involving written language (*e.g.* three-fourths) are easier than translations involving written symbols (*e.g.* $\frac{3}{4}$); and (c) the easiest translations are those that only require a student to 'read' a fraction or ratio in two written forms. (p.48).

Note that there are important differences between the Lesh et al. study and the work to be reported here. The Lesh et al. test items required translations from one type of *presented* representation to another kind of *presented* representation. No active ER construction was required by the subject. The age of the subjects and the domain (mathematics) were also different. Despite those caveats, the Lesh et al. result is interesting because it shows that, when all other factors are held constant, between-modality (*e.g.* symbols to pictures) translations are more difficult than within modality (*e.g.* written language to symbols) translations. Lesh, Post & Behr (1987) state:

> Part of what we mean when we say that student 'understands' an idea...is that: (1) he or she can recognize the idea embedded in a variety of qualitatively different representational systems, (2) he or she can flexibly manipulate the idea within given representational systems and (3) he or she can accurately translate the idea from one system to another. (p.36).

Multiple representations employed in problem *presentations* have been shown to facilitate finding solutions under some circumstances. Stone & Glock (1981) showed that, in an assembly task, directions presented both diagrammatically and propositionally resulted in fewer assembly errors than directions presented only propositionally. However, there was no significant difference between the diagram-only and diagram-plus-text conditions. The diagram-only condition produced fewer errors of parts-orientation than text only, however. When 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 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.

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? Another important issue concerns the redundancy of information presented in the two modalities - for example in the Stone & Glock (1981) study, the information in the two modalities was redundant[5] In Hyperproof, information presented diagrammatically and sententially is far less redundant.

The next chapter of the thesis presents the first empirical study — in it, some of the issues raised by the literature review are addressed. The aims of the first study are summarised below.

Summary of aims of first study

The aims of the first empirical study were to examine:

1. the variety of ERs selected and constructed by subjects across a range of analytical reasoning problems

---

[5] The same information was represented in each modality. The level of redundancy (whether same or different information is represented in the two modalities) is an important factor in heterogeneous reasoning research.

2. the effect of problem determinacy level upon representation selection and performance
3. the relationship between ER type/modality and performance
4. slips and errors in ER construction and their effects upon performance
5. the utility of particular ERs across a range of task demands
6. subjects' use of multiple representations and the effects of multiple representation use upon performance

Chapter 5

An investigation of ER use in reasoning.

Introduction

This chapter presents the results of a study in which a large corpus of workscratchings were collected and analysed. The workscratchings were spontaneously produced by subjects in the course of solving a range of paper and pencil-based analytical reasoning problems that were administered under test conditions. There were several aims. One (broad) aim was to investigate the variety and types of ERs spontaneously used by subjects across a range of different problems. Another aim was to evaluate the effectiveness of different ERs and to test predictions, derived from specificity theory (Stenning & Oberlander, 1995), regarding the relationship between ER modality and the degree of determinacy possessed by the problem.

Since each problem has 4 or 5 associated questions, and since the questions differ markedly in their task demands, it was also of interest to study, for each problem, the effectiveness of a particular ER formalism across the range of questions. It was hypothesised that subjects might respond to changing task demands by switching representations or producing additional representations.

Also of interest was the relationship between the accuracy and form of the ERs and performance (question scores).

A final, and more minor, aim was to investigate the effectiveness of the ERs recommended by the GRE crammers from which the problems were taken.

Issues of validity

Analytical reasoning test items are posed as pencil-marked multiple choice items. Hence, any residual workscratchings on the test booklet have not been produced in response to instructions to 'show your working' or with the assumption that they will perused by an examiner at a later time. For that reason, the workscratchings may be considered to be relatively valid indexes of externalised cognition. Almost every preceding study that has analysed external representations has either specifically instructed subjects to 'draw diagrams' or 'show their working'. Even where those instructions are not given, subjects in many studies are aware that their responses will be examined in detail a later time. Often, too, the experimenter is present as an onlooker. In those situations, subjects probably produce qualitatively different ERs - ones intended for an audience and not solely as private, externalised cognitive activity. Subjects in the study reported here were advised that 'it may be useful to draw a rough diagram' but were not compelled to use an ER. Hence the ERs produced by subjects can be considered to have been produced spontaneously. In this respect, the methodology of the investigation reported here yielded ERs of greater validity (as examples of externalised cognition) than most previous work on ERs and reasoning since, in much of the previous work, subjects were instructed to use particular ER formalisms.

Problem selection

Three problems were selected for detailed analysis (Appendix A). They were taken from a GRE test 'crammer' (Brownstein, Weiner & Green, 1990). 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[1].

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

---

[1] 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).

finding the solution. One problem (problem two) was indeterminate and requires the modelling of quantifier information. 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 and 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. Another aim was to examine the prediction, derived from specificity theory (Stenning & Oberlander, 1995), 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 it was hypothesised that prior experience with these formalisms would be an important prerequisite for their effective use. It was predicted that the indeterminate problem would pose considerable difficulties for subjects who were unfamiliar with appropriate representational formalisms.

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.

The three problems selected for analysis in this investigation were included, along with other analytical reasoning and verbal reasoning items, in several versions of tests developed for use in a longitudinal evaluation of a logic teaching course (Cox, Stenning & Oberlander, 1994, 1995; Stenning, Cox & Oberlander, 1995; Oberlander, Cox & Stenning, 1995a,b; Oberlander, Cox, Monaghan, Stenning & Tobin, 1996). The number of subjects responding to each of the three questions therefore varies for this reason. Forty subjects completed all 3 problems, 29 completed the ‘office’ and ‘poets’ problem, 8 subjects completed the ‘poets’ and ‘dogs’ problems. Eight subjects responded to the ‘office’ problem only, 3 to the ‘dogs’ problem only and 14 to the ‘poets’ problem only.

Procedure

GRE-like tests of analytical and verbal reasoning were administered, under test conditions, to four classes of students as part of an investigation into logic teaching. Before answering the analytical reasoning problems, subjects were instructed:

Table 5.1: 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 5.2: 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)

> In this part, each question or group of questions is based on a passage or set of conditions. In answering some of the questions, it may be useful to draw a rough diagram. For each question, circle the best answer choice given.

Results and Discussion — Workscratching Corpus

Tables 5.1, 5.2 & 5.3 summarise the ER behaviour on the three problems. They show the proportion of subjects using a particular ER who responded correctly to each question.

Table 5.3: 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

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 are provided in Appendix A. 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 A.2, A.3, A.4, Appendix A. These graphical ERs were more varied than the ordered-text ERs such as Figure A.5, Appendix A. 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 A.4, Appendix A). Three subjects drew 'minimal' plans consisting of horizontally arrayed vertical lines that represented office partitions (Figure A.3, Appendix A). Three subjects drew 'cubicles' — *i.e.* 3-sided offices with an open wall — an example is shown in Figure A.1, Appendix A.

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 5.1) 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 5.1). 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...

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 of graphics to express abstraction (*i.e.* their specificity — Stenning & Oberlander, 1995). 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 A.3, Appendix A. 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, a tutor or 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 of ERs. On the other hand, a higher proportion of subjects chose not to use an ER on problem 2 than the other two problems (Table 5.2) — 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'. The Brownstein et al. solution is shown in Figure A.21, Appendix A. The recommended ER is actually semantically incoherent, this will be discussed further in Chapter 6.

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 5.2).

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 A.7, Appendix A. Specificity theory (Stenning & Oberlander, 1995) predicts the utility of logic as an ER modality for expressing indeterminate information. The semantics of first-order logic can be used to readily represent quantifier information and permit useful inferences.

All of the set-diagram using subjects in the current study built single models of the premises — characteristic[2] diagrams (*e.g.* Newstead, 1989; Stenning & Oberlander, 1995) were more popular single models (*e.g.* use of A within B diagram to represent the premise 'All A's are B's' than the identity diagram — see Figures A.9, A.10, A.11 in Appendix A and Figure A.17 in Appendix A. Three of the 7 subjects who used set-diagrams responded incorrectly to problem 2 possibly because they constructed only one of many possible diagrammatic models of the information (see Figures A.9, A.11 in Appendix A and Figure A.17, Appendix A).

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 the most effective representations for 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

---

[2] The characteristic diagram is the diagram that represents the maximum number of types of individual consistent with the premise. 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'.

of the representation. Consider set diagram notation — one subject (Figure A.10, Appendix A) consistently drew a small circle 'E' within a larger circle 'B' to represent the first premise of problem 2 ('All those who enjoy...Browning also enjoy...Eliot'). The misconception resulted in that subject erroneously representing 3 of the 7 premises. 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. Moreover, reasoning with external representations is not necessarily totally external. Subjects who suspect that their representation is incorrect may shift their cognitive resources and rely more on internal representations as a compensatory strategy.

As in Problem 1, the second problem's questions vary widely in terms of task requirement. Hence a single ER type is not likely 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 (*e.g.* Figure A.6, Appendix A) 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. The data in table 5.2 suggest that for subjects who lack knowledge of logic (and for whom logic is therefore not an option), directed graphs are optimally useful for answering questions 1 & 2 and set diagrams are optimal for answering questions 3 & 4.

Eighteen of the 20 subjects who used text to re-represent the problem information in problem 2 (*e.g.* Figure A.8, Appendix A) 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 premises in the problem. Given 'All those who like Browning also like Eliot' (premise 1) and 'Those who enjoy Eliot despise Coleridge' (premise 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.

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. It is also unclear that prescriptive ER advice based on group data is necessarily beneficial for individual subjects, due to wide variations in prior knowledge and cognitive style.

Problem 3

The third problem was a determinate problem like problem 1, but was more difficult. 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 A.13, A.14,and Figure A.15 in Appendix A. In contrast to the study by Schwartz (1971), in this study tables *were* the most frequently chosen ER (Table 5.3).

There were, however, important differences between the *types* of tables used. Contingency tables were associated with poor performance because they are difficult to search. Contingency tables are matrix representations in which separate two dimensional tables are constructed for each possible pairing of variables in the problem. An example can be seen in Figure A.15, Appendix A. Two of sixteen subjects in the first *switchER* study (to be described in Chapter 6) also produced contingency tables. Contingency tables are difficult to read-off conclusions from - information has to be searched for in more than a single location and then resolved or unified mentally.

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. An example is provided by Figure A.12 in Appendix A. 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 A.13, Appendix A. 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. The reason for this is that the salient dimension from the point of view of responding to the questions is prize order. An interesting paper by Gilmore (1991), using the river pollution task[3] has shown that salience (factory ordered versus pollutant ordered) and type of representation (list versus table) interact. Gilmore (1991) partitions Green's (1989) cognitive dimension of 'visibility' into accessibility, salience and congruence. Accessibility and salience are static properties of the representation but congruence is dynamic and varies depending upon the extent to which the salient structure in the representation is relevant to the use being made of the representation at a particular moment.

Another example of discourse ordering is illustrated by the unusual semi-tabular 'letters & lines' representation shown in Figure A.16, Appendix A. The problem information was first discourse-ordered in adjacent columns of a table and subsequently re-ordered by means of directed lines connecting elements of the table.

Table 5.3 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 A shows examples of contingency tables (Figure A.15) and unified tables (Figures A.12 & A.14)).

On the whole, there were very few incorrectly constructed ERs on problem 3. Six subjects produced incomplete ERs — the 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

---

[3] Berry & Broadbent (1989) — the objective is to detect which of 8 factories is responsible for polluting a river via strategically testing river water samples.

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 responding[4] to items than incorrect responding[5]. 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 by Figures A.19 and A.20 in Appendix A — the first shows the use of a restricted logic notation together with a plan representation on problem 1 and the second shows a subject's representation for problem 3 in which both textual notes and a tabular representation are used.

These findings are 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) write:

> ... the act of representation tends to be *plural, unstable*, and *evolving* ... we have found that students seldom work through solutions in a single

---

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

[5] Correct responses to less than three-quarters of a problem's questions.

> 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 (p.37).

Furthermore, diSessa (1979), 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 (p.250).

More recently, the use of multiple ERs has been described as 'heterogeneous reasoning'. Barwise & Etchemendy (1992) 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'[6].

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.

---

[6] Barwise & Etchemendy's (1994) 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'.

Summary and Conclusions — Workscratching Data

In brief, the findings of study 1 can be summarised as:

- Subjects use a wide range of ERs
- A single ER is usually not equally effective for every question in a problem's set
- Subjects can use incorrectly constructed ERs successfully (to the extent that they perform better using the incorrect ER than they do if they abandon it)
- There is evidence that using a partially incorrect ER results in better performance than abandoning it completely
- Subjects sometimes use ER formalisms that they do not fully understand
- Consistency of answers with ERs is associated with correct responding
- Contingency tables yield uniformly poor performance
- Subjects sometimes use more than one representation
- Multiple ERs are effective
- There is support for specificity theory – when the expressivity of an ER is matched to the determinacy level of the problem, better performance results

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

As shown by the examples in Appendix A, 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, 1995) 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. For the two determinate problems (1 and 3), weakly expressive representations such as plans or tables were optimal.

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 A.10, Appendix A in which a single set diagram model for each premise ER was constructed. The first premise ("All those who enjoy Browning also enjoy Eliot") is represented (erroneously) by a small circle "EL" inside large circle "BR").

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.) be capable of expressing any indeterminacy in the problem, 2.) be useful computationally (Larkin & Simon, 1987) and 3.) be appropriate for meeting the task demands of the problem question (Day, 1988).

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 5.2 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.

In solving analytical reasoning problems, the processes of problem comprehension, representation selection and multiple representation *construction from scratch by the student* place very high cognitive loads upon the learner. However, information about subjects' interactions with their ERs and the time-course of those interactions is not preserved in paper and pencil workscratching records. A range of empirical questions could be addressed through the use of a methodology that permits the dynamic nature of reasoning with ERs to be studied. For example, to what extent do subjects erase and redraw elements of their representations? In what order do subjects construct the elements of their representation? How long do subjects spend at each of the stages of reasoning? Does the time spent at each stage provide a useful index of how cognitive effort is partitioned? To what extent do subjects construct multiple representions? How are multiple representations used? In what respects does the behaviour of subjects who choose not use ERs differ from those subjects who do use ERs? ,*etc.*

To address those questions, a second study was conducted, using a computer-based system (*switchERI*), and in which the dynamic, second-by-second events during each stage of reasoning with ERs were recorded. The *switchERI* study will now be presented in Chapter 6.

Chapter 6

SwitchERI - an ILE for analytical reasoning.

Introduction

An obvious disadvantage of residual workscratchings and written protocols is the fact that timing information is lost and the transitions between problem solving episodes are impossible to detect.

Study 2 was conducted, therefore, in order to examine the time-course of reasoning with ERs, to identify the major stages, to 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.

The use of time-stamped, user-system interaction logs provides a method of collecting information about timing, transitions between problem solving stages, the order in which elements of an ER are constructed, whether or not the subject uses a representation subsequent to construction and how ERs are used in the resolutions of impasses in reasoning, whether errors in ER construction are made, detected and rectified and so on. The approach also allows comparisons to be made between various kinds of ER in terms of construction time, efficiency in read-off solutions *etc.*.

This study also investigated the relationship between subjects prior knowledge of ERs and the selection, construction and use of ERs (again) in the domain of analytical reasoning. The main motivation was a series of questions regarding the potential of representation switching as an effective reasoning strategy. The term 'switching' means being able to swap between informationally equivalent (or non-equivalent) representations that differ in their modality *e.g.* logical/sentential versus circle diagrams. Koedinger & Tabachnek (1994) suggests that multiple strategy users who change strategy during problem solving are generally more successful than single strategy users. Koedinger & Tabachnek (1994) studied four kinds of solution strategy (two of which did not involve ERs). Also, Schwartz & Fattaleh (1972) showed that a third of the subjects in their experiment switched ERs from that used to present the problem to another kind when solving the problem. In contrast to the studies of Koedinger & Tabechnek (1994) and Schwartz & Fattaleh (1972), in this study we are concerned with ER switching of a different kind — we are concerned with switching during the problem solving process — that is, in situations where, during problem solution, the subject constructs a representation in one modality but subsequently switches and builds a second, different ER. To address the switching issue, a prototype of an exploratory environment (*switchERI*) was designed and implemented.

Prior knowledge

In order to assess subjects' prior knowledge of ER types and its effects upon ER selection, an ER classification task similar to that of Lohse, Walker, Biolsi & Rueter (1991) and Lohse, Biolsi, Walker & Rueter (1994) was devised. However, the classification task used in the current study employed a broad range of ERs, including natural and logical language fragments, ordered texts, tabular forms, set diagrams and other tasks that were not included in the Lohse et al. corpora.

Another reason for devising the ER classification task was to permit cluster analyses on the data from all subjects. This technique, used by Lohse et al. (1991; 1994), yields a taxonomy of ERs based on the combined subjective categories assigned to the corpus of ER stimuli. Unlike the Lohse et al., (1991; 1994) studies, however, the taxonomy used included ERs associated with reasoning tasks rather than solely with data visualisation. The ER classification task was administered as a pre-test.

Subsequently, the subjects used the *switchERI* system to solve analytical reasoning problems. Subjects' interactions with *switchERI* were dynamically recorded. This permitted subjects' perceptions and conceptualisations of ER forms to be related to both their subsequent behaviour during the *processes* of ER selection, construction and use as well as to the usual performance measures of accuracy and speed of solution on the analytical reasoning tasks. It was hypothesised that when solving the analytical reasoning problems, subjects would effectively utilise only the types of ER that they had a.) perceptually discriminated and b.) precisely labelled on the taxonomy pre-task. A further hypothesis was that subjects whose taxonomies demonstrate richness and 'deep' structure [1] would demonstrate greater ER selection accuracy and score higher in terms of correct responses to questions than 'surface' sorters.

Finally, another aim of the *switchERI* study was to acquire information in order to inform the design of a second system *switchERII* capable of providing intelligent user support.

Method and Procedure

Subjects

Two groups of subjects solved the three problems using the *switchERI* 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. All the subjects used in the study were familiar with the Apple Macintosh graphical user interface.

Taxonomy task

Subjects' prior knowledge about a wide range of ERs was assessed by means of a ER-taxonomy (card sort) task administered before they attempted the reasoning problems.

---

[1] That is, not based just on perceived similarity of surface characteristics of ERs but displaying an appreciation of dimensions such as causality, set membership, and hierarchy, for example.

The taxonomy pre-test was developed using 87 items taken from wide variety of sources such as texts on graphics, physics textbooks, fragments of computer programs, formulae, instructions, charts, plans, schematic sketches, maps, tables, music, childrens' drawings, circle diagrams, illustrations, cartoons, X-Y graphs, logic, directed graphs, maps, tables, tree diagrams, bar charts and circuit diagrams. Each representation was photocopied and mounted on 8 × 5" white index cards which were numbered on the reverse. Twenty one of the items consisted of hand-produced examples of diagrams, graphs and tables taken from paper and pencil tests of analytical reasoning in the course of an earlier, unpublished study that utilised different subjects from those in this investigation. The other items were professionally produced graphical or type-set items. The 87 stimulus items are shown in Figures A.22 & A.23 in Appendix A.

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 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.

switchERI

The *switchERI* system consists of a Macintosh Hypercard program that provided 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).

Subjects were given a practice question (Figure 6.1) which the experimenter used to illustrate the system's features. They then solved the practice problem, taking as much time as they wished. When they were ready, the subjects then attempted the three experimental problems. No time limit was imposed.

Figure 6.2 shows a *switchERI* screen display as subject 1 used the *switchERI* diagram tool to construct a set diagram ER in the course of solving problem 2. Figures 6.3, 6.4 and 6.5 show the other *switchERI* ER construction environments in use.

Subjects' interactions with the system were dynamically recorded using Farallon Inc.'s 'ScreenRecorder' utility[2] 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 or audiotaped during problem solving and the replay of the screen recordings.

Using this methodological approach, it was possible 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.

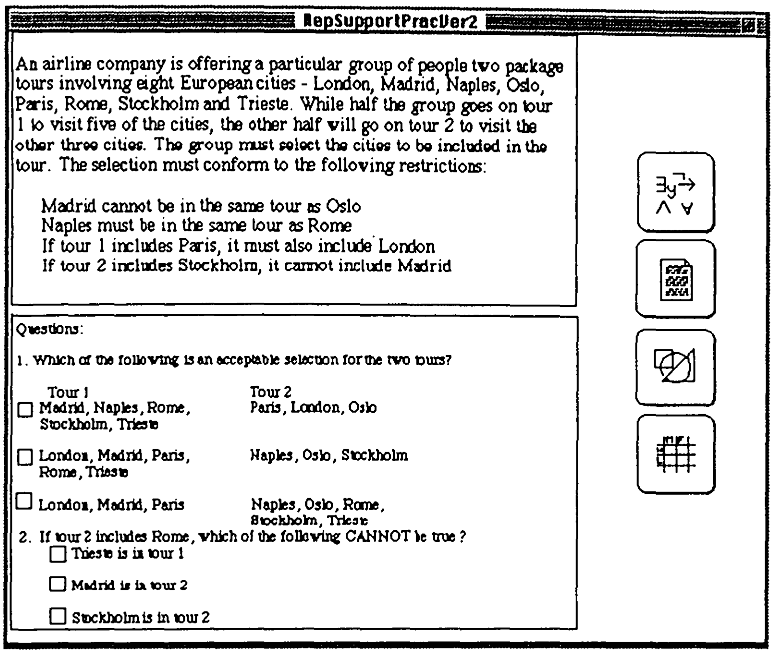
Results

Taxonomy task

The taxonomy pre-test data was subjected to two distinctly separate analyses. First, a cluster analysis was performed in order to compare a taxonomy of representations derived from this study with those of previous researchers. This can be considered to be a task validation exercise and is reported in the next section. Secondly, the taxonomy pre-test data were used as means of determining whether each subject understood (*i.e.* precisely identified) a particular representational formalism in order to related prior knowledge to subsequent analytical reasoning performance. The second analysis is reported later in the chapter in a section on the relationship between prior knowledge

---

[2] Part of Farallon Inc.'s 'MediaTracks' package.



An airline company is offering a particular group of people two package tours involving eight European cities - London, Madrid, Naples, Oslo, Paris, Rome, Stockholm and Trieste. While half the group goes on tour 1 to visit five of the cities, the other half will go on tour 2 to visit the other three cities. The group must select the cities to be included in the tour. The selection must conform to the following restrictions:

Madrid cannot be in the same tour as Oslo
Naples must be in the same tour as Rome
If tour 1 includes Paris, it must also include London
If tour 2 includes Stockholm, it cannot include Madrid

Questions:

1. Which of the following is an acceptable selection for the two tours?

	Tour 1	Tour 2
☐	Madrid, Naples, Rome, Stockholm, Trieste	Paris, London, Oslo
☐	London, Madrid, Paris, Rome, Trieste	Naples, Oslo, Stockholm
☐	London, Madrid, Paris	Naples, Oslo, Rome, Stockholm, Trieste

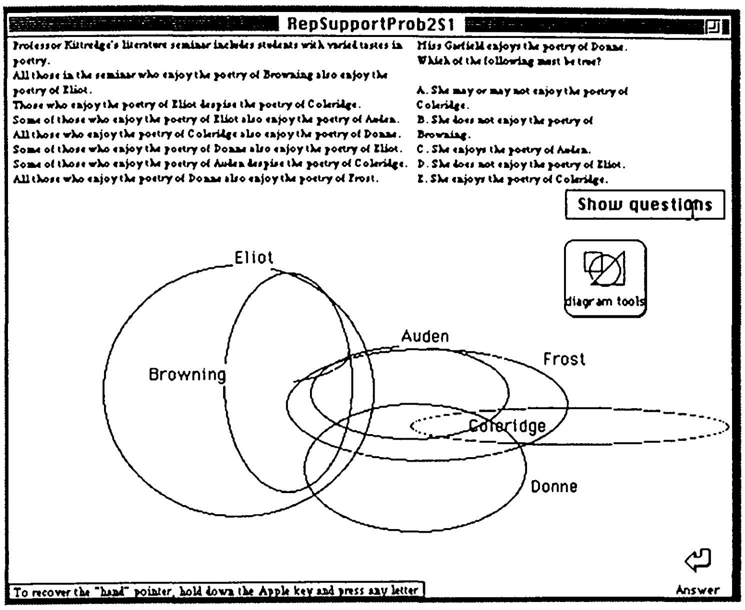
2. If tour 2 includes Rome, which of the following CANNOT be true?
   - ☐ Trieste is in tour 1
   - ☐ Madrid is in tour 2
   - ☐ Stockholm is in tour 2

Figure 6.1: Practice problem used in the *switchERI* study

and reasoning performance.

Cluster analysis The number of categories that subjects used in their sorts varied from five to twenty. The mean number was 11.9 (median 11) with a standard deviation of 4.6 categories.

Each subject's card sort was represented in an 87 by 87 matrix (there were 87 stimulus items in the corpus). Each cell of the matrix coded the relationship between one distinct pair of items. Considering any two items, the subject either placed them in the same pile or not. For example, if the subject sorted card 3 and card 46 into the same pile, then a one was coded at the cell corresponding to the intersection of row 3 with column 46. Thus if items co-occurred within a pile then a one was coded in the cell, else a zero. A matrix of ones and zeros was constructed for each subject. There were 3741 possible pairwise similarity comparisons among the 87 items. The sixteen individual subject matrices were added to form a group matrix in which the value of any one cell ranged from 0 to 16 (there were 16 subjects). The value of any one cell represented the number of subjects who put those two particular items in the same group.


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.

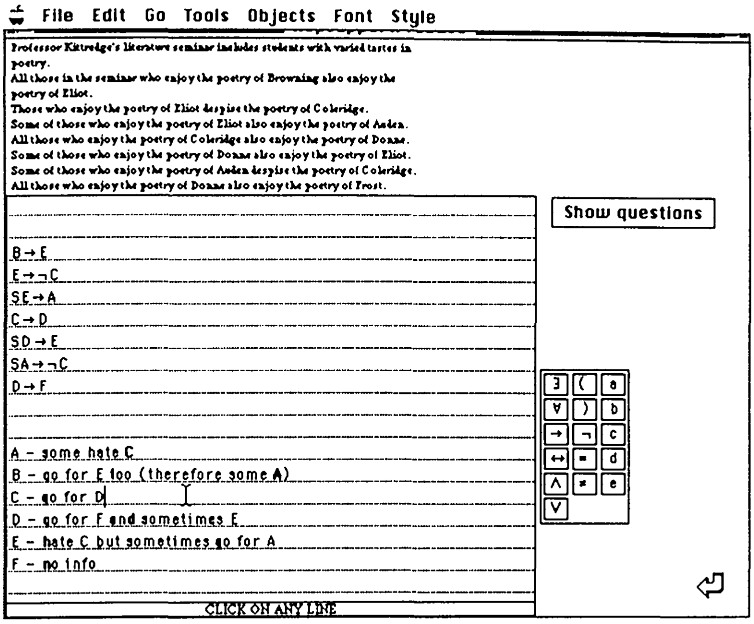
To recover the "hand" pointer, hold down the Apple key and press any letter

Figure 6.2: Subject 1 uses the *switchERI* diagram environment to construct a set diagram on problem 2.

The group matrix formed the input to SPSS [3] PROXIMITIES procedure. Since the data was of an interval scale of measurement, a Euclidean distance measure was employed (Lorr, 1983). The PROXIMITIES procedure output a similarity matrix. The similarity matrix formed the input to the SPSS CLUSTER procedure which was used to compute a multilevel, agglomerative, hierarchical cluster analysis (*e.g.* Aldenderfer & Blashfield, 1984; Johnson, 1967). Hierarchical cluster analysis organises a set of entities into homogeneous units (Lohse *et al.*, 1991; 1994). The item clusters are arranged hierarchically with individual items at the leaves and single cluster at the root. The branching factor determines the model's "bushiness" and is dependent upon the cluster method used. Single linkage methods cluster according to the rule that

---

[3] Statistical Packages for the Social Sciences



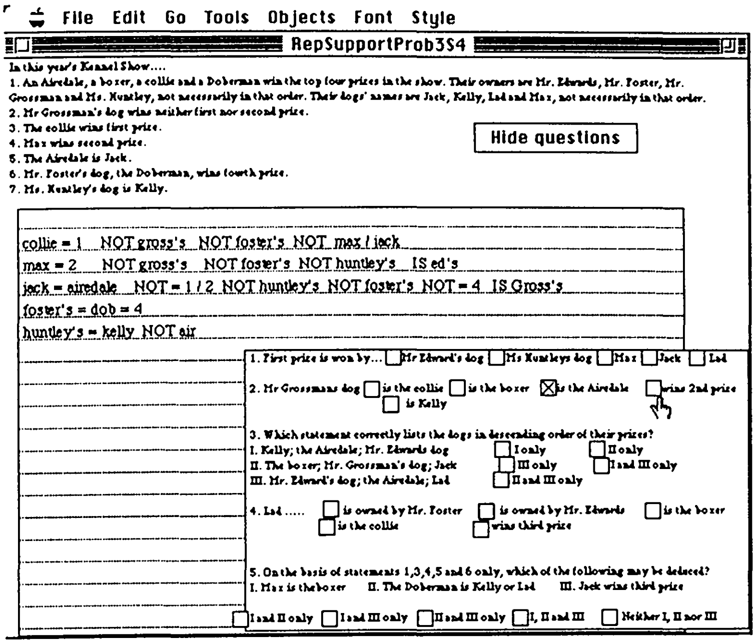
Professor Kittredge's literature seminar includes students with varied tastes in poetry.
All those in the seminar who enjoy the poetry of Browning also enjoy the poetry of Eliot.
Those who enjoy the poetry of Eliot despise the poetry of Coleridge.
Some of those who enjoy the poetry of Eliot also enjoy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

B → E
E → ¬C
SE → A
C → D
SD → E
SA → ¬C
D → F

A - some hate C
B - go for E too (therefore some A)
C - go for D
D - go for F and sometimes E
E - hate C but sometimes go for A
F - no info

Figure 6.3: Subject 6 uses the *switchERI* logic symbol palette to represent the information in problem 2.

cases will be joined to existing clusters if at least one of the members of the existing cluster is of the same level of similarity as the case under consideration for inclusion. In the case of complete linkage, a candidate under consideration for inclusion must be within a certain level of similarity to *all* members of that cluster (Aldenderfer & Blashfield, 1984). Lohse *et al.* (1991;1994) used complete linkage in their cluster analysis. Complete linkage techniques can produce high degrees of branchiness and small, tight clusters. Average linkage techniques were developed for use in hierarchical agglomerative cluster analysis in order avoid the extreme outcomes of either single or complete linkage methods (Lorr, 1983). Average linkage was therefore used in the analysis.



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.

collie = 1 NOT gross's NOT foster's NOT max / jack
max = 2 NOT gross's NOT foster's NOT huntley's IS ed's
jack = airedale NOT = 1 / 2 NOT huntley's NOT foster's NOT = 4 IS Gross's
foster's = dob = 4
huntley's = kelly NOT air

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 ☐ II only ☐ III only ☐ I and III only ☐ II and III only

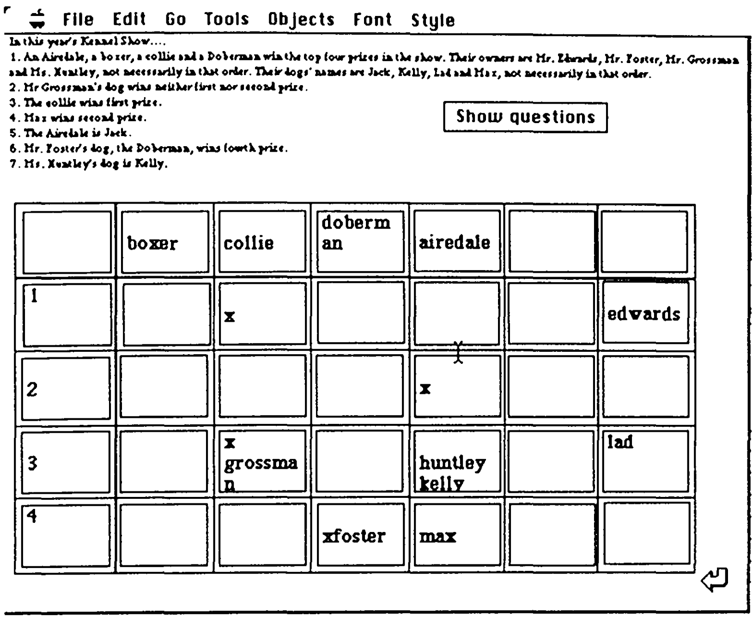
4. Lad ..... ☐ is owned by Mr. Foster ☐ is owned by Mr. Edwards ☐ is the boxer ☐ is the collie ☐ wins third prize

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

Figure 6.4: Subject 4 uses the *switchERI* text environment in her solution to problem 3.

Eighteen principle clusters emerged from the average linkage solution. They are represented in dendrogram or two-dimensional tree-diagram in Figure 6.6. The highest level distinction was between maps and non-maps. The large non-map category consisted of 15 sub-divisions:

- set diagrams, circle diagrams
- list; ordered text; tables
- x/y graphs
- tree diagram



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.

	boxer	collie	doberman	airedale		
1		x				edwards
2				x		
3		x grossman		huntley kelly		lad
4			xfoster	max		

Figure 6.5: Subject uses the *switchERI* table environment in his solution to problem 3.

- node and arc diagrams (concept maps)
- plans
- logic/mathematical formulae
- text (includes computer program listings)
- music
- depictive illustrations (pictures)
- abstract geometric forms

- puzzles
- scientific illustrations
- chessboards, objects on matrices
- instructional sequences, cartoon strips

In the case of set diagrams, tables and logical/mathematical formulae, subjects distinguished between hand-produced and printed items ('h' or 'p' in Figure 6.6). The numbers at the 'leaves' of the Figure 6.6 dendrogram correspond to the items in Appendix A.

The results were compared to those of previous researchers who have attempted to classify representations (reviewed in Chapter 1). Maps, network diagrams, music, tree diagrams, tables, lists and pictorial sequences were categories that Twyman (1979) also identified in his schema of graphical language. Lohse, Biolsi, Walker & Reuter (1994) developed a classification of visual representations based on cluster analyses of subjects' subjective ratings of graphics along 10 dimensions. They identified eleven basic categories of graphics: graphs, tables, graphical tables, time charts, networks, structure diagrams, process diagrams, maps, cartograms, icons and pictures. The results of the present study were comparable to those of Lohse et al., allowing for differences in the corpora of stimuli – for example, Lohse et al. included icons in their corpus whereas they were not included in the corpus used in the current study. The task can therefore be considered to be a reasonable one for assessing the extent of subjects prior knowledge of a broad range of ERs.

As mentioned earlier, the taxonomy data was also used in a second way. The labels that subjects applied to their sorted categories were also analysed - they were used as a means of assessing prior knowledge. For example, if a subject discriminated set diagrams from other circular ERs (*e.g.* pie charts) and labelled them precisely (*i.e.* using terms such as 'set diagram', 'Euler's circles', 'Venn diagrams', *etc.*), then this provided a reasonable indication of good comprehension of the formalism. The relationship between prior knowledge and reasoning performance is discussed later in this chapter.

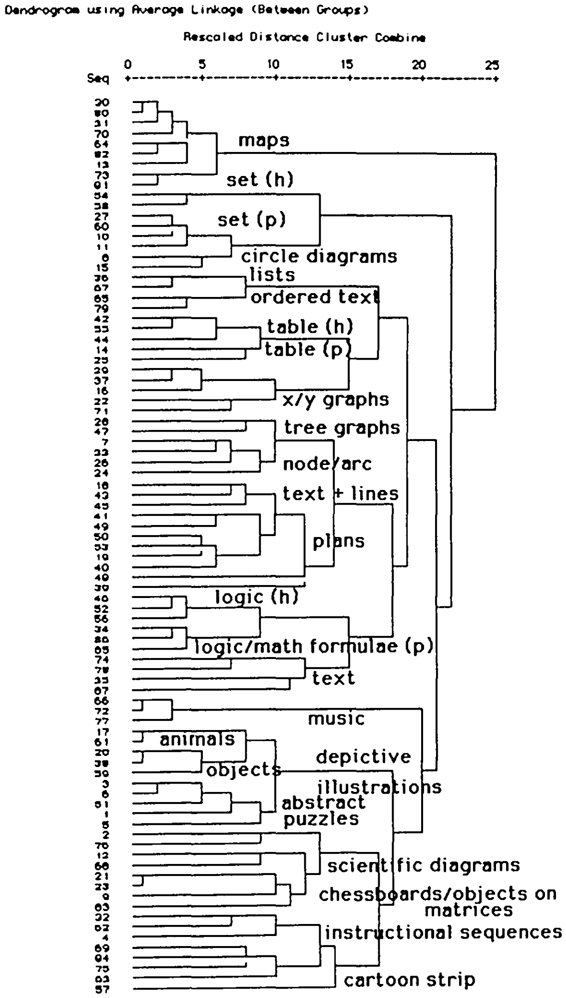


Figure 6.6: Dendrogram - note that numbers at leaves of tree correspond to item numbers of stimuli shown in Appendix 7

Table 6.1: 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	#corr. /4	Time	Rep. Used	#corr. /4	Time	Rep. Used	#corr. /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.

switchERI study

Statistical tests revealed that the two subject groups ('formal' and 'art' students) differed significantly from each other in terms of total time-to-solve on problem 2. 'Formal' students' solutions tended to be faster ($t = -1.93, df = 14, p < .05$). The two groups also differed in terms of total score on problem 3. 'Formal' students tended to score higher than 'art' students on that problem (Mann-Whitney 'U', $z = -1.99, p < .05$ — see Table 6.1). No other significant differences between the groups were found either in terms of score or times[4]. Since the group differences were not extensive, it was decided to pool the data into a single group of sixteen subjects.

---

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

The *switchERI* 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 screenrecordings were analysed in detail using Mediatracks[5]. Mediatracks is software for editing the digital screen recording made with ScreenRecorder. It allows the duration of screenrecording segments to be accurately measured and also permits annotations and notes to be added to the screenrecordings. It also permits screenrecordings to be replayed at slower-than-real-time or faster-than-real-time rates. External representation 'events' were coded, counted and recorded using that software.

Problem comprehension time was defined as the time from initial presentation of the problem to the point at which the subject selected a representation environment by clicking on one of the button icons (as shown in Figure 6.1).

The time-course analysis of the current study revealed that students typically spent 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.

A range of obstacles to comprehension were outlined in Chapter 3 — problem solvers must, at the problem comprehension stage:

- negotiate complicated syntactic structures and overcome the uncooperative aspects of the problem's discourse
- use the full problem information context (PIC)
- recognise the number of dimensions (variables) in the problem
- accurately discern the number of values along each dimension

---

[5] Farallon Computing Inc., 2000 Powell St, Emeryville, CA 94608, USA.

- understand that some information is implicitly stated
- utilize negatively-stated information
- accurately gauge the problem's level of determinacy
- detect redundancy, symmetry and critical points in the problem (Amarel, 1968)
- adopt an 'extensionalist stance' or 'puzzle mode' mind-set

It is surprising that subjects do not spend longer on this phase of problem solving, but seemingly rush headlong into an ER selection decision. Of course, subjects could re-read the problem information following ER selection and through ER construction and question answering, but it is, nevertheless, interesting that they do not reflect for longer at this point.

Even highly proficient subjects make comprehension errors. Wollman (1983) has shown that even mathematical physicists may initially make a reversal error in their algebraic representation of 'students and professors' problems. However, in those subjects, self-monitoring and checking result in the initial incorrect equation being corrected. For more typical subjects, Wollman (1983) has shown that even brief training in monitoring, comparing and checking (he calls this an 'active operation' approach) produces dramatic improvements in performance. It would be relatively straightforward to arrange for an intelligent learning environment to support 'active operations' on the part of the student and this is an interesting topic for future research.

ER selection

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, 1995). Subjects have to be familiar with relatively domain-specific ER forms such as Euler's circles or Venn diagrams and must 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 *switchERI* 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 (Figure A.18, Appendix A, Subject S11) is associated with lower efficiency in terms of score per unit time (Table 6.1, Problem 2, subjects 11 and 14).

The 'pseudo set-diagram' representation for problem 2 recommended by the 'crammer' from which it was taken (Brownstein et al., 1990) is shown in Figure A.21 in Appendix A.

The crammer tried to invent a dotted-circle notation as augmentation for set diagrams to allow the requisite abstractions to be expressed. As Cox, Stenning & Oberlander (1995) have observed, this augmentation is semantically incoherent. However, set diagrams are far from useless for problem 2. Subjects who demonstrated on the taxonomy pre-test that they understood set-diagram semantics scored well using them (Subjects

S1, S3 and S9, Table 6.1. Set diagrams can be used to identify pairs of premises (from the total of 8 premises) which form the syllogisms that, when solved, provide the answers. Indeed, these syllogisms can themselves be solved using set-diagrams with some abstraction tricks. What set diagrams cannot do is to provide a representation of all eight premises, even using dotted circles, which is what the crammer recommends.

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' premise that explicitly gives the location of 'Ms Green' in the 'office 5', the 'room with the large window'. This is a sensible strategy because the search space is pruned by early use of that heuristic. Early use of determinate information speeds the recognition of crucial problem characteristics such as whether a single unique model can be built for the information. If not, then a change or switch of representation might be called for.

As in the first (workscratching) study, discourse ordering of ER construction was associated with lower scores than target ordering and was a 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 multidimensionality (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. Some evidence in support of this contention is provided by the fact that the majority (six) of the ten switching subjects did not accurately identify the representation that they switched from on the taxonomy pre-task. 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. On problem 3, the only problem where the numbers of switchers (5) and non-switchers (10) were large enough to permit a statistical comparison, the time difference was significant ($t = 3.06, df = 13, p < .01$).

Like the workscratching data, the *switchERI* 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) 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, were a pseudo set-diagram, list structure and a tabular representation, respectively. None of those ERs were capable of representing 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 (Figure A.18, Appendix A).

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 *switchERI* 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.

The relationship between prior knowledge and reasoning performance

An important source of between subject variation is prior knowledge. The *switchERI* study highlights the importance of this factor particularly when indeterminate information has to be represented. Figure A.18 in Appendix A shows that when the subject (subject 11, problem 2, see Table 6.1) does not fully understand the representational formalism, unconventional annotations — various types of arrow — are invented 'on-the-fly'. ERs marked with an asterix in Table 6.1 indicate that the subject used an ER formalism that they identified accurately in the taxonomy task in terms of precisely categorising and labelling the representational formalism. The absence of an asterix indicates that they did not show evidence of comprehending the formalism on the taxonomy task.

Lack of prior knowledge of ER formalisms tended to result in poorer scores[6]. The effect is particularly pronounced for the indeterminate problem for which 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 ER users with prior knowledge and 1.8 for ER users who lacked prior knowledge.

---

[6] 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.

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). This result therefore provides some empirical support for specificity theory (Stenning & Oberlander, 1995).

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 were 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 & Holder (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 visualization[7] 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.

---

[7] It 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.

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 easily express crucial aspects of the problem, as shown by several subjects' attempts 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 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), possibly 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 subjects in some circumstances may be better off using their incorrect ER than guessing, if switching ERs or ER reconstruction is not an option. 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.

Summary and Conclusions — *switchERI* study

The findings of the *switchERI* study can be summarised as:

- Subjects seem, often, to 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

- Prior knowledge can be assessed using a taxonomy task
- Students do better if they fully comprehend the semantics of the ER formalism that they attempt to use in their solutions

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

The empirical results from both the 'workscratching' corpus data and the *switchERI* study strongly suggest that flexible environments are needed for supporting analytical reasoning with ERs. The *switchERI* 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.

A range of ER 'issues' that an intelligent computer-based support environment might detect and use in coaching interventions were identified, though the 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, the data yield a range of suggestions for the kinds of knowledge that a coaching system might draw upon. Some of the findings were incorporated into the *switchERII* system to be described in Chapter 7. The second version of the system was 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 would serve 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 premises may also aid comprehension by permitting related information to be juxtaposed. As an example, in the second problem, premise re-organisation would make transitivities much more salient. For example premise 4 "All those who enjoy the poetry of Coleridge also enjoy the poetry of Donne" could be placed immediately adjacent to premise 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 easily express important aspects of the problem. This can be due either to the semantic properties of the ER (*e.g.* graphical specificities), to lack of prior knowledge, 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, since 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. The learning outcomes from self-directed learning are much richer and more valuable in the long term than from approaches that foster 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, and elsewhere (*e.g.* Cox, Stenning & Oberlander, 1995), 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, 1995). A general principle for which there is empirical and theoretical support consists of 'Choose a representation with sufficient expressive power to represent the indeterminacy in the problem, but which does not have *more* expressive power than required.' (Stenning & Oberlander, 1995; Cox, Stenning & Oberlander, 1995).

Unfortunately, though, 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, as described in Chapter 2, 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). To date, however, 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. This point is explored further in Chapter 8 in relation to a proposed 'ER curriculum'.

In the current studies 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). Cox, Stenning & Oberlander (1995) also examined ER selection errors. They have shown that subjects who use a determinate ER (*e.g.* a table) in their solutions to an indeterminate problem results in worse performance than the converse *i.e.* using abstraction-expressive ERs on determinate problems.

Supporting ER Construction

A facility for highlighting the problem premise 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' words[8] might also assist ER construction. The system could ensure that explicitly given information[9] 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, perhaps 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 the most practical way to inform the system.

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 can detect ER construction sequences that reflect the order in which information is presented in the problem (*i.e.* discourse ordering). The detection of discourse ordering could 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

---

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

[9] eg 'Ms Green, the senior employee, is entitled to Office 5, which has the largest window.'

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 options.

Many of the design ideas that emerged from the *switchERI* data were incorporated into the design of a second system, *switchERII* which will be described in the next chapter.

ER Switching Two kinds of switching were apparent from the *switchERI* study data — task requirement driven versus impasse driven. The former is opportunistic and takes place when the subject judiciously switches ER because the task requirement changes and the previous ER is now computationally inefficient compared to some alternative. The latter, less adaptive, kind of switching might be termed 'impasse-driven'. This type of switching occurs as a response to uncertainty about how to proceed ('thrashing'). In the case of impasse driven switching[10], a frequent response was for subjects to switch ERs. There is much scope for debate about the merits of switching. 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.*

---

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

Foss, 1987) a limited form of thrashing can be productive. Therefore the proposal is to encourage ER switching with support to help the student avoid the more damaging aspects of thrashing.

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) offers support to the subject during the extremely demanding activity of ER construction. It has been 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 *switchERI* 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 could enable ER construction efforts 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 the subject.

On the other hand, however, the cognitive load of constructing even a single ER is very high. Perhaps it would be very reasonable to offer co-construction as a selectable option for the user?

Or 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 *switchERI* (*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 'expert' system[11].

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

---

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

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 is possible to address, at least to a limited extent.

Conclusion

It is argued, 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.

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 plus ER construction places a heavy cognitive burden on the subject which intelligent support can help to alleviate.

In summary, the results of the *switchERI* study were used to inform the design of a second system *switchERII*. In developing the *switchERII* system, a major aim was to provide a degree of support and feedback to the subject at the stages of ER selection and construction. Other objectives were to a.) further investigate the relationships between prior knowledge of ERs and reasoning with ERs and b.) to empirically evaluate the effectiveness of the *switchERII* system.

Chapter 7

SwitchERII – an intelligent ILE.

switchERII development

In developing the second system (*switchERII*), the aims were:

- to extend *switchERI* and to design an ILE capable of providing support and guidance to subjects in the selection and construction of ERs and to evaluate the efficacy of the system's feedback and interventions
- to use the system to further investigate the relationship between subjects' prior knowledge of ER formalisms and their performance in reasoning, with special reference to the case of set-diagrams
- to further investigate the issue of ER switching as a means of resolving reasoning impasses
- to further investigate stages in the *process* and time-course of reasoning with ERs
- to incorporate design features derived from the workscratching and *switchERI* studies. Some features were implemented via interface facilities and some via knowledge representation in the system (*e.g.* real-time diagram parsing during ER construction
- to use an interactive learning environment to study ER use in a quasi-naturalistic setting (as discussed in Chapter 1)

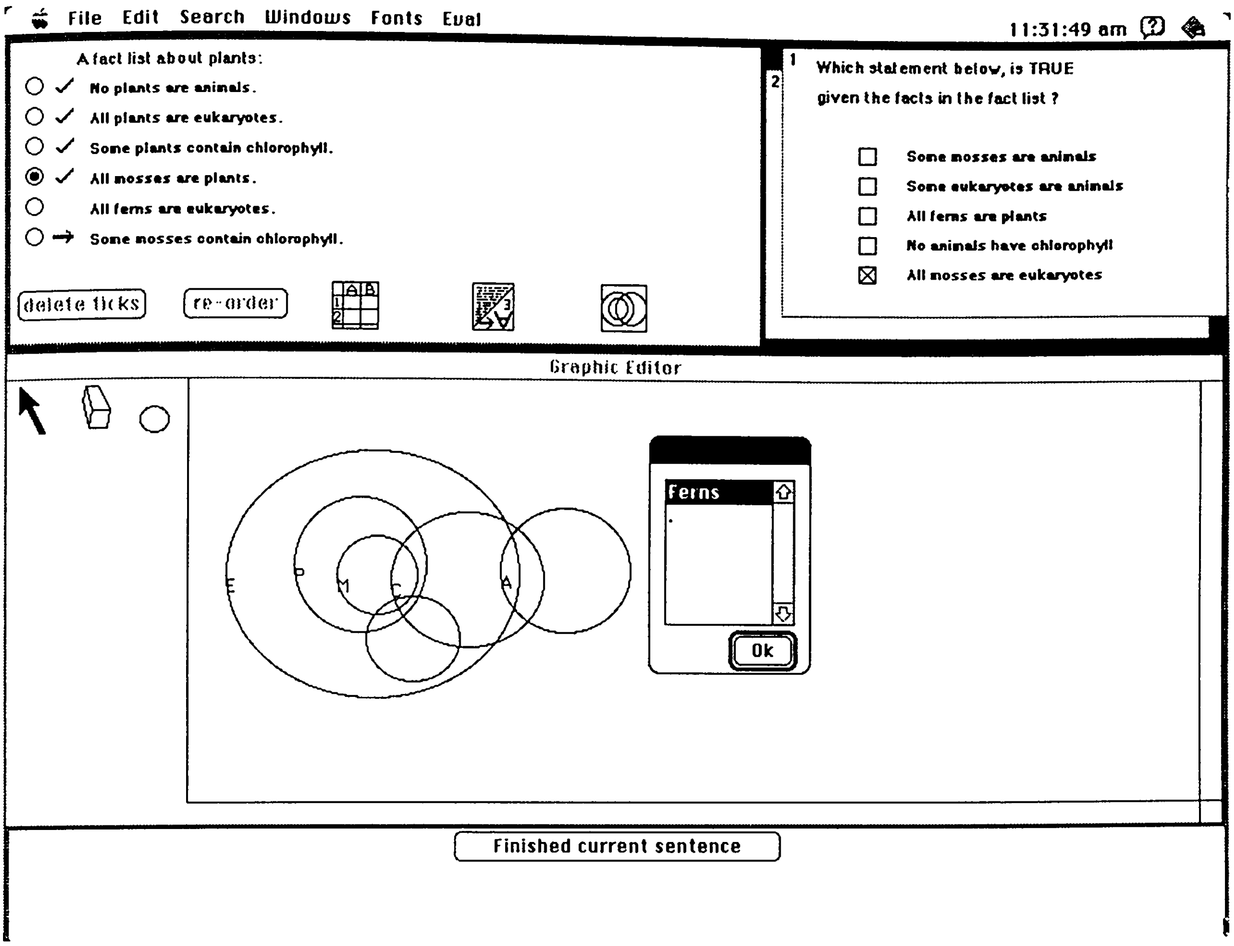
The general approach adopted in study 3 was to focus upon analysing, in detail, ER construction errors and switching in subjects' use of a set diagrams. Euler's circles were chosen as the graphical formalism since they have a well-defined, computationally tractable, underlying semantics. Also, there is a relatively large amount of experimental data on how they are interpreted (*e.g.* Newstead, 1989, 1995; Stenning & Cox, 1995) and constructed (*e.g.* Grossen & Carnine, 1990; Ford, 1995). They have also been extensively studied theoretically (Stenning & Oberlander, 1995).

The *switchERII* system incorporated ER parsing — as the subject constructed his or her diagram, the system interpreted the ER. Compared to *switchERI*, the generality of the *switchERII* system was restricted. The ER environments of *switchERI* could be used to construct any type of ER for any type of analytical reasoning problem since the system possessed no knowledge of representational semantics. In *switchERII*, however, it was decided to represent the semantics underlying Euler's circles in order to be able to parse the user's diagram dynamically during construction and offer feedback to the user in event of errors in construction.

For study 3, two problems were selected, both of which are solvable using set diagrams of the Euler's Circle (EC) kind. The first problem was a practice problem developed by the author (Figure 7.1). The experimental problem used for data collection was the 'poets' problem employed in studies one and two. The poets problem is indeterminate in that many different, but valid, diagrammatic models can be constructed of the information. Data from studies one and two of this thesis, and from Cox, Stenning & Oberlander (1995), show that the poets problem is 'brittle' in the sense that it is highly sensitive to appropriate ER selection and requires accurate ER construction.

Assessing prior knowledge of ER formalism

The adoption of the poets problem also permitted the use of a more focussed and domain-specific test of prior knowledge than was the case with the card-sort taxonomy task employed in study 2. The task is based on one used by Newstead (1989) and subsequently by Stenning & Cox (1995), Stenning, Yule & Cox (1996) and Newstead (1995). In the test, five Euler's circle diagrams were presented to the subject. Each diagram consisted of two circles one labeled 'A' and one labeled 'B'. Diagram 1 showed



A fact list about plants:

○ ✓ No plants are animals.
○ ✓ All plants are eukaryotes.
○ ✓ Some plants contain chlorophyll.
● ✓ All mosses are plants.
○ All ferns are eukaryotes.
○ → Some mosses contain chlorophyll.

1 Which statement below, is TRUE given the facts in the fact list ?

☐ Some mosses are animals
☐ Some eukaryotes are animals
☐ All ferns are plants
☐ No animals have chlorophyll
☒ All mosses are eukaryotes

Figure 7.1: Practice problem used in *switchERII* study

the identity relation (circle 'A' and circle 'B' superimposed); diagram 2 showed a small circle 'A' inside larger circle 'B'; diagram 3 showed a small circle 'B' inside circle 'A', diagram 4 showed circles 'A' and 'B' intersecting and diagram 5 showed 2 non-overlapping, disjoint circles 'A' and 'B' (Figure 7.2).

Below the diagrams, 4 premises were listed in the order ALL As are Bs, NO As are Bs, SOME As are Bs, and SOME As are NOT Bs. Adjacent to each premise were the numbers 1 to 5. Subjects were instructed :

> "Below this paragraph there are five circle diagrams labeled 1 to 5. They represent sets of objects (A's and B's). Below the circle diagrams

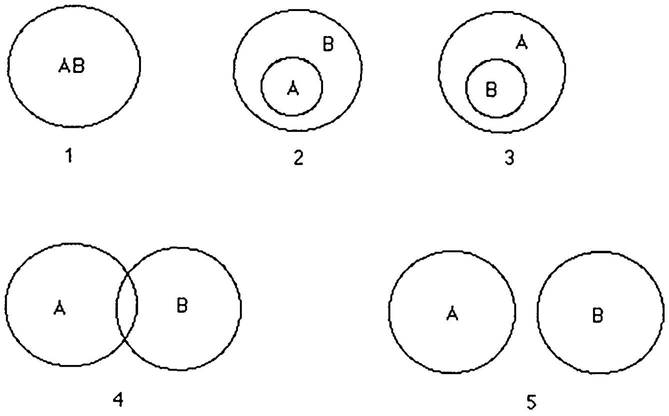


Figure 7.2: Euler's Circle interpretation task due to Newstead (1989)

> there are four statements. Please circle the number(s) of the diagram(s) that the sentence is true of. If you think 'All A's are B's' is true of diagram 3, circle 3 alongside that sentence. You may circle more than one number per statement. Please interpret "some" to mean "at least one and possibly all".

There is a substantial amount of data available on subjects' performance on this task and it can be used to detect a variety of interpretational error 'syndromes' such as conversion errors, 'Gricean' errors and 'island' responding. These error patterns are described in detail below. The diagrams selected by each subject for each statement are shown in Table A.1 in Appendix A.

switchERII features

A major aim with *switchERII* was to construct an intelligent learning environment (ILE) rather than a passive system and to inform its design by means of findings from the *switchERI* and workscratching studies.

As with *switchERI*, the aim was to build a system which presents problem stem information and associated questions and which provides a range of ER construction environments for subject to choose from and switch between.

SwitchERII was implemented in LPA MacProlog 4.5[1]. The development was made as modular as possible with the sections of Prolog code being divided at top level into a. interface, b. ER tools and c. diagram parsing routines.

SwitchERII incorporated a facility for subjects to re-order the problem stem sentences before (or at any time during) problem solving and/or ER construction. This feature was implemented because it was noted in the previous studies that subjects often re-ordered the problem information in the course of re-writing the problem information or while constructing an ER. In a sense, sentence re-ordering can be thought of as a kind of ER construction activity.

SwitchERII provides feedback to the user regarding:

---

[1] Logic Programming Associates Ltd, Studio 4, RVPB, Trinity Rd, London, SW18 3SX.

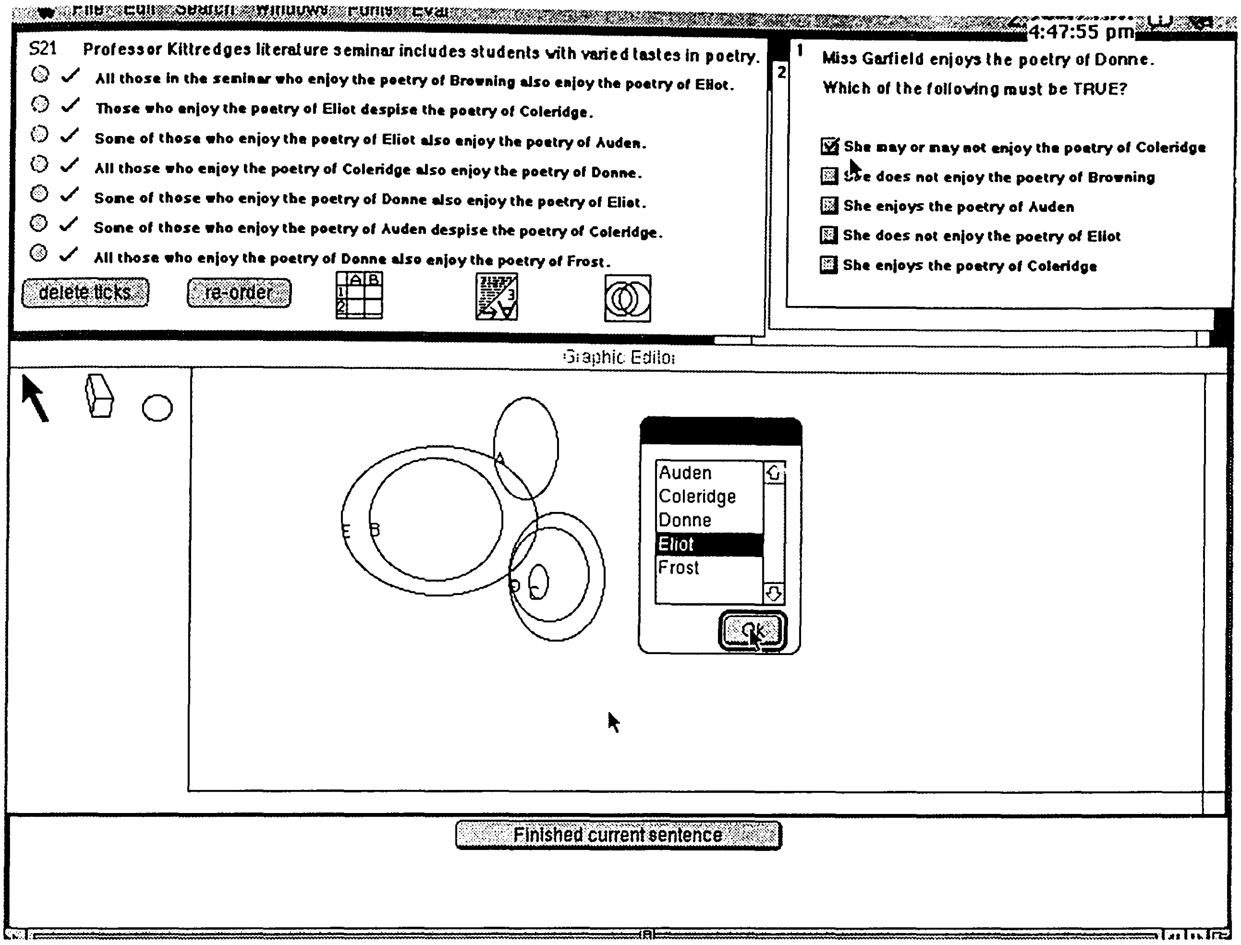
- his/her progress through the problem
- the comprehensiveness of his/her representation in terms of the proportion of problem information represented
- the accuracy and validity of his/her diagrammatic model

ER 'co-construction' was suggested in the discussion of the *switchERI* study (see Chapter 6, section on 'Implications for the Design of an Intelligent Learning Environment') and was implemented to a limited extent in *switchERII*. In *switchERII*, when the user constructs a set diagram using the graphical ER construction environment, a 'propositionalised' re-representation of the problem stem information is entered into the text ER environment on a sentence by sentence basis as the subject works through the problem. For example, if an Euler's circle representation of 'All those who enjoy the poetry of Browning also enjoy the poetry of Eliot' is constructed in the graphics window, perhaps as a small circle labeled 'B' contained by a larger circle 'E', then the sentence 'Browning liker definite Eliot liker' is written into the text environment.

The system's entries into the text environment are unobserved by the user, however, until s/he switches from the graphics environment to the text environment. It should be noted that the sentence re-representation facility is not intelligent — if the subject incorrectly models the problem information in the graphical ER environment, a correct 'propositionalised' sentence corresponding to the currently-worked-on premise is, nevertheless, entered into the text environment. Future developments of the system could feasibly implement intelligent co-construction, however.

***switchERII* ER construction environments** The three ER construction environments supported by *switchERII* are illustrated in Figure 7.3, Figure 7.4 and Figure 7.5.

- The graphics environment provides an ER construction area and tools palette. The tools consist of an ellipse/circle drawing tool, an eraser tool and selection cursor. Ellipse labels are selected from a scrolling list in a dialogue box presented whenever an ellipse is drawn and which contains a list of entity names appropriate



S21 Professor Kittredges 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.

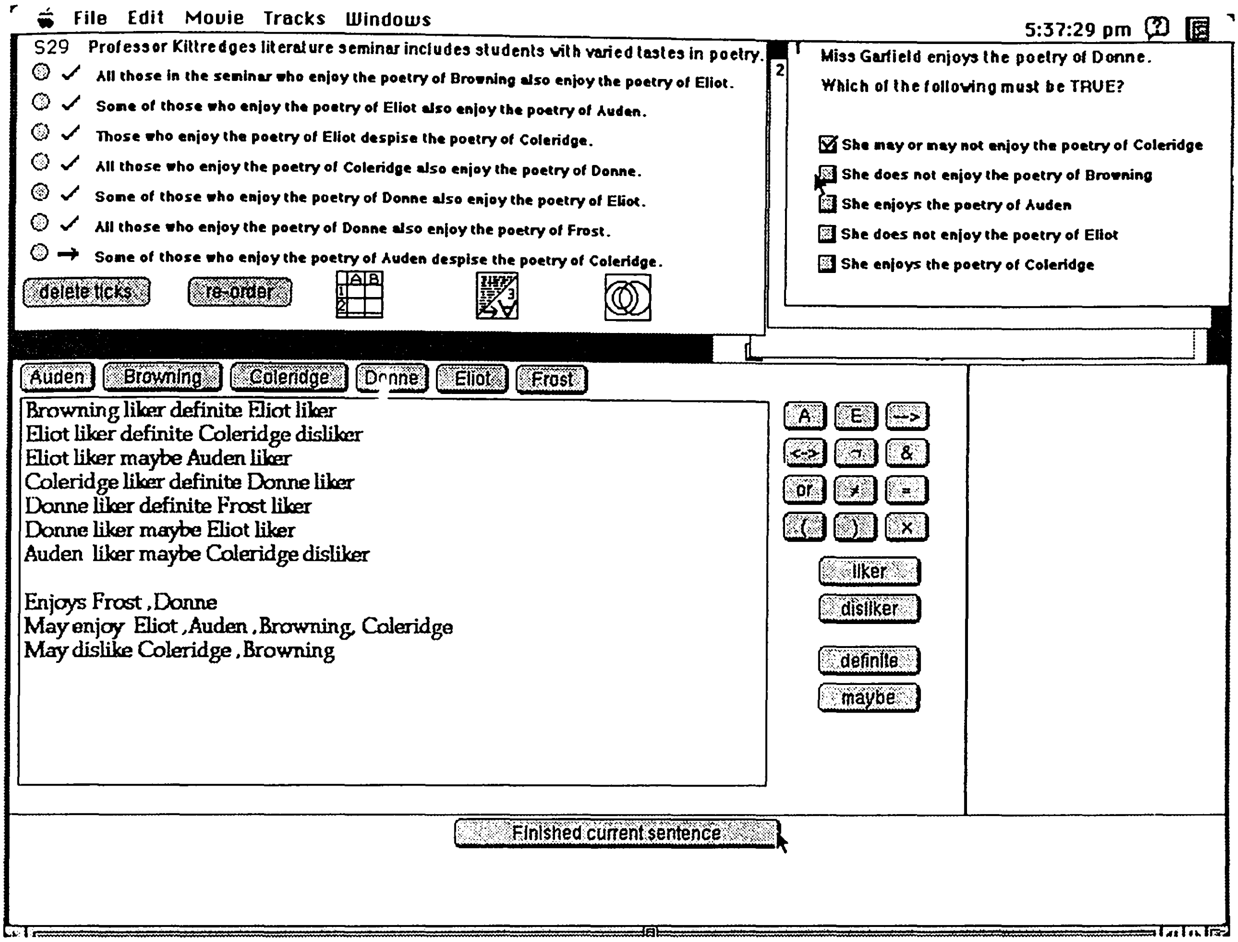
○ ✓ All those who enjoy the poetry of Donne also enjoy the poetry of Frost.

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

Figure 7.3: Subject 6 using the *switchERII* diagram environment to construct a set diagram



S29 Professor Kittredges 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.
○ ✓ Some of those who enjoy the poetry of Eliot also enjoy the poetry of Auden.
○ ✓ Those who enjoy the poetry of Eliot despise the poetry of Coleridge.
○ ✓ 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.
○ ✓ All those who enjoy the poetry of Donne also enjoy the poetry of Frost.
○ → Some of those who enjoy the poetry of Auden despise the poetry of Coleridge.

2

Miss Garfield enjoys the poetry of Donne.

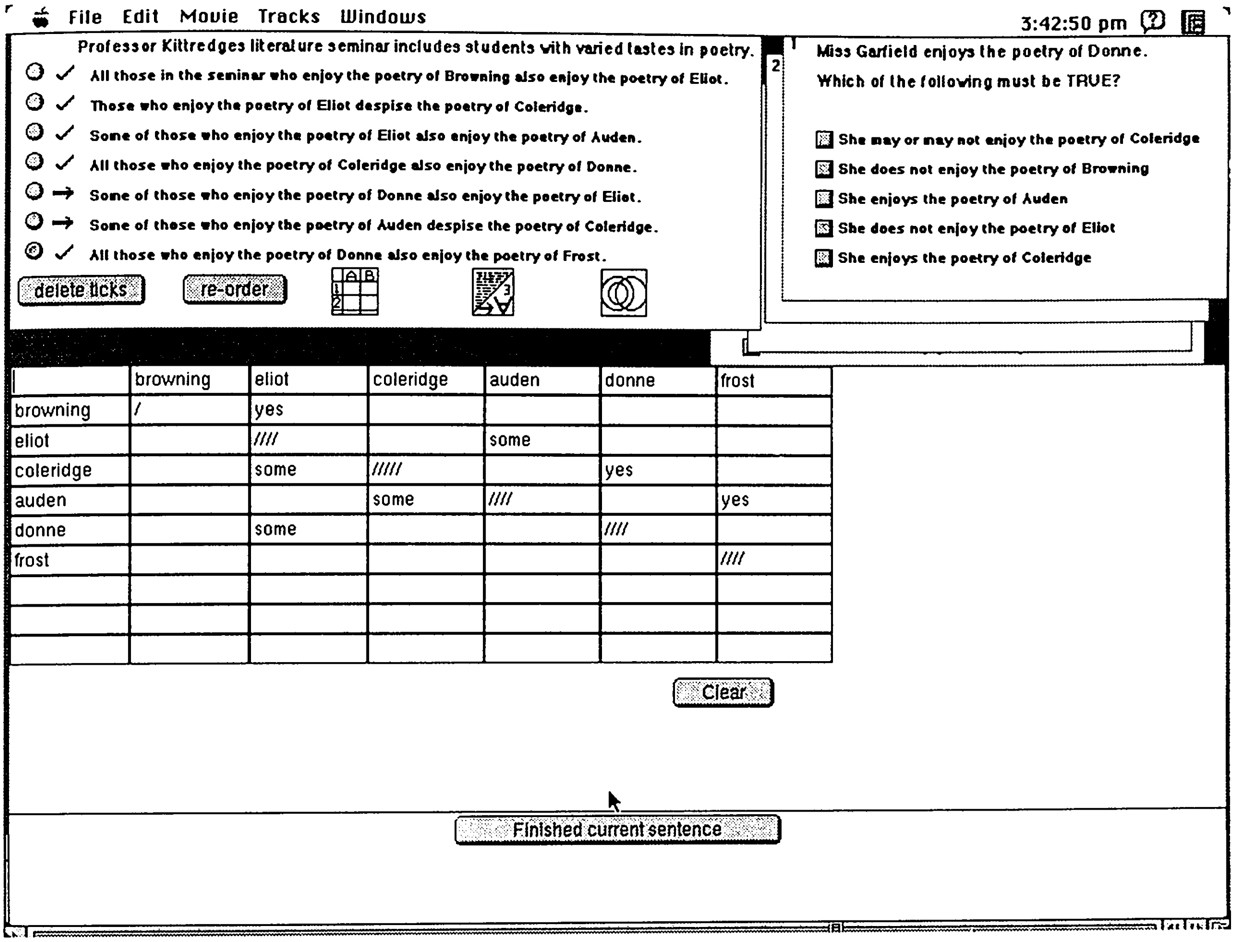
Which of the following must be TRUE?

☑ 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

Browning liker definite Eliot liker
Eliot liker definite Coleridge disliker
Eliot liker maybe Auden liker
Coleridge liker definite Donne liker
Donne liker definite Frost liker
Donne liker maybe Eliot liker
Auden liker maybe Coleridge disliker

Enjoys Frost, Donne
May enjoy Eliot, Auden, Browning, Coleridge
May dislike Coleridge, Browning

Figure 7.4: Subject 16 uses the *switchERII* text environment to re-write the problem information and to make notes.



Professor Kittredges 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?

☐ 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

	browning	eliot	coleridge	auden	donne	frost
browning	/	yes				
eliot		////		some		
coleridge		some	/////		yes	
auden			some	////		yes
donne		some			////	
frost						////

Figure 7.5: Subject 14 using the *switchERII* matrix ER environment to construct a contingency table.

to the problem (Figure 7.3)[2].

SwitchERII's graphic environment incorporates the interpretative semantics for set-diagrams (Euler circles) and detects the overlap of graphic elements such as ellipses, the containment of one graphic element by another, and non-overlap, non-containment (*i.e.* disjoint) relationships between graphic elements. *SwitchERII* dynamically parses the subject's representation *during* construction and provides feedback.

- The text environment provides basic word-processing facilities and a palette of first-order logic symbols (the symbols were also available from the keyboard). Buttons arrayed around the text area, when clicked, cause their labels to be written into the text area, thus saving typing and time for the subject... *e.g.* clicking on button labeled 'Auden' causes the word 'Auden' to be written into text area at the current cursor position (Figure 7.4).

  Other buttons correspond to 'liker', 'disliker', 'definite' and 'maybe' (Figure 7.4). The reason for providing these nouns and adjectives (as opposed to a set corresponding to the actual wording of the problem, *i.e.* all, enjoy, also, like *etc*, was to encourage what might be termed 'propositionalisation' during re-representation. For example 'If Browning liker then Coleridge liker' as a representation of 'All those in the seminar who enjoy the poetry of Browning also enjoy the poetry of Eliot' is more an analogue of a set diagram using Eulers circles than a direct re-write of the original sentence. In other words, the 'If Browning liker then Coleridge liker' form is akin to converting the original subject-predicate-object syntactic form into a propositionalised form midway between the original natural language and the more formal, single arity notation:

  $\forall x(browning\_liker(x) \rightarrow eliot\_liker(x))$

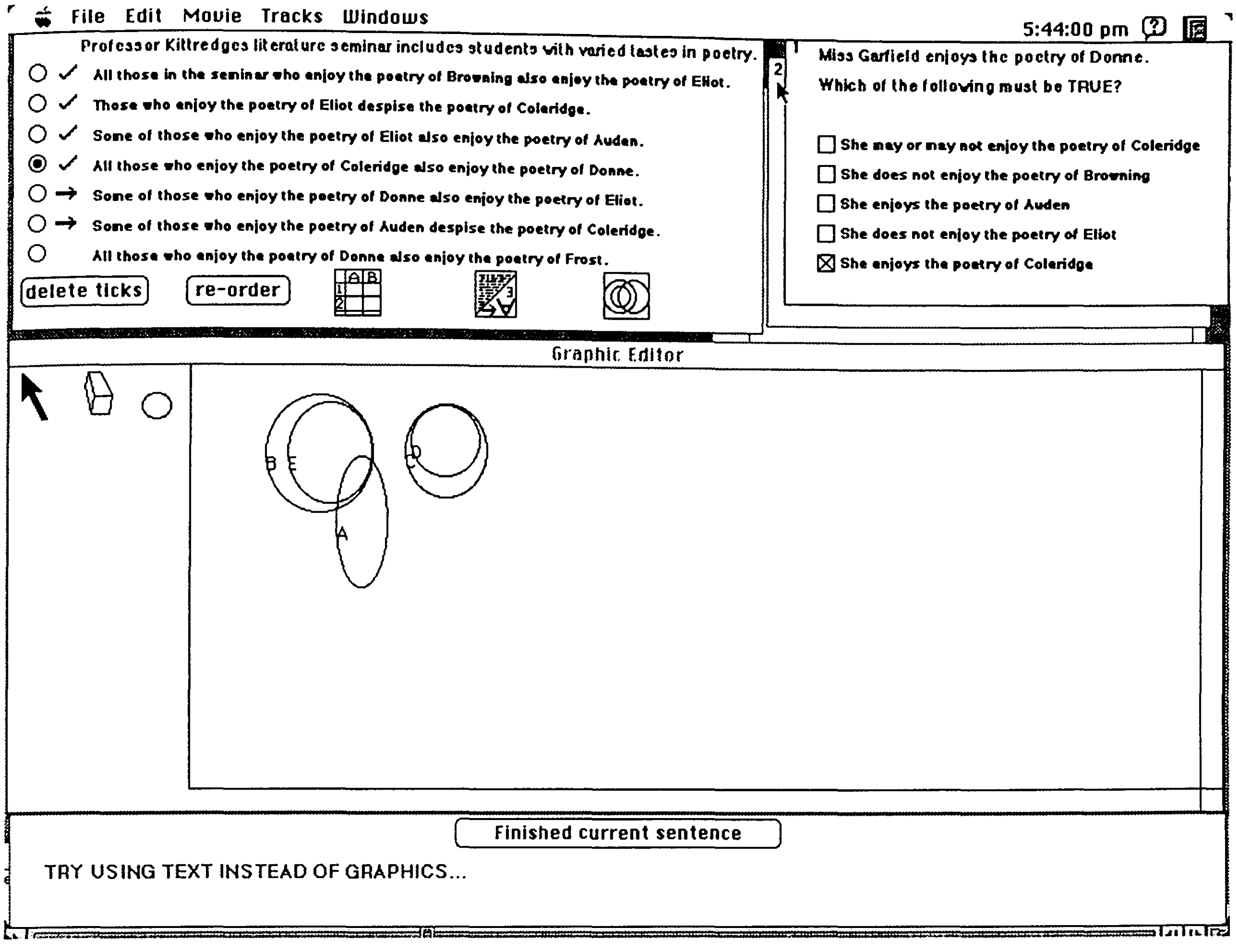
  switchERII's shorthand might be termed 'partial propositionalisation'

- Spreadsheet

  This environment consists of a 10 × 7 grid of cells. Text or numerical information can be entered into the cells (Figure 7.5).

---

[2] Note that, for illustration purposes, Figure 7.3 shows the entity label dialogue window with 5 of the 6 entity labels (poets' names) present. In actual use, only *unused* labels appear.



Professor Kittredges 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?

☐ 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

TRY USING TEXT INSTEAD OF GRAPHICS...

Figure 7.6: 'Represent-ahead' ticks and system advice to switch following three ER construction errors. Note reversal errors (BE, DC) in diagram and resultant system message in lower window (Subject 9)

User feedback provided by switchERII The subject indicates which problem sentence s/he is working on by clicking on a radio button next to the sentence (*e.g.* Figure 7.6). The radio button then becomes filled with a green dot and a tick appears in between the radio button and the sentence. The ticks therefore indicate to the subject which sentences s/he has worked on so far.

System feedback is also provided to the user in the problem sentence window. Under certain conditions, *switchERII* generates an arrow alongside a problem stem sentence, as shown in Figure 7.6. In the example shown, the subject has represented the first four sentences of the problem in the graphical (set-diagram) environment - they have represented relationships between likers of Browning, Eliot, Coleridge, Donne and Auden. However, in doing so s/he has also (implicitly at this stage) represented the fifth and sixth sentences ('Some of those who enjoy the poetry of Donne also enjoy the poetry of Eliot' and 'Some of those who enjoy the poetry of Auden despise the poetry of Coleridge'). *SwitchERII*, following a parse of the current state of the set-diagram, detects this and flags this to the subject via the arrows. The arrows are termed 'represent-ahead' arrows because they indicate a commitment to representing information ahead of the user's current position in his or her working through the problem information.

System messages appear in the message window - the narrow window at the bottom of the display — as shown in Figure 7.6. There are five system messages:

1. 'Tables are not advised for this problem - proceed anyway?' - this message was generated if the subject selected the table ER construction environment
2. 'Please indicate the sentence you are now working on' - generated if the subject attempts representational activity without nominating which sentence is current focus of representational activity via clicking radio button alongside the sentence
3. 'Maybe you weren't representing the right sentence?' This message is generated when the subject uses either the text or graphical ER environments if the problem entity labels (*i.e.* poets names) in the representation do not correspond to the currently selected sentence in the problem stem.
4. 'There may be a problem with the representation...' - this message is generated when the system detects a semantic error in the subject's set-diagram. One of the most common reasons for this feedback to be given is because the subject 'converts' during representation construction. An example of conversion is shown in Figure 7.6 — the subject has represented 'All those who enjoy Browning also enjoy Eliot' with a small circle labeled 'Eliot' inside a larger circle 'Browning' - whereas a correct representation could be a smaller circle 'Browning' contained by a larger circle 'Eliot'
5. The system message 'TRY USING TEXT INSTEAD OF GRAPHICS...' is

generated following three semantic errors in diagram construction and an explicit recommendation that the subject considers switching representational modality[3]

Third study - *switchERII*

The aims of the third study were:

- to evaluate the efficacy of system feedback and guidance provided by *switchERII*
- to assess subjects' prior knowledge of Euler's circles and to investigate the relationship between subjects' performance on a diagram *interpretation* task and their subsequent performance during diagram *construction*
- to further study the processes, stages and time course of reasoning with self-constructed ERs

Pilot study

Prior to the main experiment, a pilot study involving 3 subjects was conducted in order to evaluate system performance and to finalise the details of the experimental procedure. The pilot study proved to be very worthwhile - numerous comments from the pilot subjects resulted in changes to the procedure and system.

The instructions to subjects, for example, were modified in the light of the pilot study...one of the pilot subjects was unclear about whether the 'poets' problem's questions had a single correct answer or whether several multiple-choice options could be chosen (the instructions ask for subjects to 'select the best answer choice given').

The pilot study also allowed a bug in the diagram parsing software to be detected - as a result 'higher resolution' overlap detection was incorporated. The positions of several interface buttons were also altered in response to pilot subject suggestions.

---

[3] Evidence reported in Cox, Stenning & Oberlander (1995) suggests that where a valid diagrammatical model is unavailable or inappropriate, the more expressive, linguistic, modality (*i.e.* natural language or first-order logic) is optimal in terms of performance on indeterminate problems such as the 'Poets' problem.

Subjects

The subjects consisted of sixteen students (7 females and 9 males) from a variety of undergraduate and masters level courses at the University of Edinburgh[4] who responded to poster advertisements placed on noticeboards. The posters invited subjects to participate in a study of 'computer-based problem solving' and offered £5 as payment for a one-hour session.

Subjects were randomly assigned to two groups of eight. The two groups comprised the 'Feedback' (FB) and 'No feedback' (NFB) conditions. The FB group consisted of 6 males and 2 females and the NFB group consisted of 5 females and 3 males.

Subjects in the FB condition used a version of *switchERII* which generated feedback and guidance during ER construction. The NFB subjects used a version of *switchERII* which did not display feedback messages[5]. The *switchERII* systems used in the FB and NFB conditions were, in fact, identical except that the message-to-subject text strings in Prolog code of the version of *switchERII* used by the NFB group were replaced by empty strings. Hence the FB and NFB systems behaved identically in every way except that the NFB system did not display any of the 5 message types in the message window.

Procedure

The experimental procedure (pre-test and *switchERII* trials) took approximately 45 – 50 minutes.

Pre–test

Subjects were first administered the pre-test of knowledge of Euler's method of representing syllogistic premises. Details of the test (originally due to Newstead (1989)) were described above. The pre-test typically took 5–6 minutes to complete.

---

[4] Cognitive science, psychology, linguistics, business studies, computer science, physics, and economics.

[5] n.b. Ticks and 'represent-ahead' arrows were displayed alongside the problem stem sentences in both groups.

switchERII trials

Following the paper and pencil pre-test, subjects were introduced to the *switchERII* system via the practice example (Figure 7.1). The experimenter pointed out the various screen windows and regions — problem information, ER construction environment icons, question windows, the ER construction area, system message window *etc.* Subjects in the FB group were told that messages would appear if the system 'thought' there was a problem with their representation. They were told that they were free to act on any system suggestions, or ignore them, as they saw fit. The sentence re-ordering facility was demonstrated to all subjects by the experimenter and each ER construction environment was demonstrated. In *switchERII*, a particular ER environment is invoked by clicking on a button labeled with the environment's icon. The icon–buttons can be seen, for example, in Figure 7.3. The icons correspond to table ERs, text/logic, and graphics and their presentation order from left to right is randomly determined each time the *switchERII* program is run. The experimenter demonstrated the ER construction environments in whatever order the icons appeared, from left to right. The purpose of the icon randomisation was to reduce the effects of (inadvertent) suggestions by the experimenter that a particular ER environment might be 'preferred'. Randomisation was also intended to minimise systematic biases in ER environment selection — such as, for example, any tendency by subjects' to initially select the leftmost ER environment. During the demonstrations, information 'carry over' from the graphical environment to the text environment was illustrated. It was emphasised to the subject that, whichever ER environment they used, they should indicate which sentence they were currently working on by clicking the radio button next to it. When they had finished representing the information in the sentence, they were instructed to click on the 'Finished current sentence' button in the lower (system message) window[6].

When the experimenter had finished demonstrating the system features, the subject attempted the practice problem on their own. When they had completed the practice example, subjects were asked if they were ready to attempt the experimental problem and whether they understood the system. Further demonstrations were given at that

---

[6] The experimenter frequently had to remind subjects to do this during both practice and experimental problem solving sessions.

point, if required. When the subject indicated that they were ready to attempt the experimental problem, the experimenter told them that a screenrecorder would be run in the background but that this would not affect the operation of the system in any way. A miniature lapel microphone was attached to the subject's clothing and they were asked to 'think aloud' as they solved the experimental problem. They were further instructed that if they felt that thinking aloud was disrupting their performance they should not feel compelled to comply with the request. When they were ready, the experimental problem was run and the experimenter read out the following instructions:

> (In this problem)...'the questions are based upon a set of conditions. In answering, it may be useful to draw a diagram. For each question, select the best answer choice given.'

Subjects were instructed to begin their solution using set diagrams. They were further told, however, that if they were unhappy with using set diagrams that they could switch to an alternative ER environment at any time.

When the subject indicated to the experimenter that they were ready to begin, the screenrecorder[7] was started. The audiotape recorder was also switched on at that point. The only intervention by the experimenter from that point was to occasionally remind the subject to click the 'Finished current sentence' button. When the subject had answered the final (fourth) question, the experimenter stopped the screenrecorder. The experimenter made notes of his observations of the subject as s/he attempted the experimental problem. At the end of the session, the experimenter then questioned the subject for about 5 minutes. The conversation was audiotaped.

Results

Pre-test data — Euler's circle interpretation task

Table 7.1 shows the types of interpretation error made by each subject. Six subjects (3 in each group) demonstrated error-free performance on the Euler's circle interpretation task.

---

[7] 'Cameraman' — one of the Multimedia Utilities suite of programs, Motion Works International, San Francisco, CA. This utility produces Apple 'Quicktime' format screenrecordings.

Table 7.1: Conversion, Gricean, Island and other errors of interpretation.

	Feedback								*No feedback*							
Subject	1	3	5	7	9	11	13	15	2	4	6	8	10	12	14	16
Conversion								×					×			
GriceSO				×		×					×	×				
GriceSN				×		×				×			×			
IslandNO													×			
IslandSN												×				
OtherOmit	×				×	×		×		×		×	×	×		
OtherComit													×			
Total	1	0	0	2	1	3	0	2	0	2	1	3	5	1	0	0
ProbScore	3	2	3	4	2	2	4	3	3	3	4	4	1	3	4	4

Correct responses were defined as diagrams 1 and 2 for the premise 'All As are Bs', diagram 5 only for 'No As are Bs', diagrams 1,2,3 and 4 for 'Some As are Bs' and diagrams 3,4 and 5 for 'Some As are not Bs' (see Figure 7.2).

The results can be compared to previous data on this task reported by Newstead (1989) and Stenning & Cox (1995). The subjects in this study tended to perform at better levels than the undergraduate subjects studied by those previous researchers, though it must be born in mind that the sample size is much smaller here (n=16) than was the case for Newstead (n=40) and Stenning & Cox (n=138). The proportions of subjects correctly responding for each quantifier were as follows (S&C=Stenning & Cox, New=Newstead):

- ALL .75 (S&C=.61; New=.60)
- NO .94 (S&C=.81; New=.75)
- SOME .50 (S&C=.29; New=.33)
- SOME-NOT .62 (S&C=.38; New=.29)

The usual pattern across quantifiers was replicated here (ALL easiest, SOME most difficult).

The majority of subjects (10 out of 16) made errors of omission and/or commission. The error patterns were idiosyncratic but at least three 'syndromes' could be identified. These were errors of conversion, 'Gricean' interpretation errors and what the author terms 'island' responses. For each subject, 7 measures were derived — conversion errors, Gricean errors on 'Some As are Bs', Gricean errors on 'Some As are not Bs', island responding on 'No As are Bs', island responding on 'Some As are not Bs', errors of omission not accounted for by conversion, Gricean or island errors and, finally, errors of commission not accounted for by conversion, Gricean or island errors.

Conversion errors Conversion was deemed to have occurred when the subject chose diagram 1 alone as a model of 'All As are Bs'. That is, the subject interprets 'All As are Bs' to be equivalent to the statement 'All Bs are As', *i.e.* they 'convert' the universal quantifier. Table 7.1 shows the number and type of errors committed by each subject on the interpretation task. Two subjects converted (S10[8], S15).

Gricean errors These errors are characterised the adoption of a natural language interpretation of 'some' as excluding the possibility of 'all' (Grice, 1975).

Despite explicit instructions[9] to adopt a logical and not a natural language interpretation of 'some', 6 subjects showed Gricean errors. In this study Gricean errors on existential positive and existential negative premises have been distinguished. The former (Gricean-some) errors are defined operationally as the selection of diagrams 3 and/or 4 only as being true of premise 'Some As are Bs'. The latter (Gricean-somenot) errors are defined as the selection of diagrams 3 and/or 4 only as being true of the premise 'Some As are not Bs'.

The results suggested that the subjects in the present study were slightly less prone to Gricean interpretation errors than samples reported by Stenning & Cox (1995) and Newstead (1989). Four subjects (25 percent) demonstrated Gricean interpretations for the existential quantifier 'some' (S&C=40 percent, New= 30 percent). Four subjects (25 percent) showed Gricean errors on 'Some-not' (S&C= 38 percent, New= 27 percent).

---

[8] Table A.1 in Appendix A shows that S10 selected diagrams 4 and 5 in addition to diagram 1 as being true for the universal affirmative 'All As are Bs'. This error of commission is very unusual and S10 was therefore classified as a 'converter'.

[9] Recall that subjects were told in the instructions that 'Some' should be interpreted to mean 'at least one and possibly all'.

Island responses The term 'island' responding is coined here but describes a phenomenon noticed by the author in previous data (Stenning & Cox, 1995). It is defined operationally as a response in which the subject interprets the identity diagram (diagram 1) and diagram 2 as valid models of 'All As are Bs' - but then subsequently interprets diagram 2 as being consistent with the premise 'Some As are not Bs' — that is, interpreting diagram 2 as representing an 'island' of As in a 'sea' of Bs. In other words, the phenomenon is one of inconsistent semantics in which the subject's interpretation of spatial inclusion as metaphor for set membership holds for 'All As are Bs' but then metaphor changes when 'Some As are not Bs' is interpreted — whereas for 'All As are Bs' the small circle A is taken to represent a set containing *both* As and Bs, in an 'island' response, the small 'A' circle is interpreted to represent an 'island' of As in a 'sea' of B's. Island interpretation errors differ from conversion and Gricean errors in that they are due to inconsistent application of the 'spatial-containment-for-set-membership' metaphor across quantifier conditions.

Island-NO responses (Table 7.1) are defined as the selection of diagrams 2 and/or 3 for 'No As are Bs'. Island-SN responses are defined as the selection of diagram 2 as being true of 'Some As are not Bs'.

Table 7.1 shows that subject 8 manifested an island interpretation of 'Some As are not Bs' and subject 10 did so in response to 'No As are Bs'. For subject 8, island responding was associated with Gricean errors and for subject 10 island responding was associated with both Gricean and conversion errors.

Errors of commission versus errors of omission As Stenning, Yule & Cox (1996) point out, the psychological literature has tended to focus upon errors of commission where subjects make inferences that are invalid (as in the case of conversion errors). However, errors of omission such as failing to conclude that 'some B are A' given 'some A are B' have been studied very little. In the current study errors of omission were defined as diagram omissions not accounted for by those associated with conversion, Gricean or island responses. As table 7.1 shows, only one subject (S10) made an error of commission not accounted for by conversion, Gricean or island errors.

SwitchERII results

The *switchERII* screenrecordings were replayed by the experimenter. A 'first pass' through the data permitted the development of a coding sheet for the recording of significant problem solving and ER events. The screenrecordings were then played several more times and the experimenter recorded the incidence, frequency, type *etc.* of various events. A summary of the screenrecording event analysis data is presented in Table 7.2.

As Table 7.2 shows, 3 subjects made use of *switchERII's* problem stem sentence re-ordering facility. Re-ordering the sentences was associated with high rates of correct responding to the problem's questions — the 3 subjects who re-ordered scored 3, 4, and 4 out of 4 respectively. The row labeled 'Solution focus' in Table 7.2 indicates whether subjects built a complete model of the problem information (problem stem focus) or built partial models of the information on a question by question basis (question focus). Most subjects adopted the problem focussed approach. Two subjects reasoned mostly with a complete problem model but built question-specific models for question 3 (subject 9) and questions 1-3 (subject 16).

The next section of rows in Table 7.2 shows, for feedback group subjects, the frequency with which *switchERII* gave various feedback responses. The row headings correspond to the system messages 'Tables are not advised for this problem, proceed y/n'; 'Please indicate the sentence you are currently working on'; 'Maybe you weren't representing the right sentence'; 'There may be a problem with your representation' and 'Try using text instead of graphics...'. The label 'rep ahead' refers to the number of times that 'represent-ahead' arrows were generated by the system for that subject. Note that the two subjects for whom *switchERII* generated the highest number of 'represent-ahead' arrows (S9 and S10) performed relatively poorly in terms of correct question responses.

The section of Table 7.2 that refers to the type of ER errors made during construction requires some explanation. The notation 'revBE' for example indicates that the subject constructed a set diagram in which 'All those who enjoy Browning also enjoy Eliot' is represented by a small circle 'Eliot' contained by a larger circle 'Browning' instead of a smaller circle 'Browning' contained by a larger circle 'Eliot' *i.e.* a reversal error.

Table 7.2: *SwitchERII* screenrecording analysis - general results

Group	*Feedback*							
Subject	1	3	5	7	9	11	13	15
Re-order?	y			y				
Solution focus	stem	stem	stem	stem	stem/qn	stem	stem	stem
First sent. rep'd	1	1	2	1	1	1	1	1
Tables not advised								
Pls ind sent			1					
Maybe weren't ... sent		1	1	3	6			
Maybe problem ... rep			2		4	3	2	1
Try text instead ...					1	1		
Rep. ahead arrows	2	2	2	1	5	2	2	1
No.ER switches	0	0	2	1	0	1	3	0
No. errors in ERs	0	1	2	0	3	3	1	1
Type ER errors	- - -	$\neg for\forall CD$	revBE revCD	- - -	revBE revCD $\exists for\forall DF$	revBE $\exists for\neg CB$ revCD	revBE $\exists for\neg CB$	revDF
All sent rep'd?	y	n	y	n	n	y	y	y
Final (set) ER valid?	y	n	y	y	n	n	n	n
ER corrected?	-	n	-	-	n	n	n	y@q4
No. uses delete	2	4	7	0	6	0	8	8
ER construct	5'41"	5'02"	10'13"	12'41"	13'05"	13'54"	12'00"	6'17"
Total (to q4 ans)	18'08"	14'24"	21'47"	22'12"	21'39"	26'45"	24'08"	15'27"
q1 response	e	e	e	a	e	e	a	a
q2 response	b	b	b	b	b	b	b	b
q3 response	e	d	e	e	d	c	e	d
q4 response	c	c	c	c	c	c	c	c
Score total/4	3	2	3	4	2	2	4	3

Group	*No feedback*							
Subject	2	4	6	8	10	12	14	16
Re-order?								y
Solution focus	stem	stem	stem	stem	stem	stem	stem	stem/qn
First sent. rep'd	1	1	1	1	1	1	1	1
Tables not advised	-	-	-	-	-	-	-	-
Pls ind sent	-	-	-	-	-	-	-	-
Maybe weren't ... sent	-	-	-	-	-	-	-	-
Maybe problem ... rep	-	-	-	-	-	-	-	-
Try text instead ...	-	-	-	-	-	-	-	-
Rep. ahead arrows	2	0	2	3	7	2	2	2
No.ER switches	0	1	0	0	0	0	3	4
No. errors in ERs	2	0	0	1	3+	0	1	1
Type ER errors	$\exists for\forall BE$	- - -	- - -	revDF	revCD revDF $\exists for\forall BE$	- - -	revBE	$\neg for\exists DE$
All sent rep'd?	y	y	y	y	y	y	y	y
Final (set) ER valid?	n	noset	y	n	n	y	n	n
ER corrected?	part	y-text		n	2 fails		y/n	n
No. uses delete	5	0	5	8	4	3	7	2
ER construct	7'52"	3'04"	4'33"	14'01"	6'40"	5'02"	16'10"	16'05"
Total (to q4 ans)	13'06"	15'16"	11'19"	24'16"	20'36"	16'36"	23'49"	26'57"
q1 response	e	d	a	a	a	a	a	a
q2 response	b	b	b	b	e	b	b	b
q3 response	e	e	e	e	d	a	e	e
q4 response	c	c	c	c	e	c	c	c
Score total/4	3	3	4	4	1	3	4	4

Other entries, such as $\exists for \forall$ indicate that the subject represented 'Some' via ellipse intersection instead of 'All' via an ellipse container–contained relation in their set diagram, *etc.*

Effect of *switchERII* system feedback upon performance The effects of *switchERII's* feedback were subtle. There were substantial effects attributable to features such as the 'represent ahead' arrows that will be discussed later in this chapter in the context of a detailed case study. However, the effects of the explicit system feedback messages (*e.g.* 'There may be a problem with the representation...') were not striking. There was no significant difference between feedback and no-feedback group subjects in terms of score (median score for feedback group subjects was 3 (out of 4), for no-feedback subjects it was 3.5). Neither did the groups differ significantly in terms of time spent on the problem (mean for 8 feedback subjects = 20.5 minutes, s.d.= 4.47; for 8 no-feedback subjects, mean = 19.0 minutes, s.d. = 5.83). However, as the median score levels suggest, there were several subjects in each group who scored at ceiling level in terms of correctly answered problem questions. The effects of feedback might be expected to manifest themselves in the performance of subjects in the middle range of performance. To examine this hypothesis, the data of subjects who performed in the middle range (2 or 3 out of 4 questions answered correctly[10]) were analysed.

There was a suggestion in the data that middle-range subjects in the feedback group tended to modify their ERs more frequently than subjects in the no–feedback group. The frequency of use of the graphical delete tool was taken as an index of the number of set diagram modifications (Table 7.2. The 6 middle-performing feedback subjects averaged[11] 5 uses of the delete tool during set diagram construction compared to 3.5 uses in the 3 middle-performing no-feedback group subjects. This difference was not statistically significant, however.

Middle-performing feedback group also tended to spend longer constructing their ERs than their no-feedback counterparts (mean time for ER construction = 9 minutes , s.d. = 3.90; versus 5.75, s.d.= 2.22). This is consistent with extra use of the delete tool

---

[10] *i.e.* excluding ceiling performing subjects (S7, S13,S6,S8,S14 and S16) and the one floor-level performing subject (S10)

[11] Median.

Table 7.3: Relationship between average (median) errors on interpretation task and analytical reasoning score level - all subjects.

Reasoning score out of 4	1	2	3	4
Interpret errors (median)	5	1	1	0.5
No. of subjects	1	3	6	6

and also with responding to system feedback.

Several subjects reported that they did not attend to the lower (system message) window or that they 'didn't notice' the system messages since they were concentrating on the problem solution. This issue is discussed further below in the context of a detailed case study.

Relationship between interpretation and performance on problem questions

Table 7.3 shows the relationship between subjects errors on the interpretation task (see Table 7.1) and score (correct answers out of 4) to the Poets problem questions for all subjects. The data suggest that question performance is remarkably resilient to a wide range of interpretation errors in the sense that for only one subject (S10) were errors of interpretation on the pre-test predictive of poor reasoning performance on the Poets problem. In particular, Gricean errors do not seem predictive of performance with self-constructed ERs.

These results support other recent findings. For example, decoupling of error patterns on interpretation tasks and tasks requiring inference have recently been noted in the domain of syllogistic reasoning by Newstead (1995) and Stenning, Yule & Cox (1996).

These results extend those findings - to the author's knowledge, this is the first study to have investigated the relationship between *graphical* interpretation task errors and subsequent performance when reasoning with *graphical*, self-constructed ERs.

Conversion errors The relationship between subjects' performance on ER interpretation tasks and their performance when reasoning or making inferences with ERs is far from clear. Newstead (1989) found that different measures of illicit conversion (*i.e.* sentential and graphical) failed to correlate in predicted ways. Conversion errors have been shown by Stenning, Yule & Cox (1996) to be a predictor of *generally* poor reasoning performance but not of illicit conversion during syllogistic reasoning[12] Subjects who converted on either the sentential interpretation task or the graphical interpretation task (or on both tasks) demonstrated significantly poorer performance on 'conversion susceptible' syllogistic reasoning problems. However, this significant difference was also found on 'conversion unsusceptible' problems.

To explain the decoupling, Newstead (1995) appeals to depth of processing differences (*i.e.* processing is assumed to be deeper in inference tasks and shallower in interpretation tasks) — hence fewer than predicted errors on inference. Stenning, Yule & Cox (1996) offer an explanation based on individual differences in inference strategies.

In the present study, the task was not one of inference on sententially presented three term syllogisms but one in which subjects were required to reason with quantifiers on a puzzle type problem. The in erpretation task and inference tasks differ in terms of abstraction/realism. The interpretation task is posed 'abstractly' (*e.g.* 'All As are Bs'), whereas the inference task is 'thematic' in Newstead's (1989) terms in that the terms refer to (hypothetical) poet-liking individuals (*e.g.* 'All those who enjoy the poetry of Browning also enjoy the poetry of Eliot'). Newstead (1989) , however, showed that there is little difference in interpretation performance attributable to statement realism or concreteness. Yule (1995) has shown the same for inference tasks. The fact that the two tasks differ along that variable, therefore, does not compromise the Newstead Euler's circle interpretation task as a pre-test for the poets problem.

In the present study, subjects were instructed to begin their solutions by using Euler's Circles. The cognitive and semantic properties of diagrams, combined with the effects of representation externalisation and active construction, make reasoning with graphical ERs a profoundly different situation from that of interpreting presented representations. It is not surprising, therefore, that performance differs under the two very different conditions. However, the literature has not tended to distinguish between types of graphical reasoning task. These results indicate that the processes of *construction* and *externalisation* make reasoning with ERs under those conditions a profoundly different cognitive task from tasks that merely require the *interpretation* of

---

[12] Major and minor premises presented sententially, subject writes conclusion. Stenning, Yule & Cox used all 64 syllogisms.

ERs.

These results also demonstrate that findings in the diagrammatic reasoning literature that are based solely upon interpretative tasks are severely limited in their generality.

Reversal errors in Euler's circle ER construction The most common error in graphical ER construction (reversals), seemed to be more reliably associated with *conversion* and *island* responses on the pre-test than with Gricean errors. Both subjects who showed conversion errors on the pre-test manifested reversal errors during ER construction. This was to be expected in these subjects since they regard 'All As are Bs' to be synonymous with 'All Bs are As'. Both subjects who gave 'island' responses on the pre-test also made reversal errors. Gricean errors, though, were more or less equally divided between reversing subjects (3 Gricean responders) and non-reversing subjects (2 Gricean responders). Newstead (1995) examined whether Gricean implicatures measured on interpretation tasks are reflected in syllogistic reasoning (*i.e.* inference task) performance (Newstead, 1995). He analysed existing data from the literature but also presented data from new experiments. Newstead (1995) concludes that the relationship is very poor. He argues that the deeper processing required by the logically more complex inference tasks causes subjects to make relatively fewer Gricean errors than on interpretation tasks. The data here too suggest that Gricean errors on interpretation tasks are poor predictors of performance in reasoning tasks requiring inference.

As Table 7.2 shows, half of the subjects (8) made reversal errors during ER construction and 7 subjects (discounting S4 who didn't produce a set diagram) did not show reversal errors.

Most reversals occurred in representations of the Browning/Eliot and Coleridge/Donne universal quantifier relations. Table 7.4 shows the construction sequence for the elements in each subject's set diagram. One striking feature of the data is the significantly higher number of draw (and erase/redraw) sequences of activity in the construction activity of subjects who made reversal errors. The mean number of element-drawing events for the 8 reversing subjects was 11.75 (sd=4.68, n=8) and the mean for non-reversing subjects was 6.71 (sd=1.80, n=8). This difference was statistically significant ($t = -2.67, df = 13, p < .01$). Hence, it seems that subjects who make reversal errors also tend to delete and redraw significantly more elements of their representation than their non-reversing counterparts. This finding could not be attributed to a higher prevalence of switching out of the diagrammatic ER environment by non-reversing subjects since more of the reversers were switchers (4 subjects) than was the case for the non-reversing subjects (3 subjects). Further analyses revealed that subjects who made reversal errors spent significantly more time on their solutions (mean time from start to answering final question = 22 minutes versus 17 minutes for non-reversers; $t = -2.11, df = 13, p < .05$).

There were two particularly surprising results, however. The first was the failure of system feedback to prevent reversal errors or cause subjects to modify their ERs. Five of the 8 feedback group subjects made reversal errors during construction, despite system feedback messages indicating that there may be a problem with their representation. Several feedback group subjects remarked in the post-session interview that they did not see or attend to system messages in the lower window since they were concentrating on the upper problem, question and representation windows. Subject 7, for example, when asked by the experimenter in the post-session interview about the system message, replied '(I)...couldn't tell you what they said'. Subject 13 reported that he 'didn't notice' the system messages. Future versions of the *switchERII* system may need to incorporate more prominent system message displays.

The second surprising result was based on a comparison of the problem question scores of reversing and non-reversing subjects. A Mann-Whitney 'U' test between the question scores (out of 4) of subjects in the 2 groups was not significant.

Hence it seems that serious errors in ER construction are associated with a.) significantly more ER activity in terms of the creation and deletion of ER elements, b.) conversion and island errors of interpretation and c.) significantly longer ER construction times and total time spent on the problem.

Ultimately, though, these factors do not seem to affect performance in terms of responding to the problem's questions and the reason for this cannot be attributed to switching behaviour.

Graphical reversal and algebraic reversal

To recapitulate, graphical reversal was a common error in set-diagram (Eulers' circle) construction. Seven of the 16 subjects made at least one reversal error (Tables 7.2, 7.4).

As mentioned in Chapter 3, in the discussion of the effects of linguistic factors upon comprehension, reversals are common 'correspondence' errors (Greeno, 1977) that are observed when subjects translate word problems into algebraic formulae. A reversal error is defined as a tendency to match word order in the problem to the entity order in the representation of the problem's information. The well-known 'students and professors' types of problem provide a good example. Wollman (1983) reports that one in three college students produce the erroneous algebraic representation 6S=P for the sentence 'There are 6 times as many students as professors'. Clement, Lochhead & Monk (1981) found that 37% of engineering students and 57% of non-science students made this type of error. Kaput (1987) argues that the high error rate in word-problem-to-algebra translations (mostly of the 6S=P variety) is due to natural language overriding the rules of algebraic syntax and reference.

The screenrecording data logged in the current study permitted Kaput's (1987) translation hypothesis to be examined in the case of *graphical* reversal. The order in which subjects added the elements of their set-diagrams was examined in order to see if construction order parallelled the order in which entities occurred in the problem. Screen-recorded construction sequences were coded for each subject at the single graphical element level of granularity. The data are presented in Table 7.4. The coding notation employed is explained in the caption to Table 7.4.

Using the data in Table 7.4, the order of *first appearance* of graphic elements (*e.g.* labeled circles) was identified for each subject. The elements of an error-free, problem ordered[13] set diagram ER would appear in the order Browning, Eliot, Coleridge, Auden, Donne, Frost (BECADF). It was possible to compute, for each subject, the number of times that their element sequence departed from the 'ideal' sequence. Taking the data

---

[13] One in which the order of construction of elements of the diagram corresponds to the order in which the entities are ordered in the problem stem (assuming the stem sentences have not been re-ordered).

from table 7.4, and discounting repeated elements, the following sequences are obtained for each subject: S1 E,C,A,B,D,F (0); S2 B,E,C,D,A,F (2); S3 B,E,D,A,C (3); S4 - (excluded); S5 E,C,D,A,B,F (4); S6 B,E,C,A,D,F (0); S7 E,B,C (5); S8 B,E,A,C,D,F (2); S9 B,E,A,D,F,C (4); S10 B,E,C,D,F,A (3); S11 B,E,C,D (3); S12 B,E,C,A,D,F (0); S13 B,E,C,A,D,F (0); S14 B,E,C,A,D,F (0); S15 B,E,C,D,A,F (2); S16 B,E,C,A,D,F (0).

Computing the number of elements that were 'out of place' compared to a 'pure' problem ordering revealed a significant relationship between the number of graphical elements 'out of place' (shown in brackets for each subject following each sequence) and problem question score. For subjects who re-ordered the problem sentences prior to constructing their ER, the number of elements out of place metric was computed in relation to the re-ordered sequence. For subject 1 the revised order was ECABEDF, for subject 7 it was BECDFA, the procedure was not relevant for subject 16 who re-ordered after set diagram construction.

The median problem question score of 9 subjects who had 0,1 or 2 elements of their ER 'out of place' was 4 (out of 4), and that of 6 subjects who had 3 or more graphical elements 'out of place' was 2. The difference was significant when tested using the non-parametric Wilcoxon/Mann-Whitney test, $z(corrected) = 2.30, p < .03$ (Siegel & Castellan, 1988).

Moreover, the rank-order correlation between the number of ER elements out of place and the number of errors in the subject's ER was large and highly significant[14] (Spearman's $\rho = .82, n = 14, p < .005$).

It therefore seems that problem ordering of graphical elements during translation from word problem to diagrammatic representation is beneficial in terms of accurate ER construction and question problem solving performance. This is in contrast to translation from word problem to algebraic representations where problem ordering is associated with translation error.

A major shortcoming of the '6S=P' type of translation studies is that none have exam-

---

[14] Two subjects were excluded from the correlational analysis - S4 did not produce a graphical ER and S7 was dropped because her data was highly idiosyncratic - an outlier who manifested 5 graphical elements out-of-place but made no errors during ER construction.

ined the effect of student-professor reversal errors upon actual algebraic *reasoning*, as opposed to simple translation from language to (often ill-formed) formulae. 'Students and professors' problems are a kind of interpretation task. As we have seen in data from the *switchERII* study, interpretation task performance and reasoning task performance are often radically decoupled. Later in this section, data will be presented that relate the validity of a subject's ER (*i.e.* it's well-formedness), to answer correctness and the extent to which responses are commensurate with a direct read-off from that ER. It will be seen that even flawed ERs can be useful aids to reasoning. As long as the semantics are internally consistent for the user, and the representation is not used for communication with other individuals, idiosyncratic, personal, formalisms can function adequately. This is an interesting phenomenon which warrants further research in the domains of algebraic translation and graphical reasoning.

Validity of ERs, answer correctness and extent to which responses are commensurate with read-off from ER. The dynamic screenrecordings afforded a detailed analysis of the relationship between the validity or correctness of subjects' ERs and the extent to which their responses to problem questions were commensurate with read-off from their ERs. An analysis of the relationship between those factors for each of the four problem questions was beyond the scope of this thesis. Consequently, it was decided to perform the analysis for one of the four problem questions only. An item analysis formed the basis of the decision to select question 1 for this in-depth analysis.

An item analysis[15] revealed that, of the four questions posed by the poets problem, question 1 was the most sensitive discriminator. Question 1 divided the 16 subjects into two fairly equal groups consisting of 9 correct responders and 7 incorrect responders. Responses to question 1, therefore, were chosen to investigate whether subjects' answers were necessarily commensurate with direct read-off from their ERs – an issue that was introduced and discussed in the context of studies 1 and 2 (workscratching analysis and *switchERI* study).

The screenrecording data revealed that all 9 correctly responding subjects gave answers that were commensurate with direct read-off from their ERs *whether or not their ERs*

---

[15] Examination of extent to which question discriminates between subjects.

Table 7.4: Sequence of graphical ER construction by subjects in FB and NFB groups (*n.b.*$\neg w$=disjoint with; $\exists w$= intersects with). Note that this notation is not formally complete (*e.g.* for distinguishing between varying cases of 3 ellipse intersection) but it is sufficient to unambiguously capture all observed cases in this study.

Subj.	Grp	EC construction sequence
S1	FB	$E; C\neg wE; A\exists wE, C; BinsideE, B\exists wA;$ $DoutsideC, A, B, D\exists wE; FoutsideD, F\exists wE$
S3	FB	$B; EoutsideB; D\neg wE, B; A\exists wE, A\neg wB;$ $Ddeleted; Dredrawn, D\exists wE, A, D\neg wB; Adeleted; C\neg E, B, D;$ (incomplete)
S5	FB	$E; C\neg wE; DinsideC; AinsideE; Adeleted; BoutsideE, B\neg wC;$ $A\exists wB, E, C, D; Adeleted; AinsideB; A\exists wE; Ddeleted; DredrawninsideC;$ $FoutsideD, FinsideC; Bdeleted; BredrawninsideE, B\neg wA; Cdeleted;$ $CredrawninsideD; Fdeleted; Ddeleted; Adeleted; Aredrawn\exists wE; A\neg wB, C$ $DredrawnaroundC, D\exists wE, D\neg wB, A;$ $FredrawnaroundD, F\exists wE, F\neg wB, A;$ switch to text q1, switch back to EC q3
S7	FB	$E; BinsideE; C\neg wE, B;$ switch to text...
S9	FB	$B; EinsideB; C\neg B; A\exists wB, E; DinsideC; Ddeleted;$ D redrawn inside C;$F\exists wC, D; ansq1\&2;$ all deleted, rebuilt at q3; $B; F\exists wB; C\exists wB; A\exists wB, F; D\exists wC$
S11	FB	$B; EinsideB; C\exists wB, C\neg wE;$ $DinsideC, D\neg wB;$ switch to text
S13	FB	$B; EinsideB; C\exists wB, C\neg wE; A\exists wE, B, C; DoutsideC;$ $D\exists wB, A; FoutsideD, F\exists wA, B;$ switch to text, switch back to graphic q2; all deleted;$C; Cdeleted; B; F\exists wB; Fdeleted;$ $E\exists wB; C\neg wE, B;$ switch to text q3
S15	FB	$B; EoutsideB; C\neg wE; DoutsideC; D\neg wE; Ddeleted;$ D redrawn outside C,$D\exists wE; A\exists wE; FinsideD, F\neg wE, C;$ (corrects F at q4);$deletesF, D; redrawDoutsideC, D\exists wE;$ $D\neg wB;$ redraws F outside D, $F\exists wE, F\neg wB$
S2	NFB	$B; E\exists wB; C\neg wE, B; DinsideC; Ddeleted; DinsideC;$ $A\exists wC, D, E; Adeleted; Ddeleted;$ $DoutsideC, D\exists wE, D\neg wB; A\exists wE, A\neg wB, C, D;$ $FoutsideD, F\exists wE, F\neg wB, A$
S4	NFB	selects graphic ER environment but no construction activity - switches to text
S6	NFB	$B; EoutsideB; C\neg wB, E; A\exists wE, A\neg wB, C;$ $DoutsideC, D\neg wE, A; Ddeleted; D\exists wC, D\neg wA, E;$ $Ddeleted; Dredrawn\exists wE, D\neg wC, B, A; Cdeleted;$ $CinsideD, C\neg wE, A, B; FoutsideF, F\exists wE; F\neg wB, A$
S8	NFB	$B; EinsideB; deletesE; A\exists wB; deletesA; EoutsideB;$ $C\neg wE, B; A\exists wE, A\neg wB, C; DoutsideC, D\neg wE, A;$ $FoutsideD, F\neg wE, A; deletesD; D\exists wF, C, E, D\neg wB, A;$ $deletesC, F, D; D\exists wE, D\neg wB, A; C\exists wD, C\neg wA, E, B;$ $Cdeleted; CinsideD, C\neg wE, A, B; FinsideD, F\neg wC, E, A, B$
S10	NFB	$B; EoutsideB; CoutsideE; DoutsideC; FoutsideD; A\exists wB, AinsideE, C, D, E, F;$ $Cdeleted; Fdeleted; CinsideD, CoutsideE, A, B; FoutsideD, C, E, B, A;$ $alldeleted; B; EinsideB; CoutsideB, E; D\exists wB, DoutsideE, DinsideC;$ $AinsideE, B, D, C; FinsideC, FoutsideD, B, E, A; deletesall; B; E\exists wB;$ $C\neg wB, E; A\exists wE, B, A\neg wC; DinsideC; FinsideD$
S12	NFB	$B; EoutsideB; C\neg wE, B; A\exists wE, A\neg wB, C;$ $DoutsideC, D\neg wE, A, B; F\exists wE, D, F\neg wB, C, A;$ $Fdeleted; FoutsideD, C, F\exists wE, F\neg wA, B$
S14	NFB	$B; EinsideB; C\neg B, E; A\exists wE, B, C; DoutsideC, D\exists wA, D\neg wB, E;$ $Ddeleted; DoutsideC, D\exists wE, A, B; Ddeleted;$ $DoutsideC, A, D\exists wE, B; Adeleted; A\exists wE, B, AinsideD, A\neg C;$ $FoutsideD, A, C, F\exists wB, E;$switch to text; switch to table; switch back to diagram
S16	NFB	$B; EoutsideB; C\neg wE, B; A\exists wE, A\neg B, C;$ $Cdeleted; Credrawn\neg wB, E, A; DoutsideC, D\neg wA, E, B;$ $FoutsideD, C, F\neg wA, E, B;$ switch to text(q1); switch back to diagram(q4)

were valid models of the problem information. Five subjects read correct answers from error-free ERs. Four subjects read-off (correct) answers from ERs containing at least one error.

This result is surprising. As in the *switchERI* study, it seems that once again reading off from ones ER, even if there are errors in the ER, is a better strategy than abandoning the ER.

Six of the 7 remaining subjects gave (incorrect) responses to question 1 that were *inconsistent* with a response based on read-off from their ERs again, *whether or not their ER was a valid model of the problem information.* Four subjects gave wrong answers to question 1 despite constructing a valid ER, two gave wrong answers that were inconsistent with direct read-off from their (invalid) ER. *Only one subject gave an incorrect response to question 1 that was commensurate with a direct read-off from his (incorrect) ER.*[16]

The screenrecording data showed that 6 of the 9 subjects who answered question 1 correctly used set diagrams at the time of responding. Three used textual representations. Four of the 7 subjects who gave incorrect responses to question 1 used set diagrams. It is interesting to note, however, that three of the four subjects whose (incorrect) responses were not consistent with their (valid) ERs used textual representations. The lack of specificity of the sentential modality may have caused read-off errors in those cases. Two subjects whose set diagram ERs were error-ridden and whose (incorrect) responses were inconsistent with those (erroneous) ERs, appear therefore to have abandoned their flawed ERs in favour of internal strategies.

Switching

Subjects in the feedback and no-feedback groups did not differ in the extent to which they switched ERs during reasoning. Four feedback subjects switched at least once and three no-feedback subjects switched (Table 7.2).

Switchers tended to score more highly than non-switchers on problem questions (me-

---

[16] Subject 9 who made a reversal error in his set diagram when representing the relationship between Coleridge and Donne likers.

dian score=4 versus median score of 3). The difference, however, was not significant when tested using Mann-Whitney's 'U'.

Since switching involves the use of multiple representations, it is reasonable to expect that switchers might take longer to solve the problem than non-switchers. The groups differed significantly in terms of solution times — switchers, mean=21.9 mins, s.d.=5.34 mins; non switchers= 17.2 minutes, s.d. = 4.4 mins, t=1.91, df=14, $p < .05$. This finding replicates the results of the *switchERI* study.

Interpretation errors and switching There was a suggestion in the data that ER switching behaviour in subjects who made relatively few errors of interpretation represented a fundamentally different strategy from switching in subjects who made more interpretation errors. The data are presented in table 7.5 — 'good' performance on the interpretation pre-test is defined as 0 or 1 error and 'poor' performance is defined as 2 or more errors (see table 7.1).

A Mann-Whitney 'U' test between the problem question scores of switching and non-switching subjects who performed well on the interpretation task approached significance (p=.057). The median score of switchers in that group was 4 and that of non-switching subjects was 3. The difference between solution times was significant (Mann-Whitney[17] $p < .02$)).

The difference between the scores of switching and non-switching subjects who performed poorly on the interpretation test was not significant (median score=3 for both groups). The difference between solution times for those subjects (Table 7.5) was also not significant.

These results suggest that switching ERs during reasoning might represent very different processes in the two groups of subjects shown in Table 7.5. For those subjects who comprehend the semantics of set-diagrams and who made relatively few errors of interpretation on the pre-test, switching is associated with better performance and longer solution times. Those subjects switch judiciously, and exploit graphical and non-graphical representations for their expressive properties, in response to changing

---

[17] Used instead of 't' due to non-normal distribution and small N.

Table 7.5: Relationship between interpretation error level on pre-test and switching during problem solving, in terms of median problem question score and mean solution time.

switcher?	yes	no
interpretation test performance		
Good		
Subjects Median score (out of 4) Mean time (mins)	S5,S13,S14,S16 4 24.25	S1,S3,S9,S2,S6,S12 3 15.83
Poor		
Subjects Median score (out of 4) Mean time (mins)	S7,S11,S4 3 21.33	S15,S8,S10 3 20.00

demands of the task. They may be said to be engaging in true heterogeneous reasoning.

In contrast, the switching behaviour of subjects who performed poorly on the interpretation pre-test represents an attempt to escape from a representation whose formalism they do not understand. Switching under those circumstances is not associated with improved performance or changes to solution time.

Switching case study

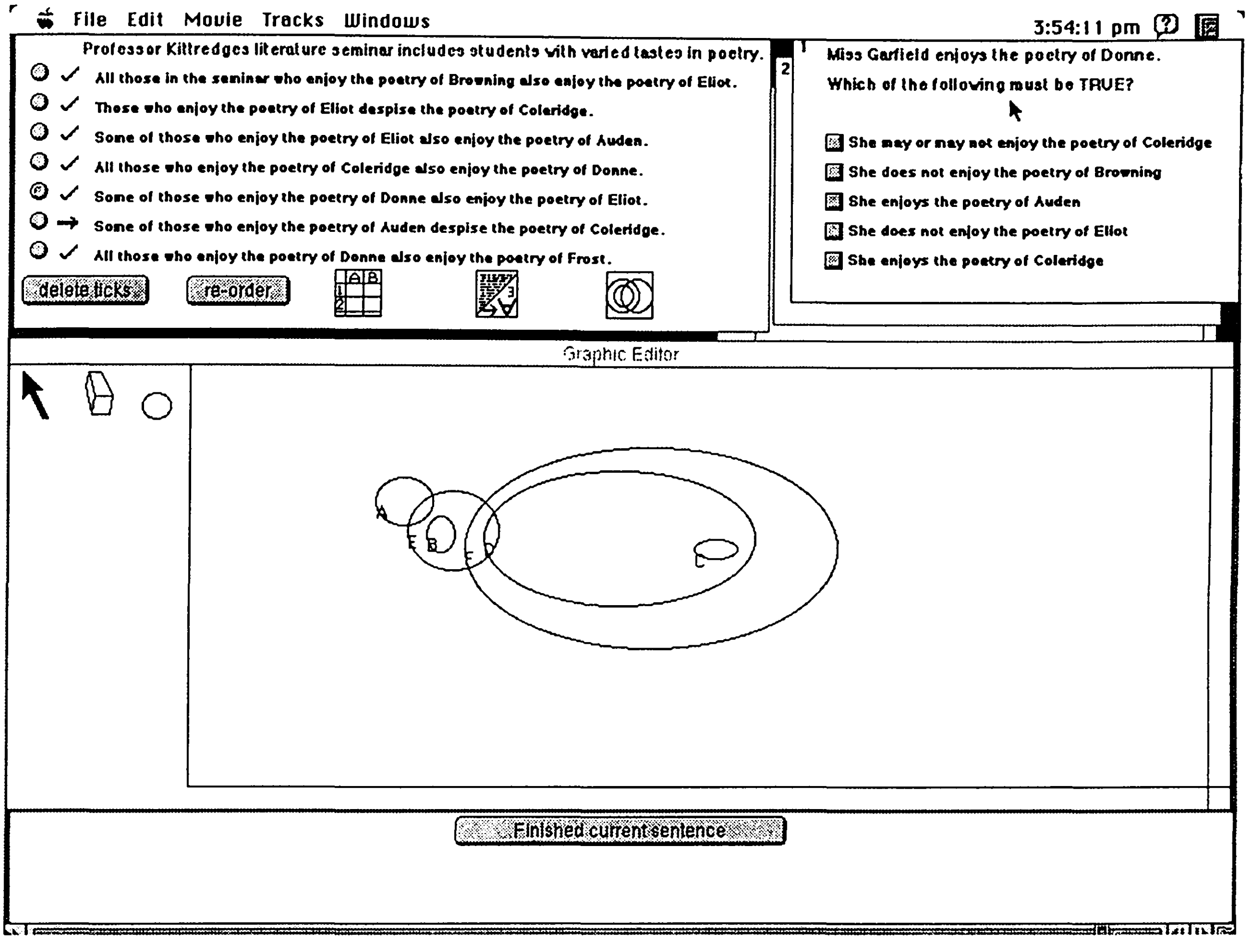
Subject 5 will be presented as a case study example of 'judicious' switching. S5 performed well on the interpretation pre-test task, demonstrating no conversion, island, Gricean or other errors.

The screenrecording reveals that he selected the ER construction environment 10 seconds after the problem was presented. He read the problem stem information for 1 minute 24 seconds before beginning ER construction activity. Then, as instructed, he began his solution by constructing a set diagram. Table 7.4 shows his ER construction sequence. He did not choose to re-order the problem information. Unusually, he began by representing sentence 2 of the problem first (he was the only subject to do so). Sentence 2 states 'Those who enjoy the poetry of Eliot despise the poetry of Coleridge.' Subject 5 drew two disjoint ellipses and labeled the one on the left 'E' and the one on the right 'C'. Next he chose to work on sentence 4 — 'All those who enjoy the poetry of Coleridge also enjoy the poetry of Donne.' He drew a smaller circle 'D' inside 'E', thereby making a reversal error. Clicking the 'Finished current sentence' button then caused the system to respond with the message 'There may be a problem with the representation...'. He did not act on this advice. Sentence 1 was represented next — 'All those in the seminar who enjoy the poetry of Browning also enjoy the poetry of Eliot'. He drew a large circle around the circle 'E' and labeled it 'B'. Circle 'B' was disjoint with circles 'C' and 'E'.

The system now placed a 'represent-ahead' arrow next to sentence 5 ('Some of those who enjoy the poetry of Donne also enjoy the poetry of Eliot'). In the then current representation, though, circles 'D' and 'E' were disjoint. Sentence 3 was chosen next ('Some of those who enjoy the poetry of Eliot also enjoy the poetry of Auden'). This time S5 drew a large ellipse 'A' that intersected with all 4 circles (E,B,C,D). He then erased 'A' and redrew it inside 'B' but intersecting with 'E'. The system responded by displaying a second 'represent-ahead' arrow alongside sentence 6 ('Some of those who enjoy the poetry of Auden despise the poetry of Coleridge') which indicated that S5 had made an implicit commitment to representing the relationship between 'A' and 'C'. Actually, 'A' and 'C' were disjoint in his diagram (*i.e.* a valid model of sentence 6). Next, circle 'D' was deleted from inside circle 'C'. It was redrawn in the original configuration almost immediately. S5 now attempted to represent the final sentence of the problem 'All those who enjoy the poetry of Donne also enjoy the poetry of Frost'. He seemed uncertain about where to place the 'F' circle, but eventually drew it around 'D' (and contained by 'C'). He then erased circle 'B' and redrew it inside 'E' but disjoint with 'A'. Circle 'C' was deleted and redrawn inside 'D' (and therefore also inside 'F').

There now followed several minutes of checking in which S5 compared his representation against the two sentences which had 'represent ahead' arrows adjacent to them (*i.e.* the sentences 'Some of those who enjoy the poetry of Donne also enjoy the poetry of Eliot' and 'Some of those who enjoy the poetry of Auden despise the poetry of Coleridge'. In its then current state, his set diagram was inconsistent with the first sentence since 'D' and 'E' were disjoint, but it was consistent with the second sentence assuming 'some' was interpreted as 'all' (circles 'A' and 'C' were disjoint).

Circle 'F' was deleted, circle 'D' was deleted, circle 'A' was deleted, 'A' was redrawn again intersecting with 'E' but with the intersection region in upper left of 'E' instead of



Professor Kittredges 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

Figure 7.7: Subject 5's set diagram representation immediately prior to switching to text.

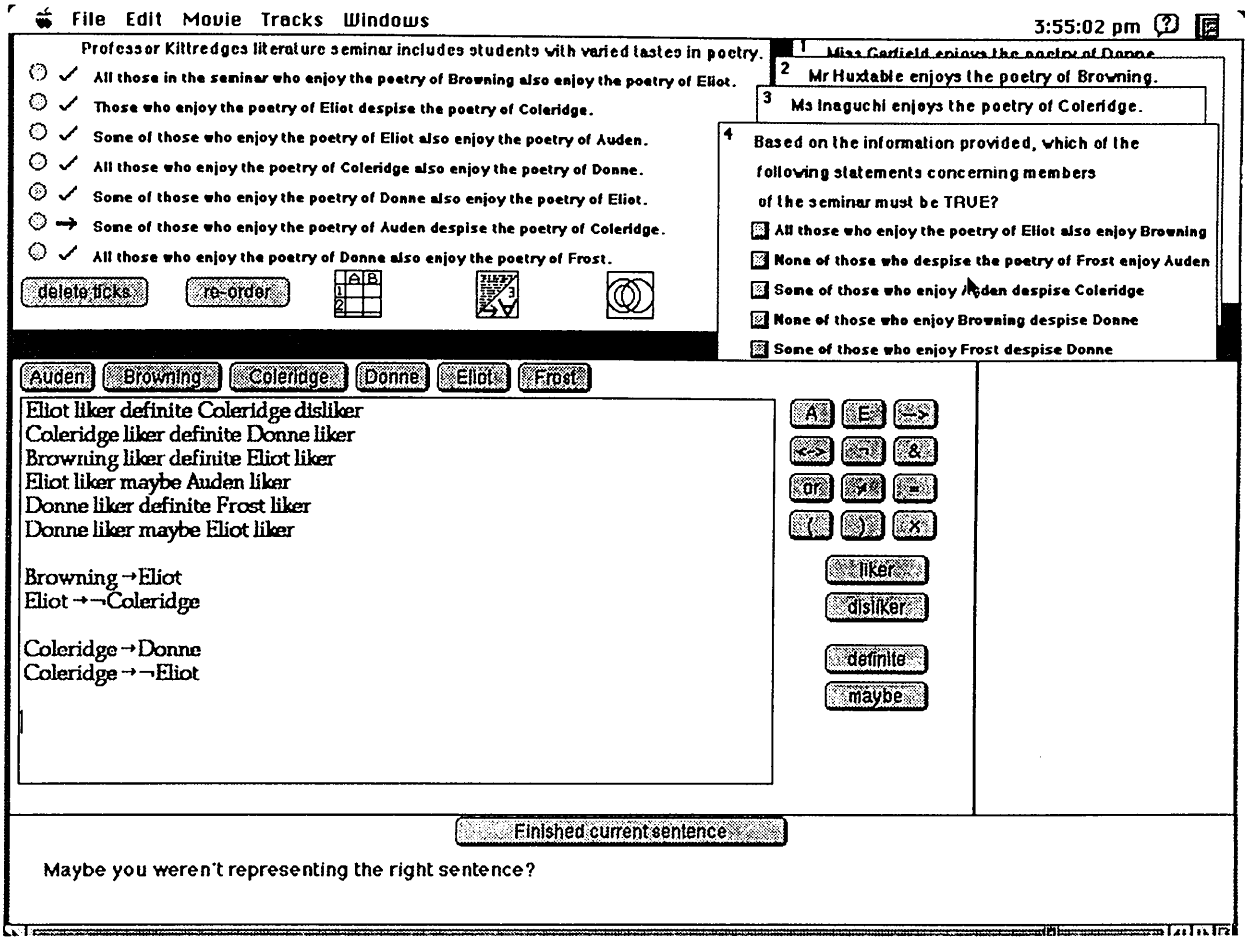
lower right where it was before — 'A' was still disjoint with 'B'. Next, 'D' was redrawn, this time intersecting with 'E' and totally enclosing 'C'. 'F' was redrawn around 'D' – note that both 'D' and 'F' intersected 'E' but did not intersect 'B' within 'E' or 'A'. The ER was now a valid model of the problem information. S5 depresses the 'Finished current sentence button' and the system responds without a feedback message (*i.e.* the model is valid.

Subject 5's ER at this point is shown in Figure 7.7.

Subject 5's attention was now turned to question 1 ('Miss Garfield enjoys the poetry of Donne, which of the following must be TRUE' (5 answer choices)). Throughout the recording S5 moved the mouse cursor around as his gaze and attention shifted from problem sentences, to ER, to question...Hence it provided a reliable indication of where his attention was being directed. At this point, S5 switched to the text ER environment and read the system generated 'propositionalised', re-written, problem sentences. They read...'Eliot liker definite Coleridge disliker'; 'Coleridge liker definite Donne liker'; 'Browning liker definite Eliot liker'; 'Eliot liker maybe Auden liker'; 'Donne liker definite Frost liker';'Donne liker maybe Eliot liker' (see Figure 7.8). Twelve minutes had elapsed up to the point of switching. The subject read the system generated sentential representation and question 1 for about one minute. He then selected the 5th answer choice to question 1 'She enjoys the poetry of Coleridge'. This is incorrect - the correct response is the first answer option ('She may or may not enjoy the poetry of Coleridge').

Subject 5 now selected the second question window 'Mr Huxtable enjoys the poetry of Browning, he may also enjoy any of the following poets, EXCEPT...' (5 answer choices). He then read the system generated propositions. For question 2, S5 added his own pseudo-logic notation to the text window, below the system generated propositions. He typed 'Browning → Eliot' , which was merely re-writing the system's propositional representation of the original problem sentence. S5 then chose the second response option for question 2 'Coleridge' (correct). At this point the subject had been working on the problem for approximately 15 minutes.

S5 now selected the third question 'Ms Iganuchi enjoys the poetry of Coleridge. Which of the following must be FALSE?' (5 response choices). Instead of proceeding with question 3, however, the subject returned to question 2 (presumably to check his answer) and added 'Eliot → $\neg$ Coleridge' below 'Browning → Eliot'. Having done this, S5 did not change his answer to question 2, but selected question 3 again. He added 'Coleridge → Donne' to the text window, read the system generated proposition 'Donne liker definite Frost liker', he added 'Donne → Frost', then deleted 'Donne → Frost', highlighted references to Donne in the system-generated propositions, highlighted 'Eliot liker definite Coleridge disliker', typed 'Coleridge → $\neg$ Eliot' and re-read his own and the system generated text for about 30 seconds. It is interesting to note that S5 used



Professor Kiltredges 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

2 Mr Huxtable enjoys the poetry of Browning.

3 Ms Inaguchi enjoys the poetry of Coleridge.

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

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

Eliot liker definite Coleridge disliker
Coleridge liker definite Donne liker
Browning liker definite Eliot liker
Eliot liker maybe Auden liker
Donne liker definite Frost liker
Donne liker maybe Eliot liker

Browning → Eliot
Eliot → ¬Coleridge

Coleridge → Donne
Coleridge → ¬Eliot

Maybe you weren't representing the right sentence?

Figure 7.8: Subject 5's sentential representation immediately prior to switching back to his set diagram at question 4.

the representation exclusively in his solutions to the questions — he did not refer to the window containing the original problem statement. He then selected answer option 5 to question 3 'She may enjoy the poetry of Eliot' (correct). Elapsed time was now approximately 19 minutes.

Subject 5's textual ER at this point is shown in Figure 7.8.

Question 4, the final question, was now selected: 'Based on the information provided, which of the following statements concerning members of the seminar must be TRUE?' Note that the final question differed from the others in that it did not propose a hypothetical person who liked a particular poet, instead it posed a different task - one of assessing the truth value of each response option assertion. The subject spent about one minute reading question 3 and then switched back to his original set-diagram representation. At this point approximately 20 minutes had elapsed.

S5 started to work down the 4th question's 5 response options, the cursor alternating between the question window and the set-diagram, mirroring S5's shifts of attention. When he reached the third response option ('Some of those who enjoy Auden despise Coleridge'), his attention shifted to the problem stem window, he checked and re-checked, noticing that the problem stem contained exactly the same statement. He almost decides to select that response option but moved on to check response options 4 and 5 against the diagram. Neither option 4 or 5 were selected as the answer, however, and S5 selected 'Some of those who enjoy Auden despise Coleridge' as the answer (correct). Total time spent on the problem was 21 minutes and 47 seconds. Subject 5 answered 3 out of 4 questions correctly.

The purpose of presenting subject 5's reasoning protocol in so much detail is to attempt to convey the richness of the screenrecorded data and to discuss the role of *switchERII's* feedback. The screenrecordings reveal a great deal about analytical reasoning with ERs. For example, when answering questions 1 to 3, subject 5 preferred a mixture of the system-generated sentences and his own pseudo-logical representations. On question 4, however, he used his self-constructed set-diagram in combination with the problem statements in the problem stem window. Note that in both cases, a degree of heterogeneous reasoning occurred. It involved two types of textual representation (propositional sentences and pseudo-logic) in the first case and two modes of representation (set diagram and original problem sentences) in the second case. For question 4, the set diagram permitted faster checking of statements against the model than the sentential representation would have.

At the end of the session, the experimenter asked S5 *inter alia*:

> Did any of the system's feedback help you?

S5 replied:

> ...yes definitely, 'coz I was misreading the 'All', the universal quantifier...I was actually interpreting it the wrong way round and that was system feedback that reminded...that showed me that...

However, the screenrecording protocol shows that S5's self-corrected errors in set-diagram construction was not prompted by the overt system message ('There may be a problem with the representation...') but by S5's reflection upon the significance of the 'represent-ahead' arrows. The represent-ahead feedback from the system indicates to the subject that their set-diagram represents information in sentences that they have not yet reached in their progress through the problem. The subject's attention is drawn to those sentences and this prompts the subject to check his model against them. This process resulted in subject 5 spotting his reversal errors. Hence, the 'represent-ahead' arrows functioned as a 'flag' (Corbett & Anderson, 1990; Schooler & Anderson, 1990), serving to draw attention to a state of affairs but not feeding back much information content or 'knowledge of results'. This type of feedback is preferred by some students since it is less disruptive when they under heavy cognitive load. As Schooler & Anderson (1990) write:

> ..., the processing of feedback ...(competes)... for limited cognitive resources. When a subject is provided with feedback, the feedback necessitates that they set new goals to process it. When they re-emerge from the feedback episode, the previous goals may have been lost, increasing the likelihood that the subject would rely on the feedback. In contrast, if the feedback processing were somewhat less disruptive, then they might return from the feedback episode with their goals intact. (p.707)

This example of the 'flag' function of the represent ahead arrows may help to explain why there were not larger differences in performance between the feedback and no-feedback conditions in study 3, since represent-ahead arrows were generated by *switchERII* for subjects in both groups. However, only feedback group subjects received overt system feedback of the 'There may be a problem with the representation..' kind. Data from the subject 5 case study suggest that when the subject is actively engaged in interpreting and discerning the significance of system feedback (as in the case of 'represent-ahead' arrows) is more effective in facilitating ER correction than overt system messages. The reflection and self-correction promoting effects of flag tutoring certainly warrant further study.

Discussion and conclusions

The *switchERII* data show that not all errors of ER *interpretation* predict performance on tasks in which subjects *construct and reason* with ERs. This finding points up the importance differences between reasoning with ERs under those two conditions - hitherto this distinction has not been made in the literature. Conversion and 'island' interpretation errors are particularly associated with reversal errors during construction.

Reversal errors during ER construction were very commonly observed and were sometimes associated with construction slips and sometimes associated with more profound semantic misconceptions.

The *switchERII* study replicated the *switchERI* finding that subjects who respond to questions in a way that is consistent with their ERs score better than subjects whose responses are not commensurate with their ER *whether or not the ER is a correct model of the information.* Abandoning ones ER and reasoning internally (as opposed to switching to a different kind of ER) is a poor strategy which results in poor performance. The cognitive load imposed by the poets problem is too great for solely internal reasoning (except for a few exceptional individuals). Considering all subjects together, switching to a different ER costs time, but preserves performance. However, two distinctly different kinds of switching can be identified within subjects. For subjects who understand the semantics of set diagrams, switching to a different ER during problem solving represents 'true' heterogeneous reasoning in which representations in both the graphical and sentential modalities are judiciously exploited for their expressive properties. Skilled switching is reflected in good performance (*i.e.* correctly answered problem questions). In contrast, switching by subjects whose understanding of set-diagram semantics is poor represents 'thrashing' — a less principled search for useful problem representations. The latter type of switching does not improve performance.

As specificity theory would predict, subjects tend to read off erroneous conclusions from valid sentential ERs more frequently than they do from graphical ones.

The poets puzzle is a difficult analytical reasoning problem which places a high cognitive load upon the subject at all phases of the solution process, especially during representation construction. During that phase of reasoning, the subject has no additional cognitive resources available for processing explicit error messages presented sententially. Other, less intrusive, less semantically content-laden forms of feedback, such as the 'represent ahead' arrows seemed to prompt more self-reflection and ER corrections than overt feedback messages. This finding is consistent with findings from 'flag' tutor experiments reported by Corbett & Anderson (1990) and Schooler & Anderson (1990).

Future work and development of *switchER* systems

The results suggest a range of improvements that can be considered for implementation on future versions of *switchER.*

An obvious improvement would be to broaden *switchERII's* ER repertoire to include the semantics of homomorphic representations such as plans and maps. This would entail programming the system to parse and comprehend graphical features such as arrays and object orderings. This would permit *switchERII* to intelligently interpret subject's diagrams for problems such as the office allocation problem. Another useful representational formalism would be network diagrams or acyclic graphs. There is a class of analytical reasoning problems that pose questions about family inheritance for which network diagrams would be useful.

Matrix ERs could also be usefully parsed — the system could use its representation of the problem information to detect the dimensions of data in the user's table. If the user constructs a table in which neither row or column headings correspond to the problem's salient dimension (*e.g.* prize ordering in the dogs and prizes problem), then system feedback could be provided.

It would be relatively straightforward for the system to diagnose and remediate ER construction mistakes such as reversal errors in real time, on-line. *SwitchERII's* feedback could easily be extended to include more detailed feedback of the form 'The relationship of Browning-likers and Eliot-likers seems to be reversed in your current representation.' The system could offer to auto-correct such reversal errors if the user wished.

Future versions could also make text and graphics concurrently available. The decision to permit only serial access to *switchERII's* ER environments was based on methodological requirements. Requiring subjects to switch between representations provided an unequivocal method of tracking their reasoning behaviour.

The current *switchERII* graphical ER environment does not support use of the identity diagram[18]. This facility could be added in the next version.

A more major extension would be to implement intelligent ER co-construction in which the semantics of the subjects graphical representation are preserved when the system builds a parallel representation in the sentential modality. In other words, the subjects original representation is replicated 'warts and all'.

There is also plenty of scope for *switchERII's* feedback to be improved. In *switchERII* all the system feedback messages indicate problems with the representation — perhaps subjects should also receive positive feedback, such as 'Your representation seems OK so far...'

There are also several issues regarding feedback that warrant further attention. The *switchERII* results suggest that the cognitive load of ER construction is too high to permit the user to read detailed system feedback messages. Flag feedback, in which the system signals a potential problem by means of a non-linguistic flag, might be more effective than detailed feedback messages.

The system could also monitor the extent to which the subject's responses to the problem questions are consistent with his or her representation. In the case of a valid model, the system could point out the inconsistency. In the case of an invalid ER, the system might ignore the inconsistency and assume that the subject had abandoned the ER and was reasoning internally.

---

[18] The diagram for 'All As are Bs' in which circles A and B completely overlap.

Chapter 8

General thesis discussion

Summary of what has been achieved

Using interactive learning environments to study the process and time-course of reasoning with ERs proved to be highly informative. The dynamic approach provided useful insights into cognitive processes at each stage of reasoning and offered numerous methodological advantages over the study of static, residual workscratchings.

The results emphasise that effective reasoning with ERs involves the interaction of at least 3 factors: (a) within-subject variables such as the subject's representational repertoire (prior knowledge) and representational modality preferences (cognitive style); (b) skill at overcoming a large variety of barriers to comprehension and an ability to discern the salient attributes and characteristics of different problem types and (c) an understanding of the semantic and cognitive properties of graphical and non-graphical representational systems coupled with an ability to match those properties to the problem's task demands.

These factors will be discussed further under 4 main headings: problem comprehension, ER selection and construction, individual differences, and constructivism.

Problem comprehension

The linguistic and structural features of analytical reasoning problems make comprehension difficult. The 'register' of the language in which the problems and questions are presented represents a major linguistic barrier to comprehension. A high degree of self-monitoring is required in order to avoid slips such as reversal errors. The manner in which problems are posed also violates several Gricean co-operative discourse maxims, most noticeably that of manner. Analytical reasoning problems tend to be obscurely expressed and information is not typically well-ordered from the problem solver's perspective. They are far from being 'maximally informative' — a general characteristic of puzzles, of course.

Analytical reasoning problems have a bearing on a significant range of issues in the classroom, especially in maths and science where students are expected to develop arguments of some sort. Because of their puzzle-like nature, analytical reasoning problems are, in many ways, an analogue of many tasks that are used in teaching. There is a sense in which some aspects of teaching consist of uncooperative discourse. Teachers do not 'give away' solutions to problems and often use exercises posed in puzzle (uncooperative) form (*e.g.* word algebra problems). Like word algebra problems, analytical reasoning problems are expressed in ways that 'real world' problems are expressed *i.e.* as verbal problems posed in English. In these respects analytical reasoning is an authentic educational domain.

Comprehension is also affected by a range of structural factors - for example, subjects have to discern the number and type of dimensions in the problem, extract implicit information and cope with negatively stated information.

Ideally, the reasoner needs to build up a full problem information context by taking time to read the questions as well as the information in the problem stem. Research in the physics domain (*e.g.* Chi, Feltovich & Glaser, 1981; Larkin, McDermott, Simon & Simon, 1980) has shown that experts tend, *inter alia*, to spend more time than novices on analysing and understanding problems, but produce faster solutions. The *switchER* subjects rarely did this - instead they tended to rush into representation selection. This may partly explain poor ER decisions such as discourse ordering the content of tables when efficient read-off from the point of view of question-answering would be better facilitated by ordering along a single salient dimension. Although, differently ordered tabular representations can be informationally equivalent, the nature of the questions posed will determine whether they are also computationally equivalent for the task of reading-off information. In analytical reasoning problems, the best way of determining the salient dimension is to read the questions as well as the problem stem, something that few students appear to do spontaneously. Most subjects in the *switchER* studies read the questions for the first time *following* ER construction, rather than before or during ER construction. Hence, for most subjects, ER construction proceeded in the absence of what Bernado & Okagaki (1994) have termed a full 'problem information context'. One reason for this might be that the cognitive load associated with overcoming the linguistic and structural barriers of the problem stem is so high that there is no spare capacity for question look-ahead. Future *switchER* systems might offer support to the reasoner in this respect by monitoring which parts of the problem and questions the subject has read and offering a suggestion to read the questions before selecting and constructing an ER.

Skilled analytical reasoners become test-wise and are quick to spot opportunities to prune the search space. The clearest example is provided by the office allocation problem - spotting determinate information such as 'Mrs Green is entitled to Office 5' and representing it first provides an 'anchor' to which related information can be efficiently linked. In that problem, for example, the only two non-smokers must flank office 5. Hence, the information about Mrs Green provides the basis for a major 'split into cases' and, in Amarel's (1968) terms, represents a 'critical point' in the problem.

One of the most important comprehension tasks confronting the reasoner is assessing the degree of indeterminacy in the problem. For the skilled problem solver, this factor heavily influences the choice of representational system, and the choice of a specific ER within the system. If more than one model of the information can be constructed, as in the case of the poets problem, then the problem is indeterminate. Generally, the weak expressiveness of graphical systems makes the linguistic modality (natural or logical language) a better choice than graphics for representing indeterminacy. However, if the subject has been taught a set-diagram formalism such as Euler's circles, then the property of specificity that diagrams share with internal (cognitive) representations makes them cognitively more tractable (Stenning & Oberlander, 1995). In addition, set diagrams, as with all graphical representations, permit subjects to use the bandwidth and computational efficiency of their visual-spatial cognitive subsystems. This permits (seemingly effortless) visuo-spatial judgments to be substituted for linguistically-based reasoning which may be more error-prone and difficult (*e.g.* Larkin & Simon, 1987).

ER selection & construction

As discussed in Chapter 3, the ER that a subject chooses depends, *inter alia*, upon:

- interpretation and comprehension of the problem's explicit and implicit information
- interpretation and comprehension of problem solving task posed
- an ability to recognise the salient features and characteristics of the problem
- a matching of the problem's salient features to an appropriate ER formalism
- an appropriate ER formalism being in the subject's repertoire
- the subjects cognitive style and/or tendency to use *external* representations

Graphical representations that possess abstraction-expressing properties, such as multiple diagrams or annotated diagrams, offer the 'best of both modalities' since a general rule, from the work reported here and elsewhere (Cox, Stenning & Oberlander, 1995), seems to be *select a representation with sufficient expressive power to capture the indeterminacy in the problem, but not more than is needed.* This rule is not symmetrical, however — generally speaking, selecting weakly expressive representations (graphics) for problems which require abstraction has more serious consequences in terms of reasoning performance than selecting expressive representations for determinate problems (Cox, Stenning & Oberlander, 1995).

Switching ERs

The questions associated with an analytical reasoning problem are typically quite varied in terms of their task demands. This means that the task requirement can change as problem solving progresses. For this reason, as the work reported here and by Cox, Stenning & Oberlander (1995) has shown, subjects' ability to select the ER recommended by the analytical reasoning 'crammer' does not predict reasoning performance very well. The skilled reasoner changes his or her ER as task demands change. Data from the three studies reported in this thesis show that subjects often utilise multiple representations in their solutions, either concurrently or serially via ER switching. Two distinctly different types of switching were observed. One kind ('thrashing') is associated with poorer performance, is impasse driven and reflects less comprehensive prior knowledge, inability to select an appropriate ER and hazy problem comprehension. Thrashing corresponds to a trial and error-based strategy for ER selection and may reflect a relatively impoverished ER repertoire (factor a. above) and/or changes in the subject's problem comprehension (factor b.).

Judicious switching, on the other hand, is opportunistic and associated with high levels of problem comprehension and skilled matching of ER properties to changing task demands. Judicious switching reflects expert behaviour in which changes in the task demands of the problem prompt the reasoner to switch representations. The reasoner skillfully matches the cognitive and semantic properties of the new representation to the new task requirements. Knowledge of a range of representations (factor a. above), a high level of problem comprehension (factor b.) and a good grasp of the ERs' representational formalisms (factor c.) are all required.

Constructivism

The Euler's circle interpretation task used in the *switchERII* study illustrated that performance under conditions in which subjects interpret prefabricated diagrams does not necessarily predict performance under conditions in which the subject reasons with an ER that they themselves have constructed. The reasons for the decoupling are complex. One part of the explanation is that subjects are under much greater cognitive load when constructing ERs than when interpreting them. During ER construction they attend to the task very closely and hence may be less prone to make slips.

Another part of the explanation is that the process of externalising cognition confronts the reasoner with his or her misconceptions or ambiguities in ways that interpreting pre-fabricated diagrams does not. As far as the author is aware, this distinction has not, hitherto, been addressed in the literature.

There are strong reasons to believe that subjects reason better with representations that they construct than with provided, 'prefabricated', ERs. To the author's knowledge, only two studies in the literature have directly demonstrated the advantages of active construction.

In the syllogistic reasoning domain, the process of constructing an external representation (Euler's circles) has been shown to result in richer learning outcomes than the use of pre-fabricated diagrams (Grossen & Carnine, 1990). Hesse et al. (1995) have shown that subjects who are permitted to manipulate diagrams perform better than subjects who use static representations.

In explaining the effectiveness of constructivist approaches, there is a need for explanations to go beyond merely attributing the superiority of 'a required diagram-drawing response' to 'deeper processing' (Grossen & Carnine, 1990, pp. 179-180). In the author's view, part of the explanation lies in the effects of constructing ERs in terms of externalising one's cognition. The effects of externalisation can be partitioned into effects akin to the 'self-explanation effect' and effects that derive from the re-presentation of stimuli to oneself via externalisation.

Roles of externalisation

Externalisation and the self-explanation effect As suggested in Chapter 1, the self-explanation effect may operate during translation *across* modalities (*e.g.* from verbal to diagrammatic or vice versa). The modality in which an ER is constructed (*i.e.* linguistic or graphical) may affect the operation of the processes that are assumed by Chi *et al.* (1989) to underly the self-explanation effect. In the case of constructing a diagram for example, the semantic properties of graphics 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, 1995). They suggest that graphical representations compel certain classes of information to be represented and that these representations are less expressive of abstraction than sentential representations. 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 (Stenning & Oberlander, 1995) provides grounds for predicting that the process of translating information *from* a linguistic representation such as natural language or logic *to* a graphical representation might be more effective than translation from one representation to another within the same modality.

Limited support for an account of this kind is provided by Lewis (1989), who 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. Studies of self-explanation during problem solving with ERs may help to shed light on to the mechanisms of both representation externalisation and self-explanation. Further support is provided by Katz & Anzai (1991)[1] who took protocols from an undergraduate as she learned to represent and solve vector arithmetic problems. They report that as her expertise increased, the role played by the vector diagrams she produced changed. Katz & Anzai (1991) report that as her expertise increased, her diagrams helped her to recognise useful calculations that she may not have otherwise discovered.

Externalisation - turning one's representations into stimuli As reviewed in Chapter 1, Reisberg (1987) sees the process of constructing an ER as a procedure for 'widening the context of understanding' and 'turning ones representations into stimuli'. ER selection and construction consist of dynamic iterations and interactions between external and mental models. This was illustrated by Hinton's (1979) cube task which demonstrates that often subjects' mental images are not fully elaborated and that attempts to mentally re-construct the image of an object (*e.g.* the patterning of corners of an imaginary cube held between forefinger and thumb) are often less effective than constructing an external representation such as a diagram. It helps to turn ones initial

---

[1] Reviewed in Chapter 3.

internal representation into an external stimulus which, upon re-processing, assists with finding a solution.

In other words, the process of externalisation helps to disambiguate ambiguous mental images. In a similar way, the process of drawing an ER such as a diagram (*i.e.* externalising a mental model) can facilitate problem solving. One interpretation of Reisberg's view is that externalisation facilitates the transfer of information between cognitive subsystems in ways that are not possible internally according to the dual-coding hypothesis (*e.g.* Paivio, 1986).

Individual differences

Prior knowledge

ER selection is affected profoundly by subjects' prior knowledge, and individuals differ widely in terms of their ER 'repertoires'. Apart from the studies reported here, few, if any, studies to date have attempted to relate subjects' prior knowledge of ER formalisms to their reasoning performance. Subjects' prior knowledge of ER formalisms was assessed in both *switchER* studies. In the *switchERI* study, the card sort task showed wide individual differences between subjects in terms of their prior knowledge and representational repertoires. For domain-specific representations, such as set-diagrams, which have an underlying formalism that must be learned, lack of prior knowledge predicts poor performance if the subject attempts to use the ER. For more ubiquitous ER forms, such as tables, the link between prior knowledge and successful use is less strong. One means of addressing this issue might be to provide direct instruction to students on a range of representational formalisms independently of particular subject-matter domains. The idea of an 'ER curriculum' is discussed further below.

In the *switchERII* study, the Euler's circle interpretation task was of Newstead (1989) was used as a pre-test of prior knowledge. A range of common error patterns were observed across subjects and there were considerable individual differences between subjects in terms of errors of omission/commission. The *switchERII* study results demonstrated that traditional error patterns (*e.g.* conversion, Gricean) on the Euler's circle *interpretation* task did not necessarily predict performance on tasks in which subjects *construct and reason* with ERs.

Cognitive style

The results of the studies reported here show that 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.

As reviewed in Chapter 1, individual differences along what can very loosely be termed the 'visualiser–verbaliser' dimension have been shown to be important in reasoning with ERs (MacLeod, Hunt and Mathews, 1978; Matsuno, 1987; Riding & Douglas, 1993; Ford, 1995; Cox, Stenning & Oberlander, 1994, 1995; Stenning, Cox & Oberlander, 1995; Oberlander, Cox & Stenning, 1994, 1995).

An implicit assumption in many studies (*e.g.* Frandsen & Holder, 1969), is that subjects who score poorly on psychometric tests of spatial visualization therefore prefer or tend to use the linguistic modality internally and/or externally when reasoning. This assumption is not justified and the relationship between internal cognitive modality preferences and the use of external representations requires much more research. For example, it would be interesting to examine whether subjects classified as highly spatial reasoners or diagrammatic modellers respond to efforts to broaden their ER repertoire by training in the use of non-graphical external representations. To the author's knowledge, this has never been demonstrated and remains an interesting topic for future research.

Riding & Douglas (1993) have shown that subjects independently classified *a priori* as 'visualisers' *produced* more graphical ERs in their responses to a reasoning problem than their 'verbaliser' counterparts. Macleod *et al.* (1978) clearly demonstrated that some subjects use a visual strategy and others use a verbal strategy on the sentence-picture verification task. Non-diagrammatic reasoners have been shown to be more prone to select weakly expressive representations for indeterminate problems (Stenning, Cox & Oberlander, 1995). Subjects report widely different types of internal imagery during reasoning - some subjects imagine depictions of scenes, whereas others reason with internal (imagined) graphical diagrams (Matsuno, 1987). Recent analyses of students reasoning with Hyperproof suggest that the idea of 'cognitive modality preference' that is implicit in many notions of cognitive style might be too simplistic - skilled performance with ERs seems to involve dynamic bi-directional translation between modalities, whereas less skilled reasoning is associated with more activity within a particular modality (Oberlander, Cox, Monaghan, Stenning & Tobin, 1996). Another interesting question that warrants further research, therefore, concerns the extent to which individual differences in reasoning with ERs reflects immutable differences in cognitive modality preference (*i.e.* cognitive style differences along some kind of 'visualiser–verbaliser' kind of dimension) and to what extent do they reflect differences in prior knowledge and are therefore malleable? The rationale for suggesting that 'ER curricula' be taught directly to students (discussed below) is predicated on the latter assumption.

Implications for theories of reasoning

An adequate theory of reasoning must be capable of accommodating individual differences — 'cognitive tractability' and 'substituted perceptual judgment' theories cannot be the whole story, since individual differences in reasoning with ERs are quite wide, as demonstrated in the studies reported here and by others.

Degree of externalisation

Another potentially important individual difference, one that is orthogonal to linguistic/graphical modality preferences, and one that has received no research attention, is the extent to which individuals *externalise* their reasoning. The results of the studies to be reported in this thesis and those of others (*e.g.* Schwartz, 1971) show that most subjects (usually > 80%) use ERs in their solutions to ER problems. In the case of the 'office allocation' problem, the study 1 data show that 7 (9%) out of 77 subjects did not produce an ER in their solutions to the problem yet performed well in terms of correct responses to the problem questions[2].

A minority of subjects, then, choose to solve some analytical reasoning problems without constructing an ER. If no ER is selected, then reasoning must proceed exclusively via mental processes. In that case, it is difficult to know whether the subject failed to use an ER because he or she didn't know which ER to select, or whether the subject recognised the type of ER required but did not have it in his or her repertoire. Alternatively, the student might be very competent and may not require an ER. However, as the results show, accurate performance in the absence of an ER is observed far more often on easy problems than on structurally more difficult ones and so task difficulty (cognitive load) is a determining factor.

In between the 'full blown' use of an ER and an exclusively internal strategy, there is a class of what might be termed 'minimal' ER strategies involving the partial externalisation of reasoning. Subjects may simply re-formulate or translate the problem information, re-arranging the premisses for example in an attempt to remediate the 'uncooperative' nature of the information presentation. Another 'minimal' strategy observed in the *switchER* studies was the use of the multiple-choice check boxes alongside each question. A few subjects initially check all the boxes and subsequently uncheck individual boxes as alternative responses are eliminated. This might be termed an 'elimination array' strategy.

Do ERs serve the same function for all subjects? As we have seen, subjects certainly differ in the modality of the ERs they use - partly due to cognitive style effects, partly due to the representational demands of the problem, and partly due to their prior experience and ER repertoire. But even in the case of subjects who build 'full blown' ERs, the ER may be more 'central' to the reasoning of some subjects than for others. ERs probably serve different functions for different people. Subjects probably vary in the way in which they partition their internalised and externalised cognition. The proportion of reasoning that is externalised by a particular subject may also vary at different stages of reasoning. Some subjects may use an ER to keep track of their progress through the problem, while reasoning internally for the most part. Other

---

[2] Thirty-four subjects (44%) used a linguistic ER and thirty-six (47%) used a graphical ER (Cox & Brna, 1995).

subjects may exploit the cognitive and semantic properties of the ER fully in their reasoning, adopting a model-based mode of reasoning.

Subjects differ too in the extent to which they interact with or operate upon their graphical representations. For fully model-based (*i.e.* diagrammatic) reasoners, the level of interaction and the amount of inter-modal translation of information is greater than for their less diagrammatically inclined counterparts (Oberlander, Monaghan, Cox, Stenning & Tobin, 1996). However, when diagrammatic reasoners are unable to use graphics[3], then their performance is degraded to a greater extent than less 'ER sensitive' subjects (Cox, Stenning & Oberlander, 1995).

The way forward - future work

Graphics curriculum

In general, if a representational formalism is capable of expressing indeterminacy, then the user requires direct instruction in order to use it effectively. This is certainly true in the linguistic modality since propositional/first-order logic and some types of natural language usage must be learned. It is also true of representations in the graphical modality — set-diagram formalisms such as Euler's Circles require training for effective use. In contrast, maps, network diagrams[4] (*e.g.* London underground map), and tabular representations (calendars, sports fixture tables, *etc.*), are ubiquitous and are not usually the subject of direct instruction.

For many students, the range of ERs that they are familiar with when they emerge from their formal education (*i.e.* their representational repertoire is quite 'hit and miss'). They may *happen* to know about and understand semantic networks because their biology teacher represented food webs in that way. They may *happen* to understand set-diagrams because they studied set theory in mathematics. Results from the taxonomy task used in the *switchERI* study show that only 25% of subjects showed evidence of accurately discriminating set diagrams from other circular representations - that

---

[3] Perhaps because the information to be represented is indeterminate and hence graphically inexpressible or because the subject is unfamiliar with an appropriate graphical formalism.

[4] Actually, network diagrams *are* capable of expressing indeterminacy - see Stenning & Inder (1995).

is they indicated in their choice of labels ('set' diagram, 'Venn' diagram *etc.*) that they understood the formalism of the representation. Why is it that so few students are capable of using such a useful formalism? The answer may lie, at least in part, in the nature of the school curriculum. In current curricula, students are typically introduced to specific ER formalisms in highly domain-dependent contexts. Often, curricula offer only vague advice to teachers and specifically mention only a narrow range of ERs. For example, the UK National Curriculum[5] suggests the use of the following representations:

> 'frequency tables', '...graphs and diagrams, including block graphs, pictograms and line graphs; ...pie charts' (maths, key stage 2)
>
> '..bar charts . line graphs, pie charts, frequency polygons, scatter diagrams and cumulative frequency diagrams' (maths, key stages 3 & 4)
>
> '..extract and interpret information presented in simple tables and lists' (maths, attainment target 4)
>
> 'Pupils should be taught to ... use a wide range of scientific and technical vocabulary and conventions, and to use diagrams, graphs, tables and charts to communicate information and to develop an argument;' (science, key stage 4 (double))
>
> 'use graphs to identify relationships between variables;' (experimental and investigative science, key stage 4 (single))
>
> 'pupils should be taught ... how distance, time and speed can be determined and represented graphically;' (science, key stage 4 (single) physical processes)
>
> 'pupils should be taught to ... make maps and plans at a variety of scales, using symbols, keys and scales...'(geography, key stage 3).

---

[5] The National Curriculum, Department for Education, London: HMSO

No specific references to other important representational formalisms such as set diagrams, network diagrams or tree diagrams could be found in the National Curriculum. For this reason, some students may emerge from the school system with unnecessarily limited ER repertoires and may not be exposed to useful ER formalisms such as set diagrams or network diagrams.

Part of the reason why ERs, especially diagrammatic ERs, have been ignored in curriculum development, might stem from negative attitudes such as those alluded to in Chapter 1, where it was noted that among teachers of logic and mathematics, graphical methods of external representation remain highly controversial. Eisenberg (1992), for example, points out that, in the mathematics community, the idea that mathematics must be communicated in a non-visual manner is 'deeply rooted'. Attitudes towards graphical representations are sometimes seemingly inconsistent. Graphical representations can be assigned different status depending upon whether they are used as a conceptual aid or as a medium of communication. Eisenberg (1992) illustrates this by citing Hilbert:

> I have given a simplified proof of part (a) of Jordan's theorem. Of course, my proof is completely arithmetizable (otherwise it would be considered non-existent); but, investigating it, I never ceased thinking of the diagram (only thinking of a very twisted curve), and so do I still when remembering it.

As also mentioned in Chapter 1, Sutherland (1995, p.80) also makes the point that:

> In our culture where sentential systems have higher status than visual systems, 'academic' is almost always synonymous with articulate.

In the logic domain, Barwise (1993) quotes the views of Neil Tennant as representative of the attitude of traditional logicians to the role of diagrams in logical proofs:

> [The diagram] is only an heuristic to prompt certain trains of inference; ...it is dispensable as a proof-theoretic device; indeed, ...it has no proper place in the proof as such. For the proof is a syntactic object consisting only of sentences arranged in a finite and inspectable array. (Tennant, 1986).

How might all students be provided with a basic ER repertoire? An 'ER curriculum' might be one answer. This could be taught in a domain-independent manner, like English. This is not the first call for such a curriculum. Recent years have seen several pleas for such training. For example, Guri-Rozenblit (1988) has written:

> Since the construction of schematic representations seems to be a notoriously difficult task to perform, and the interpretation and processing of visual displays poses problems of understanding, it seems important to include the learning and practicing of visual skills into the basic reading and writing skills in schools' curricula. The use of practice of visual language is to be learned as any other language... (p.232)

As mentioned in Chapter 3, Balchin & Coleman have written:

> It is hoped that the concepts of graphicacy and ingraphicacy will be taken up and developed by educationists, to mould the vague idea of visual aids at large into a more integrated goal of education, and to carry it down into the earliest stages to take its rightful role as one of the essential underpinnings. (p. 947)

Also Twyman (1979) states:

> .. formal teaching of graphical language appears to be limited,..., to the 'verbal/numerical' mode of symbolization ... In recent years young children have been taught how to produce simple line graphs, bar charts, and pie charts from data they have acquired themselves. At a later stage in their education, those specialising in certain fields may well learn the particular approaches to graphic language that are held to be appropriate to their specialty. On the whole, however, it is true to say that children are not taught to read the wide range of graphic language they will be confronted with in later life. Still less of course are children taught to originate information in anything like the range of approaches to graphic language presented in the

> matrix[6] (p.143-144)

Lewis (1989), too, writes that ...

> 'it seems appropriate to recommend including specific training of students' representation skills within the mathematics curricula of American schools.' (p530)

It is impossible to anticipate every reasoning situation that students might face, and individual differences preclude a curriculum that is too prescriptive about ER selection. However, an ER curriculum should include general information about the cognitive and semantic properties of representational systems and token representations within those systems. Ways of taxonomising representations could be compared (*e.g.* Twyman with Lohse *et al.*, *etc.*). The role of ERs in various kinds of task (reasoning/problem solving, data visualisation, communication) should also be considered. The advantages of heterogeneous reasoning should be stressed. The general approach should be one that aims to expose students to wide range of ER formalisms so they may select ERs on a principled basis and at the same time indulge their cognitive modality preferences.

Would such an approach work? The likely answer is 'yes' since several intervention studies provide good evidence for the effectiveness of direct instruction in ER use — Grossen & Carnine (1990); Frandsen & Holder (1969); Lewis (1989) and Lindvall, Tamburino & Robinson (1982).

As educational technology becomes more and more integrated into the curriculum and the variety and sophistication of data visualisation and external representation techniques increase as a result of information technology, the issue of an ER curriculum is likely to increase in importance. Whereas, in the era of paper and pencil and chalkboards, representations were laborious to construct, and difficult to modify, new technologies make a large range of representational formalisms easier to exploit. Animations and 3-dimensional representations can quickly be created and displayed. Graphical tools are influencing visual thinking to the same, and possibly greater, de-

---

[6] Twyman's (1979) table of ER forms which crosses 4 modes of symbolization with 7 methods of configuration.

gree that word processing tools influence writing[7]. Currently, computers do not parse graphical input to the same extent as alphanumeric input, but this situation is likely to change quickly. Technology has not yet advanced to the stage where it is capable of making intelligent decisions about the assignment of representational formalisms or modalities to information. Until it does, and probably even afterwards, students must be equipped with heuristics and principles for making representational decisions on a principled basis. They will need those skills in order to fully exploit information technology in the development and communication of their ideas, for problem solving and to fully realise the potential of information technology for augmenting cognition.

---

[7] Current examples that are being integrated into educational curricula include the Texas Instruments TI-92 graphing calculator running Cabri interactive software for geometry and the Hyperproof programme which supports a diagrammatic approach to teaching logic (Barwise & Etchemendy, 1994).

Chapter 9

Conclusion

This final chapter summarises the contributions of the thesis and lists some of the implications of the work for future research.

Thesis contributions

The contributions of this thesis are:

1. an investigation of the use of multiple representations in reasoning, particularly ER switching behaviour
2. a number of empirical results related to reasoning with ERs:
   - single ERs are not equally effective for every question in a problem's set and hence reasoning with multiple ERs is often effective
   - subjects can use incorrectly constructed ERs successfully (to the extent that they perform better using the incorrect ER than they do if they abandon it)
   - consistency of answers with read-off from ERs is associated with better performance than inconsistent responding *even in cases where the ER is a partially incorrect model of the information*
   - subjects seem, often, to allocate too few resources to problem comprehension
   - tabular (matrix) representations are usually best for determinate problems

  - target-oriented ER construction is superior to discourse-ordered construction
  - contingency tables yield uniformly poor performance
  - there were several lines of empirical support for specificity theory – when the expressivity of an ER is matched to the determinacy level of the problem, better performance results. However, the performance penalty of allocating a weakly expressive representation to an indetermiate problem is greater than using a strongly expressive representation for solving a determinate problem. Also, subjects tend to read-off erroneous conclusions from valid sentential ERs more frequently than they do from graphical ones.
  - ER switching is relatively common during ER construction
  - switching is positively related to problem difficulty
  - switching extends solution time but preserves performance
  - impasse-driven switching can result in 'thrashing'
  - opportunistic switching behaviour represents the judicious assignment of a representational system to a task
  - students do better if they fully comprehend the semantics of the ER formalism that they attempt to use in their solutions – idiosyncratic representations are associated with poor performance
  - not all errors of ER *interpretation* predict performance on tasks in which subjects *construct and reason with* ERs

3. the development of a process account of the cognitive events associated with the (5) stages of analytical reasoning with ERs

4. provision of comprehensive and structured reviews of the literatures on:

  - the psychometrics of analyical reasoning problems and their linguistic and structural properties
  - ER taxonomies and classification schemes
  - the semantic and cognitive properties of ERs
  - reasoning with ERs in a wide variety of domains

- individual differences (prior knowledge, cognitive style) in reasoning with ERs

5. development, through an iterative design process, of *switchERII* – an intelligent, interactive learning environment for reasoning with ERs
6. the elucidation of cognitive mechanisms by which the process of *externalisation* facilitates reasoning performance and is central to a constructivist account of reasoning with ERs
7. the proposal of a mechanism by which the use of diagrammatic ERs may facilitate the self-explanation effect
8. a study of the relationship between subjects' prior knowledge of ER formalisms and their reasoning with ERs
9. the development, via card sort methodology and cluster analysis, of a taxonomy of ERs used by subjects in problem solving
10. a comparison of subjects' performance under conditions in which they interpret prefabricated ERs with their performance under conditions in which they spontaneously select and construct their own ERs
11. an innovative methodology in data acquisition *i.e.* use of ILE's, dynamic user-system interaction logging techniques
12. the development of an argument in favour of directly instructing students (via an 'ER curriculum' and in a relatively domain independent way) to use a wide variety of useful representational formalisms

Of these, the most valuable contributions are probably those relating the expressive properties of representations to task characteristics, the identification of two types of switching behaviour and the demonstration of decoupling between ER *interpretation* and *construction and use.*

Studying the ways in which externalisation and self-explanation processes facilitate reasoning is also of theoretical interest and will be the subject of future work.

Future work

The future work suggested by the thesis includes:

1. extending the representational semantics of *switchERII* to include node and arc diagrams, first-order logic, topographical representations such as plans and maps, *etc.*
2. the development of a relatively domain-independent 'ER curriculum' for teaching students how to use a wide variety of representational formalisms in both sentential and graphical modalities
3. investigating individual differences in 'cognitive style', in particular the relationship between psychometrically assessed internal cognitive modality preference (*e.g.* imagistic versus sentential) and overt external representational behaviour (*e.g.* tendency to use graphical versus sentential ERs)
4. determining whether subjects identified as 'visualisers' or 'verbalisers' in terms of their cognitive modality preference can be trained to broaden their ER repertoires in their less-preferred modality - *i.e.* testing the immutability/malleability of cognitive style
5. investigating various types of user-feedback using *switchERII*
6. re-using the *switchER* user-system interaction recordings as resources for future learners
7. further investigating ways in which mappings between information presented in the problem stem and elements of the representation can be preserved for the benefit of the reasoner
8. investigating subjects' use of multiple representations - especially comparing the serial use of multiple representations (ER switching) with the concurrent use of multiple representations (heterogeneous reasoning). What effect does the severe cognitive load of information integration (in the case of concurrent multiple ERs) have upon performance?

9. investigating ways in which *switchERII* might encourage subjects to self-monitor, avoid translation slips and to generally spend longer at the problem comprehension stage prior to ER selection

10. investigating ways in which *switchERII* might check whether the subject's responses to a problem's questions are consistent with direct read-off from his/her ER. The objective would be to encourage the reasoner to respond consistently

11. implementing more intelligent 'ER co-construction' such that the system builds an ER in parallel with the user but in a different modality - the co-constructed representation should reflect the original with as much fidelity as possible, perhaps including errors

12. investigating further the proposal that skilled performance with multiple ERs involves dynamic bi-directional translation between modalities rather than merely a preference for reasoning either graphically or sententially

Most of the areas identified as topics for future work are interesting. However, some are of greater theoretical interest than others. The highest priority topics for future work would include extending the representational semantics of *switchERII*, and developing an ER curriculum, though both are fairly long-term prospects.

Pragmatically, the easiest-to-implement topics are those that involve improvements to *switchERII* — such as building in a mechanism to check whether the subject's responses are consistent with his or her ER. Another relatively straightforward study would be to investigate the effects of various types of user feedback and to improve the mappings between problem stem information and elements of the user's representation.

Finally, another short-term goal would be to conduct more research into subjects' use of multiple representations, especially the nature of information translation between modalities.

Chapter 10

References

References

Aldenderfer, M. S. & Blashfield, R. K. [1984] *Cluster Analysis.* London: Sage Publications.

Amarel, S. [1968] On representations of reasoning about actions. In D. Michie (Ed.), *Machine Intelligence: Volume 3.* Edinburgh: Edinburgh University Press.

Anderson, R.E. & Helstrup, T. [1993] Visual discovery in mind and on paper. *Memory & Cognition*, **21(3)**, 283–293.

Angoff, W. H. & Johnson, E. G. [1990] The differential impact of curriculum on aptitude test scores. *Journal of Educational Measurement*, **27(4)**, 291–305.

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.

Austin, J.L. & Howson A.G. [1979] Language and Mathematical Education. *Educational Studies in Mathematics*, **10**, 161–197.

Baddeley, A.D. [1990] *Human Memory: Theory and practice.* Hillsdale, NJ: Lawrence Erlbaum & Associates.

Balchin, W.G.V. & Coleman, A.M. [1965] Graphicacy should be the fourth ace in the pack. *The Times Educational Supplement*, November 5th, 947.

Bartram, D.J. [1980] Comprehending spatial information: The relative efficiency of different methods of presenting information about bus routes. *Journal of Applied Psychology*, **65(1)**, 103–110.

Barwise, J. [1993] Heterogeneous reasoning. In G. Allwein G. & J. Barwise (Eds) *Working papers on diagrams and logic.* Visual Inference Laboratory, Indiana University.

Barwise, J. & Etchemendy, J. [1992] Hyperproof: Logical reasoning with diagrams. In H. Narayanan (Ed.), *Working notes - AAAI Spring Symposium on Reasoning with Diagrammatic Representations*, Stanford, CA, March, 80–84.

Barwise, J. & Etchemendy, J. [1994] *Hyperproof for the Macintosh.* CSLI Lecture Notes Number 42, Centre for the Study of Language & Information, Leland Stanford Junior University: CSLI Publications.

Barwise, J. & Shimojima, A. [1995] Surrogate reasoning. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, **2(4)**, 7–27.

Bell, A.W. [1983] *A Review of Research in Mathematical Education*, Nelson: National Foundation for Educational Research.

Bernado, A.B.I. & Okagaki, L. [1994] Roles of symbolic knowledge and problem-information context in solving word problems. *Journal of Educational Psychology*, **86(2)**, 212–220.

Berry, D.C. & Broadbent, D.E. [1989] On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, **36A**, 209–231.

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.

Biron, D. & Bednarz, N. [1989] The role of children's diagrams and pictures in their solution of complex arithmetic problems. *Presented at the biennial meeting of the Society for Research in Child Development*, Kansas, April. ERIC document ED-310920.

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.

Boyle, C.F. & Anderson, J.R. [1984] Acquisition and automated instruction of geometry skills. *Paper presented at the Annual meeting of the American Educational Research Association (AERA-84)*, New Orleans.

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. *Psycoloquy* (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.

Carroll, J.M., Thomas, J.C. & Malhotra, A. [1980] Presentation and representation in design problem solving. *British Journal of Psychology*, **71**(**1**), 143–153.

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 R. Cox, M. Petre, P. Brna, & J. Lee (Eds.) *Proceedings of the AI-*

ED93 Workshop on Graphical Representations, Reasoning and Communication, August, Edinburgh.

Chi, M.T.H. & Bassok, M. [1989] Learning from examples via self-explanations. In L.B. Resnick (Ed.), *Knowing, learning and instruction: essays in honor of Robert Glaser*, Hillsdale, NJ: Lawrence Erlbaum Associates, 251–282.

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.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. [1981] Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121–152.

Clement, J. [1982] Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education*, **13**, 16–30.

Corbett, A.T. & Anderson, J.R. [1990] The effect of feedback control on learning to program with the lisp tutor. *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 796–803.

Clement, J., Lochhead, J., & Monk, G. [1991] Translation difficulties in learning mathematics. *American Mathematical Monthly*, **88**, 286–290.

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

Cox, R., 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.

Cox, R., Stenning, K. & Oberlander, J. [1995] The effect of graphical and sentential logic teaching on spontaneous external representation. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, **2**(4), 56–75.

Davis, P.J. & Hersh, R. [1981] *The mathematical experience*, Harvester Press.

Davis, R.B., Young, S. & McLoughlin, P. [1982] The roles of 'understanding' in the learning of mathematics. Urbana/Champaign: Curriculum Laboratory, University of Illinois.

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.

Duncker, K. [1948] On problem solving. *Psychological Monographs*, **58**, (Whole No. 270).

Duran, R., Powers, D. & Swinton, S. [1987] Construct validity of the GRE analytical test: A resource document. *ETS Research Report 87-11*. Princeton, NJ: Educational Testing Service.

Durkin, K. & Shire, B. [1991] Lexical ambiguity in mathematical contexts. In K. Durkin & B. Shire (Eds) *Language in Mathematical Education: Research and Practice*. Open University Press.

Educational Testing Service GRE Board, [1992] *Practicing to take the GRE General Test Number 9*. NJ: Warner.

Eisenberg, T. [1992] The development of a sense for functions. In Harel, G. and Dubinsky, E. (Eds.) *The Concept of Function*, Mathematical Association of America, MAA Notes Volume 25.

Emmerich, W., Enright, M. K., Rock, D. A. & Tucker, C. [1991] *The development, investigation and evaluation of new item types for the GRE analytical measure*, Princeton, NJ: Educational Testing Service.

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.

Ford, M. [1995] Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, **54**, 1–71.

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.

Fuson, K.C. & Willis, G.B. [1989] Second graders' use of schematic drawings in solving addition and subtraction word problems. *Journal of Educational Psychology*, **81**, 514–520.

Gardner, M. [1982] *Logic machines and diagrams*, (2nd Edition), Chicago: University of Chicago Press.

Gick, M.L. & Holyoak, K.J. [1983] Schema induction and analogical transfer. *Cognitive Psychology*, **15**, 1–28.

Gilmore, D.J. [1991] Visibility: A dimensional analysis. In D. Diaper & N. Hammond (Eds.), *People and Computers, Volume VI*, Cambridge University Press, 317–329.

Goel, V. [1995] *Sketches of Thought*. Cambridge, MA: MIT Press.

Goldschmidt, G. [1991] The dialectics of sketching. *Creativity Research Journal*, **4**, 123–143.

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.

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.

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.

Greeno, J. [1977] Process of understanding in problem solving. In N.J. Castellan, D.B. Pisoni & G.R. Potts (Eds.), *Cognitive Theory: Volume 2*. Hillsdale, NJ: Lawrence Erlbaum Associates, 43–83.

Grice, H.P. [1975] Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics: Volume 3 - Speech Acts*, New York:Academic Press, 41–58.

Grossen, G. & Carnine, D. [1990] Diagramming a logic strategy: Effects on difficult problem types and transfer. *Learning Disability Quarterly*, **13**, 168–182.

Guri-Rozenblit, S. [1988] The interrelations between diagrammatic representations and verbal explanations in learning from social science texts. *Instructional Science*, **17**, 219–234.

Guthrie, J.T. & Kirsch, I.S. [1987] Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology*, **79(3)**, 220–227.

Guthrie, J.T. & Mosenthal, P. [1987] Literacy as multidimensional: Locating information and reading comprehension. *Educational Psychologist*, **22(3-4)**, 279–297.

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.

Hayes, P.J. [1985] Some problems and non-problems in representation theory. In R.J. Brachman & H.J. Levesque (Eds.), *Readings in Knowledge Representation*, Los Altos, CA: Morgan Kaufmann.

Hesse, F., Tiberghien, A., Baker, M., Picard, E., & Reinhard, P. [1995] Computer supported collaborative learning in sciences. Presented at the LERANT symposium, *European Association for Research in Learning & Instruction (EARLI) conference*, Nijmegan, August.

Hinton, G.E. [1979] Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, **3**, 231–251.

Holyoak, K.J. & Spellman, B.A. [1993] Thinking. *Annual Review of Psychology*, 44, 265–315.

Humphreys, G. & Bruce, V. [1989] , *Visual cognition: Computational, experimental and neuropsychological perspectives*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Janvier, C. [1987] Representation and understanding: the notion of function as an example. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics*, Hillsdale, NJ: Lawrence Erlbaum Associates, 207–214.

Johnson, S.C. [1967] Hierarchical clustering schemes. *Psychometrika*, 32, 241–254.

Kaput, J. J. [1987] Towards a theory of symbol use in mathematics. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics*, Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Kintsch, W. [1989] Learning from text. In L.B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser*, Hillsdale, NJ: Lawrence Erlbaum and Associates, 25–46.

Kintsch, W. [1991] A theory of discourse comprehension: Implications for a tutor for word algebra problems. In M. Carretero, M. Pope, R-J. Simons & J.I. Pozo, *Learning and instruction: Volume 3*, A publication of the European Association for Research on Learning & Instruction, Oxford: Pergamon Press.

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. & Tabachnek, H.J.M. [1994] Two strategies are better than one: Multiple strategy use in word problem solving. *Submitted to the American Educational Research Association Conference (AERA-94)*, April.

Larkin, J. [1985] Understanding, problem representations, and skill in physics. In Chipman, S., Segal, J.W. & Glaser, R. (Eds.) *Thinking and learning skills: Volume 2: Research and open questions*, Lawrence Erlbaum, Hillsdale: NJ.

Larkin, J. [1989] Display–based problem solving. In D. Klahr & K. Kotovsky (Eds.) *Complex information processing: Twenty-first Carnegie-Mellon symposium on cognition*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J.H., McDermott, J., Simon, D., & Simon, H.A. [1980] Expert and novice performance in solving physics problems. *Science*, **208**, 1335–1342.

Larkin, J.H. & Simon, H.A. [1987] Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, **11**, 65–100.

Lepik, M. [1990] Algebraic word problems: The role of linguistic and structural variables. *Educational Studies in Mathematics*, **21**, 83–90.

Lepper, M.R. & Gurtner, J. [1989] Children and computers: Approaching the twenty-first century. *American Psychologist*, **44(2)**, 170–178.

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.

Levesque, H. J. [1988] Logic and the complexity of reasoning. *Journal of Philosophical Logic*, **17**, 355–389.

Lewis, A.B. [1989] Training students to represent arithmetic word problems. *Journal of Educational Psychology*, **81**, 521–531.

Lewis, A.B. & Toth, J. [1992] Situated cognition in diagrammatic reasoning. In H. Narayanan (Ed.), *Working notes - AAAI Spring Symposium on Reasoning with Diagrammatic Representations*, Stanford, CA, March, 47–52.

Lindvall, C.M., Tamburino, J.L. & Robinson, L. [1982] An exploratory investigation of the effect of teaching primary grade children to use specific problem solving strategies in solving simple arithmetic story problems. *Presented at American Educational Research Association annual conference*, New York, March. ERIC document ED-216904.

Lohse, G.L., Walker, N., Biolsi, K. & Rueter, H. [1991] Classifying graphical information. *Behavior and Information Technology*, 10, 419—436.

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

Lorr, M. [1983] *Cluster analysis for social scientists.* San Francisco, CA: Jossey-Bass.

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.

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.

Mason, J. [1987a] Representing representing: notes following the conference. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics*, Hillsdale, NJ: Lawrence Erlbaum Associates, 207–214.

Matsuno, T. [1987] Cognitive style and representational strategies in categorical syllogistic reasoning. *Tohoku Psychologica Folia*, **46(1-4)**, 97–102.

Mayer, R.E. [1976] Comprehension as affected by structure of problem representation. *Memory & Cognition*, 4, 249–255.

Mayer, R. E. & Sims, V. K. [1994] For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, **86**, 389–401.

McGuinness, C. [1986] Problem representation: The effects of spatial arrays. *Memory & Cognition*, **14(3)**, 270–280.

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.

Miller, R. & Wild, C.L. [1979] *Restructuring the Graduate Record Examinations Aptitude Test*, Princeton, NJ: Educational Testing Service.

Molitor, S., Ballstaedt, S-B. & Mandl, H. [1989] Problems in knowledge acquisition from text and pictures. In H. Mandl & J.R. Levin (Eds.) *Knowledge acquisition from text and pictures*, Amsterdam: North-Holland, 3–35.

Moyse, R. [1989] Knowledge negotiation implies multiple viewpoints.*CITE Report 63*, Open University, UK.

Newell, A. & Simon, H.A. [1972] *Human problem solving*, Englewood Cliffs, NJ: Prentice-Hall.

Newstead, S.E. [1989] Interpretational errors in syllogistic reasoning. *Journal of Memory & Language*, **28**, 78–91.

Newstead, S.E. [1995] Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, **34**, 644–664.

Nolder, R. [1991] Mixing metaphor and mathematics in the secondary classroom. In K. Durkin and B. Shire (Eds.) *Language in Mathematical Education. Research and Practice*, Open University Press.

Norman, D.A. [1988] *The psychology of everyday things.* New York, NY: Basic Books.

Norman, D.A. [1993] *Things that make us smart: defending human attributes in the age of the machine.* Reading, MA: Addison-Wesley.

Novick, L.R. [1990] Representational transfer in problem solving. *Psychological Science*, **1(2)**, 128–132.

Oberlander, J., Cox, R., & Stenning, K. [1995a] Proofs as discourse: an empirical study. In Working Notes of the *AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Stanford, March.

Oberlander, J., Cox, R., & Stenning, K. [1995b] Proof styles and multi-modal reasoning. In Seligman, J. & Westerstahl, D. (Eds.) *Language, Logic and Computation: The 1994 Moraga Proceedings*, Stanford: CSLI Publications, 403–414.

Oberlander, J., Cox, R., Monaghan, P., Stenning, K. & Tobin, R. [1996] Individual differences in proof structures following multimodal logic teaching. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 201–206.

Oakhill, J. & Garnham, A. [1993] On theories of belief bias in syllogistic reasoning. *Cognition*, **46**, 87–92.

O'Donnell, A. [1992] Searching for information in knowledge maps and texts. Presented at the 25th International Congress of Psychology, Brussels. Abstract in International Journal of Psychology, **27(3-4)**.

Paige, J.M. & Simon, H.A. [1966] Cognitive processes in solving algebra word problems. In B. Kleinmutz (Ed.), *Problem solving*, New York, NY: Wiley, 51–119.

Paivio, A. [1986] *Mental representations: A dual-coding approach.* Oxford: Oxford University Press.

Peterson, D. [1994] Re-representation and emergent information in three cases of problem solving. In T. Dartnell (Ed.) *Artificial Intelligence & Creativity*, Dordrecht, Holland: Kluwer, 81–92.

Petre, M. & Green, T.R.G. [1993] Learning to read graphics: Some evidence that 'seeing' an information display is an acquired skill. *Journal of Visual Languages and Computing*, 4, 55–70.

Pimm, D. [1987] *Speaking Mathematically: Communication in mathematics classrooms*, Routledge and Kegan Paul.

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.

Pylyshyn, Z. [1973] What the mind's eye tells the mind's brain. *Psychological Bulletin*, **80**, 1–24.

Pylyshyn, Z. [1979] Imagery theory: Not mysterious – just wrong. *Behavioural & Brain Sciences*, **2**, 561–563.

Reed, S.K. & Ettinger, M. [1987] Usefulness of tables for solving word problems. *Cognition and Instruction*, 4, 43–59.

Reisberg, D. [1987] External representations and the advantages of externalizing one's thoughts. In *Proceedings of the 9th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates, 281–293.

Reiser, B.J., Ranney, M., Lovett, M.C. & Kimberg, D.Y. [1989] Facilitating students' reasoning with causal explanations and visual representations. 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.

Research & Education Association [1994] *Reasoning builder for admission and standardized tests.* Piscataway, NJ: Research & Education Association.

Resnick, L.B. [1983] A developmental theory of number understanding. In H.P. Ginsburg (Ed.), *The development of mathematical thinking*, New York: Academic Press.

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.

Roberts, M.J., Wood, D.J. & Gilmore, D.J. [1994] The sentence-picture verification task: Methodological and theoretical difficulties. *British Journal of Psychology*, **85**, 413–432.

Rumelhart, D.E. & Norman, D.A. [1988] Representation in memory. In R.C. Atkinson, R.J. Herrnstein, B. Lindzey, & R.D. Luce (Eds.), *Handbook of Experimental Psychology*, Chichester: Wiley.

Schank, R.C. & Abelson, R.P. [1977] *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A.H. [1988] Problem solving in context(s). In J. Lave, J.G. Greeno, A. Schoenfeld, S. Smith & M. Butler, *Learning mathematical problem solving*, Institute for Research on Learning, Palo Alto, CA: IRL Report No. IRL88-0006, 30–44.

Schooler, L.J. & Anderson, J.R. [1990] The disruptive potential of immediate feedback. *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 702–708.

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. [1993] Measuring integration of information in multi-representational software. *Interactive Learning Environments*, **3(3)**, 177–198.

Shalin, V. & Bee, N.V. [1985] Analysis of the semantic structure of a domain of word problems. *Technical Report No. APS-20*, Pittsburgh: University of Pittsburgh, Learning, Research & Development Centre.

Shin, Sun-Joo [1994] *The logical status of diagrams*, Cambridge University Press.

Siegel, S. & Castellen, N.J. [1988] *Nonparametric statistics (2nd Edition).* New York, NY: McGraw-Hill.

Simon, H. [1981] *The sciences of the artificial.* Cambridge, MA: MIT Press.

Sims-Knight, J. & Kaput, J. [1983] Misconceptions of algebraic symbols: Representations and component processes. In J. Novak (Ed.), *Proceedings of the International Seminar on Misconceptions in Mathematics and Science*, Columbus, OH: ERIC Clearinghouse for Science, Mathematics and Environmental Education, 477–488.

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.

Stenning, K. & Cox, R. [1995] Attitudes to logical independence: traits in quantifier interpretation. In J. D. Moore & J. Fain Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum & Associates, 742–747.

Stenning, K., Cox, R., & Oberlander, J. [1995] Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. *Language and Cognitive Processes*, **10 (3/4)**, 333–354.

Stenning, K. & Inder, R. [1995] Applying semantic concepts to the media assignment problem in multi-media communication. In B. Chandrasekaran, J. Glasgow & N.H. Narayanan (Eds.) *Diagrammatic Reasoning: Computational and Cognitive Perspectives on Problem Solving with Diagrams.* Menlo Park, CA: AAAI Press/MIT Press, 303 - 338.

Stenning, K. & Oberlander, J. [1995] A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, **19(1)**, 97–140.

Stenning, K., Yule, P. & Cox, R. [1996] Quantifier interpretation and syllogistic reasoning: An individual differences account. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 678–683.

Stone, D.E. & Glock, M.D. [1981] How do young adults read directions with and without pictures? *Journal of Educational Psychology*, **73**, 419–426.

Sutherland, R. [1995] Mediating mathematical action. In R. Sutherland & J. Mason (Eds.) Exploiting mental imagery with computers in mathematics education. *Proceedings of NATO Advanced Research Workshop.* Berlin: Springer.

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.

Tennant, N. [1986)] The withering away of formal semantics. *Mind and Language*, 1, 302–18.

Treu, S. [1992] Interface structures: conceptual, logical, and physical patterns applicable to human–computer interaction. *International Journal of Man–Machine Studies*, **37**, 565–593.

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

van Essen, G. & Hamaker, C. [1990] Using self-generated drawings to solve arithmetic word problems. *Journal of Educational Research*, **83**, 301–312.

VanLehn, K., Jones, R.M., & Chi, M.T.H. [1992] A model of the self-explanation effect. *The Journal of the Learning Sciences*, **2(1)**, 1–59.

van Sommers, P. [1984] , *Drawing and cognition*, Cambridge, UK: Cambridge University Press.

Vessey, I. [1991] Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, **22**, 219–240.

Wertheimer, M. [1959] *Productive thinking*, Enlarged edition, New York, NY: Harper & Row.

Wertsch, J. [1991] *Voices of the mind: A sociocultural approach to mediated action*, Cambridge, MA: Harvard University Press.

Wertsch, J. & Toma, C. [1995] Discourse and learning in the classroom. In L.P. Steffe & J. Gale (Eds.) *Constructivism in education*, Hillsdale, NJ: Lawrence Erlbaum Associates.

White, B.Y. & Frederiksen, J.R. [1990] Causal model progressions for intelligent learning environments. *Artificial Intelligence*, **42**, 99–157.

Willis, G.B. & Fuson, K.C. [1988] Teaching children to use schematic drawings to solve addition and subtraction word problems. *Journal of Educational Psychology*, **80**, 192–201.

Wollman, W. [1983] Determining the sources of error in a translation from sentence to equation. *Journal for Research in Mathematics Education*, **14(3)**, 169–181.

Wusteman, J. [1992] Explanation-based learning: a survey. *Artificial Intelligence Review*, **6**, 243–262.

Yule, P. [1995] Logic and implementation in human reasoning: The psychology of syllogisms. *PhD Thesis*, Centre for Cognitive Science, University of Edinburgh.

Appendix A

Three examples of analytical reasoning 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 ?

Example ERs - workscratching figures

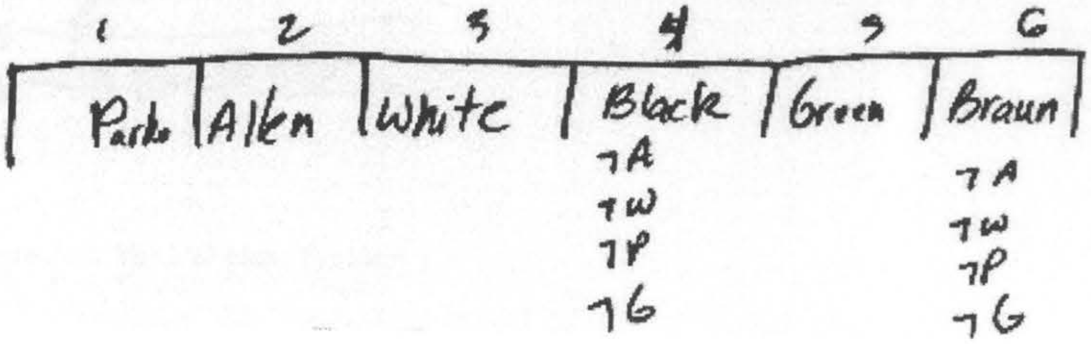


Figure A.1: Plan, Problem 1

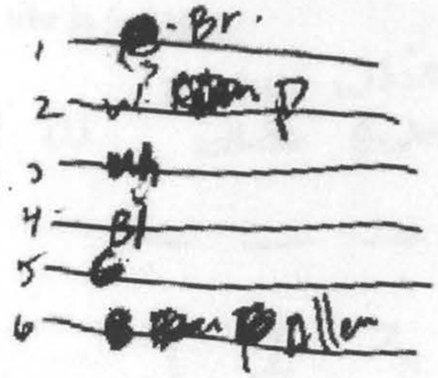


Figure A.2: Vertical plan, Problem 1

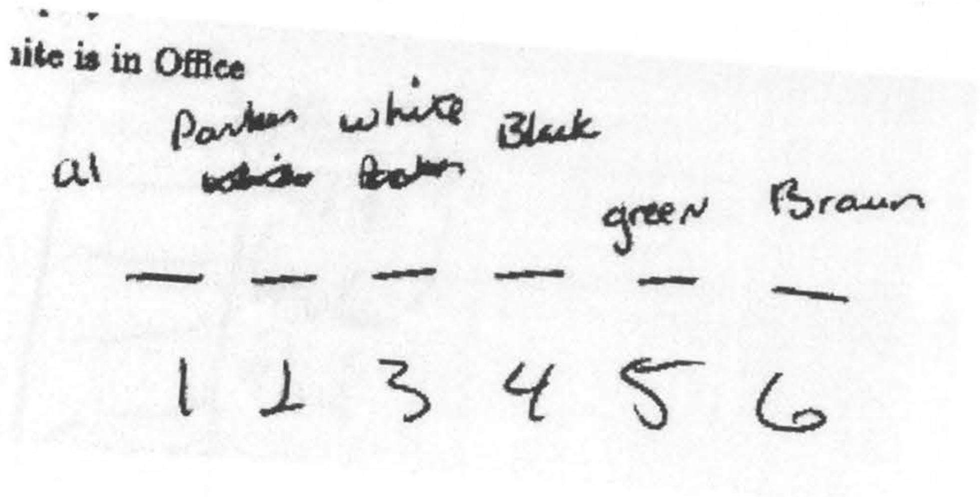


Figure A.3: Minimal plan, Problem 1

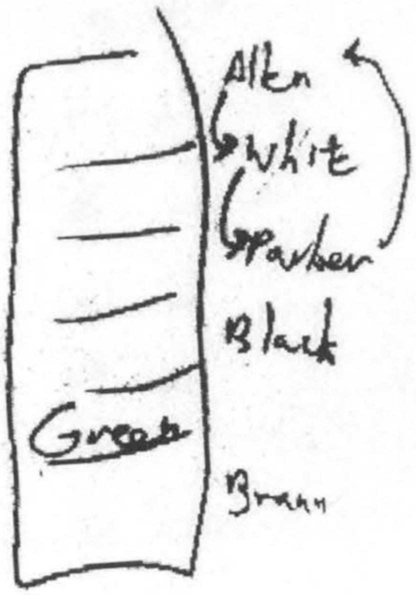
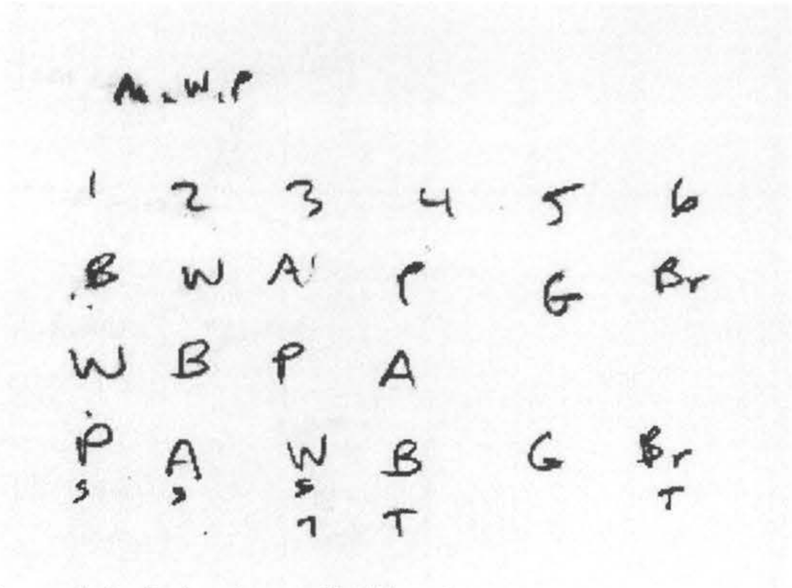


Figure A.4: Vertical plan, Problem 1



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
		7	T		

Figure A.5: Ordered text, Problem 1

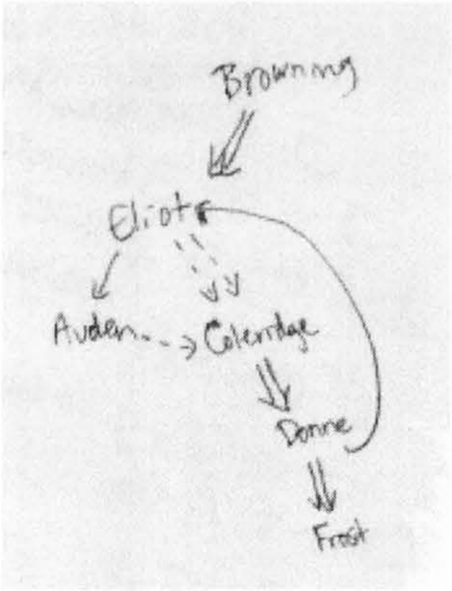
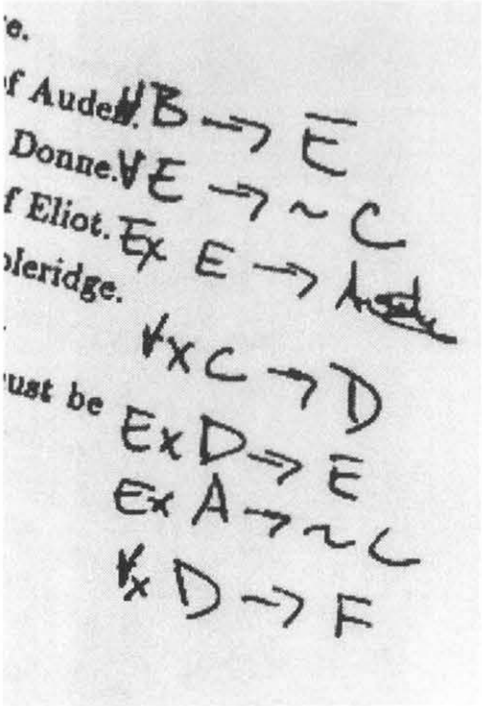


Figure A.6: Directed graph, Problem 2



e.

f Aude. $\forall B \rightarrow E$

Donne. $\forall E \rightarrow \sim C$

f Eliot. $\exists x\ E \rightarrow$ [illegible]

leridge.

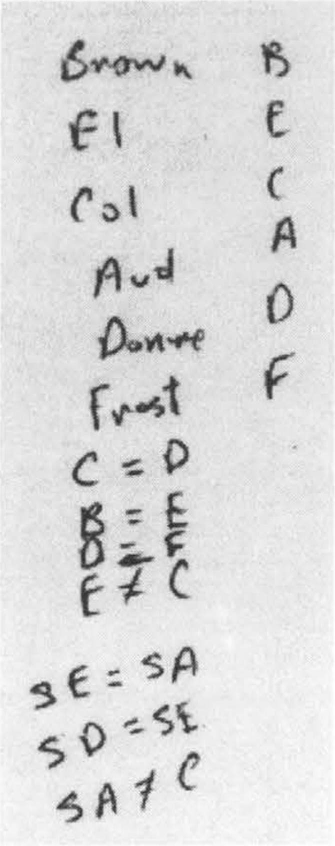
$\forall x\ C \rightarrow D$

ust be $\exists x\ D \rightarrow E$

$\exists x\ A \rightarrow \sim C$

$\forall x\ D \rightarrow F$

Figure A.7: Logic, Problem 2



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

SE = SA
SD = SE
SA ≠ C

Figure A.8: Text, Problem 2

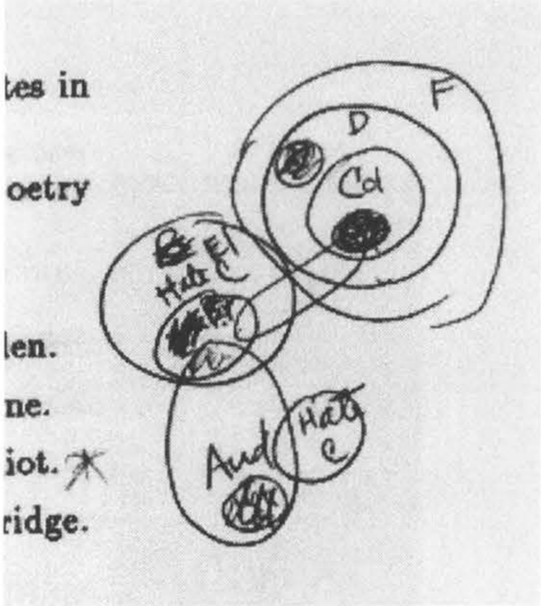


Figure A.9: Set diagram, Problem 2

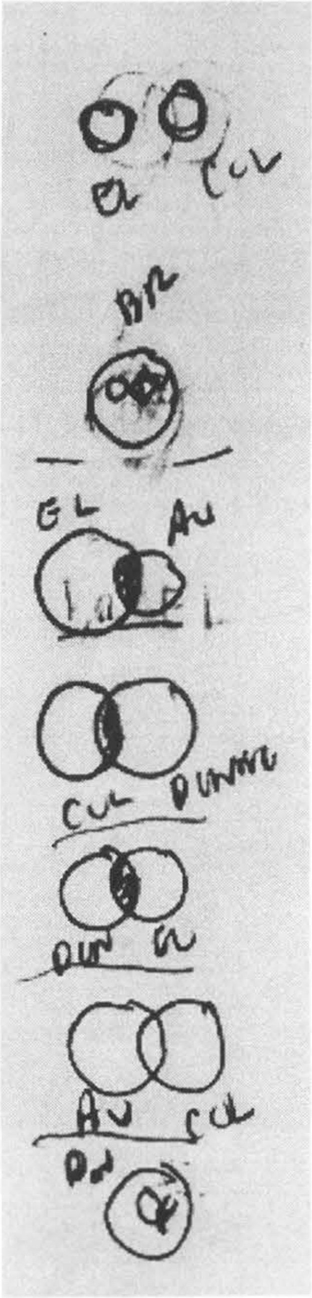


Figure A.10: Set diagram (non unified), Problem 2

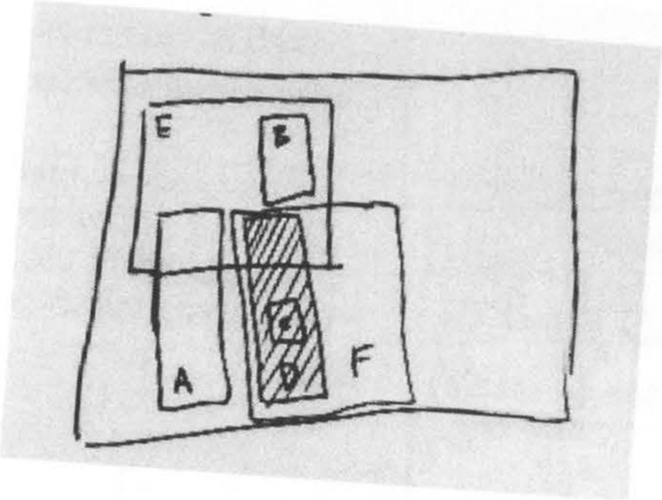
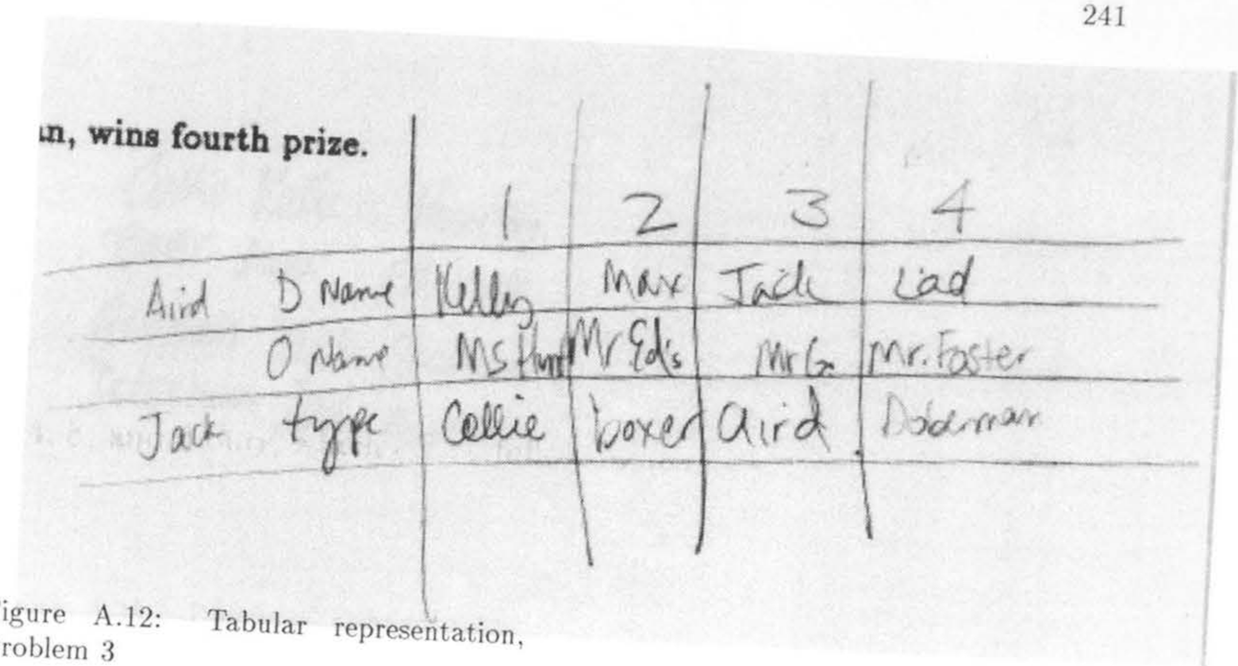


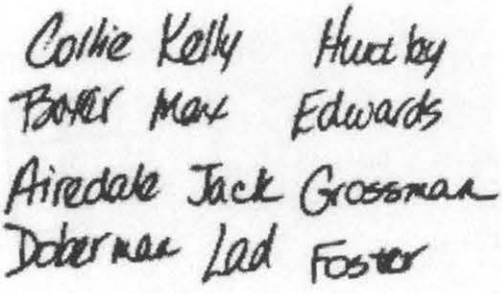
Figure A.11: Set diagram using rectangles, Problem 2



n, wins fourth prize.

		1	2	3	4
Aird	D Name	Kelly	Max	Jack	Lad
	O Name	Ms Hunt	Mr Ed's	Mr G.	Mr. Foster
Jack	type	Collie	boxer	aird	Doberman

Figure A.12: Tabular representation, Problem 3



Collie	Kelly	Huckby
Boxer	Max	Edwards
Airedale	Jack	Grossman
Doberman	Lad	Foster

Figure A.13: Discourse ordered table, Problem 3

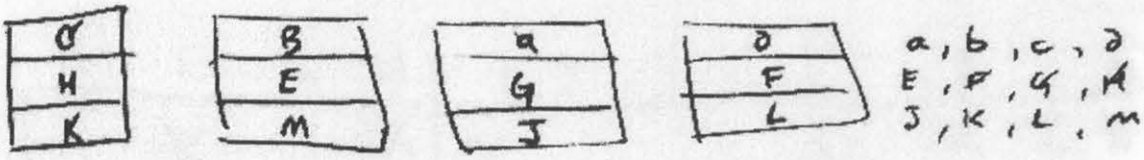
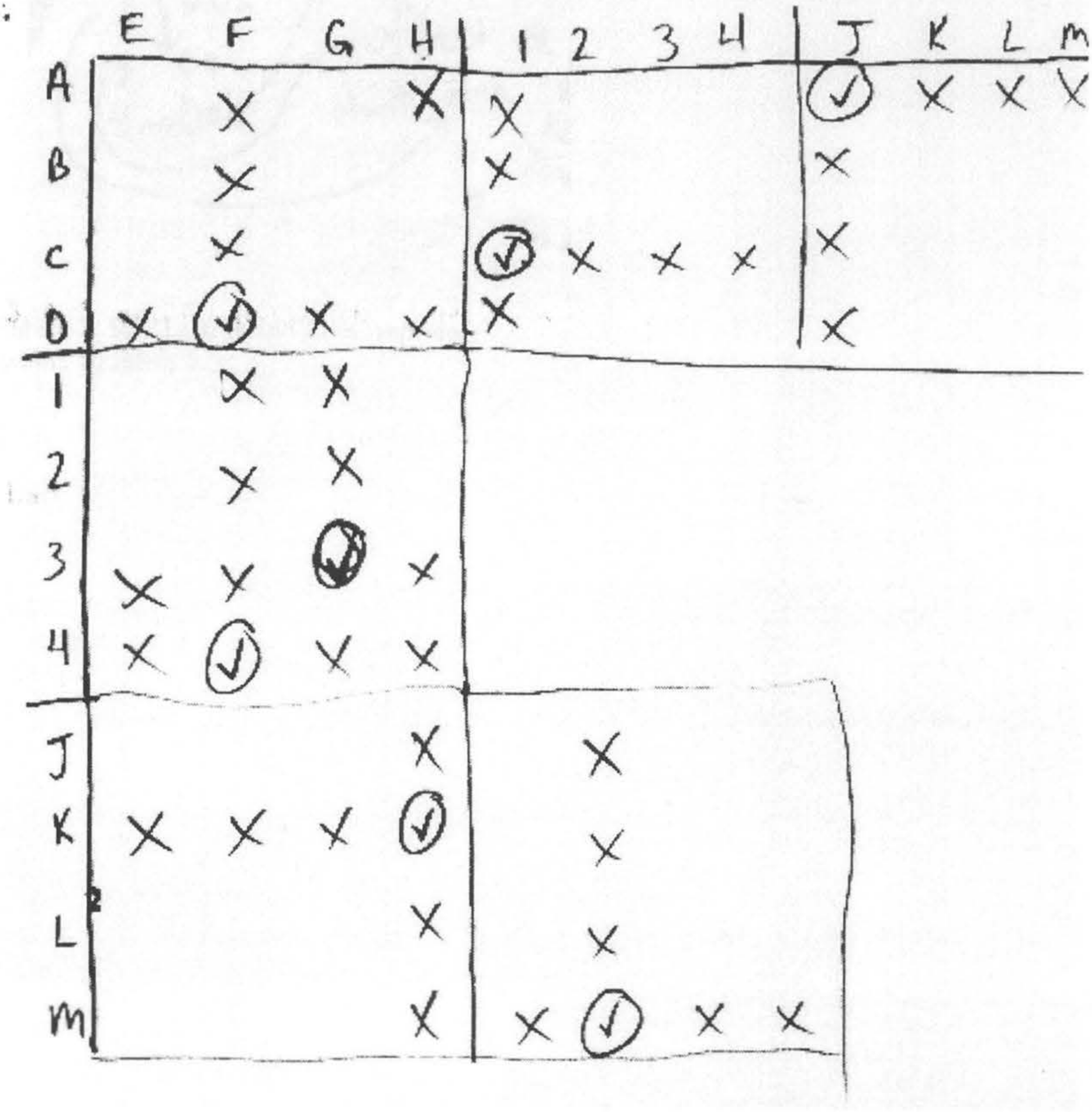


Figure A.14: Tabular representation, Problem 3



	E	F	G	H	1	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 A.15: Contingency table, Problem 3

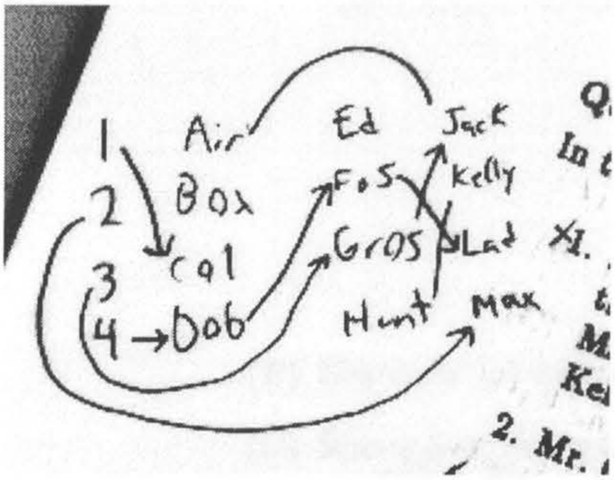


Figure A.16: 'Letters and lines' representation, Problem 3

Single Models — Ignoring Alternatives

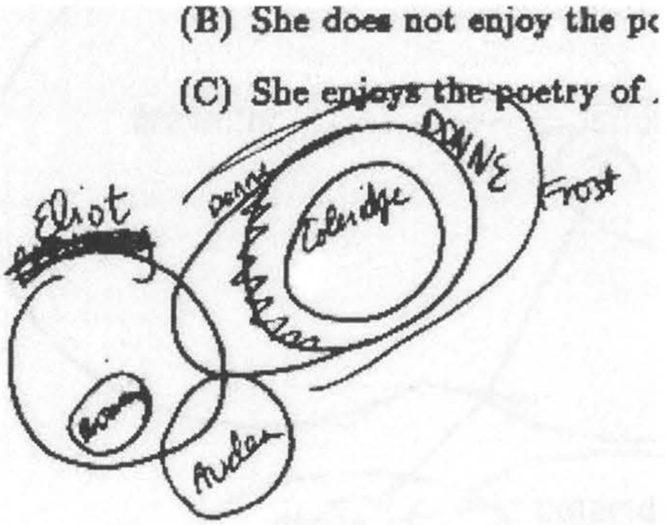


Figure A.17: Set diagram, Problem 2 — only one of many possible models of problem

‘Invented’ Annotations

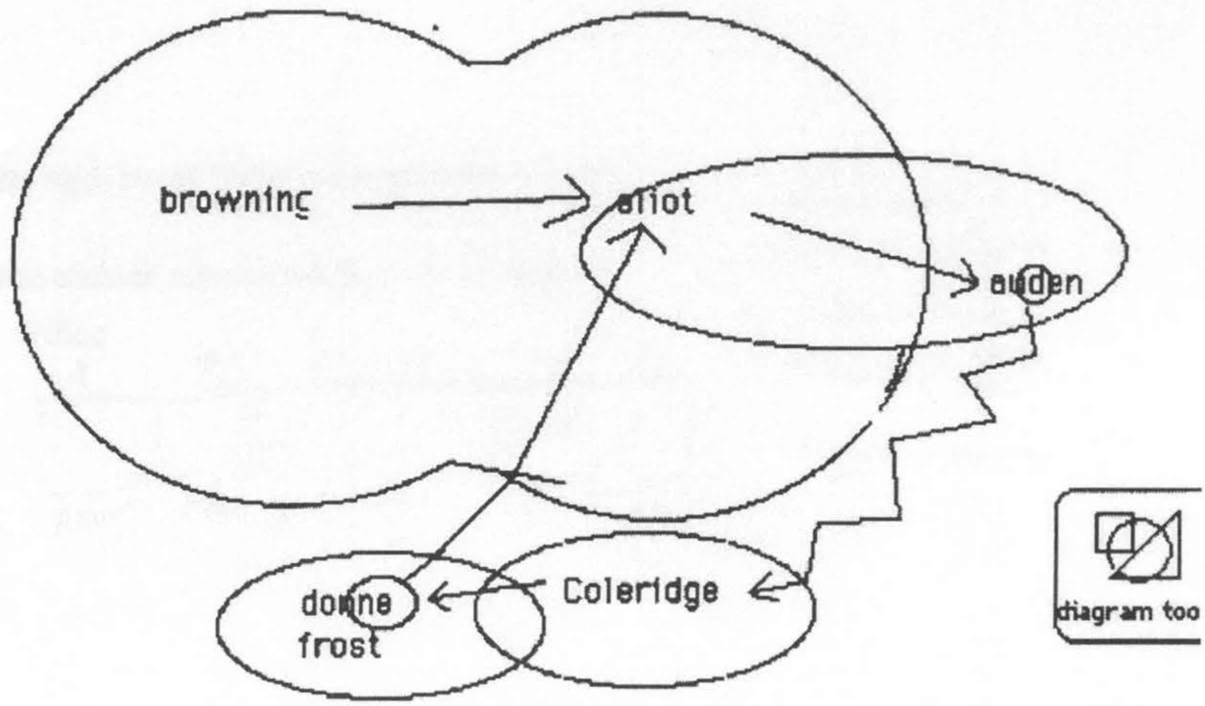
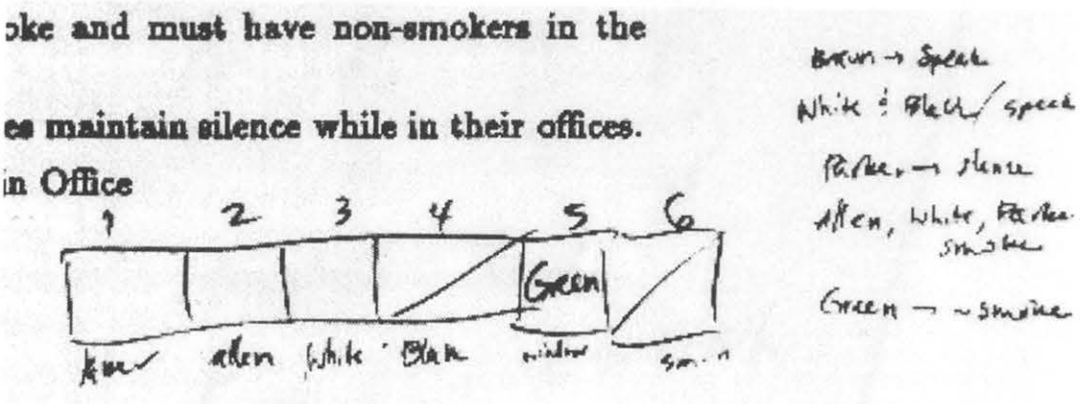


Figure A.18: Subject using *switchERI* diagram tool to construct ‘set diagram’ — annotated with various arrowed lines

The Use of Multiple Representations



ɔke and must have non-smokers in the

es maintain silence while in their offices.

ın Office

Brown → Speak

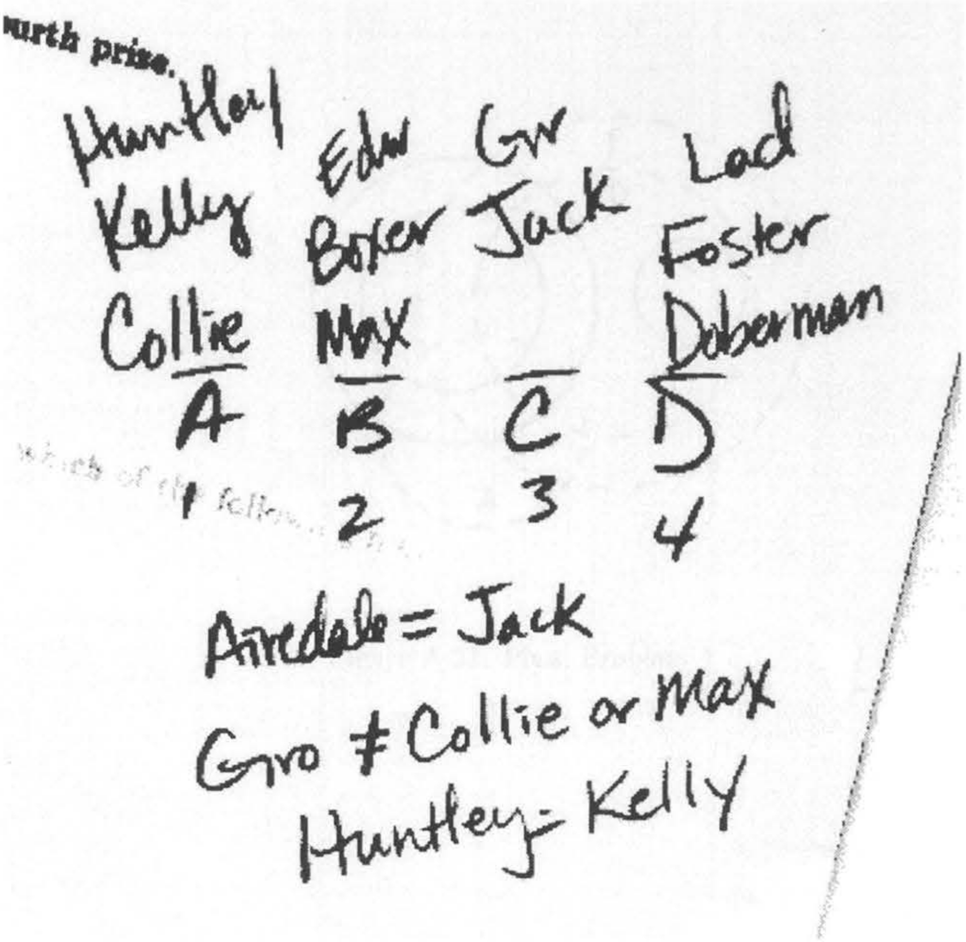
White & Black / speak

Parker → silence

Allen, White, Parker smoke

Green → ~smoke

Figure A.19: Use of multiple representations on Problem 1 — restricted logic plus plan



urth prize.

Huntley	Edw	Grv	Lad
Kelly	Boxer	Jack	Foster
Collie	Max		Doberman
A	B	C	D
1	2	3	4

Airedale = Jack

Gro ≠ Collie or Max

Huntley = Kelly

Figure A.20: Use of multiple representations on Problem 3 — textual notes plus plan

Representation recommended by Brownstein et al. for Problem 2

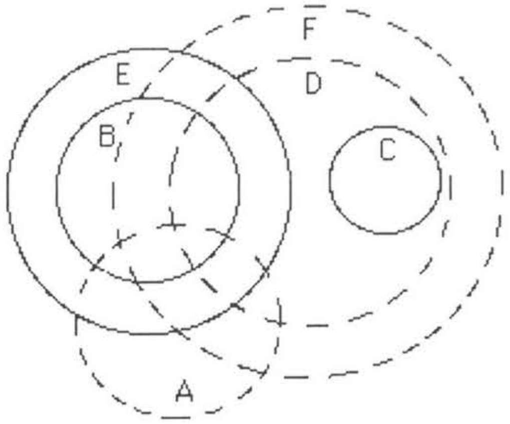


Figure A.21: Plan, Problem 1

Eighty-seven stimulus items used in the *switchERI* study taxonomy pre-task

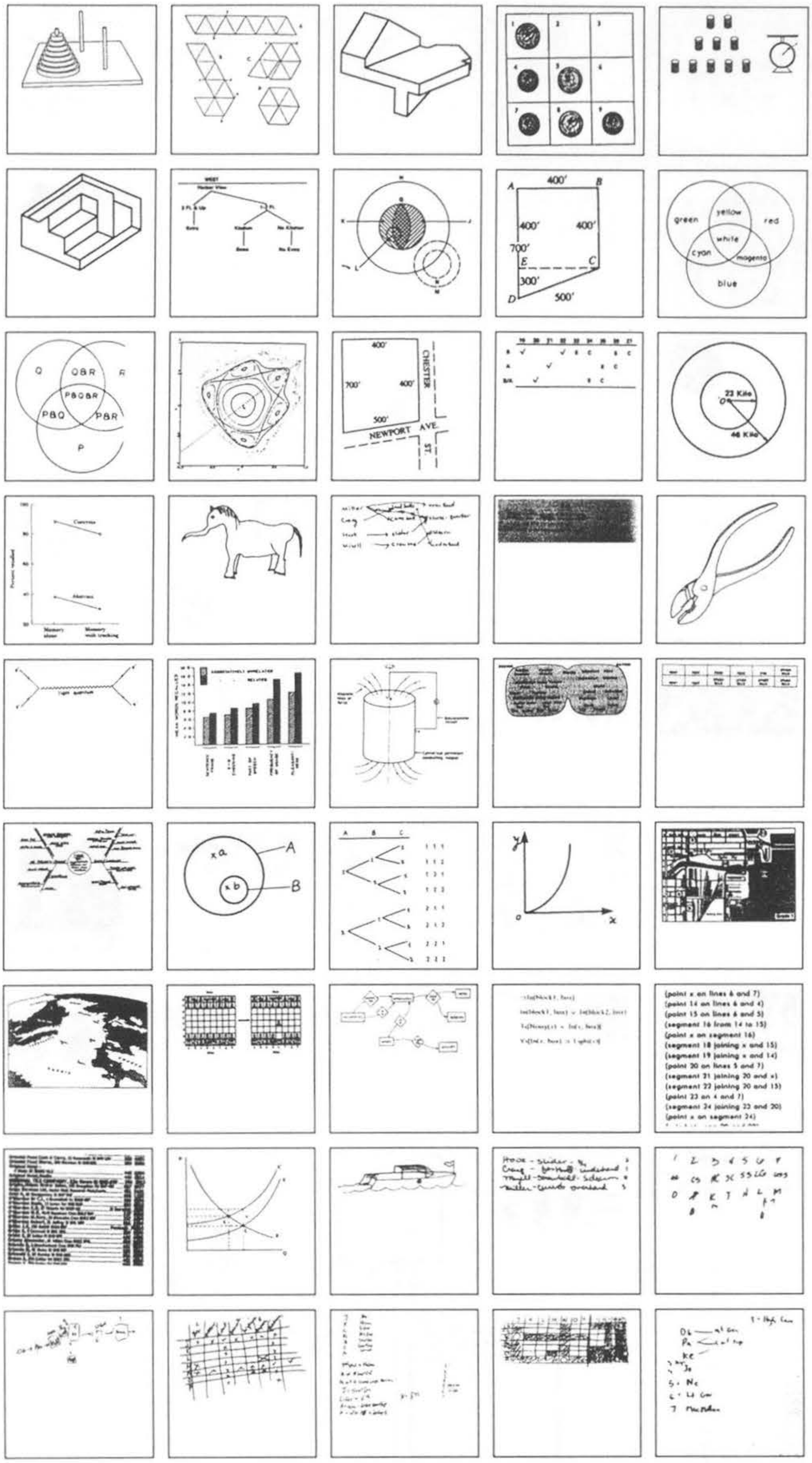


Figure A.22: Taxonomy task items 1 to 45, 1 is top left, 5 is top right, 45 is bottom right

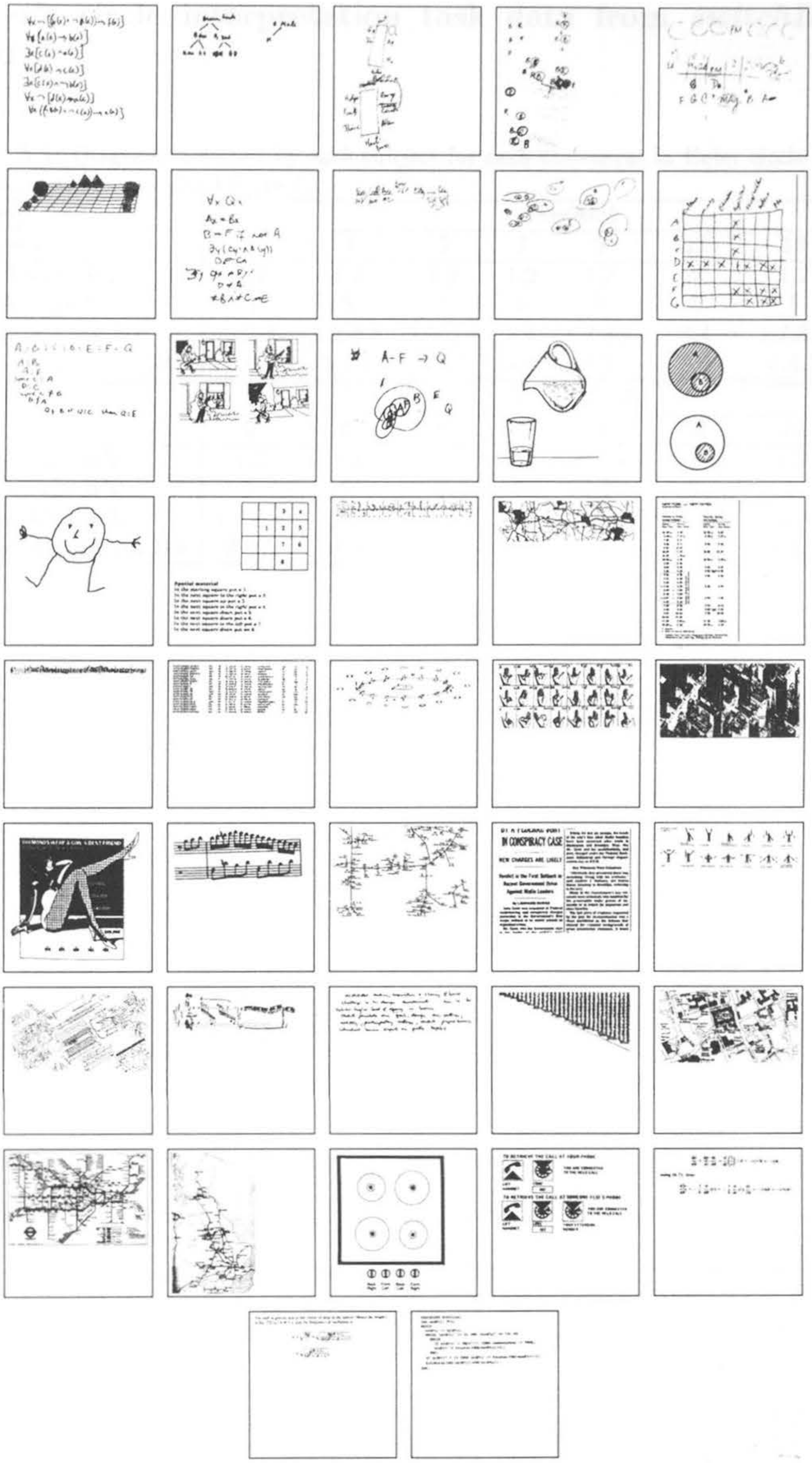


Figure A.23: Taxonomy task items 46 to 87 (item numbers correspond to numbers at 'leaves' of dendrogram shown in Figure 6.6).

Euler's circle interpretation task data from *switchERII* study.

Table A.1: Diagrams selected by each subject for each statement in Euler circle interpretation task - see also Figure 7.2.

Group	*Feedback*							
Subject	1	3	5	7	9	11	13	15
All A's are B's	1,2	1,2	1,2	1,2	1,2	2	1,2	1
No A's are B's	5	5	5	5	5	5	5	5
Some A's are B's	1,2,3	1,2,3,4	1,2,3,4	3,4	1,2,3,4	3,4	1,2,3,4	2,3,4
Some A's are not B's	3,4,5	3,4,5	3,4,5	3,4	4,5	3,4	3,4,5	3,4,5

Group	*No feedback*							
Subject	2	4	6	8	10	12	14	16
All A's are B's	1,2	1,2	1,2	1,2	1,4,5	2	1,2	1,2
No A's are B's	5	5	5	5	2,3	5	5	5
Some A's are B's	1,2,3,4	2,3,4	3,4	3,4	2	1,2,3,4	1,2,3,4	1,2,3,4
Some A's are not B's	3,4,5	3,4	3,4,5	2,4	3	3,4,5	3,4,5	3,4,5

Appendix B

Publications bound into thesis

Items marked with an asterisk are bound into the thesis following this page.

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

Cox, R., 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.

*__Cox, R., Stenning, K. & Oberlander, J.__ [1995] The effect of graphical and sentential logic teaching on spontaneous external representation. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, **2(4)**, 56–75.

Oberlander, J., Cox, R., & Stenning, K. [1995a] Proofs as discourse: an empirical study. In Working Notes of the *AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Stanford, March.

Oberlander, J., Cox, R., & Stenning, K. [1995b] Proof styles and multi-modal reasoning. In Seligman, J. & Westerstahl, D. (Eds.) *Language, Logic and Computation: The 1994 Moraga Proceedings*, Stanford: CSLI Publications, 403–414.

*__Oberlander, J., Cox, R., Monaghan, P., Stenning, K. & Tobin, R.__ [1996] Individual differences in proof structures following multimodal logic teaching. To appear in the *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates.

*Stenning, K. & Cox, R. [1995] Attitudes to logical independence: traits in quantifier interpretation. In J. D. Moore & J. Fain Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum & Associates, 742–747.

Stenning, K., Cox, R., & Oberlander, J. [1995] Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. *Language and Cognitive Processes*, **10 (3/4)**, 333–354.

*Stenning, K., Yule, P. & Cox, R. [1996] Quantifier interpretation and syllogistic reasoning: An individual differences account. To appear in the *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates.

■特集一認知における内的，外的資源　　特集論文

The effect of graphical and sentential logic teaching on spontaneous external representation

Richard Cox, Keith Stenning, & Jon Oberlander

A study of two logic courses employing different modalities of information presentation (Stenning, Cox, & Oberlander, 1995) demonstrated improvements of general reasoning ability as measured by Graduate Record Exam (GRE) type analytical ability reasoning pre- and post-course tests, as well as interactions between students' pre-course aptitudes and modality of teaching. This paper investigates the reasoning processes involved in the students' solutions of one sub-scale of the GRE problems from that study by analysing their 'work-scratchings' on analytical reasoning (AR) items. These data are used to examine changes in what representations students select; their association with correct and incorrect solutions; the changes in selection brought about by teaching different kinds of students in different kinds of courses; the association between these changes and improvements in solution performance; and the relation between intuitive teaching recommendations and a theoretically motivated taxonomy of representations.

Stenning & Oberlander (1995) present a theory of the cognitive differences between graphical and sentential representations which ascribes major cognitive properties of graphics to *weakness of expressiveness.* We apply this theory to the GRE AR problems and derive principled predictions of some constraints on the appropriateness of representations for problems. Analysis of the students' spontaneous representation selections shows that representational strategies do change differentially as a result of different teaching methods; the kinds of representation proposed by intuitive teaching recommendations as embodied in 'crammers' are globally correlated with success at solution; the theoretically based predictions of appropriate representations based on weakness of expression make rather better predictions that can be related to individual differences between students known to be important predictors of performance. These results are argued to have important practical pedagogical implications.

Keywords: problem solving, self-constructed representations, diagramatic reasoning, visualization, individual differences, spontaneous representations.

1. Introduction

The present paper is a subset of a larger study — a field evaluation of different ways of teaching elementary logic to undergraduate students.

This study had the goal of testing a theory of the contrasting cognitive properties of sentential and graphical representations. But the study also had the practical goal of evaluating a new interactive computer environment *Hyperproof* (HP) (Barwise & Etchemendy, 1994) as a method of teaching first order logic (FOL). Stenning, Cox, &

論理学に関する図と文による教示が外的表象の自発的生成に及ぼす効果, リチャード・コックス, キース・ステニング, ジョン・オーバーランダー (人間コミュニケーション研究センター, エジンバラ大学).

Oberlander (1995); Cox, Stenning, & Oberlander (1994); Oberlander, Cox, & Stenning (1994) and Oberlander, Cox, & Stenning (1995) provide accounts of the main study. In this paper, we focus upon the relationship between the expressiveness of representations and their effectiveness in problem solving.

One group of Stanford undergraduates was taught with HP. Information in HP is presented in two forms, as sentences of logic and as diagrams depicting 'blocks worlds'. A comparison class of students was taught syntactically, using a sententially-based natural deduction system (*i.e.* in the traditional way). The HP students and the comparison class students were given two tests before and after 12 week logic courses. One of these tests was a pseudo-GRE Analytical Ability test. Stenning, Cox, & Oberlander (1995) show that students in both the traditional natural deduction control course and the HP course show substantial (14%) increases in GRE AR pre- to post-test scores. The GRE is designed to be an uncoachable test. The GRE test sets problems at least superficially quite unlike the problems encountered in the courses, but which are known to correlate with general reasoning ability. We regard them as a useful, if imperfect, test of transfer of learning. The same study uncovered strong interactions between pre-test aptitudes for solving AR problems, the kind of logic course taken, and subsequent pre- to post-test score changes on the 'blocks world' (BW) test.

When doing the GRE AR test students were allowed to make whatever rough workings they chose on the test sheets. Students were advised that 'drawing diagrams might be helpful to them in answering some of the questions', just as in real GRE tests. We refer to these rough workings as 'work-scratchings'. The purpose of this paper is to use these work-scratchings from the GRE AR test to throw additional light on students' reasoning processes and the changes brought about in them by logic teaching. Whereas the main contrast in teaching interventions between HP and traditional logic courses is a contrast between representations *presented* to students, these work-scratchings are representations freely *constructed* by students. But they can be subjected to similar theoretical analysis.

Our original interest in HP was in its use of heterogeneous representations, both graphical and sentential. The HP environment provides graphical 'blocks-world' representations of models of sets of sentences of FOL, along with rules of inference for 'moving' information back and forth between diagram and sentences. We had been developing a theory of the distinctive cognitive properties of graphical representations choosing Euler's graphical method of teaching syllogistic logic as our example domain (Stenning & Oberlander, 1995). That theory develops the classical observation that many graphical systems are *weakly expressive* in the technical logical sense. That is, there are many abstractions which they cannot express, at least without indefinitely large disjunctions of diagrams. Computational theory shows why weakness of expressive power allows tractable reasoning for any reasoner, whether human or machine (*e.g.* Levesque, 1988). For an example of the implications of inexpressiveness for tractability of reasoning the reader is referred to the discussion of the semantics of HP below. Our theoretical motivation for studying HP was to extend this theory of graphics from the analysis of Euler's Circles to a larger domain using quite different graphical devices, and to study a more realistic learning situation.

Our theory provides an analysis of HP graphics which predicts that certain aspects of their semantics will be critical for determining their impact in teaching. This theory can also be applied, at least in broad outline, to the work-scratching representations spontaneously constructed by students as we will illustrate here. One of the appeals of using the GRE AR test for an investigation of graphical and sentential teaching methods, with a theory based on a logical analysis of expressive power, is that the GRE is a wholly *verbal* test in its presentation of problems, and in its collection of responses. Our analysis of the usefulness of graphics looks for the benefits in terms of *semantic* properties as opposed to simply *perceptual* ones.

The GRE analytical ability scale actually consists of two subscales: logical reasoning (argument analysis) and analytical reasoning (AR – constraint satisfaction problems). Items in both subscales are presented sententially. This paper focusses on the work-scratchings produced by subjects in the course of their solutions to items on the AR subscale. Performance on the logical reasoning (argument analysis) subscale is not the focus of this paper and will not be discussed.

The AR subscale items consist of constraint satisfaction puzzles of the kinds illustrated in Figures 2 and 3. In fact, there are two kinds of AR subscale item which we refer to as *determinate* and *indeterminate* problems. Determinate problems state sufficient constraints to determine a unique satisfying model. Figure 2 presents an example determinate problem. A set of constraints is stated which in fact determine a single unique model which satisfies them all. Because there is a unique model, it is possible to construct a diagram representing the problem information. Although this particular example is about spatial relations, the critical determinant of whether or not a diagram can be drawn is whether there is a unique model.

Despite the verbal surface of the AR subscale, it turns out that its two types of item (determinate and indeterminate) are composed of problems which differ in their model theoretic properties in just the way that our theory predicts should distinguish problems for which graphical approaches are appropriate. The students' work-scratchings therefore allow us to apply the same general concepts to the pre- and post-test performance that we apply to HP itself.

One advantage of the work-scratching data is that it allows examination of students' spontaneous selection of external representations (ERs), and their private[1] use of representations at pre-test, and effects of different teaching on their selections at post-test. Any changes in representational choice can

---

[1] Subjects in the study were unaware that their workscratchings would be seen or analysed later by the experimenters. In most previous studies of subjects' use of ERs, (*e.g.* Schwartz, 1971) subjects have been encouraged to 'show their working'.

then be related to changes in AR scores, and to students' other performances. Although there is awareness amongst the AI and mathematical communities that finding good representations for problems is critical to reasoning success (see *e.g.* Polya 1957; Simon, 1981; Kaput, 1987; Kaput, 1992), not many studies of students' spontaneous use of representations and effects of teaching on these habits have been reported. Amarel (1986, 1990) pioneered the AI study of representation selection. His conceptualisation of the problem as one of describing a space of representational *systems* from which reasoners must select, and within which they perform their reasoning, is an important contribution in itself. Our own approach to a theory of differences between graphical and sentential modalities of representation can be seen as another approach to Amarel's questions.

The difference between what a *token* representation (a particular diagram or piece of text) represents, and what the *system* of representation of which it is a member forces its users to represent, is initially subtle, but nevertheless far-reaching. In general, textual systems always *can* specify a piece of information, but they allow their users to leave information *unspecified.* Graphical systems, in contrast, often enforce the representation of some classes of information. Both the advantages and disadvantages of the modalities stem from this difference, as we shall see below. We focus not on what systems *can* represent, but on what they *must* represent. This distinction can only be drawn at the *system* level. Unless a diagram or a text is conceived of as a member of a system of possibilities, there is no basis for defining what all tokens must represent. Only when *systems* of representation are distinguished from their tokens can the issue of selecting between them emerge.

This emphasis on the system can be clarified by contrasting our approach with the well-known approach to media/modality differences of Larkin & Simon (1987). Their approach emphasises differences between what they call *informational* and *computational* equivalence of token representations. A token text and a token picture may represent the same information, but one may make it much easier to compute inferences from this information than the other. For example, graphical representations allow the parallel searching mechanisms of our eyes to find relevant items of information much more quickly and easily. We do not disagree with these observations, but suggest that they result from deeper differences between the *systems* of representation concerned. Implementations of fast parallel search are possible for graphical systems *because* these systems are logically inexpressive. For example, if a system cannot denote the same thing with alternative expressions, this enormously simplifies search. Graphics generally cannot do this: languages generally can. But if one takes very circumscribed sentential languages which cannot express these abstractions, then there will be computational implementations which allow just as facile inference to be performed on their representations as is the case with graphics. One has to know what system of representations a token is drawn from in order to know which general computational regimes will be applicable. We believe that this shift of emphasis onto reasoning as *selecting representation systems* affords insight into human behaviour as well as issues of machine design.

The plan of the paper is as follows. We begin by sketching the theoretical approach to graphics, and its application to both HP and the range of representations in the students' work-scratchings. We then describe the relevant methodology of the study and present the results of using the work-scratchings analysis to explore students' test-scores, and changes in test-scores with logic teaching. Finally, we discuss both theoretical and practical implications of these findings.

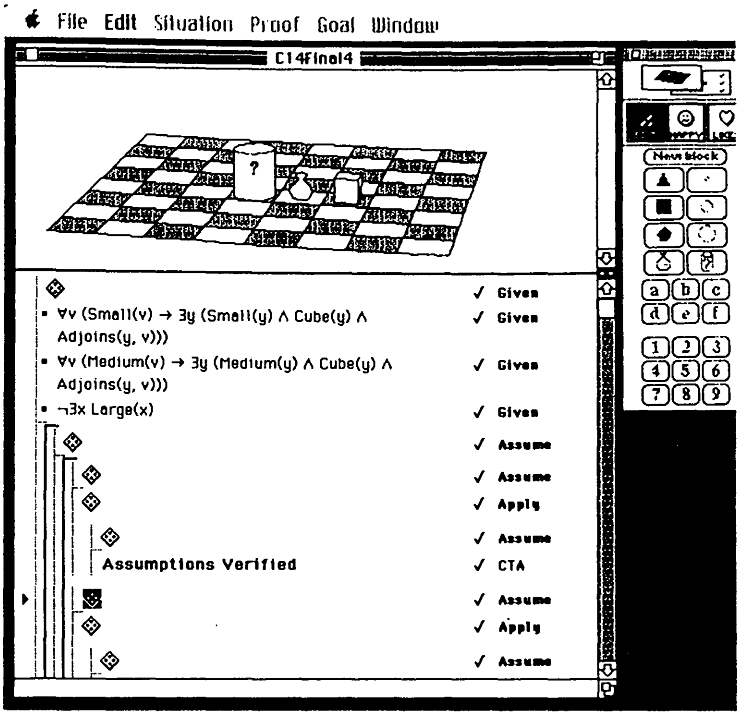
2. The semantic properties of graphical representations

We now give an illustration of the application to HP graphics of our analysis in terms of expressiveness and tractability. Stenning & Oberlander (1995) and Stenning & Inder (1995) give more general accounts. Stenning & Oberlander (1991) give an elementary application of the theory to tabular representations which is especially relevant in the current context.

The HP universe of discourse is a domain of polyhedra which have shapes (tetrahedron, cube, dodecahedron), sizes (small, medium and large), and positions on an eight by eight chequer board. Graphical representations of HP 'worlds' consist of diagrams of the chequerboard with icons differing in shape and size in the obvious way, placed singly on squares on the chequerboard. The icons may or may not have one or more labelling letters on them, but no two icons ever have the same label. Particularly interesting features of the diagramatic system of representations are some 'tricks' for expressing limited, but nevertheless useful, abstractions. Figure 1 shows an example of an HP graphic.

One type of icon is a cylinder. Note that there are no cylinders in HP worlds. Cylinder icons stand not for cylinders, but for polyhedra of unspecified size and shape. There is only one size of cylinder icon, but cylinder icons can bear badges indicating the shape of the denoted object. Another abstraction device is a paper bag icon which stands for a polyhedron of unspecified shape. Paper bag icons come in three sizes denoting the three sizes of their contained polyhedra. Another abstraction trick is an off-board area known as Tombolia in which icons may be placed. Icons in Tombolia denote objects that are not in Tombolia but at some unspecified position on the board. Finally, there is the possibility of having a set of HP diagrams which represent alternative possible HP worlds. They are essentially disjunctions—the world is as in one or another of the individual members of the set.

In the case of this specific graphical system it is easy to illustrate what is meant by the limited expressiveness of graphics as compared to the FOL sentences. The HP graphical system easily expresses some facts, say that there is a cube which is either small, medium or large in the front right hand corner, or the fact that there is a small cube somewhere. But it is not possible to express the fact that there is a small cube in the 3rd row from the front, or that there is a cube



◈ ✓ Given
- ∀v (Small(v) → ∃y (Small(y) ∧ Cube(y) ∧ Adjoins(y, v))) ✓ Given
- ∀v (Medium(v) → ∃y (Medium(y) ∧ Cube(y) ∧ Adjoins(y, v))) ✓ Given
- ¬∃x Large(x) ✓ Given

◈ ✓ Assume
◈ ✓ Assume
◈ ✓ Apply
◈ ✓ Assume
Assumptions Verified ✓ CTA
◈ ✓ Assume
◈ ✓ Apply
◈ ✓ Assume

Figure 1 The Hyperproof (HP) interface. The main window panes—graphical and calculus—are supplemented by control palettes. The situation being viewed is the fifth in the course of the proof, and corresponds to the fifth diamond-shaped 'situation' icon in the body of the proof. The graphical window pane contains three symbols of varying degrees of abstraction.

that is small or medium but not large. A particularly interesting class of abstractions inexpressible in HP graphics are indeterminacies of identity. It is not possible to state that there is something that is large in the back row, and something that is a cube in the back row, without specifying whether or not they are the same thing. Strictly speaking, these propositions can be expressed in HP graphics, but only by forming enormous disjoined sets of all possible worlds consistent with the information. Although these sets would be technically finite, they may contain many millions of full diagrams and would be quite impractical for reasoning.

Conversely, it is easy to see how the ability to express arbitrary abstractions (as FOL can) may rapidly lead to inferential intractability. For example, if there are a number of polyhedra about which we have incomplete and logically independent abstract descriptions, we may have to make immensely complex inferences about their relative identities in order to carry out simple tasks. In graphical representations, this problem does not generally arise. For many tasks which do not require the expression of these abstractions, the diagrammatic representations which do not allow them to be expressed in the first place will be much more efficient. On the other hand, if a task requires an abstraction for efficient reasoning (because a very large number of alternative cases otherwise have to be listed) then graphics will be pathological. Generally, the best representation will be one which has sufficient power for the task, *but not any more power than is needed.*

3. The GRE problems and the workscratching representations

As mentioned earlier, the GRE analytical ability scale itself is divided into two subscales— 'logical reasoning' and 'analytical reasoning' (AR)[2]. Logical reasoning items require argument analysis and verbal reasoning skills for their solution and are not the focus of this paper. The AR items consist of constraint satisfaction puzzles of the kinds illustrated in Figures 2 and 3. In fact, there are two kinds of AR item which, as we mentioned earlier, we refer to as *indeterminate* and *determinate* problems. Many of the determinate and indeterminate AR problems do not deal with spatial relations, but it is nevertheless possible and often helpful to construct diagrams for them. Set diagrams which use inclusion within closed curves to define sets are a common example.

There is also a close association between diagrams and tabular representations (see Stenning & Oberlander (1995) for an extended discussion). The row and column headings of a table enforce simultaneous representation of their classes of information. To take an example GRE problem, a table of information with dog breeds as column headings and their owners names as row headings enforces the simultaneous representation of both owner and breed on any element represented. It is not possible to express the existence of a dog of some breed without determining its owner, or *vice versa.*

The reason that the properties of having a unique model, and of being diagrammable are related is because diagrams are inexpressive. If there are many alternative models of the constraints, then a diagram will only be possible with abstraction 'tricks' which happen to capture the right set of models, and it is usually difficult and often impossible to find such tricks.

Most of the reasoning necessary to solve the problem in Figure 2 can be embodied in the process of constructing a diagram of office allocations. Reading the answers to the more straightforward questions off a correct diagram is relatively trivial. Even for the questions which demand consideration of alternative models, a representation of the unique initial model is a powerful aid.

Graphical reasoning generally places the burden of inference on the processes of representation *construction, de-construction and re-construction*[3].

AR indeterminate problems are more miscellaneous, but none of them present constraints which determine unique models, and constructing a diagram is generally impossible or at least not useful. Figure 3 presents an example indeterminate problem.

A few AR problems are intermediate between determinate and indeterminate problems in that although they consist of a set

---

[2] Logical reasoning and AR items are mixed together unidentified in the analytical ability scale of the GRE.

[3] The extent to which a representation resists modification is an important cognitive dimension proposed by Green (*e.g.* 1989).

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

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

(2) The best employee to occupy the furthest office from Mr Black would be Allen, Braun, Green, Parker or White?

(3) The three smokers should be placed in offices 1, 2, & 3, or 1, 2 & 4, or 1, 2 & 6, or 2, 3, & 4, or 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.

Figure 2 Example of a determinate AR problem and associated questions.

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

Figure 3 Example of a indeterminate AR problem and associated questions.

of constraints they do not determine unique models. Sometimes diagrams may be helpful for these problems in representing models of sub-sets of constraints and thereby guiding choice of which subsets of premisses are required to answer which questions.

So AR determinate problems are ones for which an inexpressive diagram is useful because there is a unique model solution and a diagram can represent this solution. Problems are presented in highly expressive linguistic modality and one solution method is to find a weakly expressive representation. This strategy requires judgement of which problems are determinate problems—they are not labelled in any way in the test.

To see whether a problem has a unique model, and so decide whether to construct a diagram, is not trivial. There are some problems which carry immediately accessible cues to the fact that constructing a model is impossible or inappropriate. For example, the nature of the questions may make this clear. But often the only method for deciding whether there is a unique model is to attempt to specify one. In this case, there are some useful rules of thumb as to how to go about this process. It is best to start from any determinate information. In the example, this is the information that 'Mrs. Green is entitled to Office 5'. Then any information that can be linked to this 'anchor' should be incorporated. In the example in Figure 2, the only two non-smokers must flank Green's office, and these two alternatives provide a major 'split into cases'. The further information is then used to find out which of these classes of cases must contain the correct model. The reason why it is good to use determinate information early is that it leads to a pruning of the search space in which model construction proceeds.

Because the sequence of use of information in solving a problem can be critical, the test setters frequently manipulate the surface sequence of constraints so that the best order of inference is obscured. Observing students solving these problems reveals that domination by the presented sequence where this is not a good solution sequence is a common difficulty.

A descriptive study of subjects' knowledge of representational formats (Cox & Brna, 1993) used sorting methods to empirically arrive at categories for classifying students' work scratchings. The categories emerging from that study, and from one by Schwartz (1971), were: *matrices/tables*, *set enclosure diagrams*, *sentences reordered*, *networks*, *informal grouping/ordered texts*, *logic/formal sentential notations*, *miscellaneous*, and *no representation*.

Our theory did not predict that the other categories of representation would be useless for determinate problem solution. Because the reordering of the premisses in the inference process is often critical, we would expect even re-writing the questions in certain ways to be useful for some problems. But in general we would expect them to contribute to only part of solution processes.

Normative schemes of representation for GRE problems are available in published 'crammers' (*e.g.* Brownstein, Weiner, & Green, 1990). These works do not typically provide general rules about what representation to use, but they do at least provide a representation of 'folk teaching wisdom' against which we can compare students' spontaneous constructions, and the predictions of our theoretical analysis.

In summary, earlier analysis of this study shows that that logic teaching of both a traditional kind and using HP does generally enhance students' scores on GRE AR and BW tests. The sort of representations imposed on the students by teaching regimes interact with students' pre-course aptitudes, in determining post-course changes in reasoning, and in proof-styles developed in course.

This study focusses on the relation between spontaneous constructions of styles of representation on AR items of the GRE tests, and these other performances. In particular it asks whether the same theoretical analysis of representations in terms of their expressive power is insightful for these self-produced representations.

4. Overview of the Study

Methods

Stenning, Cox, & Oberlander (1995) give a fuller account of the main study. An overview of the relevant features of the method is presented here for the reader's convenience.

Two logic courses, one based on HP and the other on traditional syntactic natural deduction teaching were given to two groups of first-year Stanford undergraduates. Complete pre- and post-course data was obtained for 16 of the 22 subjects in the HP group and for the 13 subjects in the Syntactic group. Assignment to courses could not be strictly randomised because the courses had to run in different semesters, but the course descriptions by which students chose them were not differentiated in any way by the content or method of teaching. Students could not know that one course was 'graphically oriented' ahead of signing up. To control for the motivational effects of computer-use, the 'traditional' class also used a special version of HP that had the graphics window disabled, leaving only the 'sentential window' containing the representations on which traditional teaching is based. The traditional course used Bergman, Moor & Nelson (1990) as a text: the HP course used the lecture notes which subsequently became Barwise & Etchemendy (1994).

Two tests of reasoning were developed in order to measure the effects of teaching by comparing pre- and post-course score: the GRE test and the BW test. This report focusses upon the kinds of external representations (ERs) used by subjects in their responses to determinate and indeterminate AR items of the GRE test such as the examples in Figures 2 and 3. All students took parallel forms of GRE and BW test before and after their logic course. All the tests had a time limit. Students were free to select which items they attempted within the time. Students knew that the purpose of the tests was purely to aid our research in assessing the teaching methods, and that neither their GRE scores nor their workscratchings would be available to the teachers, or affect their course assessment.

Summary of results

In brief, students were classified as *model-hi* or *model-lo* reasoners before the logic course on the basis of their scores on the GRE AR problems in the pre-test. There was overall improvement of performance on GRE determinate items from pre- to post-test for both logic courses.

The model-hi students responded to HP teaching differently from model-lo reasoners in that their rate of improvement on 'blocks world' reasoning was significantly higher. However, the same kind of students (model-hi) in the syntactically taught class actually declined in their BW test performance relative to their model-lo counterparts. The general effects of logic training upon pre- and post-test measures have been reported in detail elsewhere (Cox, Stenning, & Oberlander, 1994; Stenning, Cox, & Oberlander, 1995). Evidence from the logging software recordings made during the computer-based HP exam suggests that the individual differences are reflected at the level of the structure of logical proofs that students build. Detailed results of the proof-log analyses can be found in Oberlander, Cox, & Stenning, 1994; Oberlander, Cox, & Stenning, 1995).

5. Analysis of Workscratching data

We first describe how the workscratching data was categorised and scored and then present the results relating representations to other performance.

Method

As mentioned earlier, the categories adopted for this study were derived from Cox & Brna (1993) and Schwartz (1971): *matrices/tables*, *set enclosure diagrams*, *sentences reordered*, *networks*, *informal grouping/ordered texts*, *logic/formal sentential notations*, *miscellaneous*, and *no representation*. Of these categories, the first two are constrained in their expressiveness. Theory would expect these two kinds of representation to be particularly useful for solving determinate problems. Because they are weakly expressive, neither is generally capable of representing problems that do not have a unique model. These categories accord closely with the range of representations recommended by the crammer authors. It seems that students and authors share a common categorisation.

Since our focus of interest was whether *representation selection accuracy* (RSA) improved as a result of logic teaching, and how RSA related to reasoning performance, some normative scheme of representation selection for each problem was required. Rather than initially imposing our theoretical analysis for this purpose, we adopted the normative scheme of the crammer, and return later to compare this with our theoretical constructs. RSA score was defined as the number of correct representation selections made by the subject on 6 AR problems. Independently, two raters categorised the work scratchings on subjects' test papers. For each subject there were potentially[4] six determinate reasoning items—three on the pre-test and three on the post-test, yielding scores that range from 0 to a maximum of 3 on each sub-test for each subject. Raters were given a shuffled stack of test papers and asked to sort them into the following categories of ERs: *matrix/tables* (explicit row by column organisation of dimensions); *sentence re-write* (re-writing sentences in a different order); *network graphics/directed graphs*; *set diagrams*; *informal groupings/ordered text* (information circled, placed in close proximity, connected by hyphens, etc.); *logic/notations*; *miscellaneous* (those that did not fit into other categories) and *no representation.*

The raters were instructed to place each work-scratching into an appropriate category. They were allowed as much time as they needed to complete the sort. If more than

---

[4] Assuming that the subject attempted all of the items.

one representation was evident in a student's solution, the raters were instructed to classify the one that the subject 'probably used' in their solutions. Inter-rater agreement was 83% on the 'work-scratchings' categorisation task. The two raters disagreed on 29 items (17%). To resolve the disagreements, the raters were subsequently asked to cooperatively sort the 29 items into mutually agreed categories.

Results We first report the effects of teaching modality on RSA scores. We then report on the relationship between RSA score and reasoning success at AR problems. Finally, we turn to examine our theoretical claim about the role that weak expressiveness plays in determining the usefulness of a representation for reasoning.

To examine changes in RSA score as a function of teaching modality we categorised subjects' RSA performances on the pre-test, and on the post-test.

'Poor' RSA was defined as failure to select recommended ERs on half or less of the items attempted. 'Accurate' RSA was defined as accurate representation selection on more than half the items attempted. For each group (HP and Syntactic), the number of subjects showing a change from poor to accurate RSA from pre- to post-test was calculated together with the number of subjects who showed the reverse change.

In the HP group, 1 subject changed from accurate to poor, 7 subjects were accurate at both pre- and post-tests, 1 subject was poor on both pre- and post-tests and 7 subjects changed positively from poor to accurate. In the Syntactic group, 2 subjects changed from accurate to poor, 6 subjects were accurate at both pre- and post-tests, 2 subjects were poor on both pre- and post-tests and 3 subjects changed positively. Of the 7 positively changing subjects in the HP group, 4 were model-hi reasoners and 3 were model-lo reasoners. Of the 3 positively changing subjects in the Syntactic group, 1 was a model-lo reasoner and 2 were model-hi reasoners. A binomial (nonparametric) change test (Siegel & Castellan, 1988) revealed that a significant proportion of the HP subjects demonstrated positive change in RSA ($p < .05$). The proportion of positively changing subjects in the syntactic group was not significant ($p = .188$). Thus, for a significant proportion of HP subjects, the experience of learning logic graphically improved ER selection in a different kind of model-based reasoning domain.

Item-by-item analysis of response patterns provides some insight into the nature of these RSA changes. The proportion of each group (HP and Syntactic) that responded correctly was computed for each question associated with the problems. There were 13 questions in total on the pre-test and 13 at post-test. On the pre-test, the proportion of correct responses differed substantially[5)] between HP and Syntactic groups on only two questions (in favour of the Syntactic group). At post-test however, the patterning differences were much more marked. The HP group consistently outperformed the Syntactic group on 4 out of 5 questions on the first post-test

---

[5)] i.e. 15% or greater difference in proportions responding correctly.

AR problem[6]. Both groups responded with similar patterns across the 4 questions of the second post-test problem. On the the third problem, however, the pattern was reversed, with Syntactic group students outperforming the HP students on 3 out of 4 questions[7].

. Given that both groups improved similarly in terms of score from pre-to-post logic course, these results suggest that whereas at pre-test the scoring patterns were similar in the two groups, at post-test the groups differed in terms of the items that contributed to the post-test scores. The main impact of changes in representation selection strategies due to logic teaching appears to happen in the HP students. The Syntactic group shows improvements in reasoning which may be due to factors other than representation selection.

We now turn to consider the relation between RSA and AR reasoning success. Table 1 shows the relationship between the number of correct representations selected and score on each of 3 problems at pre- and post- tests grouped by teaching treatment (HP and Syntactic groups).

Table 1 Mean scores (out of 13) on AR items as a function of number of correct representations selected for pre- and post-test. Note that no subjects in either group used zero correct representations at either pre or post-test.

RSA	Mean	S.D.	N
Pre-test			
Hyperproof			
1	6.37	3.34	8
2	6.00	1.41	4
3	9.00	3.46	4
Syntactic			
1	6.20	1.30	5
2	7.80	2.59	5
3	9.33	2.89	3
Post-test			
Hyperproof			
1	5.33	3.79	3
2	8.86	1.68	7
3	8.67	3.20	6
Syntactic			
1	5.25	1.50	4
2	9.20	2.86	5
3	9.75	2.99	4

Table 1 suggests that there is a positive relationship between RSA and reasoning score for subjects in both groups, at both pre- and post-tests. Regression analysis of RSA score with reasoning score as the dependent variable reveals that there is a positive correlation between RSA and reasoning success if all subjects are pooled (pre-test (R = 0.40, $p < .05$); post-test ($R = 0.47$, $p < .02$)). Separate analyses of the teaching groups reveals that the strongest (and only significant) relationship in the four subanalyses is in syntactically taught students at post-test ($R = 0.59$, $p < .04$).

The model-lo subjects (both teaching groups combined) showed weak, positive, non-significant relationships between RSA and reasoning score at pre-test (R = 0.27, *n.s.*) and post-test (R = .33, *n.s.*). However, for model-hi subjects, there were strong and significant positive relationships at both pre-test (R = 0.54, $p < .03$) and post-test (R = 0.63, $p < .007$). This is not unexpected, since model-lo and model-hi subjects were *selected* on the basis of AR pre-test performance.

---

[6] That problem required a 4 by 3 tabular representation for effective solution.

[7] The third post-test problem required a seating-plan schematic and was similar to the pre-test 'office allocation' problem except that it was partially indeterminate.

RSA related to graphical semantics

Logic teaching, particularly HP teaching, does improve RSA and RSA, as scored by 'expert recommendations', is at least weakly correlated with correct solutions in GRE determinate-problem reasoning. But there remains the question of whether the expert recommendations can be related to the theory of the cognitive properties of graphics outlined in the introduction. Are weakly expressive graphical representations effective when abstraction is unnecessary, and ineffective when it is ? The expert recommendations of the crammer are more finely classified than the binary classification into strong and weak representations. To examine whether expressiveness is an important factor in efficacy, we classified the crammer's categories into weak representations incapable of abstractions, and non-weak representations capable of some abstraction. We also classified problems into those whose constraints defined a unique model (determinate problems), and those which merely defined a class of models (indeterminate problems).

Table 2 gives the median scores and numbers of subjects (HP and Syntactic groups combined) employing weak and strong representations on AR problems that they attempted. The representations defined as weak or non-weak for each problem were, for example, for pre-test problem 1 (the problem illustrated in Figure 2): weak ERs were plans (1 dimensional arrays); non-weak were ordered text and informal groupings. For pre-test problem 2 (the problem illustrateed in in Figure 3): weak ERs were set diagrams, tables; non-weak were network/directed graphs, logic, and text.

Table 2 Median reasoning scores on the six GRE AR problems classified by weakness of representation selected and by number of models defined by problem constraints.

	Median(n)	
Problem	Weak	Non-weak
Determinate		
Pre-test 1	3 (14)	3 (14)
Pre-test 3	5 (3)	2.5 (10)
Post-test 1	4 (26)	2 (3)
Indeterminate		
Pre-test 2	- (0)	3 (21)
Post-test 2	3 (2)	4 (24)
Post-test 3	2 (11)	3 (3)

Examination of the data in Table 2 reveals a general tendency for selection of a weak representation to be associated with a high score for determinate problems, and a low score for indeterminate problems. There is only one exception: pretest Problem 1 is equally successfully solved with either choice of representation. Subjects often appear to choose inappropriate representations as classified in this way.

Regression analysis of choice of representation strength against the dependent variable reasoning score revealed a significant negative slope for determinate problems ($R = -0.25$, $p < .04$) contrasted with a significant positive slope for indeterminate problems ($R = 0.33$, $p < .009$). This is evidence of the value of fitting the power of the representation to the type of problem predicted by our theory.

Unlike the analysis of RSA and reasoning score, this analysis of weak and non-weak representations used on determinate and indeterminate problems *does* reveal significant individual differences between the model-hi and model-lo groups of students, though *not* between the teaching treatments. The regression of strength of representation against determinacy of problem shows that there is a significant positive relation between model-hi subjects' reasoning scores and their weak/non-weak representation selection ($R = 0.60$, $p < .001$) on indeterminate problems. For model-lo subjects the regression of strength of representation against determinacy of problem shows a non-significant, negative relation between model-lo subjects' reasoning score and their weak/non-weak representation selection ($R = -0.15$, *n.s.*) on indeterminate problems. An examination of the score means revealed that, for model-hi subjects, the use of a weakly-expressive representation on indeterminate problems was associated with decreased scores to a much greater extent than was the case for model-lo subjects. Model-lo subjects scored equally well on indeterminate problems with either weak or non-weak representations. This suggests that for model-lo students, reasoning is less externalised — these students may be less 'ER sensitive' than their model-hi counterparts.

On determinate problems, neither model-lo students nor the model-hi students show any significant relationship between strength of representation and reasoning scores. The regression coefficients were $R = -0.33$, *n.s.* and $R = -0.13$, *n.s.* for model-lo and model-hi subjects, respectively. On the determinate problems, the score means indicate that model-hi subjects scored equally well with either weak or non-weak representations, but that model-lo subjects tended to score poorly when they attempted to use non-weak representations.

6. Discussion

Are the effects of logic teaching on general reasoning performances mediated by changes in representation strategies? And can the data of spontaneous representation selection reveal this mediation? Students' RSA as assessed by the 'crammer' scheme does improve between pre- and post-tests, more so amongst the HP students. Although this general improvement in RSA seems to be correlated with improved reasoning performance, it is not significantly related to the individual differences between subjects which are strongly related to changes in reasoning with logic teaching. However, when representation selection is assessed by the theoretically driven distinction between weak and non-weak representations, and their fitness for problem is assessed by the logical structure of the items (determinate or indeterminate), the nature of the effect of selection and the part that individual differences play in determining reactions to teaching emerge.

The scores of model-hi subjects show great sensitivity to representation strength on indeterminate problems, whereas model-lo subjects' scores show sensitivity to representation strength on determinate problems.

Why is RSA not more strongly associated with reasoning performance? A partial explanation is that the task requirements of the 4 or 5 questions associated with each AR problem vary quite widely. For example in the 'office allocation' example presented earlier, Cox & Brna (1995) have shown that subjects' use of plans, tables or no ER can be associated with good performance on questions 1 to 3. On question 4, however, subjects who do not use an ER tend to score better. Question 4 (see Figure 2) requires the subject to assess the impact of hypothetical changes to the originally given information. The task requirement is not one that is facilitated by straightforward read-off from an ER — there is a need to reason internally about alternative models and the number of individuals affected. Before the HP course was taught we predicted, on the basis of the semantic analysis of HP's abstraction 'tricks' that HP's impact on learning logic would hinge on students' abilities at using these devices to reason about models which are close relatives of specific depictions—in other words to extend completely concrete representations to express some limited abstractions. This prediction is born out in analyses of the different proof styles that emerge from HP teaching which are chiefly differentiated by use of abstraction symbols (Oberlander, Cox, & Stenning, 1995). The fact that students fail to develop ad hoc systems for representing not-quite determinate GRE AR problems within generally weak representational systems is perhaps not surprising.

Of course, there is more to reasoning than representation selection. Another source of score/representation dissociation is the correctness of the token representation. Cox & Brna (1995) found that 23% of a large sample of subjects constructed erroneous representations for the 'office allocation' problem. One error pattern accounted for a third of the errors observed. Furthermore, of the subjects who produced erroneous representations, just under half gave question responses that were fully consistent with their (wrong) representations. The remaining subjects seemed to be selective about which parts of their representation they 'believed' in. In other words, one interpretation is that those subjects showed some awareness of their representation'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) representations tended to score better than those whose answers were inconsistent with their (wrong) representations (Cox & Brna, 1995).

RSA as measured by crammer recommendations gives a less insightful analysis of reasoning scores than our theoretically motivated but simpler scheme predicated on weakness of expression interacting with determinacy of problem constraints. Several reasons might underlie this result. First, the crammer recommendations may simply be too fine grained. Functionally appropriate selections may be categorised as innappropriate simply because the crammer has a 'pet' representational choice. Secondly, crammer recommendations may actually be dysfunctional. There is evidence for both explanations. For example, there are strong semantic relations between the various categories of determinate representations, e.g. tables and set-diagrams (see Stenning & Tobin, in press), and between indeterminate representations, e.g. networks and logic (see Stenning & Inder, 1995). The evidence is that selections amongst these finer grained distinctions may not be important for these reasoning problems. Secondly, the crammer actually makes some recommendations which are hard to justify *simpliciter*. An interesting case is pre-test problem 2 (Figure 3) where the problem is quite radically indeterminate but where the crammer (Brownstein *et al.*, 1990) recommends set-diagrams. In fact, the crammer actually tries to invent a dotted-circle notation as an augmentation for set diagrams to allow the requisite abstractions to be expressed. This augmentation is semantically incoherent. This is not to say that set-diagrams need be completely useless for this problem. At a meta-level, they can be used to identify pairs of premisses (from the total of eight premisses) which form the syllogisms that when solved provide the answers. Indeed, these syllogisms can themselves be solved using set diagrams with some abstraction tricks. What set diagrams cannot do is to provide a representation of all eight premisses, even using dotted circles, which is what the crammer recommends.

The analysis of weak/non-weak representations applied to determinate/indeterminate problems reveals that what leads to really poor performance is attempting to use weak representations on indeterminate problems which require abstractions they cannot express. This is even more damaging to reasoning than trying to use non-weak representations that can express abstractions on problems that do not require them. This difference is visible in Table 2 and is reflected in the greater significance of the regression of representation selection on performance in the indeterminate problems. Some of this difference is undoubtedly a feature of the range of problem difficulty. Since these problems can quite frequently be solved by some students without any ER, and since even minimal reordering of premisses can serve as a useful representation, it is hardly surprising that non-weak representation of weak problems can prove less harmful to reasoning than trying to force an indeterminate problem into a weak representation that cannot actually express the information accurately.

The analysis of the individual differences between model-hi and model-lo subjects was revealing of the role of representations in analytical reasoning. Although model-hi students are defined by their ability on AR, the subset of these problems where their selection of representations is most strongly correlated with their reasoning success is in the partially indeterminate problems. These students appear to be adept at avoiding the use of weak representations for indeterminate problems. This pattern is consistent with the observation that these students are adept at using HP abstraction tricks to structure their HP exam proofs (Oberlander, Cox, & Stenning, 1995). Rather than thinking of them as 'visual' thinkers in the traditional folk phenomenology, we should think of them as adept at achieving some abstraction of representation, possibly by elaborations of graphical semantics.

The educational implications of taking ER selection seriously are potentially considerable. As long as representation selection is only pursued at the level of the crammer's rules of thumb, it remains hard to teach. Although the theory advanced here is coarse grained it is teachable. Determinacy of problem is well-defined, and so is weakness of expressiveness. The evidence presented here is that this coarse-grained theory is sufficient to mediate real differences in success at reasoning on these problems.

When it comes to asking what stance teachers should take to individual differences of the kind noted here, there are broadly two options. All students could be explicitly taught the same methods of representation selection or students could be encouraged to implicitly follow their existing representational modality preferences. The second position is compatible with the view that the cognitive style of the learner is relatively immutable, and that it is best to adapt instruction to style, rather than *vice versa*. This is the approach advocated by Snow based on studies of Aptitude-Treatment Interactions (cf. Snow, Federico, & Montague, 1980). To the authors' knowledge, only one study has demonstrated that the 'visualiser—verbaliser' dimension is responsive to educational intervention (Frandsen & Holder, 1969). The research presented here cannot decide between these alternatives, but it does show how further research might contribute to an answer to the question.

Perhaps a domain-independent 'graphics curriculum' should be devised and generally taught as advocated many years ago by Balchin & Coleman (1965). The authors tend towards the view that students should be encouraged to broaden their representational repertoires. We agree with Barwise (1993) that "efficient reasoning is inescapably heterogeneous (or 'hybrid') in nature" and with diSessa (1979, p.250), who, in a paper on learnable representations of knowledge, has written:

> 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.

Acknowledgements

The support of the Economic and Social Research Council for HCRC is gratefully acknowledged. The work was supported by UK Joint Councils Initiative in Cognitive Science and HCI, through grant G9018050 (Signal: Specificity of Information in Graphics and Natural Language); and by NATO Collaborative research grant 910954 (Cognitive Evaluation of Hyperproof). The third author is supported by an EPSRC Advanced Fellowship. Special thanks to John Etchemendy, Tom Burke and Mark Greaves for the irreplaceable part they played in making this research possible.

References

Amarel, S. (1986). On representations of reasoning about actions. In D. Michie (Ed.), *Machine Intelligence*, Volume 3, Edinburgh: Edinburgh University Press.

Amarel, S. (1990). Problem solving. *Technical Report LCSR-TR-152*, Department of Computer Science, Rutgers University, New Brunswick, NJ.

Balchin, W.G.V. & Coleman, A.M. (1965). Graphicacy should be the fourth ace in the pack. *The Times Educational Supplement*, Nov. 5th, 947.

Barwise, J. (1993). Heterogeneous reasoning. In G. Allwein & J. Barwise (Eds.), *Working papers on diagrams and logic*, Indiana University Logic Group Preprint No. IULG-93-24, May, 1-13.

Barwise, J. & Etchemendy, J. (1994). *Hyperproof.* CSLI Lecture Notes Number 42, CSLI Publications, Cambridge University Press.

Bergman, M., Moor, J., & Nelson, J. (1990). *The Logic Book.* New York: McGraw-Hill.

Brownstein, S. C., Weiner, M., & Weiner Green, S. (1990). *How to prepare for the* GRE. Barron's Educational Series, New York.

Cox, R. & Brna, P. (1993). The relationship between prior knowledge of external representations and analytical reasoning performance: Implications for the design of a learning environment. *Research Paper RP646*, Department of Artificial Intelligence, University of Edinburgh.

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

Cox, R., Stenning, K., & Oberlander, J. (1994). Graphical effects in learning logic: reasoning, representation and individual differences. In *Proceedings of the 16th Annual Meeting of the Cognitive Science Society*, Hillsdale,NJ: Lawrence Erlbaum Associates, 237-242.

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.

Frandsen, A.N. & Holder, J.R. (1969). Spatial visualization in solving complex verbal problems. *The Journal of Psychology*, **73**, 229-233.

Green, T.R.G. (1989). Cognitive dimensions of notations. In A. Sutcliffe & L. Macaulay (Eds.), *People and Computers V*, Cambridge University Press.

Kaput, J. J. (1987). Towards a theory of symbol use in mathematics. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Kaput, J.J. (1992). Linking representations in the symbol systems of algebra. In D.A. Grouws (Ed.). *Handbook of research on mathematics teaching and learning*, New York, NY: Macmillan.

Larkin, J.H. & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, **11**, 65-100.

Levesque, H. J. (1988). Logic and the complexity of reasoning. *Journal of Philosophical Logic*, **17**, 355-389.

Oberlander, J., Cox, R., & Stenning, K. (1994). Proof styles in multimodal reasoning. In J. Seligman & D. Westerstahl (Eds.) *Language, Logic and Computation: The 1994 Moraga Proceedings*. Stanford: CSLI Publications, 403-414.

Oberlander, J., Cox, R., & Stenning, K. (1995). Proofs as discourse: an empirical study. In Working Notes of the *AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Stanford, March.

Polya, G. (1957). *How to solve it: A new aspect of mathematical method.* Princeton, NJ: Princeton University Press.

Schwartz, S.H. (1971). Modes of representation and problem solving: Well evolved is half solved. *Journal of Experimental Psychology*, **91(2)**, 347-350.

Siegal, S. & Castellan, N.J. (1988). *Nonparametric Statistics.* New York: McGraw-Hill.

Simon, H. (1981). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Snow, R. E., Federico, P-A., & Montague, W. E. (Eds.) (1980). *Aptitude, learning and instruction - Volume 1: Cognitive process analyses of aptitude.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Stenning, K. & Oberlander, J. (1991). Reasoning with words, pictures and calculi: Computation versus justification. In J. Barwise, J.M. Gawron, G. Plotkin, & S. Tutiya (Eds.), *Situation theory and its applications (Vol. 2)*, Chicago: Chicago University Press.

Stenning, K. & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, **19(1)**, 97–140.

Stenning, K. & Tobin, R. (in press). Assigning information to modalities: comparing graphical treatments of the syllogism. In *Action, Language and Cognition.* In S. Lindström & E. Ejerhed, (Eds.), Proceedings of the International Conference on Dynamic Semantics, Umeå, Sweden. Kluwer.

Stenning, K., Cox, R., & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, **10(3/4)**, 333–354.

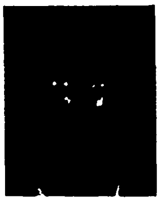
Stenning, K. & Inder, R. (1995). Applying semantic concepts to the media assigment problem in multi-media communication. In B. Chandrasekaran, J. Glasgow & N.H. Narayanan (Eds.) *Diagrammatic Reasoning: Computational and Cognitive Perspectives on Problem Solving with Diagrams.* Menlo Park, CA: AAAI Press/MIT Press, 303–338.

Wason, P. C. (1977). Self-contradictions. In P. N. Johnson-Laird & P. C. Wason (Eds). *Thinking: Readings in Cognitive Science.* Cambridge University Press, 114–128.

(1995 年 5 月 29 日受付)
(1995 年 9 月 12 日採録)

Richard Cox

Richard Cox studied psychology and education in the UK and Australia and is currently a Research Fellow in the Graphics and Language Working group of the the Human Communication Research Centre. His main interests lie in the areas of A.I. in education (e.g. intelligent learning environments) and reasoning with external representations. He is currently studying the efficacy of representation switching as a means of resolving impasses in reasoning.



Keith Stenning

Keith Stenning studied philosophy and psychology at Oxford before doing a PhD with George Miller at Rockefeller University in New York. His interests are in human discourse processing, and the relation between formal and psychological theories in understanding reasoning. Recently he has worked on a theory of the cognitive properties of graphics (as contrasted with language and other media/modalities), especially in the teaching of formal subjects such as logic, but also in HCI and software engineering. He is currently Director of the Human Communication Research Centre (HCRC) at Edinburgh University.

Jon Oberlander

Jon Oberlander is currently an EPSRC Advanced Fellow at the ESRC Human Communication Research Centre, at the University of Edinburgh. He graduated with a BA in Philosophy from Pembroke College, Cambridge, in 1983, and gained his PhD in Cognitive Science from the University of Edinburgh in 1987.

His main interests lie in the intersection of computational linguistics, cognitive science, and knowledge representation. He has published work on natural language discourse structure, non-canonical syntactic constructions, the automatic generation of temporal discourse, and on diagrammatic reasoning and communication. Currently, his primary aim is to develop a cognitively-motivated computational and formal theory of media allocation in discourse generation systems. There are three sub-goals—the intention is to make substantial progress towards solving inter-related problems concerning: (i) the logical and computational properties of media; (ii) the fundamental relations assumed in theories of discourse structure; and (iii) the communicative functions of multimodal discourse.