The effect of graphical and sentential logic teaching on spontaneous external representation

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.

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).

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, &

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

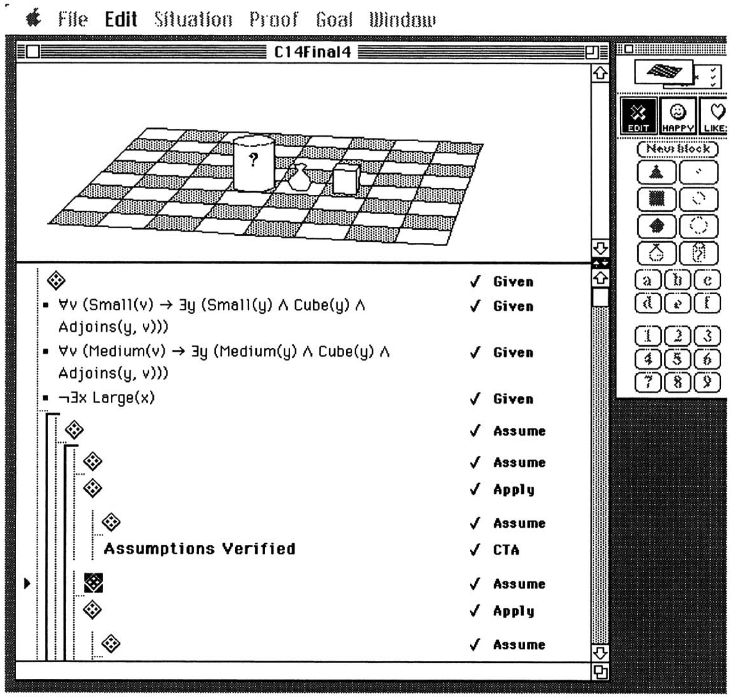

Figure 1 The Hyperproof (HP) interface. The main window panes—graphical and calculus—are supplemented by control palettes. The situation being viewed is the fifth in the course of the proof, and corresponds to the fifth diamond-shaped 'situation' icon in the body of the proof. The graphical window pane 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 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 ($R = 0.59$, $p < .04$).

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

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.