Thinking with Diagrams

Edited by

Alan F. Blackwell

Computer Laboratory, University of Cambridge, UK

Reprinted from *Artificial Intelligence Review*
Volume 15, Nos. 1–2, 2001

Springer-Science+Business Media, B.V.

A C.I.P. catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-5695-5 ISBN 978-94-017-3524-7 (eBook)
DOI 10.1007/978-94-017-3524-7

Printed on acid-free paper.

Originally published by Kluwer Academic Publishers in 2001

Table of Contents

Artificial Intelligence Review **15:** v–viii, 2001.

About the Authors

Alan Blackwell is a University Lecturer in the University of Cambridge Computer Laboratory. His research deals with the use of notational systems to create and maintain abstractions, funded by a research grant from the Engineering and Physical Sciences Research Council. His PhD in psychology was completed at the Applied Psychology Unit in Cambridge, under the supervision of Thomas Green, and his MSc in computer science at Victoria University of Wellington with Peter Andreae. His first degree, in Electrical Engineering, was taken at the University of Auckland. He has worked as an engineer for the Advanced Software Centre of Hitachi Europe Limited, for Author D. Little Cambridge Consultants, and for Progeni Systems Limited. http://www.cl.cam.ac.uk/~afb21/

Paul Brna is Reader in Interactive Learning Systems at the Computer Based Learning Unit, the University of Leeds. He received his M.A. degree in Mathematics from Cambridge University, and his Ph.D. degree in Artificial Intelligence from the University of Edinburgh. Paul Brna has worked in several areas connected with the applications of Artificial Intelligence to Education, including research into the interpretation and use of external representations, and learning environments for program construction and debugging. The joint research with Richard Cox and Judith Good reflects his interest in the development of environments for learning using multiple representations.

Peter Cheng's interests are in the nature of problem solving, discovery, learning and expertise with a particular emphasis on the role of representational systems in science, engineering and mathematics. His work encompasses: psychological studies of problem solving and learning with different representations; the development and evaluation of computer software systems and tools to support learning, knowledge acquisition and other information intensive tasks; the computational modelling of the processes of learning and discovery; the conceptual and cognitive analysis of existing representations and the design of novel representations for substantive domains. Amongst the novel representational systems he has invented and evaluated are large scale diagrammatic systems for probability theory, electricity and particle collisions. Dr Cheng has degrees in Mechanical engineering, philosophy and Artificial Intelligence. He is currently Senior Research Fellow and Deputy Director of the ESRC Centre for Research in Development, Instruction and Training, at the University of Nottingham.

Richard Cox is a Reader in the School of Cognitive and Computing Sciences at the University of Sussex. He received his PhD from the Department of Artificial Intelligence, University of Edinburgh. His research interests lie in the areas of AI in Education, diagrammatic reasoning and medical informatics. Richard has over 60 publications in these areas. Currently he holds grants for 3 projects: 1. multimedia support for medical research & training, 2. multiple representations in Java programming and 3. social and representational aspects of problem-solving in secondary schooling.

Ellen Yi-Luen Do holds a PhD in design computing from the College of Architecture at the Georgia Institute of Technology. She holds a bachelor of architecture (Honors) from National Cheng-Kung University in Taiwan, and a master degree from Harvard Graduate School of Design. She is an Assistant Professor of Architecture in the Design Machine Group at the University of Washington, Seattle, USA. She is developing computation tools that support design drawing and idea exploration, image perception, and information retrieval. Besides drawing and sketching, her research interests include visual note taking and knowledge representation. (ellendo@u.washington.edu)

Judith Good obtained her PhD in the Department of Artificial Intelligence, University of Edinburgh. Her first degree was in psychology, and she worked in the field of clinical psychology and family therapy before pursuing an interest in the cognitive aspects of learning and education. Her main interests lie in the use of graphical representations, both in programming for novices and in other educational contexts, in cognitive modelling, and in the application of artificial intelligence techniques to educational contexts. She is currently an honorary fellow of the University of Edinburgh and is pursuing free-lance work.

Mark Gross holds a PhD in design theory and methods from the School of Architecture and Planning at MIT. He currently directs the Design Machine Group in the Architecture department at the University of Washington, where projects focus on future tools and techniques for computer assisted architectural design. Gross's research interests center on how design is done and how computational tools might help people do design better. He has worked on constraint based design support systems, schemes for diagram recognition and retrieval of design information from freehand diagrams. (mdgross@u.washington.edu)

Oliver Lemon is a Senior Research Engineer at the Center for the Study of Language and Information (CSLI) at Stanford University. He holds a PhD from the Centre for Cognitive Science, Edinburgh University, for work in dialogue modelling, logic, and dynamic semantics (1996). He was then a research associate in the AI group at Manchester University, working on spatial logic, modal logic, diagrammatic reasoning, and formal semantics. After teaching AI and computational linguistics at Trinity College, Dublin, he came to CSLI to work on multimodal dialogue systems. His current research interests lie in dialogue modelling, implementing dialogue systems, spatial reasoning, logic, and formal semantics. http://www-csli.stanford.edu/~lemon

Ric Lowe is Associate Professor of Instructional Design and Learning Technologies at Curtin University, Australia. His major research interest is the human interpretation of diagrammatic information as it applies to learning and the design of technology-based learning materials. He is currently investigating the role of perceptual and cognitive processes in the way learners extract information from animated displays. This work focuses upon the influence of visuospatial and dynamic contrast in the extraction and retention of information from animations, particularly the way such contrast is reflected in 'dynamic field-ground' effects that are analogous to the field-ground effects found for static depictions.

Patrick Olivier is a lecturer in the Department of Computer Science at the University of York. His research interests broadly fall within the area of intelligent user interfaces, primarily: spatial and diagrammatic representation and reasoning, graphical presentation planning and the representation and processing of spatial expressions.

Marian Petre is an Advanced Research Fellow funded by the Engineering and Physical Sciences Research Council. She studies how expert hardware and software designers actually structure and use tools such as representations in order to devise solutions and design systems. Her empirical studies have addressed: cognitive issues in notation; programmer's mental imagery; expert strategies for interpreting and laying out graphical representations; and representations used for ideas capture. She was formerly Senior Research Fellow and Director of the Centre for Informatics Education Research.

Mike Scaife is a founder and Research Director of the Interactive Media & Virtual Environment Lab in the School of Cognitive & Computing Sciences at the University of Sussex. His research is concerned with developing a theory, of 'external cognition' to (i) understand how children and adults use representations (such as diagrams, animations and multimedia) for learning and (ii) to develop virtual environments for playing and learning.

Atsushi Shimojima received his Ph.D. in 1996, from the philosophy department at Indiana University, where he was a member of the Visual Inference Lab directed by Jon Barwise and Gerry Allwein. Shimojima's areas of interests include comparison of different representation systems as to their representational/inferential efficacy, and semantic/pragmatic analysis of human verbal communications involving graphical representations.

Keith Stenning holds a personal chair of Human Communication at Edinburgh University and is an ESRC Senior Research Fellow. His undergraduate training was in philosophy and psychology. He did his PhD in New York with George Miller; and came to Edinburgh (via Liverpool) to teach in the Centre for Cognitive Science. He is currently a member of the board of the Cognitive Science Society.

He has done research on the structure and pragmatics of natural language discourse, on its relation to logical analyses, and on the implementation in human working memory of verbal and graphical reasoning and communication. A major focus of interest has been on the cognitive effects of assigning information to different media and modalities, especially in the domain of educational communication. Studying the learning/teaching of logic and the range and uses of different kinds of representations provides a unique opportunity to analyse the impact of modalities on communication, reasoning and learning. The long term goal is a theory which can integrate cognitive, social and affective factors in educational communication.

Kirsten Whitley recently received a PhD in computer science at Vanderbilt University. Her dissertation researched the cognitive impact of visual programming languages and presents evidence in favor of visual representation on program comprehensibility. Contact her at whitley@computer.org.

Artificial Intelligence Review **15:** 1–3, 2001.

Introduction
Thinking with Diagrams

ALAN F. BLACKWELL
Computer Laboratory, University of Cambridge, Cambridge, U.K.
(E-mail: alan.blackwell@cl.cam.ac.uk)

One of the central insights offered to cognitive science by artificial intelligence research is the importance of problem representation when creating effective implementations of intelligent behaviour. This is mirrored in experimental psychology by studies demonstrating that the form in which a problem is presented can make structurally identical problems either very easy or very difficult to solve (Hayes and Simon 1977). Diagrams are an interesting artefact for this reason – their purpose is purely to modify the representation of problem situations. Furthermore, diagrams are not easily amenable to the methods that have been used to investigate other varieties of human markings. They are not linguistic in the way that speech and written text tend to be. Neither are they pictorial representations. This means that neither linguistic nor perceptual theories are sufficient to completely explain their advantages and applications.

Meanwhile, diagrammatic representations are becoming more common in everyday human experience. Bit-mapped computer displays have encouraged the use of diagrams in human-computer interaction. Improved publication technologies, especially the PostScript language, have provided the means for standardised reproduction of diagrams. Modern thought has already been greatly influenced by the ability to publish conventional pictorial illustrations in books (Ivins 1953; Ferguson 1992), and it seems that the widespread facility to create and interact with diagrams will encourage new styles of literacy in a similar fashion.

Despite this observation, there is substantial scepticism regarding the value of diagrams. The Speaker of the British House of Commons, Betty Boothroyd, rebuked an M.P. in 1994: "I have always believed that all Members of this House should be sufficiently articulate to express what they want to say without diagrams" (The Guardian 1994). Similar suspicion was directed toward analytical mathematics 200 years ago (Mehrtens et al. 1981), and to symbolic logic 100 years ago (Mineau et al. 1993). Ironically, many logicians and mathematicians are now in their turn sceptical regarding the formal status of diagrams, even if they have advantages when used for teaching or creative exploration. Much of this scepticism may be attributable to the fact that diagrams themselves are usually regarded as a tool, rather than a useful object of study in their own right (a statement that was true of mathematics itself at one time). Of course diagrams are not universally beneficial; many are badly designed or badly used. This is a further reason why the study of diagrams is overdue – as a theoretical contribution to the practical questions faced by information designers.

The papers presented in this special issue address these questions of representation, reasoning and application of diagrams. The nature of diagrams is considered by Shimojima, who reviews the philosophical positions that have been proposed regarding the distinction between diagrams and text, and by Stenning and Lemon, who discuss the logical and psychological properties that result from this nature. Olivier provides a complementary view, considering the status of diagrams when used as a machine representation rather than as a mental representation. Cheng, Lowe and Scaife present a range of studies considering the effects on human performance that result from using diagrams, and this approach is extended in the remaining papers into three investigations of specific areas of activity: Blackwell, Good, Whitley and Petre on the use of diagrams in computer programming, Brna, Cox and Good on diagrams in educational contexts, and Do and Gross on the use of diagrams by architects.

These review articles follow from work that has been presented at two AAAI symposia on Diagrammatic Reasoning in the USA, and three meetings on the topic of Thinking with Diagrams in the United Kingdom. The authors also draw on far broader academic traditions, however – in philosophy, computer science, education, architecture and many other disciplines. A useful introductory collection of early work, that defines many central concerns in thinking with diagrams, has been published by AAAI press (Glasgow et al. 1995). It is also possible to contact current research groups through the Diagrammatic Reasoning web site, accessible at http://uhavax.hartford.edu/Diagrams/ and mirrored at http://www.hcrc.ed.ac.uk/gal/Diagrams/.

Acknowledgements

I am very grateful to all of the authors, whose patience with the process of assembling such a comprehensive review of the field has resulted in a valuable resource in this special issue.

References

Ferguson, E. S. (1992). *Engineering and the Mind's Eye*. MIT Press: Cambridge, MA.

Glasgow, J., Narayanan, N. H. & Chandrasekaran, B. (eds.) (1995). *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. AAAI Press: Menlo Park, CA.

Hayes, J. R. & Simon, H. A. (1977). Psychological Differences Between Problem Isomorphs. In Castellan, N. J., Pisoni, D. B. & Potts, G. R. (eds.) *Cognitive Theory*. Erlbaum.

Ivins, W. M. Jr. (1953). *Prints and Visual Communication*. Routledge & Kegan Paul: London.

Mehrtens, H., Bos, H. & Schneider, I. (eds.) (1981). *Social History of Nineteenth Century Mathematics*. Birkhauser: Boston.

Mineau, G. W., Moulin, B. & Sowa, J. F. (eds.) (1993). *Conceptual Graphs for Knowledge Representation: First International Conference on Conceptual Structures*. Springer-Verlag: Berlin.

The Guardian (7 December 1994). Quoted in M. Scaife and Y. Rogers. External Cognition: How Do Graphical Representations Work? *International Journal of Human Computer Studies* **45**(2): 1996.

Artificial Intelligence Review **15:** 5–27, 2001.

The Graphic-Linguistic Distinction

Exploring Alternatives

ATSUSHI SHIMOJIMA
Japan Advanced Institute of Science and Technology, 1-1 Asahi, Tatsunokuchi, Nomi, 923-1292 Ishikawa, Japan
(E-mail: ashimoji@jaist.ac.jp)

Abstract. What properties, if any, distinguish graphical representations from linguistic representations? This paper looks for answers in the literature of philosophy, logic, artificial intelligence, and cognitive psychology, and extracts seven alternative binary classifications of representations that may characterize the graphic-linguistic boundary. We assess each alternative by two standards: (a) whether it extensionally fits the graphic-linguistic distinction, and (b) how far it explains the properties commonly attributed to graphic representations but not to linguistic ones.

Keywords: graphic representation, linguistic representation, diagrammatic reasoning, analogical representation, sentential representation

1. Introduction

On the common conception, a Venn diagram for a barbara's premises, a bar chart of Scotland's annual exports and imports, a geometry diagram illustrating the Pythagorean Theorem, a state map of the United States, and a picture of Mount Fuji by Hokusai are all *graphical* representations, while a set of first-order formulas describing a barbara's premises and Mishima's sentences describing Mount Fuji are *linguistic* representations. Generally, pictures, images, and diagrams are graphical representations, while sets of sentences are linguistic representations. This much seems clear and obvious. But what distinguishes graphical representations from linguistic representations? What exactly is the boundary? Once the question is generalized to this extent, the common-sense conception gives us no clear answer. The conceptual boundary between graphical and linguistic representations seems to be there, but we are not prepared to tell where.

Fortunately, a search in the literature of philosophy, logic, artificial intelligence, and cognitive psychology reveals several candidate answers, i.e. several binary classifications of representations that might be used to characterize the boundary in question. The aim of this paper is to present each of these candidates as clearly as possible, in order to give an accurate picture of what we have known so far as to the boundary. The candidate distinctions to be presented are:

A. "Analog" versus "digital" systems of representation;
B. Representation with "sequential" structures and ones with "non-sequential" structures;
C. Representations with "relation symbols" and ones with "object symbols";
D. More and less homomorphic systems of representation;
E. Representation systems with "exploitable" limitations on expressivity and ones without;
F. Representations obeying both intrinsic and extrinsic constraints and ones obeying extrinsic constraints only;
G. Representations that projects nomic constraints and ones that do not.

As we will see shortly, not all of these distinctions were originally proposed to capture the boundary between graphic and linguistic representations. Contrast A has been proposed for a rather remote typological concern; Contrasts B–F for related, but possibly different typological distinctions, such as: "analogical" versus "Fregean" (C and D), "analog" versus "propositional" (C and F), "graphical" versus "sentential" (E), "diagrammatical" versus "linguistic" (B, D, and F). Thus, in many cases, what we discuss is simply the *applicability* of the cited distinction *to* the graphic-linguistic issue, and the proposals considered are often not the views explicitly held by our predecessors, but the results of translating them into the form of proposal on the present issue.

Before we start presenting and assessing each candidate, let us make clear what we would count as an adequate answer to the graphic-linguistic issue. First of all, for the purpose of drawing a line between graphic and linguistic representations, we do not have to find a *necessary and sufficient* condition for a representation to be graphical. It would be sufficient if we could find a property shared by all graphical representations but by no linguistic representations. (This also amounts to finding a property shared by all linguistic representations but by no graphical representations.) If we find such a property P, we can use the presence of P as a proof of a representation's being graphical, and the absence of P as a proof of its being linguistic, given that the representation is either graphical or linguistic.

Secondly, although we do not demand P to be a sufficient condition for being graphical, we do demand P to be an *explanatory* property of graphical representations, in the sense that P accounts for other properties that are commonly attributed to graphical representations but not to linguistic representations. To illustrate this point, suppose Bill Clinton had a mysterious, but acute sense on "graphic" and "linguistic" and can classify all graphics as graphics without classifying any linguistic representations as graphics. Let P_h be the property of being-classified-as-a-graphic-by-Clinton. Although P_h

satisfies our first criterion, it fails our second test: the fact that the president of the United States classifies a class of representations as graphical explains no other characteristic properties of the class of graphics. P_h is hardly explanatory.

Of course, the existence of P that satisfies the above conditions is not guaranteed, and hence there may be no substance to our pre-theoretical distinction of graphic and linguistic distinction. However, there seems no a priori reason to preclude the existence of P and to deny substance to this intuitive classification scheme. It is this absence of a priori proof for either positive or negative result that makes an *actual* search for P still more worthwhile.

2. Digital versus Analog

Following Goodman (1968), let us say that a representation system is *analog* if (1) it uses an infinite class of states of affairs to indicate an infinite class of information and (2) each class is dense, namely, its members can be ordered in the way that between each pairs of elements there is another element. Let us say a representation system is *digital* if it uses a discrete class of states of affairs to indicate a discrete class of states of affairs.[1] For example, the analog speedometer on an automobile affords an analog representation system about the speeds of the vehicle since there is a dense class of states of affairs (positions of the pointer) that can hold in the meter and this class indicate a dense class of possible speeds of the vehicle. In contrast, the light on the dashboard that registers oil pressure affords a digital representation system because there are only two states of affairs (on and off) that indicate information (high and low) about the oil pressure. Here the class of indicating states and that of indicated states are both discrete.[2] Of course, a single representation system can be both analog and digital with respect to different subsets of information within its coverage.

One may be tempted to use this distinction between analog and digital to draw a line between graphical systems and linguistic systems, although Goodman himself never intended to do so. Thus:

Proposal A

1. A linguistic representation system is digital with respect to the entire set of information it covers.
2. A graphical representation system is analog with respect to at least a subset of information it covers.

In fact, it is easy to find a graphical system that is partially or entirely analog in this sense. For example, different states of an analog speedometer can be considered graphical representations of a vehicle's speeds. Also, a class of all possible line drawings of a man's ways of raising his right hand constitute a dense class – between any pair of drawings that indicate two different heights of his arm, you find a possible drawing that indicates an intermediate height of his arm. On the other hand, it is not as easy to find a class of linguistic representations that constitute a dense class. *Prima facie*, natural languages such as English do not appear to have sufficiently fine-grained vocabularies to afford a dense class of sentences in the above sense.

However, it is not *impossible* to find or construct a dense class of linguistic representations. Think of the class of sentences of the form "This car runs at x" where "x" denotes a real number. This class is clearly dense. For any pair of sentences "This car runs at y" and "This car runs at z," there is a sentence "This car runs at w" such that "w" denotes an intermediate real number between x and y. (Use, for example, the description "$(x + y)/2$" for "w".) Generally, a language has a finite number of lexical items, but it affords an infinite number of definite descriptions. Hence the possibility that a language affords a dense class of sentences.

Furthermore, there seem to be graphical representation systems that are partly or entirely digital, as pointed out by Goodman (1968, p. 68):

> Diagrams, whether they occur as the output of recording instruments or as adjuncts to expository texts or as operational guides, are often thought – because of their somewhat pictorial look and their contrast with their mathematical or verbal accompaniments – to be purely analog in type. Some such as scale drawings for machinery, are indeed analog; but some others, such as diagrams of carbonhydrates, are digital; and still others, such as ordinary road maps, are mixed.

In addition, it is easy to imagine systems of Venn diagrams, flow charts, and bar charts that are entirely digital. Thus, some graphical systems are not analog (not even partly) and some linguistic systems are not digital. Proposal A suffers from counter-examples in both directions.

3. Sequential versus Two-Dimensional

Within the terminology of Larkin and Simon (1987), a sentential representation is a "data structure in which elements appear in a single sequence," while a diagrammatic representation is a "data structure in which information is indexed by two-dimensional location" (p. 72). Thus, in a sentential representation, "each element is 'adjacent' only to the next element in the list," while in a diagrammatic representation, "many elements may share the same location, and each element may be 'adjacent' to any number of other elements" (p. 107). Bertin (1973) seems to have the same contrast in mind when he calls a system of mathematical notations "linear" while characterizing a graphical system as utilizing three variables, namely, the variation of marks and the two dimensions of the plane (p. 3).

Presumably, these authors offered this distinction merely to *define* their own terms of "sentential," "graphical," and "diagrammatic," rather than to *analyze* the exact properties that we pre-theoretically denote with these terms. Still, for the purpose of this paper, it is worthwhile to see if this terminological distinction is applicable as a genuine analysis of graphicality and linguisticity. When so applied, the distinction would amount to the following claim:

Proposal B

Define a representation *s* to be *sequential* if and only if every (information) item in *s*'s semantic content is specifiable on the basis of one-dimensional positions of *s*'s constituents.[3] Then:

1. Every linguistic representation is sequential.
2. No graphical representation is sequential.

Indeed, it is hard to imagine any linguistic representation the specification of whose semantic content requires something beyond the reference to the one-dimensional arrangement of its constituents. So perhaps, B (1) is probably true, and sequentiality is a necessary condition for linguisticity.

Proposal B (2) is faced with an obvious counter-examples, however. Hammer (1995) has already given an counter-example (p. 2). For slightly different one, let us consider so-called "position" diagrams, frequently used to solve a GRE-style problem concerning the seating of people on linearly arranged chairs. We may use, for example, a representation of the kind in Figure 1 to mean that Amy is at the leftmost seat, Mary is at the second from left, nobody is at the middle seat, Kelly is at the second from right, and there may or may not somebody at the rightmost seat.

This representation is clearly sequential: we can specify the syntactic structure of this representation in terms of the positions of symbols "A," 'M," "_," "K," and "X" in a one-dimensional arrangement and can determine the semantic content of the representation on the basis of that syntactic specification. There seems no reason not to call this a diagrammatic, and therefore graphical, representation. Generally, the existence of any "linear" diagrams jeopardizes the truth of B (2).

Stenning and Inder (1993) appear to hold a version of Proposal B when they say, "The essential property of pure linguistic modalities that sets them off from graphical ones is that the only inter-word relation which is interpreted is concatenation" (p. 319). But the similarity of their view to Proposal

Figure 1. A sequential diagram.

B may be only on the surface. They claim that because of "this paucity of interpreted temporal/spatial relations in language," a linguistic system needs an "abstract syntax" that specifies a richer syntactic structure (such as a tree structure, presumably) behind a sentence and thus enables "more than a single uniform interpretation of concatenation" (ibid). In the case of a graphical system, a representation typically contains diverse spatial relations to be interpreted, and there is no need of such an extra level of syntactic specification. Thus, the interpretation of a graphical representation is "direct", while that of a linguistic representation is not, and this is where Stenning and Inder see a difference between linguistic systems and graphical systems. So their theory is certainly more elaborate, and less vulnerable, than Proposal B. Unfortunately, as in their 1995 paper, the theory is not sufficiently detailed to be treated under a separate section here, but I expect that an updated version will be found in Stenning and Lemon's paper in this volume.

4. Relation Symbols and Object Symbols

Russell (1923) indicates that in sentences, "words which mean relations are not themselves relations" while in maps, charts, photographs, and catalogues, "a relation is represented by a relation" (p. 152). For example, the word "precedes" in the sentence "A precedes B" is not a relation, but an individual object, although it "means" a relation. In the case of a map, however, "the fact that one place is to the west of another is represented by the fact that the corresponding place on the map is to the left of the other; that is to say, a relation is represented by a relation" (ibid.). In a similar vein, Sloman (1971, 1995) distinguishes "analogical systems" of representations from "Fregean" systems by indicating that analogical representations use "properties of and relations between parts of the representing configuration" to represent "properties and relations of parts in a complex represented configuration" (Sloman 1971, p. 216) without recourse to "explicit symbols" for properties and relations (Sloman 1995, p. 13). Palmer (1978) seems to have the same distinction in mind when he says, "Propositional representations are simply those in which there exist relational elements that model relations by virtue of themselves being related to object elements" (p. 294).

When applied to the graphic-linguistic issue, this contrast amounts to the following proposal:

```
Left_of(a,b)
(∀x)Left_of(c,x)
(∃x,y)(Left_of(x,b) ∨ Left_of(a,y))
```

Figure 2. Sentences in *L*.

```
ab
(∀x)cx
(∃x,y)(xb ∨ ay)
```

Figure 3. Sentences in *L*′.

Proposal C

1. A linguistic representation utilizes a special symbol for a property or relation to express the property or relation holding in the target.
2. A graphical representation utilizes no such “relation symbols.”

Note that the Proposal C, or the original distinction made by Russell, Sloman, and Palmer, is not committed to the view that any representation, even linguistic one, can represent a property or a relation in the target world *just* by containing a corresponding relation symbol. Even when a representation has a relational symbol, the symbol must still stand in a certain relationship to other symbols (often object symbols) in order to express the fact that objects in the represented situation stand in a certain relationship. As Wittgenstein (1921) points out, it is not that the complex sign “*aRb*” says that *a* stands to *b* in the relation *R*, but that that “*a*” stands to “*b*” via “*R*” says that *aRb* (3.1432). Thus, the proposed contrast is not that one kind of representations represent a relation by a relation, while the other kind represent a relation by a relation symbol. The contrast is rather that one kind of representations represent a relation by a relation among object symbols *only*, while the other represent a relation by a relation among object symbols *plus* a relation symbol.

However, one may object to the first part, C (1), of this proposal in the following way. Consider a first-order language *L* with only one predicate symbol `Left_of`. In *L*, the strings of symbols in Figure 2 are all well-formed sentences. Now imagine modifying *L* slightly into another language *L*′, where instead of using a predicate symbol `Left_of`, we use a simple concatenation of individual terms to mean what `Left_of` means in *L*. Thus, in *L*′, the strings of symbols in Figure 3 are all well-formed sentences.

Interestingly, this shift from *L* to *L*′ does not seem to make our language “non-linguistic,” although *L*′ uses no relation symbol. If someone objects that “∀”,” “∃,” “∨” can be taken as relation symbols in a broad sense, then we

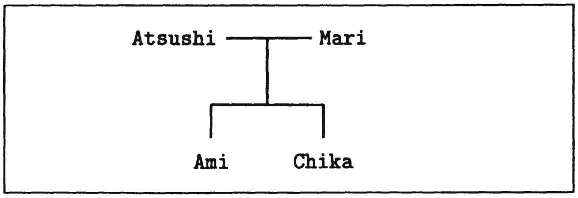


Figure 4. A diagram with relation symbols.

think of a quantifier-free, connector-free version of L'. Although it would be a extremely poor language, it would still be a system of representation, and it would be still linguistic. This example seems to suggest that the use of relational symbol is not essential in our conception of linguistic system.

Nor does the *absence* of relation symbols seem to be essential to graphical representations. Typically in tree diagrams, flow charts, or graphs in general, the edges that connect the nodes correspond to the relations holding among the objects denoted by the nodes. For example, the horizontal edge connecting Atsushi and Mari in the family tree in Figure 4 represents the relation of marriage and the cranked edge connecting `Atsushi` and `Ami` represents the relation parenthood. Thus, a graphical representation with relation symbols seems ubiquitous, contrary to what C (2) claims.

5. Homomorphism

It has been often suggested that graphics, especially pictures, "resemble" what they represent. For example, Sloman at one point claims that in the case of analogical representations, there must be "some correspondence" between the structures of the representation and its target, whereas in the case of Fregean representations there need be no correspondence (Sloman 1975, p. 433). In a similar vein, Barwise and Etchemendy write (1990, p. 22):

> Another advantage of diagrams ... is that a good diagram is isomorphic, or at least homomorphic, to the situation it represents, at least along certain crucial dimensions ... By contrast, the relationship between the linguistic structure of a sentence and that of its content is far more complex. It is certainly nothing like a homomorphism in any obvious way.[4]

Barwise and Hammer (1995) go further and spell out a notion of "homomorphism" claimed to hold between representations and the targets in a diagrammtic representation system (pp. 71–72).

Homomorphism conditions

A representation system is more or less *homomorphic* in virtue of having more or fewer of the features listed below, and also in virtue of having stronger or weaker versions of them.

1. Objects in the target situations, "target objects," are denoted by objects in the representations, "icon tokens," with different types of objects represented by different types of tokens.[5]
2. If a representation s is true of a target situation t, then:
   a) If icon tokens in s stand in some relevant relationship R, then there holds in t a relationship R', represented by R, among the target objects that they denote.
   b) The converse holds as well.
   c) If a grammatical relationship R among icon tokens has some structural property (such as transitivity, asymmetry, irreflexivity, etc.), then this same property must hold of the target relation R' represented by R.
   d) The converse holds as well.
   e) If a token k of some type T has some special property P in s, then the target object k' is of the corresponding type T' having the corresponding property P' in t.
   f) The converse holds as well.
3. Every representation is true in some target situation.

As applied to the issue of the graphic-linguistic boundary, this homomorphism criterion amounts to the following proposal:

Proposal D

Define homomorphism of a representation system as above. Then:

1. A representation system is more graphical according as it is more homomorphic.
2. A representation system is more linguistic according as it is less homomorphic.

Note that the homomorphism criterion offered by Barwise and Hammer is not meant to provide a "cut-and-dried, definitive definition" of diagrammaticality (Barwise and Hammer 1995, p. 71), but a *metric* with which we measure the degree in which a given system is more or less diagrammatic. Proposal D inherits this *gradualism*, and allows continua between purely linguistic systems and purely graphical systems. Moreover, Proposal D advances *pluralism* on graphicality and on linguisticity, according to which there is no single feature that accounts for the graphicality or the linguisticity of a given representation system. Rather, a variety of features, such as the ones in the above list, are responsible, and there are several notions of graphicality and linguisticity corresponding to these features.

In fact, Proposal D seems to well capture our intuition about the degrees in which a system is graphical or linguistic. For example, the system of Venn diagrams would be fairly graphical on this proposal, since the circles that appear in a Venn diagram denote classes in the target situations (homomorphism condition 1). The relation of disjointness between two classes is denoted by the relation of having-the-overlapping-region-shaded, which is symmetric just as disjointness is symmetric (conditions 2c, 2d). The system of Euler diagrams would be even more homomorphic than the system of Venn diagrams because, in addition to the fact that each circle stands for a class in the target, set-inclusion and set-disjointness are denoted by circle-inclusion and circle-disjointness, and while set-inclusion is transitive and set-disjointness is symmetric, circle-inclusion is transitive and circle-exclusion is symmetric. Furthermore, it is impossible for an Euler diagram to present an inconsistent set of information (condition 3). Perhaps, photographs will be classified as still more graphical; sentences of a first-order languages will be one of the least graphical.

It is not clear, however, how the homomorphism conditions are related to other properties commonly attributed to graphical representation systems but not to linguistic systems. In particular, it has not been explained how the homomorphism conditions explain the observed efficacy and inefficacy of graphics as representations of information. Certainly, we do have some intuitive answer: a representation in a highly homomorphic system on this criterion would let us "see' the structure of its target "through" the structure of itself. But it is not a trivial question what exactly is nice about this "seeing through" and exactly which of the features listed in the above criterion are responsible for this capacity. Barwise and Hammer (1995) themselves note that the "close relationships" between representations and the represented structures "allow one to deductively establish facts usually obtainable only model-theoretically" (p. 51). Presumably, this capacity stems from condition 3 in the homomorphism criterion. Features 2a and 2b may be also responsible. The details are yet to be provided.

Thus, although Proposal D seems to do justice to our intuitions about "graphical" and "linguistic," it does not seem to stand alone as a solution to the graphic-linguistic issue. The particular collection of proposed homomorphism conditions must be justified by some supporting theories that relate them to the observed capacities of graphical representations.[6]

6. Content Specificity

Stenning and Oberlander (1995) appeal to the notion of "specificity" to contrast "graphical" and "sentential" representation systems. They characterize specificity as the feature of a representation system that "compels specification of information, in contrast to systems that allow arbitrary abstraction" (p. 99). Stenning and Oberlander then identify specificity "as the feature distinguishing graphical and linguistic representations, rather than low-level visual properties of graphics" (p. 98).

We may use this idea to make the following proposal on the graphic-linguistic boundary:

Proposal E

1. Every graphical representation system has some expressive limitation that prevents representations from presenting certain sets of information without expressing certain other.
2. No linguistic representation system has such limitation on expressivity.

Indeed, it is our everyday experience that graphical representations tend to be specific in their information content. Thus, while a description of a man may well "fail to mention whether or not the man is wearing a hat," a picture of this man "has to go into details" (Dennett 1969, p. 135); it is "unreasonable use of the word 'image'" to speak of an image that does not exhibit content specificity (Pylyshyn 1973, p. 11);[7] "the expressive generality of a system is often incompatible with its capacity for being diagrammatic" (Barwise and Hammer 1995, p. 47); Fregean systems are superior to analogical systems because "the structure (syntax) of the expressive medium need not constrain the variety of structures which can be represented or described" (Sloman 1971, p. 217).

Does specificity, then, account for other properties that we commonly attribute to graphical systems but not to linguistic systems? Stenning and Oberlander claim that although the specificity of a representation system leads to the expressive weakness of the system, it also "aids processibility" (p. 98) of the information represented in the system. Again, we often notice some "trade-off" between expressive generality and inferential efficiency in graphical and linguistic representations. For example, Sloman (1971) notes that the price of the expressive generality of a Fregean system is the lack of capacity of dealing efficiently with specific problem-domains; Laindsay (1988) discusses the trade-off between the "applicability" of a system and its power of reducing "computational complexity of inference" (p. 130). So, if one could offer a clear explanation of how the specificity of a graphical system makes the information easy to process in a way linguistic systems do not, then the theory would be indeed attractive, and perhaps offer a solution to the graphic-linguistic issue.

Unfortunately, Stenning and Oberlander's explanation of the way specificity of a representation system aids processibility uses linguistic systems as model cases. They cite Levesque's work on first-order languages in which specificity seems to aid processibility (Levesque 1988), and point out that the syntactic constraints on graphical systems that are responsible for their specificity are "very similar" (p. 107) to syntactic constraints on Levesque's first-order languages. However, what we want is a theory that differentiates the way in which the specificity of a graphical system leads to inferential processibility from the ways in which the specificity of a linguistic system does so.

Stenning and Inder (1995) try to supplement this incompleteness of the Stenning-Oberlander theory, by means of the notion of "cognitive availability of the limits of expressive power" (p. 304). On the basis of the aforementioned works of Levesque and of Stenning and Oberlander, Stenning and Inder assume that as a general fact, a representation system with a limited expressive power affords more tractable inferences. But they take an additional step, and note that how much a user knows about the scope of a given representation system is a "critical determinant of cognitive properties of the system" for the user (p. 314). Even when a system is expressively weak, and has a potential for easier processibility of the information represented, "exploiting this fact relies on being aware of it" (p. 318); "availability determines whether the weakness of the representation can be exploited" (p. 304).

Stenning and Inder then use the difference in cognitive availability of the limits of expressive power to contrast graphical systems and linguistic systems: "the difference between the graphical and linguistic systems lies in the discoverability of the limitations on expression and the necessary methods of exploiting them in inference" (p. 325). According to them, once the user understands the core fo the interpretation of graphical representations, then the user can infer "quite intricate meta-logical properties" about the system, concerning what limitations are there on the expressiveness of the system. In Stenning and Inder's view, the inferences of this sort rely on the "diagrams" geometry/topology" (p. 318) and arise from "graphical constraints" (p. 334). In the case of a linguistic system, however, the inferences of this sort do not arise due to the paucity of syntactic structure of linguistic representations. Thus, generally, meta-logical facts about expressive limitations are easier to discover in graphical systems than in linguistic systems.

This consideration leads to the following modification of Proposal E:

Proposal E modified

1. Every graphical representation system has some expressive limitation that prevents representations from presenting certain sets of information without expressing certain other, while this limitation is accessible to and inferentially exploitable by users.
2. No linguistic representation system has limitations on expressibility that are as easily accessible and exploitable as those of graphical systems.

It is a quite plausible and interesting suggestion that the ways we infer metalogical facts about the expressive capacity of a system are crucially different in the cases of graphical and linguistic systems. Stenning and Inder, however, have not shown *how* this difference of cognitive availability of expressive capacities accounts for differences in inferential tractability. Assuming a user has a piece of knowledge *k* about the limits of expressivity of a system, how does the user go about exploiting *k* to make efficient inferences? For instance, which step of the user's inference is spared by the existence of this metalogical knowledge *k*, provided that an inference is the kind of process that can be divided into steps. As it is, Stenning and Inder's theory is silent about this point, and it is not clear whether this knowledge is really crucial in any instances where a graphical system appears to afford more efficient inferences than a linguistic system does.

Thus, despite Stenning and Inder's extensive treatment, we are still in a half way to a satisfactory account of how the specificity of a graphical mode of representation, as opposed to that of a linguistic mode, leads to efficient processibility of the information presented in that mode.

7. Intrinsic and Extrinsic Constraints

We saw earlier that Palmer (1978) tries to contrast "analog" and "propositional" representations in terms of the presence and absence of relation symbols with semantic significance. Apart from this contrast based on "surface manifestation" of representations, Palmer offers another conceptual distinction that he claims to capture the analog-propositional distinction.

Palmer sees a representation and its target as two worlds, the "representing world" and the "represented world" (p. 262). Let us call these worlds *s* and *t* respectively. Each of *s* and *t* comprises objects that are related in particular ways. Objects in *t* are "denoted" by objects in *s*, and the ways the latter objects are related in *s* model the ways the denoted objects are related in *t*. Thus, he is assuming some semantic correspondence at the level of relations, namely, from the relations holding in *s* to the relations holding *t*. Not all relations holding in *s* correspond to a relation holding in *t*; conversely, not all relations holding *t* have corresponding relations holding in *s*.[8]

In Palmer's words, the contrast between propositional and analog representations consists in the fact that "whatever structure there is in a propositional representation exists solely by virtue of the extrinsic constraints placed on it" (p. 296) while "whatever structure is present in an analog representation exists by virtue of the inherent constraints within the representing world itself" (p. 297). Here, "intrinsic constraints" means the structural constraints on the relations in a representing world *s*, such as irreflexivity, asymmetry, transitivity of the *above* relation, and "interdimensional constraints" such as the determination of the area of the rectangular from the lengths of their sides (p. 273). In contrast, "extrinsic constraints" are ones "imposed from outside" (p. 271) on the relations on *s* in order to make *s* conform to the represented world *t* (p. 296).

If we straightforwardly apply this idea of Palmer's to the issue of the graphic-linguistic boundary, we obtain the following proposal:

Proposal F

Define *inherent constraints* on a representation *s* as natural constraints, such as topological and geometrical constraints, imposed on the relations that possibly hold in *s*. Define *extrinsic constraints* on *s* as those conditions that *s*'s structure must satisfy in order to present only accurate information about the target. Then:

1. Every graphical representation is so structured as to obey extrinsic constraints *as well as* inherent constraints.
2. Every linguistic representation is so structured as to obey extrinsic constraints and only extrinsic constraints.

Unfortunately, this straightforward application of Palmer's ideas does not work for our purpose, for obvious reasons. First, according to Palmer's definition, all representations that present inaccurate information about their targets fail to obey extrinsic constraints, and thus are excluded by Proposal F from the classes of graphic and linguistic representations. This is obviously wrong, given the existence of many graphic or linguistic representations that present inaccurate information.

Secondly, since all representations in the physical world are so constructed as to obey natural laws, there are no physical representations, graphic or linguistic, whose structures do not obey intrinsic constraints in Palmer's sense. Thus, the sheer compliance to intrinsic constraints can hardly be the distinguishing character of graphical representations as Proposal F claims, or for that matter, not of any class of physical representations.

Thus, Palmer's definition of extrinsic constraint is simply too strong, and Proposal F consequently drives too many representations out of discussion. On the other hand, his definition of intrinsic constraint is too weak so that we cannot use the compliance to it as distinguishing characteristics of any classes of representations. Nevertheless, Palmer's distinction between the representing world and the represented world and his explicit attention to natural constraints governing the representing world seem to contain some important insights into the graphic-linguistic boundary.

In fact, several authors have suggested that a *matching* between natural constraints on representations and constraints on represented objects accounts for graphicality of at least some graphical systems. Barwise and Etchemendy (1990) is particularly explicit about this point. They say:

> Diagrams are physical situations. They must be, since we can see them. As such, they obey their own set of constraints ... By choosing a representational scheme appropriately, so that the constraints on the diagrams have a good match with the constraints on the described situation, the diagram can generate a lot of information that the user never need infer. Rather, the user can simply read off facts from the diagram as needed. This situation is in stark contrast to sentential inference, where even the most trivial consequence needs to be inferred explicitly. (p. 22)

This idea is partly reflected in 2c and 2d in the homomorphism criteria discussed in section D, where Barwise and Hammer (1995) require the relations holding in representations and those holding in their targets coincide in their structural properties, such as transitivcity, asymmetry, and irreflexivity. (Note that structural properties of relations can be considered special cases of constraints on representations and their targets.)

Stenning and Inder (1995) also seem to have the same criteria in mind, when they propose a "correspondence between the logical properties of the representing and represented relations" (p. 316) as the characteristic of the class of least expressive representation systems. Recall that in their view, limitations of expressiveness is connected to inferential tractability. Moreover, Stenning and Inder suggest a connection of this matching with the so-called "self-consistency" property of representation systems, namely, the inability of a system to express the self-contradictory set of information. Given that Stenning and Inder suggest that the properties of expressive limitation, inferential tractability, and self-consistency are shared by many graphical systems, their theory could be interpreted as an attempt to capture graphicality of a system from the standpoint of a matching between natural constraints on representations and constraints on their targets.

In Shimojima (1996), I take the constraint-matching between representations and their targets quite seriously, and build a theory on graphicality and linguisticity entirely on that idea. Thus, my theory is a synthesis and formalization of the intuitions that have been expressed by Palmer, Barwise and Etchemendy, Barwise and Hammer, and Stenning and Inder in varying degrees of explicitness. Let us now turn to this theory.

8. Projection of Nomic Constraints

The proposal on the graphic-linguistic boundary in Shimojima (1996a) is formulated in the conceptual framework of situation theory (Barwise and Perry 1983) and its descendent theory of information (Barwise and Seligman 1996). Here I will present the central ideas with minimal formal details, to make them accessible to those unfamiliar with these frameworks.

In the same spirit as Palmer's distinction between representing worlds and represented worlds, let us view a representation as a *situation* s that we create to present information about a particular target *situation* t. When a representation s targets at a situation t, we say that s *signals* t. There are various states of affairs σ, σ', …holding in s, and these states of affairs *indicate* states of affairs θ, θ', …possibly holding in the target situation t. We say that a representation s *presents the information* θ *about* the situation t if s signals t, and there is a state of affairs σ holding in s that indicates θ. Thus, s is a *true* representation of t if for every state of affairs θ, if s presents the information θ about t, then θ holds in t.

Now, instead of Palmer's distinction between "intrinsic" and "extrinsic" constraints on representing worlds, we posit the distinction between "nomic" and "stipulative" constraints that govern states of affairs holding in a representation. Let Σ, Σ' be set of states of affairs. If Σ cannot hold in a representation s without at least one member of Σ' holding in s, we say that a *constraint* $\Sigma \vdash \Sigma'$ *holds on* s. If a constraint $\Sigma \vdash \Sigma'$ is due to natural laws, such as topological, geometrical, and physical laws, we call it a *nomic* constraint. Thus, our notion of nomic constraint is a re-construal of Palmer's intrinsic constraint. If $\Sigma \vdash \Sigma'$ is due to stipulative rules on s, such as syntactic well-formedness conditions, we call it a *stipulative* constraint. Note that a stipulative constraint is a condition to be satisfied for a representation to be simply well-formed in the given system. It differs from an extrinsic constraint in Palmer's sense, which is a condition to be satisfied for a representation to present only accurate information about its target. We think of constraints $\Theta \vdash \Theta'$ on target situations in the same vein, except that we do not have to distinguish "nomic" and "stipulative" constraints for our purpose.

Roughly, my proposal was that in the case of a graphical representation system *S*, there is a nomic constraint $\Sigma \vdash \Sigma'$ governing all representations in *S* that regulates the information possibly expressed in *S*, while in the case of a linguistic representation system, there is no such nomic constraint. More precisely, this claim can be formulated in the following way:

Proposal G

1. If *S* is a graphical representation system, then there is a nomic constraint $\Sigma \vdash \Sigma$ governing all representations in *S* such that each member of Σ and Σ' indicates some state of affairs about the target situations.
2. There is no such nomic constraint if *S* is a linguistic representation system.

Figure 5 visualizes the property attributed to graphical systems in G (1), where Θ and Θ' are respectively the sets of states of affairs indicated by the states of affairs in Σ and Σ'. I called the property the "projection of a constraint" in Shimojima (1996a), and presented four main arguments to show that the property of constraint projection explains the expressive capacity or incapacity attributed to graphical representations but not to linguistic representations.

Argument 1
Consider a special case in which a representation s obeys a nomic constraint $\Sigma \vdash \Sigma'$, where Σ' is a singleton $\{\sigma'\}$. Let Θ be the set of states of affairs indicated by members of Σ, and θ' be the state of affairs indicated by σ'. Then, when we present the information set Θ by realizing the states of affairs Σ in s, we *must* realize σ' in s and thereby present the information θ' in s. If θ' is in fact a consequence of Θ (i.e. the constraint $\Theta \vdash \{\theta'\}$ holds on the target t of s), this explains a case of free ride in a valid inference, which we often enjoy in using graphical representations in reasoning, but not in using linguistic representations. (The phenomenon is noted by Sloman 1971, Funt 1980, Larkin and Simon 1987, Lindsay 1988, and Barwise and Etchemendy 1990, and explicitly analyzed by Shimojima 1996b, c.)

Argument 2
Consider another special case, where Σ' in the nomic constraint $\Sigma \vdash \Sigma'$ has more than one member. Suppose further that neither of the states of affairs Θ' indicated by members of Σ' is a consequence of the information set Θ indicated by Σ. In this case, we cannot present the information set Θ by means of Σ, without thereby realizing in s at least one member of Σ'. Thus we are *forced* to present at least one piece of unwarranted information in

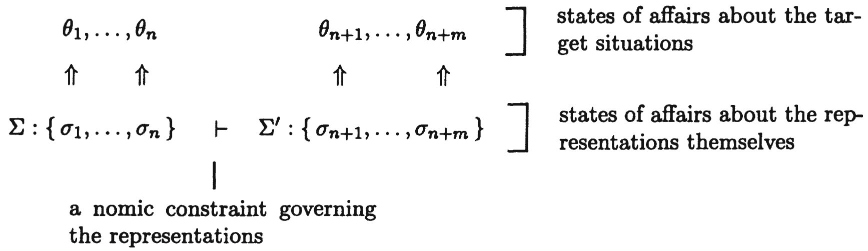


Figure 5. Projection of a constraint (finite case).

Θ'. This explains the over-specificity of graphical representations and in turn the expressive generality of linguistic representations. (Dennett 1969; Sloman 1971; Pylyshyn 1973; Barwise and Hammer 1995; we saw earlier that Stenning and Oberlander 1995 and Stenning and Inder 1995 appeal to this fact to distinguish graphical and sentential representations.) This also explains the "accidental features" of certain graphical representation, such as geometry diagrams, which have been noted since Berkeley (1710) and Hume (1739) or perhaps even before.

Argument 3

Proposal G explains a particular kind of trade-off between "inferential tractability" and "expressive generality" (Levesque and Braachman 1985; Levesque 1988) as it is applied to graphic and linguistic representations (Sloman 1971; Lindsay 1988, and especially, Stenning and Oberlander 1995 and Stenning and Inder 1995). According to Proposal G, linguistic representations project no nomic constraints. Precisely because of this lack, linguistic representations do not afford natural free rides but, at the same time, they are not over-specific in presenting any information set, unless there are syntactic stipulations that force them to be. On the other hand, all graphical representations project some nomic constraints, and hence they tend to provide natural free rides. Due to these nomic constraints, however, graphical representations cannot present some information set in a certain way without presenting one of the alternative pieces of information. No matter whether the latter information is a consequence of the former information set, this property makes graphical representations inflexible in the selection of information set to be presented. This explains the tendency of linguistic systems to be expressibly flexible *but* not supportive to efficient inference, and the inverse tendency of graphical systems. Thus, the currency in the trade-off between expressive generality and inferential efficiency is nomic constraints projected by the given representation system. As a representation system projects more nomic constraints, it obtains more inferential efficiency through free rides, but less expressive generality due to over-specificity. The opposite also holds.

Argument 4
Consider yet another special case, where the relevant nomic constraint is $\Sigma \vdash \emptyset$. This means that Σ is an inconsistent set of states of affairs, unrealizable in our representation *s*. Now, if the information set Θ indicated by Σ is also inconsistent (i.e. $\Theta \vdash \emptyset$ holds on the target *t*), this means that the nomic constraint $\Sigma \vdash \emptyset$ prevents us from presenting the inconsistent information Θ in *s* by means of Σ. If, further, a representation obeys $\Sigma \vdash \emptyset$ *whenever* Σ indicates an inconsistent information set, this means that we cannot use *s* to present any inconsistent information whatsoever about the target situation *t*. The representation *s* is *self-consistent*, so to speak. This in turn means that we can use the representation as a positive test for the consistency of a given information set Θ – if we can present Θ in the representation, then Θ is consistent. Since linguistic representation systems project no nomic constraints, they do not have this capacity. (Perhaps, this is one of the things that Barwise and Hammer 1995 had in mind as "model-theoretic" capacities of diagrammatical representations; Sloman 1995 and Lindsay 1988 also note on this capacity, and Gelernter's Geometry Machine 1959 utilizes this capacity of geometry diagrams for theorem-proving.)

As it stands, Proposal G may appear to be committed to the existence of an absolute threshold between graphical systems and linguistic systems, as well as the existence of a single feature that characterizes the threshold, and hence to directly contradict gradualism and pluralism reflected in Proposal D. Indeed, Proposal G is anti-pluralistic, in that it advances the projection of a nomic constraint or its absence as the fundamental feature that accounts for the graphicality or linguisticity of any representation system. However, Proposal G is not strictly anti-gradualistic, since it allows representation systems *more or less* graphical, according to the relative numbers of nomic constraints projected by the systems. We could even define a partial order on representation systems with a common target domain, on the basis of whether the nomic constraints projected by a system are all included in those projected by another system. Obviously, there are several more orderings possible, and they can be taken to characterize different kinds of continua between strongly linguistic systems and strongly graphical systems.

Is Proposal G the final word then? Presumably, Proposal G is the claim that has been most boldly made and most explicitly argued for, concerning the graphic-linguistic issue. My original work (Shimojima 1996a) even tries to show that Proposal G handles certain borderline cases of graphical and linguistic representations (such as those alleged counter-examples to Proposals A–C discussed earlier) in a satisfying manner. On the other hand, these arguments only show that Proposal G can be pushed in a *certain*

Table 1. Summary

	Graphic	Linguistic
Proposal A	analog systems	at least partially digital systems
Proposal B	non-sequential represntations	sequential representations
Proposal C	representations with no relation symbols	representations with relation symbols
Proposal D	more homomorphic systems	less homomorphic systems
Proposal E	systems with "exploitable" limitations on expressivity	systems with no "exploitable" limitations on expressively
Proposal F	representations obeying inherent constraints	representations obeying no inherent constraints
Proposal G	systems that project nomic constraints	systems that do not project nomic constraints

distance without encountering counter-examples and that *some* of the properties attributable to graphical representations can be explained by the notion of constraint-projection. It is yet to be shown *how much further* Proposal G can be pushed and *how many more* of those properties are explainable. Thus, Proposal G is still a conjecture whose plausibility may increase or decrease as the result of further testing.

Conclusion

Table 1 summarizes the candidate answers to the graphic-linguistic issue considered in this paper. The table makes it clear that some proposal take the properties of "graphic" and "linguistic" as properties of individual representations, while other take them as properties of entire representation systems. A proposal of the second type should be considered to classify individual representations as "graphic" or "linguistic" *derivatively*, depending on whether the system they belong to is graphic or linguistic.

No doubt the readers have reached varying conclusions on the status of the graphic-linguistic issue after looking at these candidates. One may find a particular option highly plausible or at least worth pursuing further. Another may find none promising and feel the need of an entirely different approach to the issue. Still another may find the issue itself unsolvable or ill-founded, and decide the conceptual distinction between "graphic" and "linguistic" to be useless in scientific research. We just hope that our exposition has helped better inform the readers' judgment, either to optimism or pessimism.

Acknowledgments

I am grateful to my anonymous reviewers for their comments, which led to numerous improvements on my treatment of the works cited in this paper, among others.

Notes

[1] Goodman also requires "differentiability" of these discrete members, but we ignore this extra condition in the following discussion.

[2] These examples are found in Dretske's book *Knowledge and Information Flow* (1981, p. 136). However, the notions of analog and digital developed later in Dretske's book are different from Goodman's.

[3] This proposal presupposes that the system in question appropriately determines what counts as a "constituent" of a well-formed representation of the system.

[4] The intution expressed in these quotes differs from the view that the resemblance constitutes the very relation of representation between graphics and their targets. The former uses the resemblance simply as a characteristic feature of graphical representations, while the latter uses it to account for what it is for a graphic to represent its target. Goodman (1978, p. 4) attacks the latter view forcefully, but the former is certainly not subject to Goodman's criticism.

[5] Barwise and Hammer assume that types, properties, and relations of target objects are represented by properties and relations of icon tokens.

[6] Sloman (1995) offers his own set of criteria for structural correspondence, but it also has the same limitation in application to the graphic-linguistic issue.

[7] Here, Dennett and Pylyshyn talk about images and pictures as *external* representations, although they eventually use these observations to support their view about *internal* representations.

[8] For the record, Palmer's definition of "representation" requires that at least one relation holding in s must correspond to a relation holding in t, and that no relation should hold in s that corresponds to a relation absent from t. In other words, s must carry at least one piece of accurate information about t, and s must not carry any misinformation about t.

References

Barwise, J. & Etchemendy, J. (1990). Visual Information and Valid Reasoning. In Zimmerman, W. (ed.) *Visualization in Mathematics*, 9–24. Washington, DC: Mathematical Association of America.

Barwise, J. & Etchemendy, J. (1995). Heterogeneous Logic. In Glasgow, J. I., Narayanan, N. H. & Chandrasekaran, B. (eds.) *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, 211–234. Menlo Park, CA: AAAI Press.

Barwise, J. & Hammer, E. (1995). Diagrams and the Concept of Logical System. In Barwise, J. & Allwein, G. (eds.) *Logical Reasoning with Diagrams*, 49–78. Oxford: Oxford University Press, 1996.

Barwise, J. & Perry, J. (1983). *Situations and Attitudes*. Cambridge, Massachusetts: The MIT Press.

Barwise, J. & Seligman, J. (1996). *Information Flow in Distributed Systems*. Manuscripts. (A significantly revised version has been published as *Information Flow: the Logical of Distributed Systems*, Cambridge, UK: Cambridge University Press, 1997.)

Berkeley, G. (1710). *A Treatise Concerning Principles of Human Knowedge*. Indianapolis: The Bobbs-Merrill Company, Inc., 1957.

Bertin, J. (1973). *Semiology of Graphics: Diagrams, Networks, Maps*. Translated by William J. Berg. Madison, WI: The University of Wisconsin Press.

Dennett, D. C. (1969). *Content and Consciousness*. London: Routledge and Kegan Paul.

Dretske, F. I. (1981). *Knowlege and the Flow of Information*. Cambridge, MA: The MIT Press.

Funt, B. V. (1980). Problem-Solving with Diagrammatic Representations. In Glasgow, J. Narayanan, N. H. & Chandrasekaran, B. (eds.) *Diagrammatic Reasoning: Cogitive and Computational Perspectives*, 33–68. Menlo Park, CA: AAAI Press, 1995.

Gelernter, H. (1959). Realization of a Geometry-Theorem Proving Machine. In Feigenbaum, E. A. & Feldman, J. (eds.) *Computers and Thought*, 134–152. New York: McGraw Hill.

Goodman, N. (1968). *Languages of Art: an Approach to a Theory of Symbols*. Indianapolis: the Bobbs-Merrill Company, INC.

Hammer, E. (1995). *Logic and Visual Information*. Stanford: CSLI Publications.

Hume, D. (1739). A Treatise of Human Nature. *The Philosophical Works of David Hume*. Boston: Little, Brown and Company, 1854.

Larkin, J. H. & Simon, H. A. (1987). Why a Diagram Is (Sometimes) Worth Ten Thousand Words. In Glasgow, J., Narayana, N. H. & Chandrasekaran, B. (eds.) *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, 69–109. Menlo Park, CA: AAAI Press, 1995.

Levesque, H. J. (1988). Logic and the Complexity of Reasoning. *Journal of Philosophical Logic* **17**: 355–389.

Levesque, H. J. & Brachman, R. J. (1985). A Fundamental Tradeoff in Knowledge Representation and Reasoning (Revised Version). In Brachman, R. J. & Levesque, N. J. (eds.) *Readings in Knowledge Representation*, 42–70. Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Linsay, R. K. (1988). Images and Inference. In Glasgow, J. I., Narayanan, N. H. & Chandrasekaran, B. (eds.) *Digrammatic Reasoning: Cognitive and Computational Perspectives*, 111–135. Menlo Park, CA: AAAI Press, 1995.

Palmer, S. E. (1978). Fundamental Aspects of Cognitive Representation. In Rosch, E. & Llyod, B. B. (eds.) *Cognition and Categorization*, 259–303. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Pylyshyn, Z. W. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imaginery. *Psychological Bulletin* **80**: 1–24.

Russell, B. (1923). Vagueness. In Slater, J. G. (ed.) *Essays on Language, Mind and Matter: The Collected Papers of Bertrand Russell, Volume 9*, 145–154. London: Unwin Hyman, 1988.

Shimojima, A. (1996a). On the Efficacy of Representation. Ph.D. Thesis. The Department of Philosophy, Indiana University, Bloomington, IN.

Shimojima, A. (1996b). Operational Constraints in Diagrammatic Reasoning. In Barwise, J. & Allwein, G. (eds.) *Logical Reasoning with Diagrams*, 27–48. Oxford: Oxford University Press.

Shimojima, A. (1996c). Reasoning with Diagrams and Geometrical Constraints. In Westershåhl, D. & Seligman, J. (eds.) *Language, Logic and Computation, Volume 1*, 527–540. Stanford: CSLI Publications.

Sloman, A. (1971). Interactions between Philosophy and AI: the Role of Intuition and Non-Logical Reasoning in Intelligence. *Artificial Intelligence* **2**: 209–225.

Sloman, A. (1975). Afterthoughts on Analogical Representations. In Brachman, R. J. & Levesque, H. J. (eds.) *Readings in Knowledge Representation*, 432–439. Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Sloman, A. (1995). Musings on the Roles of Logical and Nonlogical Representations in Intelligence. In Glasgow, J., Narayanan, N. H. & Chandrasekaran, B. (eds.) *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, 7–32. Menlo Park, CA: AAAI Press.

Stenning, K. & Inder, R. (1995). Applying Semantic Concepts to Analyzing Media and Modalities. In Glasgow, J., Narayanan, N. H. & Chandrasekaran, B. (eds.) *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, 303–338. Menlo Park, CA: AAAI Press.

Stenning, K. & Oberlander, J. (1995). A Cognitive Theory of Graphical and Linguistic Reasoning: Logic and Implementation. *Cognitive Science* **19**: 97–140.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*. Second Edition. Translated by D. F. Pears & B. F. McGuinness. London: Routledge & Kegan Paul.

Artificial Intelligence Review **15:** 29–62, 2001.

Aligning Logical and Psychological Perspectives on Diagrammatic Reasoning

KEITH STENNING[1] and OLIVER LEMON[2]
[1]*Human Communication Research Centre, Edinburgh University;* [2]*Center for the Study of Language and Information, Stanford University*

Abstract. We advance a theoretical framework which combines recent insights of research in logic, psychology, and formal semantics, on the nature of diagrammatic representation and reasoning. In particular, we wish to explain the varied efficacy of reasoning and representing with diagrams. In general we consider diagrammatic representations to be restricted in expressive power, and we wish to explain efficacy of reasoning with diagrams via the semantical and computational properties of such restricted 'languages'. Connecting these foundational insights (from semantics and complexity theory) to the psychology of reasoning with diagrams requires us to develop the notion of the *availability* (to an agent) of *constraints* operating within representation systems, as a consequence of their direct semantic interpretation. Thus we offer a number of fundamental definitions as well as a research programme which aligns current efforts in the logical and psychological analysis of diagrammatic representation systems.

Keywords: diagrammatic reasoning, logic, psychology, efficacy, formal semantics, complexity, constraints, availability, direct interpretation

1. Introduction

A theory of diagrammatic reasoning (DR) is a natural meeting point for psychology and logic, combining computational and representational issues from both fields. Recent advances in the logical understanding of diagrammatic representation systems (Barwise and Shimojima 1995; Shimojima 1996; Hammer 1995; Lemon and Pratt 1997a, 1997c); Shin and Lemon (1999); Lemon, de Rijke and Shimojima (1999) and in the psychological theory (Stenning and Oberlander 1995; Stenning and Inder 1995; Stenning and Yule 1997) point towards a common agenda. We shall develop a logical and empirical research programme for the investigation of DR. Our goal is a conceptual framework for explaining the *efficacy* of diagrammatic representations (DRs) for varied users engaged in varied tasks. This theory should provide a basis for predicting and comparing diagrammatic performances with analogous performances using, for example, sentential representation systems. In fact there are two types of efficacy of a representation system, that one should be careful not to conflate; *computational* efficacy (i.e. low complexity of inference) and *expressive* efficacy (eg. semantical properties such as consistency, or a restriction on the class of representable structures). A full theory should explain how users come to create, to interpret, and to deploy systems of diagrammatic representation. Three concepts are central to such an account, and shall be explicated in the development of the paper: i) *constraints* on representations and their domains, ii) *direct interpretation* of representing relations, and iii) the *availability* of constraints.

The central idea is that diagrams are generally *inexpressive* in a technical sense which takes its meaning from logic and computer science. Inexpressiveness in representation systems generally leads to *tractability* of inference. Conversely, it is the power to express abstractions which gives rise to large inferential spaces and thereby intractable reasoning. As an illustrative example, most diagrammatic systems enforce the representation of all identity relations between represented objects, whereas sentential languages are, in general, expressive enough to abstract over identity relations. For example, the evening star may or may not be the morning star, but drawing a diagram which remains agnostic on the issue is difficult. Enforcement of identity relations enormously simplifies inference, as anyone will intuitively discover when holding a conversation using two names for which identity is actively unknown. But these restrictions on the expression of abstraction are by no means the only constraints on the representational power of diagrams. It is our thesis that the expressive restrictions on DRs arise from an interaction between topological and geometrical constraints on plane surfaces, and the ways in which diagrams are interpreted.

An important issue for this theory will be the extent to which properties of diagrammatic representations are to be explained in terms of their logical expressiveness, and how much in terms of the visual nature of the medium. It is our contention that it is always the way that the medium is *interpreted* which gives rise to the cognitive properties of representations. (Note that we intend to address *cognitive*, rather than *perceptual* properties, such as layout and presentation, of DRs here.) We will mention examples where diagrammatic representations are interpreted highly expressively and are not clearly efficacious. We do this as a way of casting doubt on the role of the visual medium in explaining the cognitive properties of graphics except in combination with the style of interpretation of the medium. For instance, written text is *visual*, but it is not *directly interpreted* (see section 2.1). Thus it is the nature of interpretation of the medium, rather than the medium itself, which gives rise to the real differences between representation systems.

A note on the role of logical analysis may be appropriate here. The function of logic in the analysis of diagrammatic representation and reasoning is not to supplant psychological study but to provide a conceptual framework and an abstract analysis of *what is computed*, which should serve as a basis for empirical investigation of *how* it is computed. A bad competence theory can be highly misleading – logic misread is a dangerous tool. But a good competence theory can nevertheless make all the difference between the success and failure of an empirical programme. An analogy from the study of visual perception may help. Assuming that we perceive distance by unconsciously proving triangulation theorems was a disastrous competence theory which produced little useful empirical research into visual computation. Gibson (1950) showed that expansion rates of retinal images were a far more direct guide to understanding distance perception – the catching of balls and the avoidance of walls. But contrary to some superficial readings, Gibson did not give up geometry – he studied it more carefully.

1.1. *Availability of constraints*

Our account provides a logical framework for describing diagrammatic inexpressiveness which furnishes a point of entry for a psychological theory through the notion of the *availability* of semantic constraints to users. This attendant psychological theory must explain whether or not a user with certain competances and knowledge may learn to exploit the constraints on expressiveness inherent in the intended interpretation of a diagram. With diagrammatic systems, some critical meta-properties of the domain are revealed even to a naive user with only a simple grasp of their core semantics.

Of course, there is a paradox involved in explaining cognitive differences between sentential and graphical modalities in terms of expressiveness when expressiveness is analysed in terms of sentential logical systems (as complexity theory does). We will argue that the resolution of this paradox lies in the degree to which different representational systems exhibit constraints on expressiveness which are "available" to users in various contexts. Roughly, constraints on the expressiveness of diagrams are often available to a user who has only a simplified grasp of their semantics, whereas sentential systems provide no clues to their representational and inferential capacities, unless the user has extensive knowledge of the interpretation.

Thus, to Shimojima's "constraint hypothesis"[1] (Shimojima 1996), we add the "availability hypothesis"; that agents may or may not have full knowledge of the constraints which operate within a representation system, and that some representation systems have more "obvious", accessible, or available constraints than others.

Our theory of diagrammatic efficacy thus rests on the following notions, where it is understood that diagrams function as parts of *systems* of representation, consisting of a target domain (that which is to be represented), a representation 'language', and an interpretation.

1. formal semantics of representation systems;
2. constraints in representation systems;
3. direct semantic interpretation of representing relations;
4. complexity theory for DRs;
5. availability of constraints in DRs.

1.2. *Outline*

The remainder of the paper is structured as follows. After specifying the class of representation systems in which we are interested (section 2) we shall argue for requirements on a satisfying theory of diagrammatic reasoning (section 3). Perspectives on current logical and psychological research are then presented (sections 4 and 5), and some instructive examples of efficacy and inefficacy phenomena (sections 6.1 and 6.2) are canvassed. We then present the theoretical framework in section 7, which we argue does justice to the preceding considerations. Next, in section 8, we employ our account in an analysis of a simple diagrammatic system (of "tilings" for reasoning about set intersection) and then suggest ways of extending logical and psychological research programmes so as to locate a satisfying account of diagrammatic reasoning (sections 9 and 10).

Ultimately we provide a theory of DR systems which could help researchers locate representation languages with respect to meta-logical properties. We also explore the possible construction of appropriate diagrammatic representation languages via the notion of "constraint matching" with their target domains.

2. What Are Diagrams?

A general concern must be the range of data which we wish our framework to cover. While no definition is likely to appease everyone's intuitions, it is necessary to delineate a class of representation systems whose properties our framework is intended to explain. We make no commitment on the issue of the "location" of such representations (we know that they exist externally to human cognition, but leave open the possibility of computationally similar representations being implemented as internal mental imagery).

Note that by calling a DR system a *language* we mean only that it is a system of representation and communication. In particular, diagrammatic "languages" ought not to be thought of as having a syntax in the way that sequential languages do.[2] Sentential languages contrast in whether they have abstract syntax. For example, finite state languages have no abstract syntax. Their concatenation relation is directly semantically interpreted, usually in terms of some sort of temporal relation. So a sentence 'abc' means that a happened before b before c. More generally, wherever two symbols X and Y are in the relation of "Y is concatenated to the right of X", that means that the event which Y stands for happened after the event which X stands for. In contrast, in a phrase structure grammar generated language, which does have an abstract syntax, immediate concatenation between symbols has *no* uniform semantic interpretation. The semantic relations between adjacent symbols is mediated through their syntactic relations. Just as sentential languages contrast in this way, we claim that constrained diagrams are like finite state languages in having no abstract syntax. We find it somewhat misleading to talk of diagrammatic systems having even an impoverished 'syntax'. What formal constraints they have are generally by way of a "reflex" from their intended semantic domains (e.g.: lines in circuit diagrams do not cross, as part of a well-formedness condition, but only because circuits do not cross.) In addition, there are physical constraints on possible diagrams which ought not to be thought of as syntactic (i.e. it is impossible to draw certain configurations of regions in two dimensions).

2.1. *Properties of diagrammatic representations*

Insights such as the following, which illuminate the importance of spatial relations in DRs, and intuitions about their low complexity, deserve a careful logical treatment;

> Diagrams can build the logic of what they represent into the physical logic of their grammar (Eric Hammer 1995, p. vii.)

> . . . visual information is *inherently* more tractable than unrestricted linguistic information. (Hector Levesque 1986, p. 99)

Here we describe the main properties of diagrammatic representations which we think a theory of diagrammatic reasoning should capture.

Our basic observations are that:

1. Diagrammatic representations often exploit non-trivial spatial structure[3] in representation. The price they pay is that they must obey the mereological, topological, and geometrical constraints of the plane.
2. Constraints in DRs can be more or less available to users of the representations.
3. Diagrams are restricted in representational power and are thus potentially computationally tractable.
4. Representing relations between diagrammatic tokens are "directly" semantically interpreted.

The force of "directly" can best be seen by way of contrast with sentential languages. Sentential languages exploit the temporal or spatial properties of their media only in terms of a concatenation relation which bears no *direct* semantic interpretation.[4] Concatenation is such an omnipresent but basic feature of sentential languages that it is easy to forget.[5] Without a determination of the precise details of concatenation, text is uninterpretable. So concatenation clearly has semantic import. But the fact that two words can be in exactly the same concatenation relation, yet their meaning relation be quite diffferent (because the overall syntactic structure of the sentence is different) shows equally clearly that concatenation has no *direct* interpretation, but only one mediated through syntax.

Below we contrast the styles of semantics of sentential languages, a directly interpreted node-and-link diagram, and an indirectly interpreted node-and-link diagram (see Figure 1).

Sentential languages are typically constructed using a vocabulary of symbols:

$$P, Q, R, \ldots$$
$$\&, \vee, \ldots$$
$$(,), \ldots$$

Along with some rules of combination:

If P is a sentence, and Q is a sentence, then P & Q is a sentence.

To be strict, rules of combination are about a spatial relation (concatenation) and how it forms strings of symbols. If a "frown" ($^\frown$) is used to denote concatenation then a complex formula might look like the example below. If it continued over a line break, the concatenation relation would have to be defined to take this into account.

$$(^\frown P^\frown \&^\frown Q^\frown)^\frown \vee^\frown R$$

Semantically then, there are rules of interpretation which operate over these syntactic structures, for example:

P & Q is true just in case P is true and Q is true.

Contrast this with the way the diagrams in Figure 1 are interpreted. In the lefthand network the spatial relation (connection) is directly interpreted and has a uniform meaning (say "... loves ..."). But in the righthand network the links between the logical operator $\vee$ and the other nodes have a different semantic significance. So again, it is an abstract syntax that is being interpreted in the latter.

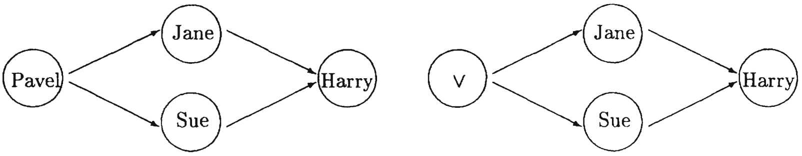


Figure 1. Direct interpretation vs. abstract syntax.

So far we have stressed the directness of semantic interpretation which characterises DRs, and we have defined directness in negative terms – without intervening abstract syntax. A further striking feature of many *effective* diagrams, which is a desirable consequence of directness, is that the spatial relations between their tokens share structural properties with the (not necessarily spatial) relations between denoted objects in the target structure. This is what is often loosely expressed as diagrams being "analogical" representations, but we shall describe it as "constraint preservation", using the terminology of (Barwise and Shimojima 1995; Shimojima 1996). The issue of direct interpretation of diagrammatic relations is thus closely related to the presence of a "matching" of structure between representing and represented relations in effective representation systems. Where no such matching occurs, representing relations may still be directly interpreted, but the representation system as a whole shall exhibit semantic flaws.

The classic example of representational efficacy arising from constraint matching is in the representation of proper set inclusion (a transitive, irreflexive, asymmetric relation) by proper spatial inclusion in the plane, in the system of Euler's Circles. Since proper inclusion is a transitive, irreflexive, and asymmetric relation, its efficacy in representing set membership is obvious. In fact, *under certain conditions*, this constraint matching will give a DR system the property of being *self-consistent*. This property is quite striking; it will not be possible to draw inconsistent diagrams. This means that every diagram in the system will have a model (in the technical but intuitive sense of model). Note that this is not the case for sentential systems, where it is generally possible to construct sets of sentences which have no models. The converse property (where every model has a corresponding diagram) is also important (it is termed "representability" in (Lemon and Pratt 1997a), and some diagrammatic systems fail to exhibit it (see section 6.2 for an example).

As we have noted then, it is not possible to draw an Euler diagram of inconsistent premisses such as *all A are B and some A are not B*. This is because the representing relations (inclusion and overlap on connected 2D regions) preserve the properties of the represented relations (set theoretic intersection and subset). Note that even though Euler's system is self-consistent, Venn's system which uses the same basic representational relation of inclusion in closed curves to mean inclusion in sets, is not (see Stenning and Tobin (1997) for a description of the differences, which turn on the addition of notations and a consequent change in the interpretation of the spatial relations). So it is evident that our caveat *under certain conditions* is significant. Often the preservation of logical properties across the represented and representing relations is much less tight than in the Euler case. Lemon (1997b) and Gurr (1996) discuss what happens when constraint matching is loosened up. We shall have more to say about the preservation of constraints later.

Thus we find that it is a combination of the direct semantic interpretation with the nature of the plane which can explain many of the properties of diagrammatic representations. As for a definition of DRs, consider the following suggestion, where "representing tokens" are the icons, points, lines, words, or regions of a diagram, and the "target structure" is the structure of which the diagram is supposedly a representation. (We define *effective* diagrammatic representations in section 7.)

DEFINITION 1. *A Diagrammatic Representation (DR) is a plane structure in which representing tokens are objects whose mutual spatial and graphical relations are directly interpreted as relations in the target structure.*

We leave open the issue of whether there are diagrammatic elements in the interpretation of texts (layout, certain interpretations of temporal order etc.) – but we see good reason to distinguish these from the core of linguistic semantics which interprets an abstract syntactic structure.

Some apparently diagrammatic representations (e.g. some highly expressive node-and-link-formalisms) should be interpreted as only using 2-D topology for defining 2-D concatenation. These formalisms have an abstract syntax interposed between diagrammatic relations and semantic interpretation. Hence they should be seen as non-diagrammatic in our sense. See Stenning and Inder (1995) for a more detailed discussion.

Of course, there are certainly *non-spatial* representing relations in many diagrammatic systems (such as hue and saturation[6]). One example[7] is of a collection of outlines of animals in which blue ones are reptiles and red ones are mammals. This example is interesting for at least three reasons. Firstly, such a diagram *does* interpret spatial relations – if shape represents species and colour represents mammal/reptile status, then the binding of status to species identity is done by the spatial relations of coloured regions. Even if shape is replaced by species names (thus removing one spatial dimension), it is still the fact that the name occurs *on the colour patch* which indicates that the animal bearing the name is of the status denoted by the colour. Secondly, if shape denotes species, even though this is a spatial dimension, the nature of its semantics is more like lexical semantics in a sentential language than the directly interpreted relational semantics of DRs. Thirdly, note that the colour dimension, though non-spatial, is directly interpreted. One or more linear dimensions in the graphic are mapped directly onto dimensions in the target domain.

Further, note that it is a consequence of definition 1 that the representation: "Earth Moon Mars", is a DR since the sequential structure of its tokens represents topological information about the solar system, and does so directly (without interposition of abstract syntax). The representation: "Earth—Moon————Mars" is a DR since it employs the topological and metric structure of the medium in its (approximate) representational work. The diagrammatic representation: "0—o————x" can also be directly interpreted, and employs graphical tokens. Such examples show our definition 1 to be a plausible one.

3. Desiderata on a Theory of Efficacy

On these simple notions (i.e. constraints, availability, directness of interpretation) we seek to build a theoretical framework for the classification and investigation of systems of DR and their efficacy. Broadly, the framework should account for:

- The efficacy and inefficacies of reasoning with diagrams
- The varied semantical properties of diagrams
- Human reasoning with diagrams; representation system selection, construction, manipulation, and interpretation

Thus we seek a theory of representation and complexity general enough to cover diagrammatic reasoning. In more detail, such a theory should:

1. describe the relation between a representation, interpretational conventions, and the structure which is being represented (cf. Palmer 1978);
2. account for the restrictions in expressive power of representational systems which are due to the interaction of their particular representational media with the style of interpretation of the medium;
3. systematically describe the differing meta-logical properties of different representation systems (eg: consistency, completeness);
4. describe the semantic effects of various transformations of the representations;
5. be sufficiently formal to admit of logical and mathematical analysis;
6. provide points of entry for a psychological theory of the performances of different users, with different knowledge, doing different tasks with the representations.

We now consider briefly the accounts currently offered by logic and psychology, with respect to the above desiderata.

4. A View from Logic

Quite recently there have been a variety of attempts to provide logical analyses of diagrammatic systems (e.g.: (Allwein and Barwise 1993; Hammer 1995; Shin 1995), although these approaches do not focus specifically on the efficacy of DRs. Typically, these analyses amount to a specification of diagrammatic *syntax* in terms of first-order logic, a set of *rules* stating how the diagrams may be manipulated, and a formal *semantics* with respect to which the validity of rules and representations may be established (via a completeness result). The essential idea here, then, is to 'translate' diagrammatic representations into logical representations, and to use standard logical techniques in their investigation. However successful these enterprises are, given their own remit, it can be seen that the approach does not yet do justice to the rich representational capacity of diagrammatic representations. – For example, the standard first-order analyses do not tackle the structural properties which are so important in diagrammatic reasoning (e.g. acyclicity and transitivity of inclusion for regions, planarity and symmetry of overlap relations over 2D regions). Further, these analyses do not provide us with a detailed enough account of representation *systems* (although that is not their immediate concern); that is – they do not account for the fine-grained relationships between a representation, interpretational conventions, the representational medium, and a target domain. In particular, various possibilities for representational error are not described. Consider a map, for example. Strictly speaking it is false (because it only approximates reality, and contains ommissions), but this is not the answer we require from a formal semantics of maps. Sure enough, standard formal semantics tells us about the truth or falsity of representations with respect to interpretation and domain, but the case of diagrams raises this more detailed issue of verisimilitude (see Lemmon 1997b; Lemon and Pratt 1998b). – As Barwise and Seligman argue (Barwise and Seligman 1993) of representations such as photographs and radar screens, diagrammatic representations may exhibit imperfections while nevertheless succeeding in representation. Consider again, for example, the (approximately true) representation: "Earth—Moon————Mars". The "logic" of such a representation relies on a notion of approximate truth, as well as the structure of spatial relations. Thus the standard logical approaches must be extended to cover cases where representation is a more complex phenomenon than we encounter in sentential systems.[8]

Returning to the point about spatial representing relations, a first-order analysis of Euler diagrams, for instance (Hammer 1995), tells us nothing about the topological properties of the denoting expressions, and how they are used in representation. Certainly, it describes how the diagrams relate to set-theoretic objects, but fails to account for the structural properties of representing relations in the diagram. A consequence of this omission is that the proffered analysis fails to account for important representational and computational aspects of the efficacy of the representations – issues which we consider to be central to any account. In short, various considerations should lead us to augment, rather than to discard, the classical logical framework. We shall argue for an enrichment of standard logical approaches, so that they may capture structural constraints in representation, as well as provide a more fine-grained account of representation systems than that available in traditional formal semantics.

Given our thesis about restricted languages and complexity, a further important omission in the application of logic to diagrammatic reasoning, to date, is the lack of a suitable complexity theory for diagrammatic systems. Explanations of computational efficacy of diagrammatic representations require the application of existing techniques in complexity theory to this new domain. Some relevant results already exist. In particular, the work of Grigni et al. (1995) on "topological inference", describes the complexity of certain systems for reasoning with regions of the plane (see Lemon and Pratt 1997c). We shall return to this point later.

A final point to note, from our perspective of an alignment between logical and psychological research, is that the existing logical analyses offer little purchase for psychological theory.

5. A View from Psychology

Of the several distinct literatures relevant to the psychology of diagrammatic reasoning, the one we intend to focus on here is the literature on verbal reasoning, especially syllogistic reasoning. The reason for this apparently Quixotic choice is that this literature has been one of the few to choose representation as its central focus, and most of its efforts have been directed to attempting to distinguish representation systems, some of which are diagrammatic. In fact, the name *verbal* reasoning refers only to the input and output form of premisses and conclusions. This field has been concerned with *internal* mental representations, but has given rise to several external diagrammatic and notational systems which are of interest in their own right. For our present purposes they have the great advantage that the logical and computational relations between the systems are now well understood. These representations arising from theories of mental reasoning offer a unique opportunity for computational analysis of proposed mental machinery. We believe that quite general implications can be drawn for the psychology of both internal and external representations.

For the most part, the literature on the psychology of reasoning, especially that part about deductive reasoning, has taken as its research goal finding the "one true mental representation system" in which people solve problems. Disagreement has been presented as about what kind of representation is used, but the idea that there is one fundamental representation system is shared by 'mental modellers' (e.g. Johnson-Laird 1993) and the 'mental logic' theorists (e.g. Braine 1978; Rips 1994). An earlier version of a related controversy is that between linguistic and spatial accounts of transitive reasoning (see Clark 1969; Huttenlocher 1968).

The fundamental distinction between the theories of mental representation proposed has been between model-based theories (here we include graphical ones), and sententially based theories. But the issue has been presented as a choice between reasoning being *semantic* or *syntactic* respectively. The claim that reasoning is semantic has been motivated by the observation of content effects in reasoning. However, the model-based systems proposed are all content independent formal theories of reasoning (graphical proof theories in logical terms). Whether one regards them as 'syntactic' will depend on whether that term is reserved for the abstract syntax of sentential systems (we think it should be). But if the model based theories are not syntactic they are completely formal and no more 'semantic' than sentential systems. At the same time all the experimental observations brought in support of either theory have been of reasoning within finite domains where any semantic method can be fully emulated by a syntactic one.

A more plausible interpretation of the issue at stake is that model-based theories propose an inexpressive representation system whereas linguistically based theories assume that representations are not limited in this way. Talk of the 'analogical' nature of mental models fits with this interpretation. This interpretation explains the features that mental models are assumed to share with spatial representation more generally (see Gärdenfors (1996) for an example of the "spatial turn" in cognitive science). On our account, spatial representations are data-reductions from the complexities of the surface of texts in expressive languages, down to representation systems which resolve all co-references. Logically, as we shall see, the latter might be thought of as languages with conventions of unique naming and no quantification. This interpretation unites this account of internal representations with our analysis of graphics, which stresses inexpressiveness rather than the visual medium.

Stenning and Oberlander (1995) provide a review of the literature on text comprehension and verbal reasoning in these terms.

Recent equivalence results have gone further and shown that, in the most important domain for this literature (that of syllogisms), these emulations are not complex or hidden but absolutely direct. Each operation in the mental models competence algorithm is mirrored by a graphical operation in a suitably formalised graphical algorithm derived from Euler (Stenning and Oberlander 1995); and by a sentential operation in a suitably formulated sentential system (Stenning and Yule 1997). These results mean that as far as reasoning *within* one of these systems is concerned (and all the theoretical accounts offered are about reasoning within a single system), the accounts cannot be distinguished computationally. This does not mean that the differences between these precisely formulated systems cannot play a role in behaviour through their differences for reasoning *external* to the systems. For example, the Euler system is, as we have mentioned, self-consistent. The other two are not. This property might have a considerable impact on a reasoner who was selecting (or constructing) the system of representation to reason within.

Failure to specify what representational difference the disagreement is about does not, however, mean that there are no such differences in the mental representations which people use in reasoning. Recent research on extended reasoning as taught in logic classes has added to much earlier research documenting the large individual differences between students in how they respond to teaching in the graphical and sentential modalities (Stenning et al. 1995) and recording that these differences extend to self-generated external representations produced in untutored problem solving (Stenning et al. 1995). Because these studies observe the use of external representations, and collect far richer data than is conventional, they provide strong evidence that students differ in their reasoning processes. In fact, the data can be analysed to reveal contrasting student reasoning styles (Oberlander et al. 1996).

It is a moot point whether the individual differences observed in these studies of real teaching are differences in *internal* mental representations. Characterisation of the strategies of proof indicates that the students who would be characterised as 'visualisers' on conventional psychological approaches (they respond well to diagrammatic teaching) do not differ from the 'verbalisers' (who respond well to sentential teaching) in virtue of a preference for the graphical modality. The evidence is rather that they are adept at strategically choosing when to translate between modalities (from sentential to graphical or from graphical to sentential). In fact, the 'verbalisers' are characterised by a tendency to translate immediately into the graphical modality, and to fail to translate in the opposite direction appropriately.

The alternative research program that these findings suggest is one that views human reasoning as dominated by the issue of the *choice* (or construction) of representations for reasoning. People may be expected to make different choices for different tasks: different people may choose different representations for the same task. Empirical study of representations will require sufficiently rich data to discriminate alternative implementations of the same logics. The results from the domain of syllogisms stand as a warning that the sparse data of input premisses and output conclusions probably won't differentiate alternative representations.

Formal analysis of the contrasting properties of graphical and sentential systems offers considerable purchase for a psychological theory of varied mental representation. Our analysis suggests that the critical differences may be at the level of metalogical properties and the availability of constraints to different reasoners.

The existing psychological results can offer a word of caution to the formal analyst. Proofs that graphical systems are constrained and their constraints available to naive users can make it seem obvious that they are preferable to sentential systems (at least for the novice user). One frequently sees the same kind of intuition advanced by designers of 'visualisation' technology if on rather more vague grounds. Where a careful empirical test is conducted, it often transpires that the graphical system is 'better' than the sentential counterpart for some users but frequently that it is worse for others. A psychology of diagrammatic reasoning will have to be able to accommodate both kinds of result. We believe that the area in which one must look for explanation is in the knowledge of interpretation which users with 'sentential' preferences bring to the task.

One might sloganise this general view of the field as follows: "people reason by finding a representation in which the problem presented is trivial. If they can do so they succeed. If they can't they give up". Clearly there are counterexamples to this slogan, and the notion of a "trivial problem" needs to be analysed logically. This claim is somewhat paradoxical from a computational point of view because the complexity of searching for representation systems looks so much worse than that of reasoning within systems. But we believe that this is an artefact of our current ignorance of how people choose and construct representations.

6. Cases Studies in Diagrammatic Reasoning

Below we present related case studies in the use of diagrams in representation and reasoning which illustrate our claims about restricted languages, availability of constraints, and efficacy. These studies are instructive, because they illustrate both the computational pay-offs and pitfalls that may accrue in thinking with diagrams. We give examples which exhibit computational efficacy (i.e. low complexity) and representational efficacy (i.e. self-consistency, representational power), and then an example where representational inefficacy arises due to topological restrictions on combinations of convex regions of the plane (section 6.2). The latter shows that if the constraints on the graphical system are *stronger* than those of its target domain, *in efficacy* (in the sense of lack of appropriate expressive power) results.

6.1. *Efficacy and inefficacy in psychological studies*

One kind of study that the framework suggests is comparison of alternative graphical and non-graphical representations of the same domain, ideally as a prelude to empirical study of users' performance with them. Because so much experimental work has been done on syllogisms, and because there are several 'rival' theories based on apparently distinct representations, the syllogism is an obvious domain.

As mentioned above, Euler's system directly interprets the inclusion of regions by closed curves in the plane to represent inclusion of members in sets. The congruence of logical properties of the representing and represented relations means that the system is self-consistent. So, some critical meta-properties of the domain are revealed to even a naive user with only a simple grasp of the core semantics. For instance, since all points in the plane are classified as included or excluded by all closed curves, it is 'available to the user' (who merely knows this much about the representation and about geometry) that the system cannot represent partially specified types of individual. It happens that the syllogism is a constrained fragment of logic in which no inferences ever require the representation of such partially specified types (see Stenning and Oberlander (1995) for an extended discussion).

Euler's system may be compared with a number of others, most obviously the standard sentential treatment. Stenning and Tobin (1997) compare it with Venn, another 'circle diagram' system, as well as with several other graphical systems devised for illustrative purposes. Stenning and Yule (1997) also compare a sentential system especially formulated to reveal the commonalities with Euler. One important observation of these studies is that even a system such as Venn's, based on exactly the same core semantics as Euler, has quite different expressiveness constraints. Venn uses notations (there are various systems differing in detail) in the regions and on the edges of a constant circle-diagram. These notations mark minimal regions, or combinations of minimal regions, as empty, non-empty, or of unknown status. The notations may be augmented by linkings (see Shin's extension (Shin 1995) of Venn). The most extended systems can express the whole of monadic predicate calculus – a vastly larger fragment than the syllogism. The notations override the diagrammatic constraints of Euler, but there is nothing graphical to stop conflicting notation of the same region. This is why the system is not self-consistent.

This observation focusses attention on what the user knows about the interpretation of the representation system in use. A user who knows nothing of Venn's notations may assume that the diagrams have the constraints that operate in the Euler system. Conversely an overcautious, perhaps logically trained, user of Euler might be reluctant to exploit Euler's constraints for lack of knowledge of what notations may be operating.

Further, some of the other graphical systems considered by (Stenning and Tobin 1997), notably a network based one, have no constraints available in respect of the interpretation of diagrammatic relations.

Comparing any of these graphical systems with the sentential ones requires us to consider what users know about the language of the syllogism. Here we are faced by the much harder problem that the syllogism is a fragment of the natural languages with which users are well acquainted. It is not nearly so clear how they might become aware of the logical constraints of the particular fragment. It is striking that the metalogical property of the syllogism which Euler's system reveals (case identifiability – see Stenning and Oberlander 1995) is nowhere to our knowledge discussed in the extensive logical literature based on sentential presentations of this logic.

There are interesting differences between the kinds of constraints operating in some of these other systems, even though these do not arise through direct semantic interpretation. For example, the 'network' system presented represents the four quantifiers *all, some, non, some . . . not* by two different kinds of link (solid and dashed) and whether the links are symmetrical or have single arrow heads at one end. In contrast, the standard sentential systems uniformly represent quantifiers by words initial in their sentences.

Sentences are inherently one-dimensional: networks two-dimensional. Because simple links are inherently symmetrical, and two of the four quantifiers are semantically asymmetrical (the order of their two arguments matters), a network system must have a method of marking asymmetry. Because sentences are inherently asymmetrical, this requirement is in-built. But appreciating whether a quantifier is semantically symmetrical or not is one of the major problems for naive syllogistic reasoners. The main fallacies of commission and omission turn on appreciating these properties. The network system is free to mark it in the graphical symmetry/asymmetry of its links/arrows (as the one described in Stenning and Tobin (1997) does), and this might be helpful to a learner. But the system is not *graphically constrained* to do this. It could use all asymmetrical arrows just like the all asymmetrical sentences. This is a good example of a "conventional constraint" as opposed to one resulting from direct semantic interpretation.

The comparative empirical study of teaching syllogistic reasoning using these alternative representations is much less developed. Dobson (Dobson 1997) has made some preliminary studies comparing Venn and Euler. His findings highlight the need to ensure that students are operating with an adequate interpretation of the system taught, and also that the "eventual destination" of the teaching must be borne in mind. His results suggest that Venn may be easier to teach than Euler, and that there are large differences between 'arts' and 'science' secondary school students in how they respond. There is some evidence that the teaching intervention did not succeed in teaching the correct interpretation of Euler. A subsequent study of interpretation (on a different sample of students) showed that the Euler interpretation of circle diagrams was 'more natural' than the Venn interpretation. Much empirical work remains to be done, but this small microcosm illustrates how empirical and formal research need to interact to make progress in this area.

6.2. *A case study in inefficacy*

As we have mentioned, it is important to realize that diagrammatic representations are generally restricted in their representational capacity – due to the presence of non-conventional spatial constraints upon representing relations in DRs. In (Lemon and Pratt 1997a) there is an instructive example, closely related to the system of Euler circles, of how using certain diagrammatic representation schemes can lead to inferential errors.

Let's suppose that a reasoner solves certain logical problems in the predicate calculus by drawing regions of the plane representing the extensions of various unary predicates. We suppose here that (as for Euler's Circles) these regions representing atomic properties are *convex* (and hence connected). Now each region represents a possible type of individual.

Drawing such regions seems a natural way to reason about combinations of properties, but it is inadequate, in general. The reason being that the chosen representation cannot express all the set-theoretic configurations that it is supposed to. The reason for this is a result of convex topology known as Helly's theorem.[9] As an application of the theorem, consider for instance 4 convex regions in the plane, each trio of which has an intersection. Then (by

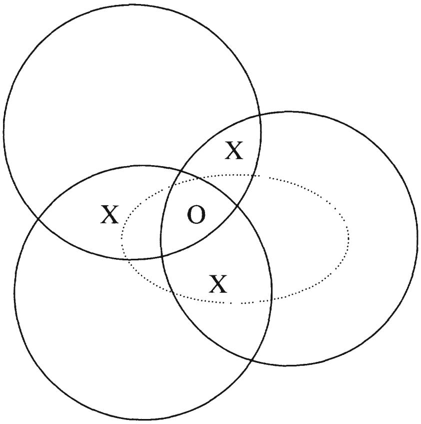


Figure 2. The Helly constraint for convex regions in two dimensions.

the theorem) there must be a quadruple intersection too (see Figure 2). This means that there is no way to add a fourth convex region (e.g.: the dotted ellipse) overlapping each pairwise intersection (X) of the other three, without also producing a quadruple intersection (in O).

According to the theorem then, no matter how you draw regions (provided they are convex), some intersections will unavoidably turn out non-empty, even though it is logically possible that the combinations of properties which they represent might not occur. This property of the representation scheme would force you to draw *unwarranted conclusions* in some cases. (Similar results are proven using non-planar graphs for the case of non-convex connected regions in two dimensions, in Lemon and Pratt (1997a, 1997c). A similar set of results can be developed (see Lemon and Pratt 1998a) for the linear diagrammatic system proposed for syllogistic reasoning of Englebretsen (1992).

Of course, the expressive limitations could be bypassed here by the introduction of some new notational device, but the important point to notice about the problem is the extent to which it relies on the spatial nature of the representations involved. Helly's theorem identifies a constraint on the representational system of convex regions which does not arise from logic alone. That is why this representation scheme is bound to yield incorrect inferences about sets, because it *cannot represent* some logically possible situations. Moreover, this example shows that a diagrammatic representation scheme may "force" representations which are spatially, rather than logically, necessary. Such a representation scheme is "information enforcing" or "over-specific" to use the terminology of Shimojima (1996) or Stenning and Inder (1995) respectively. (Of course, in some contexts this property may be an asset.) This type of problem is quite general, and crops up in other diagrammatic representation schemes. Ramifications of such results for hypotheses in cognitive science are discussed in Lemon and Pratt (1997a).

Note that there are many other restrictions on spatial representation systems, depending on which spatial relations in the plane (or in nD space) are employed in a representational capacity. For example (see Lemon and Pratt 1998b), if equal distance between points is to be used in representation (say, in representing political opposition), then the following *equidistance* constraint applies:

- there are at most n + 1 mutually equidistant points in an n-dimensional space.

Thus, for example, one could not represent (by way of equidistant points in the plane) more than 3 political parties all being equally opposed to each other. Similar constraints on "nearness" and connection relations are noted in the context of logics of spatial relations (see Lemon 1996; Lemon and Pratt 1997b).

Thus, attention to the details of possible spatial arrangements of regions in a diagram, and their semantics, may reveal that certain diagrammatic systems cannot do all the representational work that we might require of them. Having discussed the efficacy of various diagrammatic systems, we present a framework which we think allows an explanation of such phenomena.

7. The Theoretical Framework

As outlined earlier, our account of representational and computational efficacy properties is based on the notions of the *availability* of constraints, constraint *preservation*, direct interpretation, and the processing of restricted representations.

As noted in the introduction, the presence of constraints cannot be the whole story in an account of efficacy. For one may construct tightly restricted sentential languages, and yet have no idea of how to reason with them, until one is presented with an efficient theorem prover tailored to the language. In contrast, constraints in diagrammatic systems are often "available" to reasoners.

Thus the following definition of an *effective* diagrammatic representation system (for an agent) forms the cornerstone of our account.

DEFINITION 2. *An effective Diagrammatic Representation system (for an agent) is a diagrammatic representation system in which graphical and spatial relations between representing tokens are directly semantically interpreted as relations between objects in the target domain. Furthermore,* *i) the constraints on represented relations match those of their corresponding graphical or spatial relations in the diagrams;*

ii) these constraints are available to the agent;

iii) inference with the representations is tractable.

This definition incorporates conditions on representational efficacy (i) and computational efficacy (iii), and acknowledges that efficacy of a diagrammatic system is relative to an agent (ii). The point of directness of interpretation is that "constraint preservation" (Shimojima 1996) enables a reasoner with only meagre knowledge of the core semantics to infer what the semantics of a representation are from the "surface form" of the representation itself. Again, the contrast with the sentential cases is that the relation of concatenation has no direct semantic interpretation since the concatenation relation does not share any structure with represented relations. Such observations lead us to claim that in the diagrammatic case, constraints are "available" to agents reasoning with effective diagrams. Thus the notions of constraints, direct interpretation, availability, and low complexity together allow us to explain the efficacy of diagrams.

Note that it is a consequence of our definition that reasoning with Euler's Circles (convex sets in the plane) is not efficacious for 4 or more sets (due to the result of section 6.2, since constraints are not matched).

The notion of "availability" of constraints has much to do with the abilities and assumptions a user brings to a representation scheme. For example, that inclusion over regions is transitive is, we think, the sort of structural knowledge of constraints that is available to nearly all agents in their interpretation of representing relations in diagrams. Similarly, it seems to us that most users of graphical representations expect there to be a uniqueness restriction on tokens in the representation; that one token stands for one object in the target structure, and that distinct tokens stand for distinct objects. That multiple representation is not conventionally used in graphical representations reflects an assumption of a constraint that a user might bring to their interpretation of diagrams. Another such "assumption of a constraint" that users may bring to diagrams is that of planarity – for example, that arcs representing relations do not cross. If they do cross, of course, some further semantics is needed for the intersection points. Further, the use of convex regions might also be such a conventional or assumed constraint on diagram construction. Of course, empirical work would have to be done to establish such claims.

Given this explanatory framework, somewhat more technical questions arise. For instance, what is the computational complexity of reasoning with regions of the plane? What is the expressive power of various systems for combining different regions, lines, and points?

As we have seen, depending on which spatial relations are of representational import, different constraints operate on possible representations. A future theory of efficacious representation selection and construction will describe how to match such restrictions to the restrictions inherent in the target domain. Thus a general theoretical focus for DR research shall be on appropriate representation selection, or construction, for a given problem rather than the invention or discovery of a universal representation language.

8. A Sample Analysis

In order to explore the explanatory potential of our framework, we present here an analysis of a possible diagrammatic system which might be used for a fragment of syllogistic reasoning (cf. Englebretsen 1992). The analysis we offer is intended to illustrate the predictive and explanatory power of our proposed "alignment" between logic and psychology in this domain.

8.1. *"Tile" diagrams for set intersection and inclusion*

Suppose that the following diagrammatic system were proposed, in order to represent and reason about set theoretic statements of the form "Some A are B, all B are C, no A are C" and so on.

DEFINITION 3. *(Tile diagrms)*
Let each set be represented by a unique connected polygonal region of the plane. (These polygonal regions are the "tiles"). If two such regions share some portion of a boundary, this represents that the intersection of their respective sets is non-empty. In addition, if the intersection of two regions is empty, then their corresponding tiles must not share any portion of a boundary line. Further, if one tile is surrounded by another, this represents that the set represented by the outer tile contains the set represented by the inner tile. No two tiles may overlap (i.e. they partition the plane).

Thus, a reasoner is to represent set intersections by the drawing of tiles in the plane which meet along appropriate boundaries. See diagrams 3 and 4 for some examples. Inference with the representations is somewhat trivial (as is the hallmark of effective diagrammatic systems) – once the diagrams are drawn, one simply reads off boundary contact relations in order to infer the presence of non-empty set intersections and set inclusions. One might conjecture that the representation system fails to generate any interesting inferences. As our analysis will show, however, this representation system in fact generates too many inferences (some which are logically unsound).

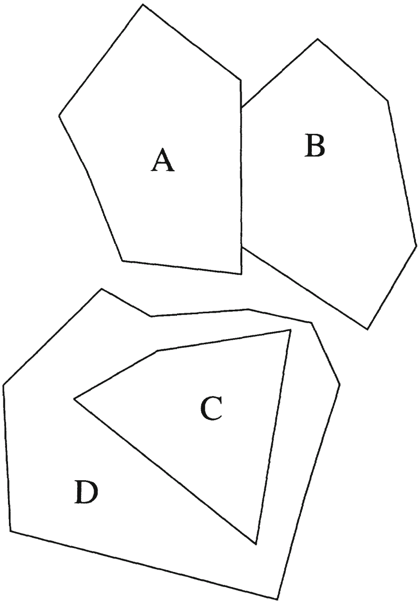


Figure 3. "Some A are B, no B are C, no A are C, no A are D, no B are D, all C are D".

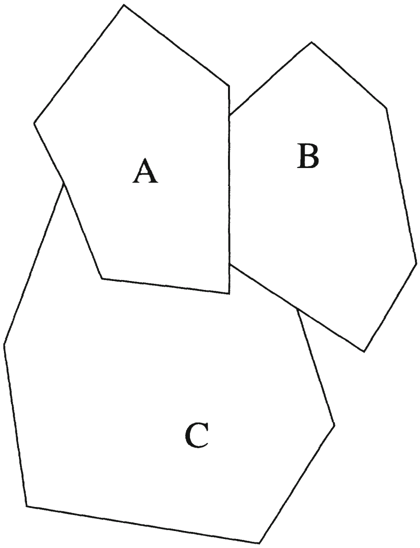


Figure 4. "Some A are B, some B are C, some A are C".

What predictions does our account make about such a proposal, and what kind of analysis does it offer?

8.2. *The analysis*

The simple tiling system is restricted in its expressive power, since it only uses two spatial relations in its representational work. A few constraints are immediate; specificity of the polygonal regions is the basic restriction, and then symmetry of the relation of boundary contact, which is directly (semantically) interpreted as set intersection, is the another obvious constraint. Further, it is clear that the "surrounds" relation (directly interpreted as set inclusion) on the tiles is transitive and acyclic. Such trivial constraints are the sort of restriction that we might reasonably expect to be "available" to almost any user of this diagrammatic system. Thus the tiling diagrams meet many of our criteria in definition 2 for efficacy of a diagrammatic system. So far, then, the proposed system looks promising.

In terms of complexity too, our analysis predicts that the system is efficacious, since the problem of reasoning about "realization of explicit topological expressions" in "medium resolution without overlap" of Grigni et al. (1995) is precisely the complexity of drawing these tiling diagrams, and is known to be polynomial. (We shall discuss the relevance of the results of Grigni et al. (1995) in section 9.)

So far so good, but what does our analysis tell as about semantic properties of the proposed representations? Here is an instance where a formal analysis of representational power reveals a (possibly unavailable) constraint. This time the *planarity* of the proposed system forces there to be further constraints on the diagrams – ones which, we expect, would not be available to many users of the system (those with some knowledge of topology excepted). Indeed, we believe that planarity and convexity constraints are not available to (average) reasoners. Our evidence for this is circumstantial at the moment; people are surprised when they discover the planarity and convexity problems, and the complexity of those constraints makes it difficult to imagine them being available. In short, it is difficult to know what access people have to these constraints, and the prima facie evidence is that whatever access they have is very weak.

In fact, planarity of the diagrams means that the system cannot reliably be used to reason about more than 4 sets! To see this, try to construct a tile representation of the following set of sentences (S): "Some A are B, some B are C, some C are A, some D are A, some D are B, some D are C, some A are E, some B are E, some some C are E, some D are E." (Note that S is a consistent set of sentences.) Figure 5 shows such an attempt – note that it fails to represent that some B are E, and that any attempt to draw such a

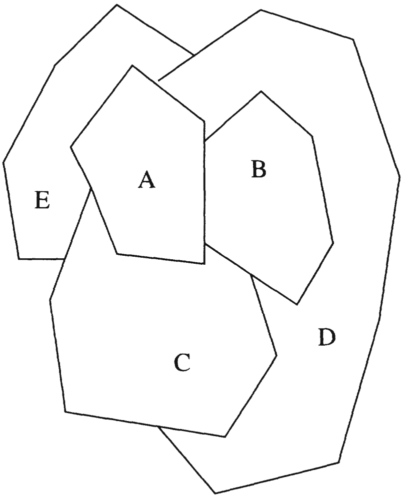


Figure 5. An attempt to realize situation S.

diagram will, of necessity, fail to represent one of the (logically possible) set intersection statements in the above description (S). Note also that reasoning with the system would thus lead a user astray; one would infer from a diagram such as Figure 5 that "No B are E", which does not follow from the problem description.

Thus our analysis predicts that the system is efficacious for reasoning about 4, or fewer, sets only. Of course, few of the possible representations involving more than 4 sets will fall foul of planarity problems in practice, so it would be useful to investigate something like the "confidence measure" that a user could reasonably have in using the tiling system, as opposed to some other representation scheme. Finally, our analysis suggests a way of bypassing the semantic problem. Using three-dimensional solids rather than tiles would overcome the representability limitations, and preserve the availability of constraints and their direct interpretation.

We now turn to some concrete implications of our framework for research directions in logic and psychology.

9. Extending the Logical Approach

Formal logic has been implicated in three major areas in the preceding discussion;

1. formal semantics of representation systems;
2. logics of spatial and graphical relations;
3. complexity theory for diagrammatic systems.

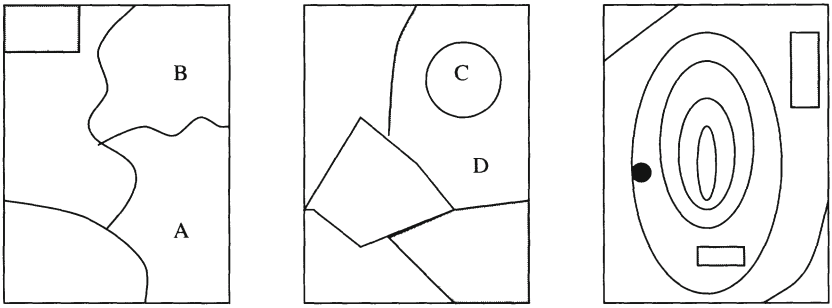


Figure 6. Some simple maps.

We shall say something about the prospects in each of these areas.

First, as noted above, standard logical approaches do not deal adequately with diagrammatic phenomena of approximate representation. An appropriate formal semantics should thus incorporate a description of failures in representation, and degrees of representational adequacy. A start has been made on this more detailed theory of representation (where the representation relation is not an "all-or-nothing" affair), in Lemon and Pratt 1998b; Lemon 1997b) (using Channel Theory (Barwise and Seligman 1997)). Second, in applications involving visual information it seems that we require formalisms which allow us to encode and process the spatial (and graphical) relationships between representational primitives. In this vein, (qualitative) spatial logics have been investigated, and some complete spatial description languages developed (Lemon and Pratt 1997b; Pratt and Schoop 1998). So we claim that there is a tight connection between QSR and our understanding of visual languages, such as maps (cf. Haarslev 1995). Visual languages exploit the spatial structure of the graphical medium in their representational work, so that part of uncovering the meaning of a representation in a visual language *is* qualitative spatial reasoning. Thus a logic for representing spatial situations (and reasoning about them) will also provide us with a formal language for expressing spatial configurations of tokens in graphical representations. The semantics of such a language will be a simple *formal semantics* for those descriptions. Consider the maps in diagram 6 for example. Part of their meaning is qualitative and spatial; that region A touches region B, that D surrounds C, and so on. Such representations are commonplace in Geographical Information Systems (GIS) (see Worboys (1995) for an introduction), yet there is currently no formal theory of their representational adequacy, or inadequacy.

However, more work needs to be done to investigate logics for qualitative spatial reasoning which are appropriate from the point of view of diagrammatic representation. In this connection note that a well-developed logical theory of relational structures already exists. *Modal logics* are structural descriptors par excellence (van Benthem 1984), and their computational properties are well-explored too. That there is some connection between modal logics and graphical systems has quite often been remarked upon. As shown by Lemon and Pratt (1997b; Lemon 1996), some effective logics of planar spatial relations may be sought amongst the *extended* modal logics $\mathcal{L}(\mathcal{D}, \Diamond)$ (de Rijke 1992; Gargov and Goranko 1993), for these extended systems allow us to distinguish all finite non-isomorphic relational structures. Extended modal logics also allow a restriction to constants,[10] thus incorporating specificity of diagrammatic tokens. Moreover, modal logics are said to take an "internal perspective" on relational structures (as opposed to the "god's eye view" of classical first-order logic) and they employ operators which evaluate information by "locally scanning" points which are accessible to them. Such properties make extended modal languages natural from the point of view of visual processing. Further investigation of the "hybrid languages" of Blackburn and Seligman (1995) seems particularly promising in this regard.

In addition, non-spatial *graphical relations* (such as relative saturation) which are often employed in visual languages, may be encoded in this framework (e.g. for saturation, using the extended modal logic defining a dense linear ordering of de Rijke (1992)). These systems, and the promise of multimodal systems, in which different operators may express spatial and graphical relations between representing tokens, remain to be fully explored.

Various researchers have proposed a similar analysis of graphical languages. Levesque (Levesque 1986, 1988), for instance, investigates "vivid" knowledge bases. These are variable-free fragments of first-order logic (FOL) with distinct constants, restricted quantification over distinguished predicates, no disjunction, and a closed world assumption. Determining entailment in such knowledge bases is shown to be tractable (Levesque 1988). Indeed, Levesque also speculates that,

> . . . perhaps the main source of vividly represented knowledge is *pictorial information.* (Levesque 1986, p. 97)

Interestingly, Howell (1976) and Sober (1976) make similar proposals while considering the relation between logics, diagrams, and mental representations. For the development of the theory, it is important that the formal properties of (something like) this class of languages be established.

Finally, coming to the prospect of a complexity theory relevant to diagrammatic reasoning, recall our motivation for investigating diagrammatic languages as spatially restricted logical languages. We wish to establish the *computational efficacy* of DRs by way of the complexity properties of the logics which they embody. The thesis is that some diagrammatic representation systems may be successfully analysed as relational fragments of FOL, perhaps of low complexity. This analysis shall make precise the various claims about their tractability.

To canvass just a few relevant results here, it is well known that certain fragments of FOL enjoy polynomial satisfiability. For example, satisfiability of the Horn fragment of FOL is in **P** (see Papadimitriou 1994, p. 79). In connection with spatial logics, Bennett's intuitionistic logic for qualitative spatial reasoning has been shown to be a polynomial time fragment (Nebel 1995) (actually, the fragment is in **NC**; efficiently solvable on parallel machines.) However, it seems we need to delve deeper than this if we are to gain complexity results relevant to diagrammatic systems. The results of Grigni et al. (1995) on "topological inference", although not directly concerned with diagrammatic reasoning, are a promising starting point here.

In the terminology of Grigni et al. (1995) "explicit topological expressions" are those for which a spatial relation is specified for every pair of regions and "medium resolution" is a level of description involving the relations "overlaps" and "contains" (between connected regions of the plane). Thus determining whether a specified diagram of non-convex (2D) Euler's Circles can be drawn or not corresponds to the problem of "realizability" (whether a description of a set of regions can be drawn in the plane) for explicit topological expressions in medium resolution. This problem is shown to be NP hard. Interestingly the same problem for "GIS-like"representations ("medium resolution" representations with boundary contact instead of overlap) can be solved in polynomial time (see our "tiling" example of section 8). These GIS-like representations consist of elements of partitions of the plane – regions which may not overlap each other, but only meet at boundaries or exhibit inclusion relations.

The results just mentioned are for the *unrestricted* non-convex cases (i.e. where there may be enough regions involved for planarity problems to arise). However, the restricted systems have also been investigated (again, not in the context of DR). They are what Grigni et al. (1995) call the *constraint satisfaction* problems, since they amount to computing various path-consistency algorithms over relation composition tables. They are all solvable in *polynomial* time. The upshot here, then, is that as long as the number of regions is restricted so as to avoid spatial difficulties (see Lemon and Pratt 1997c), reasoning with them is of low complexity. This type of result for the complexity of restricted diagrammatic systems illustrates the interaction between the representational and complexity aspects of efficacy; certain DR systems are computationally efficacious only when they are restricted so as to avoid their difficulties with representational efficacy.

Such results promise further progress in the application of computational complexity to diagrammatic reasoning (see Lemon and Pratt 1997c). We now return to the prospects that our framework suggests for the psychological studies.

10. Extending the Psychological Theory

Starting out to develop a formal framework based on expressiveness has lead to a focus on the processes of coming to understand (or construct) an interpretation for diagrams. Simultaneously a shift has occured from thinking about the one true mental representation system to thinking about how users adopt the representations they do from the indefinitely large space possible. According to this point of view, it is no accident that diagrams should figure so strongly in the teaching of new domains – in many domains more strongly in the teaching than in the practice. This is one focus of attention for developing psychological theory.

This focus will demand, and can exploit, comparisons between sentential and diagrammatic systems (as well as between different diagrammatic systems). In many fields, large individual differences occur between users which are in some (as yet poorly specified) way related to the difference between visualisations and verbalisations as representation systems. While this enormously complicates the psychologists' task, it also offers a methodological approach. If two groups of users are shown to contrast in how they respond to using different kinds of external representation of the same information, then these global differences can be used to pinpoint differences in underlying processes. Finding the differences in style can help construct computational models of alternative mental processes (see e.g. Oberlander et al. 1996a, 1996b). Lack of process accounts of these individual differences has been what has most retarded advances in their understanding. This quest for a characterisation of what it is to be a 'visualiser' or a 'verbaliser' needs a foundation in semantic analysis of the differences between different external representations. We have put much emphasis on the fact that sentential systems differ from diagrammatic ones in the degree to which their constraints may be available to users with different knowledge. But they also differ in what linguists call their 'information packaging' (see e.g. Vallduví 1992; Vallduví and Engdahl 1997). Sentences distribute information according to their speaker's beliefs about the hearer's prior knowledge. Packaging may be manipulated by lexical, syntactic, and prosodic means according to different systems in different languages. Diagrammatic systems have no such systematic information packaging. In many domains, these differences may be critical. Diagrams are frequently helpful because they abstract away from the information packaging habits of our natural languages in teaching formal systems which have no such packaging. So in teaching elementary logic, much of what a student has to learn is that the subject/predicate distinction in their natural language is quite different from the function/argument distinction made in logical calculi. This difference is intimately bound up with the differences in social relations between communication as exposition as opposed to communication as derivation. This serves as one example where empirical investigation of the systematic impact of the differences in information packaging between diagrams and sentences would be fruitful (see Stenning (1996) for an extended discussion).

As we have seen, making directness of semantic interpretation the essential feature of DRs actually classifies some graphical representation systems with abstract syntax as non-diagrammatic. The best examples are expressively interpreted node-and-link formalisms (see e.g. Schubert 1976; and for a discussion from the present perspective (Stenning and Inder 1995)). Empirically, these systems provide an important field for a programme of investigation into representations. Here is a kind of diagram which can be given interpretations of a complete range of expressiveness from completely concrete wiring diagrams to the lambda-calculus. If expressiveness is an important determinant of cognitive properties, here is an opportunity to investigate the role of expressiveness of interpretation while holding the diagram constant. Of course the complexities of empirical investigation arise for the same reason as they do in studying how people can learn to exploit constraints in sentential languages – the constraints are implicit in the knowledge of interpretation which users bring to the task. The evidence such as it is at present of the use of expressive node-and-link formalisms in computer science is mixed. This is what our framework suggests. Their usefulness will depend on subtle issues about what users know about implicit constraints (i.e. availability of constraints). For a review of recent empirical work, see Whitley (1997).

Finally, the kinds of semantical and computational analyses discussed above need to be extended to as many different kinds of representation as possible. For example, we have recently looked at how semantic analysis might be applied to the distinction between evanescent and persistent graphical media (static diagrams vs animation; Stenning 1995). Typically, interpretation of the temporal dimension of animated media is direct in the same way as diagrams interpret space directly. This has some of the same consequences for the cognitive properties of animation, though memory is implicated in a rather different way. There is in principle no reason why these very general semantic distinctions cannot be applied to any representation system. For the healthy development of theory, it is vital that they should.

11. Conclusion

We have argued for a conceptual framework on which to base a program of research into DRs jointly between logicians and psychologists. The account centres on expressiveness, constraints on expressiveness, and the availability of constraints to users with different knowledge. Our aim has been to show how the concerns of the two disciplines are inseparable in an account of thinking with diagrams. They are in symbiosis rather than in competition, and need each others' results in order to steer their own research.

Like all good research programmes, this one leaves many things out. In particular, it leaves out what the study of diagrams has mostly concentrated on so far – the study of the sometimes subtle differences in design which make different diagrammatic renderings of the same information easier or harder to use. These discussions originated from the craft knowledge of expert graphic designers. This territory constitutes a large, legitimate, and practically important domain for any student of "thinking with diagrams". There is a related and growing field concerned with how to enable machines to perceive and produce diagrams. Again this is a hard problem of legitimate scientific interest and practical application.

Beside these areas of concern, our programme looks strange indeed. Our comparisons between diagrammatic systems and fragments of logical languages are like comparison of hawks and handsaws as compared with the 'psychometrics' of graphic design. We do not wish to be exclusive. All these kinds of research are needed, and each has something to say to the others. But we do not find this strangeness suprising. A comparison with the development of the understanding of natural languages as representation systems may be helpful. Stylistics is a very old discipline which embodies the accumulated wisdom of writers. AI researchers study hard problems in the perception and production of speech and writing. But there is a mostly philosophical tradition which gave rise to the fundamental study of linguistic semantics and which began by asking rather strange questions about the relative meanings of different 'toy' sentences. Most of what we now know about the psychology of what people are doing when they use language and how they achieve it had to be grounded on these rather esoteric foundations.

Acknowledgements

The support of the Economic and Social Research Council UK (ESRC) and the Leverhulme Trust (grant #FAQ/120/AQ) is gratefully acknowledged. The work was partly supported by the ESRC funded Human Communication Research Centre. We also wish to acknowledge partial support from the following grants: Grant #L127251023 from ESRC; Cognitive Engineering Grant #GRL 86930 from the EPSRC; MNA Initiative and Grant #GRL 37953 from the EPSRC. We would also like to thank Tom Addis, Alan Blackwell and colleagues who contributed to the ESRC/EPSRC sponsored 'Thinking with Diagrams Workshop' 1997 at which a version of this paper was originally presented. Fruitful discussions also took place with Ian Pratt and Jon Shapiro in Manchester, and with Corin Gurr, John Lee, Jon Oberlander, and the Graphics and Language Group at the HCRC.

Notes

[1] "Representations are objects in the world, and as such they obey certain structural constraints that govern their possible formation. The variance in inferential potential of different modes of representation is largely attributable to different ways in which these structural constraints on representations match with the constraints on targets of representation" (Shimojima 1996, p. 13).

[2] Indeed, it seems that sentential or "list-like" languages require the interposition of abstract syntax in order to increase their expressive power (beyond that of a finite state 'language').

[3] Whereas spoken sentences, for example, are "acoustic objects", which exploit only temporal/sequential structure.

[4] Sentential languages may make direct interpretation of some features of 'layout' such as itemisation by bullet point, but these are 'graphical' features superimposed on a fundamentally indirect semantics which interprets an abstract syntax.

[5] One way of reminding ourselves is to remember that there have historically been quite different concatenation practices in written language than the ones we use today. At one time Greek was written right-to-left and left-to-right on alternate lines and without spaces between words.

[6] We refer to these as "graphical" relations. Thus spatial and graphical relations together make up the diagrammatic relations.

[7] Raised in the "Thinking with Diagrams" discussion by Yuri Engelhardt.

[8] Of course, we do not mean to exclude the possibility that verisimilitude may also be an issue for sentential representations.

[9] Helly's Theorem: Let $X_1, \ldots, X_N$ be convex regions in n-dimensional Euclidean space, $N \geq n + 1$, such that each $n + 1$-membered collection of the $X_1, \ldots, X_N$ has a nonempty intersection. Then $X_1, \ldots, X_N$ has a nonempty intersection.

[10] This is not possible in standard modal logics.

References

Allwein, G. & Barwise, J. (eds.) (1993). *Working Papers on Diagrams and Logic*, volume IULG-93-24 of *Indiana University Logic Group Preprint Series*. Indiana University, Visual Inference Laboratory.

Barwise, J. & Seligman, J. (1993). Imperfect Information Flow. In *Proceedings of the 8th Annual IEEE symposium on Logic in Computer Science*.

Barwise, J. & Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems.* Cambridge University Press: Cambridge. Tracts in Theoretical Computer Science 44.

Barwise, J. & Shimojima, A. (1995). Surrogate Reasoning. *Cognitive Studies: Bulletin of Japanese Cognitive Science Society* **4**(2): 7–27.

Blackburn, P. & Seligman, J. (1995). Hybrid Languages. *Journal of Logic, Language, and Information* **4**: 251–272.

Braine, M.D.S. (1978). On the Relationship Between the Natural Logic of Reasoning and Standard Logic. *Psychological Review* **85**: 1–21.

Clark, H.H. (1969). Linguistic Processes in Deductive Reasoning. *Psychological Review* **76**: 387–404.

de Rijke, M. (1992). The Modal Logic of Inequality. *Journal of Symbolic Logic* **57**(2): 566–584.

Dobson (1997). *Information Enforcement and Learning: Improving Syllogistic Reasoning Skills.* PhD thesis, Open University, Milton Keynes. (in preparation).

Englebretsen, G. (1992). Linear Diagrams for Syllogisms (with Relationals). *Notre Dame Journal of Formal Logic* **33**(1): 37–69.

Gärdenfors, P. (1996). Mental Representation, Conceptual Spaces, and Metaphors. *Synthese* **106**: 21–47.

Gargov, G. & Goranko, V. (1993). Modal Logic with Names. *Journal of Philosophical Logic* **22**: 607–636.

Grigni, M., Papadias, D. & Papadimitriou, C. (1995). Topological Inference. In *International Joint Conference on Artificial Intelligence (IJCAI '95).* AAAI Press.

Gurr, C.A. (1996). On the Isomorphism (or Otherwise) of Representations. In *International Workshop on the Theory of Visual Languages (AVI'96 Post-Conference Workshop).*

Haarslev, V. (1995). Formal Semantics of Visual Languages using Spatial Reasoning. In *IEEE symposium on Visual Languages*, 156–163. IEEE Computer Society Press.

Hammer, E.M. (1995). *Logic and Visual Information.* Studies in Logic, Language, and Computation. CSLI Publications and FoLLI: Stanford.

Howell, R. (1976). Ordinary Pictures, Mental Representations, and Logical Forms. *Synthese* **33**: 149–174.

Huttenlocher, J. (1968). Constructing Spatial Images: A Strategy in Reasoning. *Psychological Review* **75**(6): 550–560.

Johnson-Laird, P. N. (1983). *Mental Models.* CUP.

Larkin, J. & Simon, H. (1987). Why a Diagram is (Sometimes) Worth 10,000 Words. *Cognitive Science* **11**: 65–99.

Lemon, O. (1996). Semantical Foundations of Spatial Logics. In Aiello, L.C., Doyle, J. & Shapiro, S.C. (eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, 212–219. Morgan Kaufmann Publishers: San Francisco, CA.

Lemon, O. (1997a). Review of "Logic and Visual Information" by E.M. Hammer (CSLI Publications). *Journal of Logic, Language, and Information* **6**(2): 213–216.

Lemon, O. (1997b). Theories of Representation. In Childers, T., Kolar, P. & Svoboda, V. (eds.) *Logica '96: Proceedings of the 10th International Symposium*, 120–135. Filosofia Academic Publishing: Prague.

Lemon, O. & Pratt, I. (1997a). Logical and Diagrammatic Reasoning: The Complexity of Conceptual Space. In Shafto, M. & Langley, P. (eds.) *19th Annual Conference of the Cognitive Science Society*, 430–435. Lawrence Erlbaum Associates: New Jersey.

Lemon, O. & Pratt, I. (1997b). On the Incompleteness of Modal Logics of Space: Advancing Complete Modal Logics of Place. In Kracht, M., de Rijke, M., Wansing, H.

& Zakharyaschev, M. (eds.) *Advances in Modal Logic*, 113–130. CSLI Publications: Stanford.

Lemon, O. & Pratt, I. (1997c). Spatial Logic and the Complexity of Diagrammatic Reasoning. *Machine GRAPHICS and VISION* **6**(1): 89–108. (Special Issue on Diagrammatic Representation and Reasoning).

Lemon, O. & Pratt, I. (1998a). On the Insufficiency of Linear Diagrams for Syllogisms. *Notre Dame Journal of Formal Logic* (to appear).

Lemon, O. & Pratt, I. (1998b). Putting Channels on the Map: A Channel-Theoretic Semantics of Maps? In Moss, L. (ed.) *Logic, Language, and Computation, Volume 2*. CSLI Publications: Stanford (to appear).

Lemon, O., de Rijke, M. & Shimojima, A. (1999). Efficacy of Diagrammatic Reasoning (Editorial). *Journal of Logic, Language, and Information* **8**.

Levesque, H.J. (1986). Making Believers Out of Computers. *Artificial Intelligence* **30**: 81–108.

Levesque, H.J. (1988). Logic and the Complexity of Reasoning. *Journal of Philosophical Logic* **17**: 355–389.

Nebel, B. (1995). Computational Properties of Qualitative Spatial Reasoning: First Results. In Wachsmuth, I., Rollinger, C.-R. & Brauer, W. (eds.) *KI-95: Advances in Artificial Intelligence*, 233–244. Springer-Verlag: Berlin. (19th German Conference on Artificial Intelligence).

Oberlander, J., Cox, R., Tobin, R., Stenning, K. & Monaghan, P. (1996a). Individual Differences in Proof Development Strategy Following Multimodal Logic Teaching. In *Cognitive Science Society of America*. La Holla, CA.

Oberlander, J., Stenning, K. & Cox, R. (1996b). Hyperproof: The Multi-Modal Moral. In *Information Theoretic Approaches to Logic Language and Computation*. Regent's College: London.

Palmer, S.E. (1978). Fundamental Aspects of Cognitive Representation. In Rosch, E. and Lloyd, B.B. (eds.) *Cognition and Categorization*, 259–303. Lawrence Erlbaum Associates: Hillsdale, N.J.

Papadimitriou, C. (1994). *Computational Complexity*. Addison Wesley: New York.

Pratt, I. & Schoop, D. (1998). A Complete Axiom System for Polygonal Mereotopology of the Real Plane. *Journal of Philosophical Logic* (to appear).

Rips, L.J. (1994). *The Psychology of Proof*. MIT Press: Cambridge, MA.

Schubert, L. (1976). Extending the Expressive Power of Semantic Networks. *Artificial Intelligence* **7**(2): 163–198.

Shimojima, A. (1996). *On the Efficacy of Representation*. PhD thesis, Indiana University.

Shin, S.-J. (1995). *The Logical Status of Diagrams*. Cambridge University Press: Cambridge.

Shin, S.-J. & Lemon, O. (1999). *Diagrams*. Stanford Encyclopedia of Philosophy.

Sober, E. (1976). Mental Representations. *Synthese* **33**: 101–148.

Stenning, K. (1995). Distinguishing Semantic from Processing Explanations of the Usability of Representations: Applying Expressiveness Analysis to Animation. In Lee, J. (ed.) *Proceedings of Intelligent Multi-Modal Interface Workshop*. AAAI Press. in press.

Stenning, K. (1996). Embedding Logic in Communication: Lessons from the Logic Classroom. In *Logic and Argumentation. Proceedings of the Colloquium on Logic and Argumentation, Royal Netherlands Academy of Arts and Sciences*, 227–240. North Holland, Amsterdam.

Stenning, K., Cox, R. & 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.

Stenning, K. & Inder, R. (1995). Applying Semantic Concepts to Analyzing Media and Modalities. In Glasgow, J., Narayanan, N.H. & Chandrasekaran, B. (eds.) *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. AAAI Press/The MIT Press: Cambridge, MA.

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. (1997). Assigning Information to Modalities: Comparing Graphical Treatments of the Syllogism. In *Logic, Action and Cognition: Essays in Philosophical Logic*, 211–228. Kluwer (in press).

Stenning, K. and Yule, P. (1997). Image and Language in Human Reasoning: A Syllogistic llustration. *Cognitive Psychology* (in press).

Vallduví, E. (1992). *The Informational Component*. Garland: New York.

Vallduví, E. & Engdahl, E. (1997). The linguistic Realisation of Information Packaging. *Linguistics* (to appear).

van Benthem, J. (1984). Correspondence Theory. In Gabbay, D. & Guenthner, F. (eds.) *Handbook of Philosophical Logic*, volume 3. Reidel: Dordrecht.

Whitley, K.N. (1997). Visual Programming Languages: The Empirical Evidence for and Against. *Journal of Visual Languages and Computing* **8**(1): 109–142.

Worboys, M. (1995). *GIS: A Computing Perspective*. Taylor and Francis: London.

Artificial Intelligence Review **15:** 63–78, 2001.

Diagrammatic Reasoning: An Artificial Intelligence Perspective

PATRICK OLIVIER
Department of Computer Science, University of York, Heslington, York, YO10 5DD, UK

Abstract. A common motivation for developing computational frameworks for diagrammatic reasoning is the hope that they might serve as re-configurable tools for studying human problem solving performance. Despite the ongoing debate as to the precise mechanisms by which diagrams, or any other external representation, are used in human problem solving, there is little doubt that diagrammatic representations considerably help humans solve certain classes of problems. In fact, there are a host of applications of diagrams and diagrammatic representations in computing, from data presentation to visual programming languages. In contrast to both the use of diagrams in human problem solving and the ubiquitous use of diagrams in the computing industry, the topic of this review is the use of diagrammatic representations in automated problem solving. We therefore investigate the common, and often implicit, assumption that if diagrams are so useful for human problem solving and are so apparent in human endeavour, then there must be analogous computational devices of similar utility.

Keywords: diagrammatic reasoning, knowledge representation and reasoning

1. Introduction

Within artificial intelligence, systems that have claimed to comprise some degree of diagrammatic reasoning capability have tended to be restricted to two problem domains: geometry theorem proving and discovery, and reasoning about physical systems. Despite their highly spatial nature, the two domains exhibit some interesting areas of contrast. In the geometric domain we encounter well defined problems, less emphasis on ontological choices, and therefore a clean discussion of the integration of the sentential and diagrammatic representations. In the physical systems domain, debate focuses more on ontological issues and the soundness of the reasoning. In addition to these domains, as we will discover in section 3, the property of concretization provided by diagrammatic representations has been much exploited in the field of representation and processing of spatial expressions.

For the benefit of facilitating an initial foray into the utility of diagrammatic representations, consider an example of a very common, and successfully exploited, diagram given in Figure 1 representing the ancestry relations between a collection of ten objects labelled with the letters A–J. The arrows

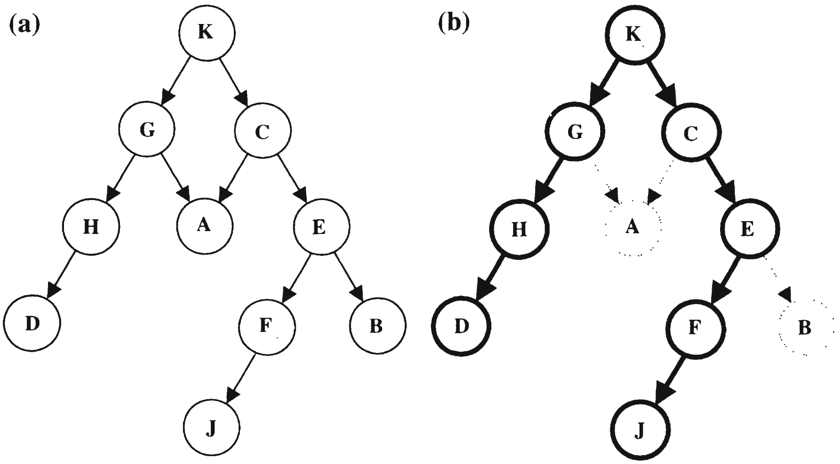


Figure 1. A family tree and examples of perceptual Gestalts that are a natural consequence of the layout conventions. As a result of direct inspection K is deemed an ancestor of J through its membership of the same perceptual cluster and not as a result of transitive inference.

correspond to the direct ancestry relations, so, for example, object G is the direct descendant of object K, and object H is the direct descendant of G. Such diagrams have layout conventions, for example, that each object appears only once and that sequences of descendants form visually contiguous segments. Figure 1(b) shows two such visually continuous segments (these segments are shown in bold).

If, for a moment, we consider the human (for whom this particular diagrammatic convention was designed) as the processor of this external representation, it is apparent that the structure of the diagram exploits a number of aspects of our visuospatial perception to facilitate both the extraction of certain information and performance of inferences. That is, amongst other considerations such as aesthetics and attention limitations, the design conventions for such diagrams exploit a human bias to visually parse the elements of images into groups of features known as perceptual Gestalts (for a review of Gestalt psychology see Gordon (1989), chapter 3). For example, the layout of successive ancestry relations are such that the resulting segments of the diagram follow the Gestalt grouping laws of adjacency and good continuity. This grouping of particular collections of relations has the consequence that certain queries can be resolved particularly efficiently. For example, whether K is an ancestor of F can be resolved without reference to the intermediate relations, but simply on the basis of the membership of K and F in the same group and the relative heights of K and F. The common ancestor between D and J corresponds to the object in the intersection of the

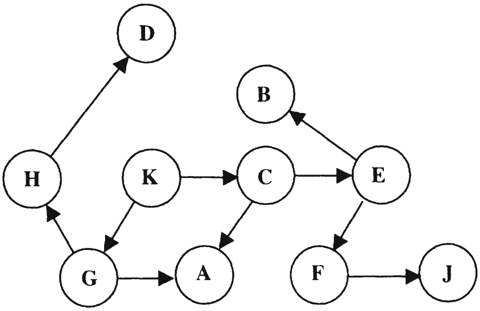


Figure 2. A family tree without length and position constraints.

perceptual group to which D and J belong. Thus the common ancestry of D and J can be computed without reference to the successive direct ancestry relations of each.

The full complement of binary kinship relations may be computed on the basis of generation (that is, relative heights of objects) and commonality of ancestry, without specific recourse to intermediate reasoning as to direct descendants and direct ancestors. The role of the diagram layout conventions in easing the resolution of such queries is easily illustrated by removing the constraints by which they are realised. Figure 2 illustrates such a diagram in which both the length of the links and the positions of the objects are unconstrained, and thus the perception of continuity and adjacency of objects is no longer allied to ancestry relations. Consequently, in such an unconstrained diagram, the lack of perceptual groupings that are meaningful with respect to the domain, means that ancestry queries must be resolved by incremental progression along the links.

In addition, the diagram facilitates qualitative comparison of features of the network of family relations. For example, since the nodes are of equal area and are equally spaced, the relative sizes of different sections of a family tree may be easily estimated by contrasting visual areas. In fact, the observation that the diagram aids the performance of inferences is problematic, since it is not transparent whether the diagram aids the performance of certain inferences or actually compiles them (relative to our "visual" processing capabilities) into the diagram in a manner that makes inference simply a matter of extraction. The distinction between the performance of inferences in the perception of a diagram and the design of a diagram such that inferences can be simply extracted from it, is not so clear cut (and we return to it in section 4), however, the adoption of either of these positions puts significant emphasis on the construction process, or rather the relationship between the domain captured by the diagram and layout conventions of the particular class of diagrams.

Summarising this initial foray into the use of diagrams as a representational device we have observed that a diagram's rules of composition are geared towards the exploitation of structure in our visual perception. However, whilst we are not explicitly interested in the cognitive status and processes involved in diagrammatic reasoning, we cannot avoid the fact that many of the approaches reviewed in the following sections are either explicit cognitive models or approaches that have been motivated by insights from human reasoning. Indeed, some approaches are directly interested in claims as to the nature of human internal representations, and many other computational formulations emphasise an analogical relationship with theories of human internal representation (Glasgow 1993; Narayanan, Suwa and Motoda 1995; Schwartz and Black 1996). In contrast, our goal is considerably less ambitious, and in the sections that follow we characterise three dimensions with respect to which we believe automated reasoning with diagrammatic representations is best characterised: knowledge indexing; concretization; and inference by inspection and transformation.

2. Knowledge Indexing

In a seminal paper, that spurred much of the recent interest in diagrammatic reasoning, Larkin and Simon distinguished a diagrammatic representation from a sentential representation as follows.

- In a *sentential* representation the expressions form a sequence corresponding, on a one-to-one basis, to the sentences in a natural language description of the problem. Each expression is a direct translation into a simple formal language of the corresponding natural language sentence.
- In a *diagrammatic* representation, the expressions correspond, on a one-to-one basis, to the components of a diagram describing the problem. Each expression contains the information that is stored at one particular locus in the diagram, including information about relations with the adjacent loci (Larkin and Simon 1987).

They considered, in some detail, two highly spatial problems, reasoning about pulley systems, and reasoning about geometric theorem proving, and contrasted the computational properties of a purely sentential representation and a diagrammatic representation. Thus, their diagrammatic representation comprised a sentential system in which each predicate is indexed by the location of the referents of the predicate arguments.

Figure 3 depicts Larkin and Simon's pulley system problem, and Table 1 and Table 2 give a particular statement of the problem, and the relevant domain knowledge, in a sentential representation. The domain knowledge is in the form of four productions which, given the class of pulley system

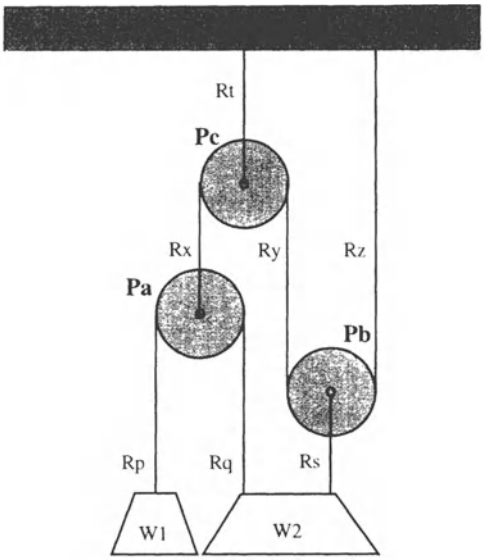


Figure 3. Diagram of the pulley problem (adapted from Larkin and Simon (1987)). The problem is to determine the tension and the value of weight W2 given the value of weight W1.

Table 1. Elements and relations fo the pulley problem (Larkin and Simon 1987)

Elements	Relations		
(Weight W1)	(Rope Rx)	(hangs W1 from Rp)	(hangs Pb from Rt)
(Weight W2)	(Rope Ry)	(pulley-system Rp Pa Rq)	(hangs Rt from c)
(Rope Rp)	(Rope Rz)	(hangs W2 from Rq)	(hangs Rx from c)
(Rope Rq)	(Pulley Pa)	(hangs Pa from Rx)	(hangs Rs from Pc)
(Rope Rs)	(Pulley Pb)	(pulley-system Rx Pb Ry)	(hangs W2 from Rs)
(Rope Rt)	(Pulley Pc)	(pulley-system Ry Pc Rz)	(value W1 1)

depicted in Figure 3, is adequate for reasoning about the tension of the ropes and ultimately the weight of W2.

The diagrammatic representation is an indexed set of predicates describing the elements of the scene and the relations. Thus the weights, ropes, pulleys and the ceiling are given unique location identifiers. Larkin and Simon construct a topological graph of the location identifiers, connecting identifiers where the corresponding elements coexist in a relation. This coexistence is termed adjacency, and it is this graph and indexing of the elements of the pulley problem and the relations, in combination with the sentential state-

Table 2. Domain knowledge for the pulley system problem (Larkin and Simon 1987)

P1. **Single-string support.**

Given a weight of known value ⟨n⟩ and a rope ⟨R⟩ from which it hangs, if there is no other rope from which it hangs (indicated by the symbol ~), then the supporting rope also has value (tension) ⟨n⟩ associated with it.

```
IF     (Weight ⟨Wx⟩) AND (Rope Rx) AND (Rope ⟨Ry⟩)
AND    (value ⟨Wx⟩ ⟨n⟩)
AND    (hangs ⟨Wx⟩ ⟨Ry⟩)
AND    ~(hangs ⟨Wx⟩ ⟨Rx⟩)
THEN   (value ⟨Ry⟩ ⟨n⟩)
```

P2. **Ropes over pulley**

If a pulley ⟨P⟩ has two ropes ⟨R1⟩ and ⟨R2⟩ over it, and the value (tension) associated with ⟨R1⟩ is ⟨n1⟩, then ⟨n1⟩ is also the value associated with ⟨R2⟩.

```
IF     (Pulley P) AND (Rope ⟨R1⟩) AND (Rope ⟨R2⟩)
AND    (pulley-system ⟨R1⟩ ⟨P⟩ ⟨R2⟩)
AND    (value ⟨R1⟩ ⟨n1⟩)
THEN   (value ⟨R2⟩ ⟨n1⟩)
```

P3. **Rope hangs from or supports pulley**

If there is a pulley system with ropes ⟨R1⟩ and ⟨R2⟩ over it, and the pulley system hangs from a rope ⟨R3⟩, and ⟨R1⟩ and ⟨R2⟩ have the values (tensions) ⟨n1⟩ and ⟨n2⟩ associated with them, then the value (tension) associated with ⟨R3⟩ is the sum of ⟨n1⟩ and ⟨n2⟩.

```
IF     (Pulley ⟨P⟩) AND (Rope ⟨R1⟩) AND (Rope ⟨R2⟩)
AND    (pulley-system ⟨R1⟩ ⟨P⟩ ⟨R2⟩)
AND    {(hangs ⟨R3⟩ from ⟨P⟩) OR (hangs ⟨P⟩ from ⟨R3⟩)}
AND    (value ⟨R1⟩ ⟨n1⟩)
AND    (value ⟨R2⟩ ⟨n2⟩)
THEN   (value ⟨R3⟩ (+ ⟨n1⟩ ⟨n2⟩))
```

P4. **Weight and multiple supporting ropes**

If a weight ⟨W1⟩ hangs from both ropes ⟨R1⟩ and ⟨R2⟩, but hangs from no other ropes, and the values (tensions) ⟨n1⟩ and ⟨n2⟩ are associated with ⟨R1⟩ and ⟨R2⟩, the value (weight) associated with ⟨W1⟩ is the sum of ⟨n1⟩ and ⟨n2⟩.

```
IF     (Weight ⟨W1⟩) AND (Rope ⟨R1⟩) AND (Rope ⟨R2⟩) AND (Rope ⟨R3⟩)
AND    (hangs ⟨W1⟩ ⟨R1⟩)
AND    (hangs ⟨W1⟩ ⟨R2⟩)
AND    ~(hangs ⟨W1⟩ ⟨R3⟩)
AND    (value ⟨R1⟩ ⟨n1⟩)
AND    (value ⟨R2⟩ ⟨n2⟩)
THEN   (value ⟨W1⟩ (+ ⟨n1⟩ ⟨n2⟩))
```

ment of both the problem and the domain knowledge, that constitutes the diagrammatic element of the representation of the problem.

The sentential and diagrammatic representations primarily diverge in how control is exercised during inference. Efficient inference with the sentential form may be effected through the inclusion of any number of control strategies, as at any particular point in the reasoning there are many possible instantiations of each of the four productions capturing the physics of the system. With the diagrammatic case, Larkin and Simon allow attention to be focused at a location and deem the elements accessible at a location to fall within the scope of attention. Attention may be shifted to locations indexed by an element within the current focus of attention. Thus attention may shift between adjacent nodes of the graph and it is shown that for the particular domain this control strategy yields highly directed and efficient inference. This cuts to the core of Larkin and Simon's claims as to diagrammatic representations, that one of the principal benefits of such an approach is that "diagrams can group together information that is used together thus avoiding large amounts of search for the elements needed to make a problem solving inference". Larkin and Simon were primarily interested in human problem solving abilities, and though their exemplification of the efficiency yielded by an attention mechanism constrained by visual perception (as captured within the adjacency constraint on attention shifts) within the framework of rule-based reasoning, is relevant to a discussion on the merits of diagrammatic representation, it was not claimed to be a original insight into reasoning in general. Indeed, efficient inference through knowledge indexing is a enterprise that has for many years occupied automated reasoning researchers, and the impact of Larkin and Simon's insight relied on the very acceptance that perceptual bias causes a "natural instantiation" of the indexing inherent in the diagram of the pulley systems.

Larkin and Simon's consideration of the power of localising knowledge in sentential form was a means of investigating the computational advantages afforded by diagrammatic representations. However, variations on such diagrammatic indexing schemes have been used to some considerable effect in a number of other application contexts. For example, in their POLYA system McDougal and Hammond (McDougal and Hammond 1992) use symbolic descriptions of the elements of a diagram to index proof plans. Also, in the geometry theorem proving domain, Koedinger and Anderson's Diagram Configuration (DC) model (Koedinger and Anderson 1992) implements a hypothesis that experts organise their knowledge according to diagrammatic schemas, and they use the recognition of different diagrammatic schemas in a geometry as the basis for planning the proof of a geometry theorem. Similarly, diagrammatic representations have been used in combination with case-based reasoning systems where the diagram has been used explicitly as the indexing scheme for the case concerned (Anderson and McCartney 1996).

3. Concretization

Though knowledge indexing is the most accessible element of the utility of diagrammatic representations, in that it is a core aspect of any formulation of a reasoning problem in terms of a sentential representation, concretization is the most apparent. Concretization focuses on the fact that for certain classes of problem it is particularly useful to be able to check the consistency of a set of predicates, through the construction of a corresponding instance. Thus, concretization is the process of constructing an instance of a theory in some "concrete" representation, and thereby checking the consistency of the theory. Probably the earliest application of diagrammatic reasoning, and a highly effective example of the use of concretization, was Gelernter's geometry theorem proving machine (Gelernter 1963). Gelernter's approach involved a backward search from a hypothesised theorem, using the definitions, postulates and theorems of geometry as the search space operators. Concretization was thereby exploited as a pruning heuristic, that is, paths that became implausible on the basis of an inability to construct a diagrammatic example were rejected.

Whilst concretization is a powerful mechanism for consistency checking in abstract mathematical domains such as geometry, it has been argued that for problems such as reasoning about natural language descriptions of spatial location, sentential representations alone are inappropriate, and must be exchanged for either a diagrammatic representation or a hybrid framework comprising both propositional and diagrammatic representations (Langacker 1988; Narayanan et al. 1994; Ioerger 1994). For example, consider the use of spatial prepositions, such as "in front", to specify the position of one object, the located object (LO), relative to a reference object (RO). There are a number of components to the meaning of even this simple spatial expression.

Firstly, "in front" falls into the class of projective prepositions, and thus the nature of the spatial constraint is both directional (i.e. "in front" implies some degree of alignment of the LO with a "front" direction) and proximity of the LO to the RO (Herskovits 1986). The direction of the constraint is always relative to one of a number of possible reference frames (Retz-Schmidt 1988). The intrinsic reference frame is centred on the reference object itself and is aligned with the intrinsic axes of the object. The deictic reference frame is centred at the speaker and extends from the speaker in the direction of the reference object. Other reference frames include the geocentric (also known as the extrinsic or allocentric) reference frame which is defined by an environmental feature or some object in the environment other than the RO or LO (Retz-Schmidt 1988). Even this rather complex picture of reference frames is a gross over-simplification, as most real spatial dialogue involves at least two speakers with different vantage points on a scene giving rise to yet more reference frames (Schober 1995; Levinson 1996). More relevantly to the concrete nature of diagrammatic representations, we can also observe that due to occlusion and other restrictions on each speaker's field of view different vantage points can give rise to the situation that for different speakers the scene can comprise different sets of visible objects.

To capture the richness of meaning of this class of language within a sentential representation would require the axiomatization of a mereology, topology, and geometry. This hardly seems a plausible strategy, and in this vein Latecki and Pribbenow (Latecki and Pribbenow 1992) pointed to a commentary by Davis on the use of general purpose representational devices for space: "[t]he very richness of geometric theory make it essentially hopeless to expect useful results from applying a general purpose geometric theorem prover to arbitrarily constructed sentences in a geometric language ... [i]t is generally necessary to restrict very tightly the kind of information allowed in a knowledge base and the kind of inferences to be made, and then to devise special purpose algorithms to perform these inferences" (Davis 1990, p. 246). In relation to spatial language, concretization is one means of overcoming this limitation, that is, the use of a representation that is analogous to the elements of the domain which it represents.

As regards the characterisation of the semantics of spatial prepositions, and locative expressions in general, purely sentential approaches have been attempted (Aurnague and Vieu 1993) but increasingly, approaches originating from (Waltz and Boggess 1979) are pursued, in which a procedural semantics of prepositions is grounded in an analogical model of the scene. As Figure 4 illustrates, such models may be discrete as in the case of Ludlow (1992), who interpreted localisation in terms of the assignment of the located object to a cell in a grid (see also Glasgow (1993), for a grid-based representation of space), or continuous, in which case the degree of spatial constraint is captured using an applicability function, and located object placements are achieved through constraint satisfaction (Yamada et al. 1993; Olivier, Maeda and Tsujii 1994). For a review of different cognitive and computational approaches to spatial expressions see (Olivier and Gapp 1998). Regardless of the degree of continuity of the diagrammatic representation, concretization exploits the fact that the ability to express constraints on properties such as proximity and occupancy are inherent to the representation. However, the penalty for concretization is specificity, and thus it is not possible to use a

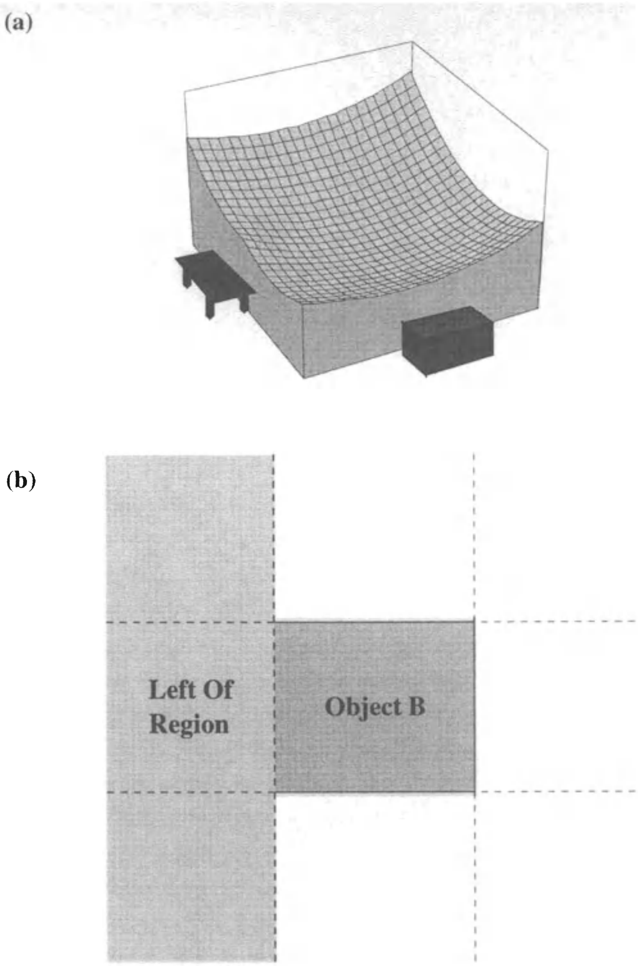


Figure 4. (a) Continuous (Olivier 1994) and (b) discrete (Ludlow 1992) concretizations of space used to capture the degree of spatial constraint implied by the use of projective prepositions.

diagram to express some predicate over a universally quantified variable, nor to capture the qualitative ambiguity of a spatial expression (i.e. capture the ambiguity in a single diagram).

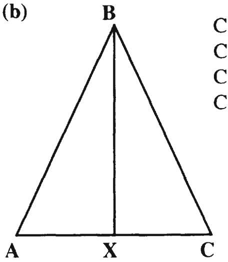
4. Inference by Inspection and Transformation

The concrete properties of diagrams are most apparent when considering the nature of inference using diagrammatic representations. We have already

(a)

POINT(A), POINT(B), POINT(C)
SEG(a, A, B), SEG(b, B, C), SEG(c, C, A)
TRIANGLE(A, B, C),
POINT(X), ONSEG(X, a),
SEG(x, C, X),
EQU(length(b), length(c)),
PERPEND(a, x)

(b)



COORD(A, 0, 0),
COORD(B, 2, 0),
COORD(C, 1, 2),
COORD(X, 1, 0)

Figure 5. Inference by inspection (adapted from Kulpa (1994)).

illustrated how a diagrammatic representation can be utilised where it is takes the form of an indexing of a sentential representation of a domain. However, inference itself may be effected by essentially visual processes alone, that is, by either inspection of the spatial elements or by transformation and inspection. One aspect of inference by inspection has already been introduced, that is, the notion of model checking. Where a diagram can be constructed that satisfies some set of universally quantified predicates, but not another, then the conjunction of the two may be deemed inconsistent, and is used to prune the search space in Gelernter's geometry theorem prover (Gelernter 1963).

A more "productive" view of inference by inspection is what Koedinger and Anderson term the "emergent properties" of diagrams (Koedinger and Anderson 1992). Contrast the sentential and diagrammatic geometric predicates given in Figure 5, adapted from (Kulpa 1994). Figure 5(a) shows the sentential description of the situation in which there are three points, A, B and C: there are three segments connecting the points; the points A, B and C form a triangle; there is a point X on the segment connecting A and B, and so on. Figure 5(b) shows an instance of a diagram that satisfies all of these sentential constraints. Koedinger identified that the advantage of the diagram in this case is that a number of properties, which would otherwise require significant sentential inference, can be simply read off the diagram. Of these the topological properties (e.g. that the mid-point of AX lies in the triangle ABC) are guaranteed to be true, whereas the geometric properties can be dependent on arbitrary metric choices in the construction of the diagram.

Inference by transformation is an extension of inference by inspection, by which the static properties of the diagram are not the only representational primitives, but phenomena in the problem domain, such as the application of a force, are mapped to transformations of the diagram. Funt's WHISPER system (Funt 1980) was an early example of this. WHISPER discretized space using a retinal array and objects which persisted in retinal array it could rotate and translate. This depictive component worked in cohort with a high level reasoner, which, using the emergent properties extracted from the retinal array (such as the symmetry and contacts between the objects contained within it), predicted the motion of the objects. The retinal array then implements the predicted motion as an incremental transformation of the position of the object. Through this cyclic interaction between the retinal array and the high level reasoner WHISPER could reason about the stability of stacks of blocks. Diagrammatic models similar in spirit have been developed for domains much more problematic than that of rigid body mechanics, such as strings (Gardin and Meltzer 1989), liquids (Decuyper, Keymeulen and Steels 1995) and structural analysis (Tessler, Iwasaki and Law 1995).

The success of systems that rely upon inference by transformation and inspection is often restricted by the ease with which transformation and inspection can be performed. Just as in the case of knowledge indexing, where the diagram can be used to focus the attention of the inference to the relevant sentential knowledge, it would be highly desirable if the application of the inspection operator for a diagrammatic representation only occurs at locations where it is likely to yield the feature of interest. Similarly, the computational cost of applying a transformation only to that part of an object or diagram where the transformation is likely to be significant, is to be preferred to the "blind" application of a transformation to a whole object or image. The focus of attention for inspection has been implemented in a number of diagrammatic reasoning systems, for example, Glasgow (Glasgow 1993) uses a hierarchy of arrays in the spatial component of her proposal for an architecture for computational imagery. Focused inspection and transformation are a feature of the KAP system (Olivier, Nakata and Ormsby 1995) which uses a hierarchical representation of mechanical components to reason about the behaviour of higher pairs (e.g. cam-follower pair or meshing gears). By rotating or translating the low resolution representation of the components, regions of the diagram where the component will not collide with other components can be identified and ignored in the next level of the spatial decomposition. In this manner only regions of the image where a collision is feasible are subject to inspection and transformation. Figure 6 depicts the region of a cam follower pair that falls within the scope of transformation and inspection at different levels in the hierarchical decomposition (to illustrate this the cam-follower is depicted at maximum resolution within this region).

So far we have only considered inference by transformation within the context of diagrams that are direct analogues of the physical world. This need not be the case, and there are situations where the problem itself may be inherently diagrammatic in that it is more appropriate to express the problem diagrammatically and subject it to diagrammatic transformations. In his BITPICT system, Furnas developed an approach in which entities are represented as elements of a raster image and are incrementally rewritten subject to

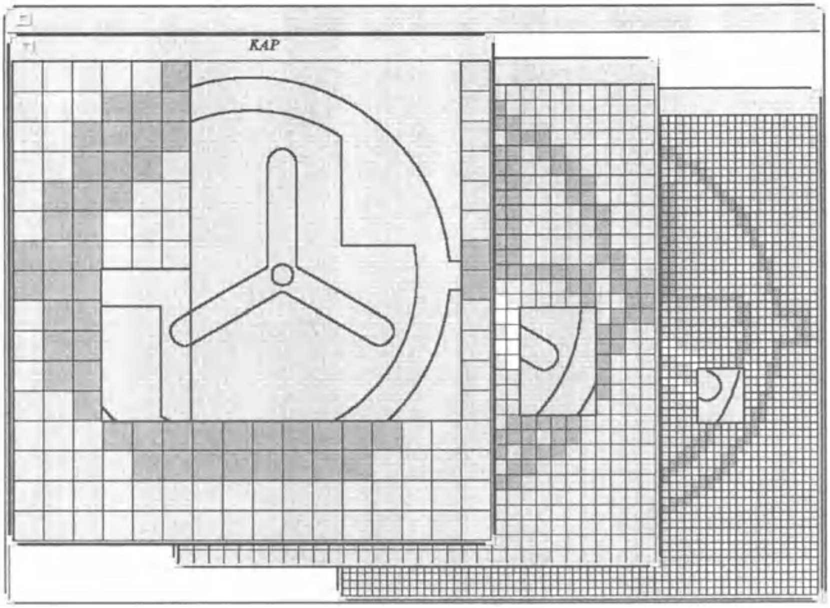


Figure 6. Focusing the scope of transformation and inspection in KAP. As the resolution of the representation increases the scope of application of the rotation transformation narrows as the regions of the image where collision between two objects might occur becomes more localised.

an ordered set of rewrite rules (Furnas 1992). The rules are specified simply as a mapping from one raster pattern to another. By selecting an appropriate set (and ordering) of rewrite rules, Furnas demonstrated how a number of interesting computations can be performed. For example, Figure 7, illustrates the initial rewrite rules for a "diagrams only" computation of the number of bifurcating trees in a tangled forest (the full set of rules rewrites the trees as a Roman Numeral count of the number of trees).

5. Closing Remarks

The preceding sections have mapped out some of the ways in which diagrammatic representations might be used in automated problem solving. However, this has been attempted without a thorough characterisation of either their representational or inferential adequacy. Indeed it would be far from true to claim that the case has been made for diagrams as a genuine representation choice in a problem solver. What might be easier to defend is that the inherently spatial nature of some classes of problem mean that a diagrammatic representation maintains the inherent spatial indexing of information which

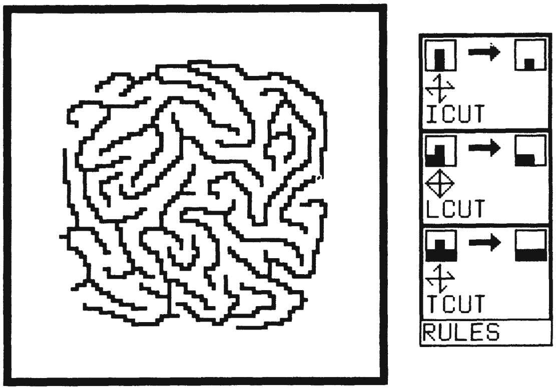


Figure 7. Counting trees through diagrammatic transformation. The three initial rules incrementally reduce each tree to a point, other rules then translate the points to the bottom of the image and the number of them rewritten in Roman Numerals.

these classes of problem can require to facilitate efficient reasoning. The fact that reasoning about physical systems is a recurrent problem domain for diagrammatic reasoning systems is therefore no great surprise, as much of our experience of physical causation is inherently local in its spatial character (gravity being a notable exception).

As regards less obviously spatial domains, diagrammatic representations are yet to prove their worth. Returning to the domain of kinship relations, topological connectivity is an analogue for the notion of inheritance, and through the appropriate layout conventions the metaphor affords inferential efficiency in the context of human perception. However, it is not so apparent what the candidates for the automated equivalent of ancestry inference (by inspection) would be. This is the nub of the problem in defining the "diagrammaticity" of a representation. The properties that we have reviewed so far, knowledge indexing, concretization and inference by inspection and transformation, are not necessarily properties of a representation that is essentially visuospatial (as we consider "diagrams" to be when we mean external representations to aid human problem solving). Indexing of knowledge need not be spatial, and more commonly is not. Similarly there is nothing inherently spatial about concretization, or inference by inspection and inference by transformation. To over emphasise this point might, however, be to incorrectly minimise the role of the human (i.e. the knowledge engineer) in automated problem solving. Indeed, current interest in diagrammatic representations is vigorously fuelled by their immediacy and accessibility, that is, their utility as a medium for encoding knowledge.

References

Anderson, M. and McCartney, R. Diagrammatic Reasoning and Cases. In *Proceedings of the 13th National Conference on Artificial Intelligence*. Portland, Oregon.

Aurnague, M. & Vieu, L. (1993). A Three-Level Approach to the Semantics of Space. In Zelinsky-Wibbelt, C. (ed.) *The Semantics of Prepositions: From Mental Processing to Natural Language Processing*, 393–440. Mouton de Gruyter: Berlin.

Davis, E. (1990). *Representations of Commonsense Knowledge*. Morgan Kaufmann Publishers: San Mateo, CA.

Decuyper, J., Keymeulen, D. & Steels, L. (1995). A Hybrid Architecture for Modeling Liquid Behavior. In Glasgow, J., Narayanan, N.H. & Chandrasekaran, B. (eds.) *Diagrammtic Reasoning: Cognitive and Computational Perspectives*, 731–751. AAAI Press: Menlo Park, CA.

Funt, B.F. (1980). Problem Solving with Diagrammatic Representations. *Artificial Intelligence* **13**: 201–230.

Furnas, G.W. (1992). Reasoning With Diagrams Only. In *Reasoning with Diagrammatic Representations: Papers from the 1992 Spring Symposium. Technical Report SS-92-02*, 115–120. AAAI Press: Menlo Park, CA.

Gardin, F. & Meltzer, B. (1989). Analogical representations for Naive Physics. *Artificial Intelligence* **38**: 139–159.

Gelernter, H. (1963). Realization of a Geometry Theorem Proving Machine. In Feigenbaum, E.A. & Feldman, J. (eds.) *Computers and Thought*, 134–152. McGraw-Hill: New York.

Glasgow, J. (1993). Imagery: Computational and Cognitive Perspectives. *Computational Intelligence* **9**(4): 309–333.

Gordon, I.E. (1989). *Theories of Visual Perception*. Wiley: London.

Herskovits, A. (1986). *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Cambridge University Press: New York.

Ioerger, T. (1994). The Manipulation of Images to Handle Indeterminacy in Spatial Reasoning. *Cognitive Science* **18**: 551–593.

Koedinger, K.R. & Anderson, J.R. (1992). Abstract Planning and Perceptual Chunks: Elements of Expertise in Geometry. *Cognitive Science* **14**(4): 511–550.

Kulpa, Z. (1994). Diagrammatic Representation and Reasoning. *Machine GRAPHICS & VISION* **3**(1/2): 77–103.

Langacker, R.W. (1988). A View of Linguistic Semantics. In Rudzka-Ostyn, B. (ed.) *Topics in Cognitive Linguistics*, 49–90, Benjamins: Amsterdam.

Larkin, J.H. & Simon, A.S. (1987). Why a Diagram Is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* **11**: 65–99.

Latecki, L. & Pribbenow, S. (1992). On Hybrid Reasoning for Processing Spatial Expression. In *Proceedings of the European Conference on Artificial Intelligence (ECAI-92)*, 389–393.

Levinson, S.C. (1996). Frames of Reference and Molyneux's Question: Crosslinguistic Evidence, In Bloom, P., Peterson, M.A., Nadel, L. & Garrett, M.F. (eds.) *Language and Space*. MIT Press: Cambridge, MA.

Ludlow, N.D. (1992). *Pictorial Representation of Text: Converting Text to Pictures*. Ph.D. Thesis, Department of Artificial Intelligence, University of Edinburgh, August, 1992.

McDougal, T. & Hammond, K. (1992). A Recognition Model of Geometry Theorem Proving. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 106–111. Lawrence Erlbaum Associates: Hillsdale, N.J.

Narayanan, A., Ford, L., Manuel, D., Tallis, D. & Yazdani, M. (1994). *Animating Language*. Workshop on Integrating Natural Language and Vision, 12th National Conference on Artificial Intelligence (AAAI-94), Seattle, USA.

Narayanan, N.H., Suwa, M. & Motoda, H. (1995). Diagram-Based Problem Solving: The Case of an Impossible Problem. *Proceedings of the 17th Annual Conference of the Cognitive Science Society, Lawrence Erlbaum Associates*, 206–211.

Olivier, P. & Gapp, K.-P. (eds.) (1998). *Representation and Processing of Spatial Expressions*. Lawrence Erlbaum Associates: Hillsdale, N.J.

Olivier, P., Nakata, K. & Ormsby, A.R.T. (1996). Occupancy Array-Based Kinematic Reasoning. *Engineering Applications of Artificial Intelligence* **9**(5): 541–549.

Olivier, P., Maeda, T. & Tsujii, J. (1994). Automatic Depiction of Spatial Descriptions. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-94)*, 1405–1410. July, Seattle, WA.

Retz-Schmidt, G. (1988). Various Views of Spatial Prepositions. *AI Magazine* (summer) **9**(2): 95–105.

Schober, M.F. (1995). Speakers, Addressees and Frames of Reference: Whose Effort is Minimised in Conversations About Locations? *Discourse Processes* **20**: 219–247.

Schwartz, D.L. & Black, J.B. (1996). Analog Imagery in Mental Model Reasoning – Depictive Models. *Cognitive Psychology* **30**(2): 154–219.

Tessler, S., Iwasaki, Y. & Law, K. (1995). Qualitative Structural Analysis Using Diagrammatic Reasoning. In *Proceedings of 14th International Joint Conference on Artificial Intelligence*, 885–891. Montreal, Canada.

Waltz, D.L. & Boggess, L. (1979). Visual Analog Representations for Natural Language Understanding. In *Proceedings of the 6th International Joint Conference on Artificial Intelligence (IJCAI-79)*, 926–934. Tokyo, Japan.

Yamada, A. (1993). *Studies on Spatial Description Understanding Based on Geometric Constraint Satisfaction*. Ph.D. Thesis, Department of Information Science, Kyoto University, January.

Artificial Intelligence Review **15:** 79–94, 2001.

Cognitive Science Approaches To Understanding Diagrammatic Representations

PETER C.-H. CHENG[1], RIC K. LOWE[2] and MIKE SCAIFE[3]
[1]*ESRC Centre for Research in Development, Instruction and Training, School of Psychology, University of Nottingham, Nottingham, NG7 2RD, UK;* [2]*Faculty of Education, Curtin University of Technology, GPO Box U1987, Perth, Western Australia, Australia 6001;* [3]*Interactive Media and Virtual Environment Lab, School of Cognitive and Computing Sciences, Sussex University, Brighton, BN1 9QH, UK.*

Abstract. Through a wide variety of approaches cognitive science has given us various important insights into the nature of diagrammatic representations. This paper surveys the findings, issues and approaches to diagrammatic representations in cognitive science. Important current issues that are highlighted include: the relation between the parts of the representational system that are internal to the mind and in external visual media that presents the diagram; the use of multiple representations which is typical of real contexts of diagram use; the benefits of diagrams in terms of (i) computational offloading, (ii) re-representation and (iii) graphical constraining.

Keywords: diagrammatic representations, cognitive science, external cognition, complex information processing

1. Introduction

Cognitive science is a diverse field encompassing many different perspectives for the investigation of a great variety of human cognitive phenomena. The study of reasoning, problem solving and thinking with diagrammatic representations (diagram use) is also diverse, ranging from work on the analysis of the characteristics of diagrams in themselves to studies of mental imagery. This paper reviews research on the nature of diagrammatic representations and what makes them effective, with a particular focus on the issues that are current in the area.

To set the scene consider three diagrams, which will be considered occasionally throughout the paper. We can all recognize that Figure 1 is a weather map for Australia. Most readers will know that the thick contour lines are isobars, or lines of equal pressure. Except for those trained in meteorology, none of us will be able to forecast the weather from the map nor say how the area of pressure will change over time, but such predictions can be made using

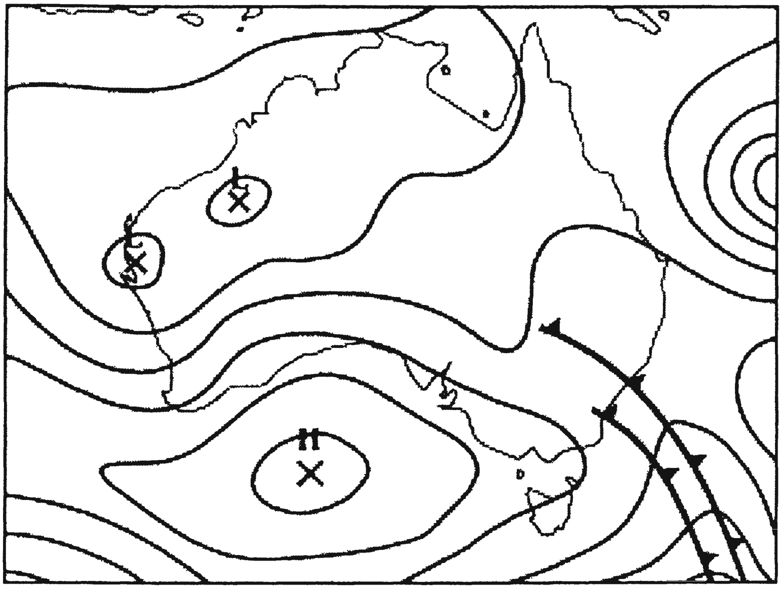


Figure 1. A weather map for Australia.

the map. There is clearly more going on here in cognitive terms than can be explained by simply saying that the map is a pictorial image and interpreted as such.

Similarly, consider a secondary school child trying to understand the path of carbon through the environment. Typically this is depicted in textbooks as a cyclical representation, involving text, pictures/schematics and a set of conventional notations such as arrows, lines or boxes (e.g. Figure 2). Children – and indeed adults – often have great problems in understanding this kind of representation at other than a superficial level, despite the inclusion of pictures (icons) and text (cf. weather map). Why is this the case and how can a cognitive science approach improve on the situation?

Figure 3 shows diagrams used by early physicists to discover the conservation of momentum and energy, in the context of head-on collisions of particles moving in a straight line. In each diagram, the labelled lines denote properties of the domain: the initial velocities (*U*), final velocities (*V*) and masses (*m*) of two bodies (subscripts *1* and *2*) for a single collision. The orientation and lengths of the *U* and *V* lines represent the direction and speed of the bodies, and the relative lengths of the *m* lines are in proportion to the masses of the bodies. There are obvious fundamental visual differences between these diagrams and the weather map and Carbon cycle diagram, such as their geometric character. But more interestingly, what can be said in cognitive

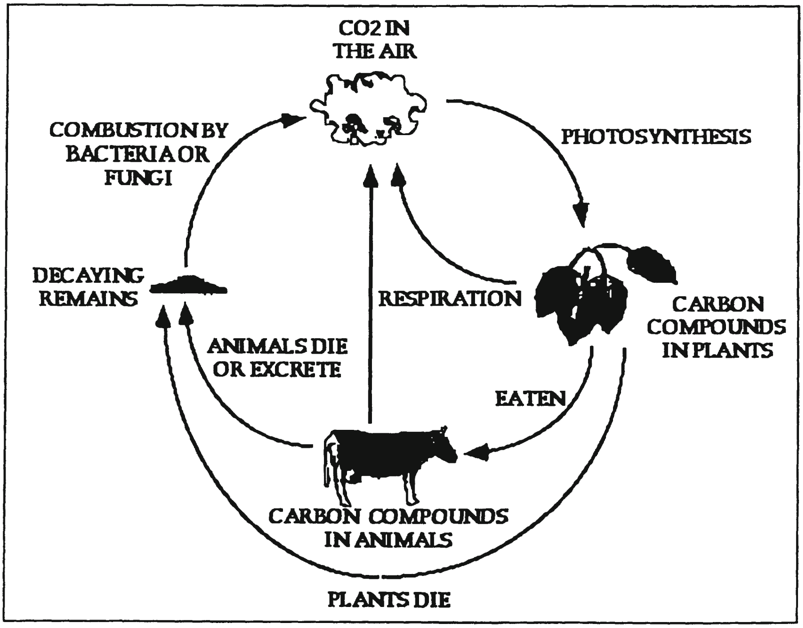


Figure 2. Diagram of the carbon cycle.

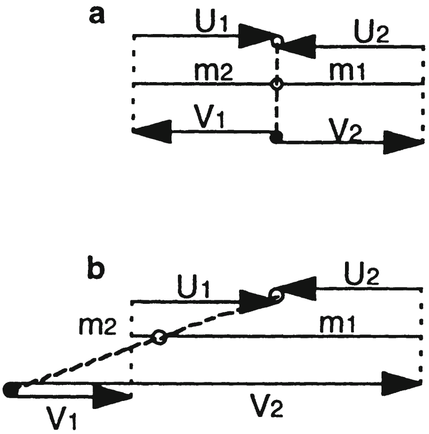


Figure 3. Diagrams for elastic collisions.

terms about their similarities that will allow us to understand how diagrams are used and what makes effective diagrams for learning and training?

In three main sections the paper considers (§3) the fundamental nature of diagrams, (§4) aspects of cognition with diagrams, and (§5) what makes diagrams effective. Within each section the issues are addressed from two different but complementary perspectives. The first perspective is a relatively general one, which considers diagrammatic representations in terms of high level characteristics. The second, at a more specific level, considers the nature of the complex information processing, CIP, that occurs with diagrammatic representations. A brief explanation of how CIP theory applies to diagrammatic representations is given in the next section, before we turn to the main sections.

2. Applying Complex Information Processing Theory To Diagrams

Although new perspectives have developed in cognitive science, the traditional approach that characterizes cognition as complex information processing in terms of physical symbols systems using heuristic search (exemplified by Newell and Simon 1972) is still as relevant and productive today as it has always been. Since the early days of the theory of complex information processing (CIP) systems, in which a narrow range of relatively simple puzzles and games were studied (Newell and Simon 1972), the scope of CIP theory has expanded to cover diverse phenomena, from low level perceptual-motor skill learning, through to design and invention, even collaborative scientific discovery, and of course the use of diagrams.

The heart of the CIP theory is the characterization of human cognition in terms of the procedures that process information in the form of expressions, assemblies of symbols. In diagrams the symbols are visual features, such as the shape, size, orientation of graphic objects. Consider the domain of elastic collisions between two bodies moving in a straight line. Each diagram in Figure 3 is an expression, an arrangement of visual symbols standing for properties of the referent objects involved in the collisions.

Typically, sets of permissible expressions, problem state spaces, are large and organised hierarchically. However, the human information processing system operates in a mainly serial fashion, thus for effective problem solving heuristics are used to guide the search through the state space. Characterizing the nature of the symbol expressions, operators and heuristics for particular tasks or domains is central to the understanding of problem solving. The information processing system searches for expressions, states of knowledge, that will achieve the goals of a given task or problem by using operators or rules to create, modify, reproduce and destroy expressions. For example, to transform the "default" collision diagram, Figure 3a, into a new diagram, Figure 3b, representing the collision between bodies of very different mass, requires operators that modify the lengths of the lines within given constraints. For example, the lengths of m_2 is shortened and m_1 lengthened, whilst keeping the total length of the two lines constant. Other geometrical operators are applied to change length or position of v_1 and v_2.

An important consequence of the CIP approach is the provision of a criterion on which to base judgements of the relative merits of different representations (e.g., diagrams vs. sentences). This is the notion of *information equivalence*, which recognises two representations as equivalent when all the information (content of expressions) in one is also inferable from the other, and vice versa (Palmer 1977; Larkin and Simon 1987). The application of this criterion permits the investigation of representations with different overall formats but that are at a fundamental level the same. Thus, it is legitimate to attribute any benefits of a representation to its cognitive or computation properties rather than merely because it contained more information in the first place.

Along with general perspective of high level aspects of diagrammatic representations, the CIP perspective will provide a basis for the consideration of the main issues of cognition with diagrams, in the next three sections.

3. Nature of Diagrams

One reason for the persistent interest in diagrams in diverse fields including computer science, education and psychology comes from the common intuition that diagrams are often more effective than propositional representations for whatever purpose they are put. The study of diagrams in cognitive science challenges this intuition and some of the basic distinctions that are held about diagrams.

Diagrams are (sometimes) better

Claims in the literature that diagrams are better, a priori, than other representations should be treated with caution. One finds in surveys of the research in computer science, psychology and education claims about the benefits of diagrams, and visual representations more generally, that seem to be motivated largely by intuitions and that are only weakly supported by rigorous empirical evaluations or any consistent attempt to derive generalisable theories.

For example, consider the case of information representation using multimedia, which seems to offer the possibility of improvement over conventional alternatives in displaying diagrammatic information. There is a wide-spread belief that representations rendered as, say, computer animations have distinct advantages over their paper-based equivalents. However, without any understanding of what makes external representations effective design will continue to be – as it is now – driven by slogans such as ‘a picture is worth 10000 words’, that ‘more is more’ or that the ‘sum is greater than the parts’ (e.g. Lopuck 1996). Such beliefs are, at best, unsupported as a general claim and often seem to rest on unproven and naive assumptions about the way that external visual representations 'produce' internal models (Scaife and Rogers 1996). Current orthodoxies about the intrinsic benefits of visualisation of information, which are grounded on the assumption that it makes the information more accessible, need to be examined far more critically.

It is noteworthy that the cognitive scientists Larkin and Simon (1987) included a qualification (in parentheses) in the title of their seminal paper, which will be considered below – "Why a diagram is (sometimes) worth ten thousand words". As will be seen below, a cognitive science perspective on the use of diagrams reveals a highly complex phenomenon with many facets, from which it is not possible to simply derive straightforward general claims about the benefits of diagrams.

Diagrams and propositional representations may not be so different

One reason why it is not possible to claim that diagrams are generally better than other representations is that they are not so different from other representations. A distinction is often made between diagrammatic and propositional representations, such as logic and mathematics. The main characteristics of diagrams is their use of space and spatial properties (location, topology, geometry, etc), but this is not exclusive to diagrams, because propositions also use "diagrammatic" properties to encode information, although to a lesser degree (e.g., in the formula '$x = y + z$', it matters whether the '$+z$' term is to the left or right of the equals sign). By the same token diagrams are not purely diagrammatic, because they contain propositions, as in the Carbon cycle diagram (Figure 2). One may consider all representations as falling at different places in a continuum from little use of diagrammatic properties to encode information through to substantial use of such properties. Thus, strong claims about the difference between diagrams and other representations should be treated with caution.

Considering the nature of the information that is being processed there is an *a priori* reason to treat diagrams as a distinct class of representational systems. Under the CIP approach in cognitive science, the assumption is that differences between representational systems can best be understood in terms of their respective symbol expressions, operators and heuristics. In this fashion, Larkin and Simon's (1987) paper compared diagrams with equivalent sentential representations, explaining that the cognitive benefits of many diagrams reside in the way that information, symbol expressions, are indexed by spatial location rather than by symbolic labels.

Diagrams are not a unitary class

Another reason why diagrams cannot be assumed to be generally better than some other class of representation is that diagrams are not a well-defined unitary class. There are many different types of diagrams, for example, just compare the map, flow/cycle diagram and the geometric diagram in this paper (Figures 1, 2, 3). The claims in the literature about diagrams in general, which are derived from studies that examined just a single type of diagram should be treated with caution.

Again, by considering the nature of the information that is being processed, the variety of diagrammatic representational systems are more clearly distinguished, because of the focus on symbol expressions, operators and heuristics. For example, compare the diagrams above. In Figure 3, lines for properties of bodies (symbols) are related by the geometric structure of the diagram (expressions), and a diagram may be modified using geometric rules (operators). In Figure 2, icons stand for the location of CO_2 in different entities and the labelled lines represent processes (symbols), and the combination of the arrows between two icons shows the transfer of CO_2 (expression). The diagram might be modified to include fossilization and the burning of hydrocarbon fuels, by adding more icons and connecting them together with appropriately labeled arrows (operators). Similar, analysis of Figure 1 is left for the reader. Such comparisons demonstrate the huge variety of diagram types, arguably more diverse than propositional representations.

4. Aspects of Cognition With Diagrams

Cognitive science has taken a number of perspectives when studying the use of diagrams, which focus on different aspects of the relation between the diagram user and the nature of the diagram itself.

At a fundamental level, a diagrammatic display can be regarded as an arrangement of various graphic elements in space. Perceptual similarities and differences between these elements allow them to be grouped or distinguished according to visuospatial characteristics of the particular display. For example, in the weather map diagram shown in Figure 1, we could group together the series of concentric curves in the northeast because of their similarity in shape, graphic treatment and location. These are readily distinguished from the bold lines bearing triangular barbs in the southeast of the diagram. Being able to configure diagram elements into groups or discriminate between them in this way is an important precondition for proper interpretation and is a key requirement of a well-designed diagram.

An explanation for why this is so comes from the consideration of the nature of information processing involved. Larkin and Simon (1987) suggested that it is the use of locational indexing in diagrams that often makes them more effective than informationally equivalent sentential representations. This form of indexing means that information that tends to be needed for the same inference can usually be found in adjacent locations in a diagram, so reducing the amount of search required to find the information. Further, perceptual inferences with diagrams allow the power of the highly parallel human visual system to replace more cumbersome serial logical inferences. Such inference do not pose a fundamental problem for diagrams under the CIP approach, as they may be treated as operators. Seeing that the lengths of the lines in Figure 3a are equal is like testing for the equality of the values assigned variables in a mathematical representation. Studies of geometry problem solving (Koedinger and Anderson 1990) and reasoning with electrical circuit diagrams (Egan and Schwartz 1979), for example, illustrate how perceptual inferences can be dealt with under the CIP approach in cognitive science.

The raw perceptual information that a diagram provides for the user by way of visuospatial cues must be modulated by knowledge about the individual and collective meanings of the graphic elements. An over-reliance on the visuospatial characteristics of the markings making up a diagram can be highly misleading. This can be demonstrated by two other sets of elements shown in Figure 1. The sets of roughly concentric markings in the southwest and southeast corners of the diagram respectively are widely separated and so, in purely perceptual terms, appear to be quite distinct. When beginning students of meteorology (novices) were asked to group the elements on this diagram, they typically distinguished between the sets of markings in these two corners of the diagram (Lowe 1993). In contrast, professional meteorologists (experts) configured these two sets as a single grouping. It seems that a major factor determining a viewer's capacity to make effective use of a diagram is how much that person already knows about the sort of subject matter depicted in the diagram and the specific method of depiction.

Using a given diagram effectively requires the viewer to think about that diagram in quite particular ways. Because diagrams are highly specialised depictions that differ substantially from more realistic pictures, the cognitive approaches that we habitually use for interpreting our everyday visual environment are inappropriate. The skills required for using diagrams effectively must be learned and appear to be highly domain-specific. There are some generic aspects that influence diagram use (such as the need to treat these as abstract rather than literal depictions).

An implication that can be drawn from this research is that diagrams, in and of themselves, do not 'contain' all the information that a viewer needs to use them properly. Rather, the background knowledge that the viewer brings to the diagram plays a critical role in whether or not it can be processed satisfactorily. This would mean that good diagram design can only go so far in determining whether a diagram is likely to be an effective way for depicting particular information. For this reason, current orthodoxies about the intrinsic benefits of visualization of information (on the assumption that it makes the information more accessible) need to be examined far more critically.

The discussion of Figure 1 has so far focused upon it as a representation of a particular state and shown the importance of domain-specific knowledge in effective diagram use. However, diagrams are frequently used in more sophisticated ways that involve mental processes such as inference and prediction. For example, a weather map diagram for a particular day can be used to make a prediction about the weather pattern that is expected on the following day. In this case the diagram is the basis for generating new information rather than simply depicting the present situation. Similarly, by modifying Figure 3a for symmetrical collision between bodies, one may, for instance, explore possible asymmetrical configurations, such as Figure 3b. The cognitive processes involved in this type of task require the creation of a suitable mental model that can be 'run' to make predictions or inferences, in the case of the weather map this may even be backwards as well as forwards in time.

Similarly, multimedia design provides a strong challenge for any general theory of external representations but also emphasises these issues which are central to understanding cognition with diagrams. Firstly there is the possibility that being able to *interact* with multimedia representations in ways not possible with single media (i.e. books, audio, video), can lead to easier learning, better understanding, and increased motivation. This is certainly the case but leads to the question of how users interact with any kind of diagrammatic representation. This is often overlooked in studies of paper-based diagrams but is surely important. For example marking the paper or making other annotations are a central feature of geometry student progress (e.g. Koedinger and Anderson 1990). Thus we need to recognise the important role of *constructing* external representations (Reisberg 1987), which is normally such an integral part of learning or problem-solving, e.g. underlining, making notes separately, re-representing text-based ideas in various diagrammatic forms, sketching etc.

Multimedia also affords novel access to *multiple* representations of information. An example is multimedia encyclopaedias which have been designed on this principle, providing a variety of audio and visual materials on any given topic. However, the issue of the benefits of multiple representations is also present for paper-based products – consider even the embedding of pictures within text. To produce effective designs it is necessary to understand how learners integrate information arising from different representations of the same and different information. This requires analysing how people learn to *read* and comprehend the significance of the content the diagram, for example how they develop an understanding of canonical diagram forms, and how this is assimilated to their current understanding of the domain.

These kinds of issues underpin the need for a more general account of diagrams, *qua* external representations, than the case-based approach which has dominated the research literature. There have been a number of approaches, albeit different, that seem highly promising in this regard. One is the work of Stenning and colleagues on paper- and computer-based representations (e.g. Stenning and Oberlander 1995; Stenning and Lemon 2001; Stenning 1999) who argue for the need to distinguish between 'expressive' and 'processing' explanations for the cognitive usability of diagrams. The former has to do with semantic constraints on the space of diagrammatic interpretations, the latter to do with perceptual and/or mnemonic limitations due to the way that the particular diagram (or other representational form) is constructed. Another approach is that of Zhang and colleagues (e.g. Zhang and Norman 1994; Zhang 1997) who emphasise the mappings between rules and the structure of the problem space, both internal and external. Finally there is the work of Green and colleagues (e.g. Green and Petre 1996) who stress the value of high-level abstractions to convey important characteristics of external representations, such as the complex interactions between parts of the representational system that lead to 'viscosity' – a resistance of any part of the representation to local change.

Scaife and Rogers have an approach they label 'external cognition', that focuses on how different representations are processed when performing different activities (Scaife and Rogers 1996). The emphasis here is on the interactions between internal and external representations considered together (cf. Larkin 1989; Norman 1993; Vera and Simon 1993). Their belief is that the process by which different external representations are used in learning or problem-solving is complex, involving an interaction between internal processes and different aspects of external representations at different stages of a task. For example, reading and abstracting knowledge from a diagram requires making connections between different elements of the display in a temporal sequence. Such a 'take' may be contrasted with accounts that either emphasise the primacy of internal representations and/or ignore the way they are co-ordinated with external ones.

At the information processing level the same set of issues is cast in terms of the relation between the aspects of the representation that are internal to the mind and those that are in the external environment. For example, Figure 3 is an external physical notation but the geometric rules to manipulate the diagrams are usually held in the user's memory. All but the most trivial problems require iterative cycles of (i) visual interpretation of the external diagrams, (ii) internal recognition of applicable operators, (iii) modification of the drawing, and (iv) further visual interpretation of the new diagrams. Larkin (1989) and Zhang and Norman (1986) consider how the distribution of representations between the mind and external environment may reduce working memory loads and lessen cognitive demands. Tabachneck and Simon (1994) present a model of how internal images, in the "mind's eye", could be processed.

Under the CIP view there is a recognition that both the users and uses of diagrams should be considered. Users of a notation in a particular domain who are, for example, more expert will engage different operators and heuristics. In effect they possess quite a different representational system to novices in the same domain, even though they may share a common external notation. In the same vein, different tasks have distinct goals, which will be satisfied by different information. The search for goal expressions may require alternate operators and heuristics to process the notation, and may even be considered to constitute different representations, in some cases. Cheng (1996) discusses some of the variety of tasks or *functional roles* that diagrammatic representations may support.

If so much of the capacity to use a diagram effectively is bound up with what the viewer already knows about the subject matter, what options are available for improving diagram use? This question should be of particular interest to educators who provide novices in a domain with diagrams on the assumption that they will make the subject matter more accessible. The problem seems to be one of 'boot-strapping'; without a certain minimum knowledge of the domain, an individual is unlikely to be able to use a domain-specific diagram effectively. One approach for addressing this issue is to help novices develop the sorts of basic knowledge structures that could support appropriate cognitive processes. This type of approach has been explored recently by providing meteorological novices with computer-based animations designed to act as external models that could help them to build mental models of weather map systems that are more consistent with those used by experts in the field (Lowe 1997).

As mentioned with respect to Figure 1, when beginning students of meteorology (novices) were asked to group the elements on this diagram, they typically distinguished between the sets of markings in these two corners of the diagram (Lowe 1993). In contrast, professional meteorologists (experts) configured these two sets as a single grouping. Further investigations indicated that the experts' knowledge of the wider context of Australian weather systems allowed them to relate these sets of markings meteorologically as two sections of a much larger-scale feature that connected them beyond the scope of the diagram (Lowe 1994). In contrast, the novices were unable to invoke this type of domain-specific knowledge and so appeared to be reliant solely on visuospatial information (Lowe 1996).

Comparisons of meteorological experts and novices suggest that the superior quality of experts' predictions of weather map patterns is related to particular characteristics of the mental model they construct from a given weather map (Lowe, in press). Not only do they appear to be able to construct more extended and detailed mental models of the depicted situation, they also rely on a rich store of knowledge about the properties and behaviour of the various meteorological features. Once again, a key factor in using a diagram effectively is what the viewer brings to the diagram (rather than what the diagram brings to the user).

This is consistent with Koedinger and Anderson's (1990) work on the differences between novice and expert geometry problem solvers. They discovered that the problem solvers search a space of perceptual chunks comprising meaningful diagrammatic configurations, so performed better than novices who deal with the visual elements of the same diagrams in a piecewise fashion.

5. Properties of Effective Diagrams

In principle, at a basic level, it is obvious that a well-designed diagram should allow the user to make a relatively straightforward mapping between the diagrammatic depiction and the situation it represents. This means that it should be a simple matter to compare each component in the represented situation with its corresponding component in the diagram ("This is Australia"). It should also be easy to compare the corresponding arrangements of these components between situation and diagram ("These are concentric isobars over Australia").

However, given the richness of the nature of cognition with diagrams discussed in the previous section, considerations of what makes a diagram more or less useful must take a broader view. Clearly the effective properties will vary with the particular diagram and situation of use but Rogers and Scaife (1999) identify at least the following kinds of 'computational offloading' – the ways in which different external representations reduce the amount of cognitive effort required to solve informationally-equivalent problems (e.g. Larkin and Simon 1987): (i) Re-representation – This refers to how different external representations, that have the same abstract structure, make problem-solving easier or more difficult and how they are selected (e.g. Zhang and Norman 1994); (ii) Graphical constraining – This refers to the way elements in a graphical representation are able to limit the range of inferences that can be made about the represented concept (e.g. Stenning and Oberlander 1995); (iii) Temporal and spatial constraining – This refers to the way different representations can make relevant aspects of processes and events more salient when distributed over time and space (e.g. the use of canonical cyclical diagrams, as in the carbon cycle of Figure 2).

Clearly other properties could be identified as the advantages of diagrammatic over other kinds of representation but a major task remains that of understanding how these properties are actually realised, which will necessitate a better understanding of the mechanisms relating internal and external representations.

At the information processing level, the same issues can be addressed, but it is first necessary to consider what the appropriate bases are for the comparison of different representational systems. For representations that are informationally equivalent (see above), Larkin and Simon (1987) demonstrated that diagrams often have computational advantages over sentential representations, as already noted. However, if diagrams that are not informationally equivalent are to be studied, the information processing approach provides other bases for comparisons of representations. At a low level, comparison can be made between the form, number and complexity of the symbols, expressions and operators of different representational systems. At a higher level, comparisons can be made in terms of the overall size and/or complexity (e.g., breadth and depth) of the problem state space for the representations. For example, Cheng and Simon (1992) showed that in the inductive discovery of the law of momentum conservation the overall space of expressions is smaller for a diagrammatic representation (similar to Figure 3) than it is for an algebraic notation.

6. Conclusion

This review has examined cognitive science approaches to understanding diagrammatic representations. Below the surface of common but somewhat naive claims about the benefits of diagram over other representations lie various complex cognitive issues that inform us about the nature of human understanding, problem solving and thinking more generally. Diagrams are sometimes, perhaps often, better than other representations, but the reasons why are complex. To conclude, a summary of the main issues covered at various points throughout the paper is presented.

1) Claims in the literature that diagrams are better, a priori, than other representations with respect to presenting information should be treated with caution.
2) Diagrams are not a unitary class of representations but (i) are similar in some important respects to propositional representations and (ii) come in a wide variety of forms which may have quite different implications for cognition.
3) Properties that make diagrams effective are shared by many other representations.
4) There are diverse uses for diagrams which may have quite different implications for cognition.
5) Diagrams are hardly ever found in isolation, so the way that multiple representations are simultaneously used for reasoning and learning is an important issue.
6) The study of diagram use should examine the cognitive processes involved in diagram interpretation and understanding and not just the perceptual properties of graphic displays.
7) Similarly, the interactive processes of diagram construction and modification should be considered in addition to the interpretation of diagrams.
8) There are internal and external aspects of diagram use that need to be explained, including the role of background knowledge and the role of diagrammatic conventions – learning to recognise canonical forms.
9) The contrast between expert and novice users of diagrams is an effective way to learn about what makes diagrams effective or not.
10) Some of the properties that can (sometimes) make diagrams particularly effective representations have been identified in terms of their effectiveness in promoting computational offloading, for example:
    i) The locational indexing of information.
    ii) Re-representation by selection of more powerful operators or redistribution of the internal/external distribution of the elements of the representations.
    iii) Graphical constraining in limiting the size and complexity of the search space.
    iv) Temporal and spatial constraining making processes and events more salient when distributed over time and space.

References

Cheng, P.C.-H. (1996). Functional Roles for the Cognitive Analysis of Diagrams in Problem Solving. In Cottrell, G.W. (ed.) *Proceeding of the Eighteenth Annual Conference of the Cognitive Science Society*, 207–212. Lawrence Erlbaum: Hillsdale, NJ.

Cheng, P.C.-H. & Simon, H.A. (1995). Scientific Discovery and Creative Reasoning with Diagrams. In Smith, S., Ward, T. & Finke, R. (eds.) *The Creative Cognition Approach*, 205–228. MIT Press: Cambridge, MA.

Egan, D.E. & Schwartz, B.J. (1979). Chunking in Recall of Symbolic Drawings. *Memory and Cognition* **7**(2): 149–158.

Green, T.G.R. & Petre, M. (1996). Usability Analysis Of Visual Programming Environments – A Cognitive Dimensions Framework. *Journal of Visual Languages And Computing* **7**: 131–174.

Koedinger, K.R. & Anderson, J.R. (1990). Abstract Planning and Perceptual Chunks: Elements of Expertise in Geometry. *Cognitive Science* **14**: 511–550.

Larkin, J.H. (1989). Display-Based Problem Solving. In Klahr. D. & Kotovsky, K. (eds.) *Complex Information Processing: The Impact of Herbert A. Simon* 319–341. Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Larkin, J.H. & Simon, H.A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* **11**(1): 65–100.

Lopuck, L. (1996). *Designing Multimedia*. Peachpit Press: Berkley, CA.

Lowe, R.K. (1993). Constructing a Mental Representation from an Abstract Technical Diagram. *Learning and Instruction* **3**(3): 157–179.

Lowe, R.K. (1994). Selectivity in Diagrams: Reading Beyond the Lines. *Educational Psychology* **14**: 467–491.

Lowe, R.K. (1996). Background Knowledge and the Construction of a Situational Representation from a Diagram. *European Journal of Psychology of Education* **11**: 377–397.

Lowe, R.K. (1997, August). Effects of Interactive Animation in Complex Visual Learning. *Paper Presented at the Seventh European Conference for Research on Learning and Instruction*. Athens, Greece.

Lowe, R.K. (in press). Domain-Specific Constraints on Conceptual Change in Knowledge Acquisition from Diagram. In Vosniadou, S., Carretero, P. & Schnotz, W. (eds.) *New Perspectives on Conceptual Change*. Erlbaum.

Lopuck, L. (1996). *Designing Multimedia*. Peachpit Press: Berkley, CA.

Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Prentice-Hall: Englewood Cliffs, NJ.

Norman, D. (1993). Cognition in the Head and in the World. *Cognitive Science* **17**(1): 1–6.

Palmer, S.E. (1977). Hierarchical Structure in Perceptual Representation. *Cognitive Psychology* **9**: 441–474.

Reisberg, D. (1987). External Representations and the Advantages of Externalizing One's Thoughts. *Proceedings of the IX Annual Conference of the Cognitive Science Society*. Seattle, WA/Hillsdale, NJ: LEA.

Rogers, Y. & Scaife, M. (1999). How can Interactive Multimedia Facilitate Learning? *Proceedings of First International Workshop on Intelligence and Multimodality in Multimedia Interfaces*. Menlo Park, Ca: AAAI Press (in press).

Scaife, M. & Rogers, Y. (1996). External Cognition: how Do Graphical Representations Work? *International Journal of Human-Computer Studies* **45**: 185–213.

Stenning, K. (1999). Distinguishing Semantic from Processing Explanations of the Usability of Representations: Applying Expressiveness Analysis to Animation. *Proceedings of*

First International Workshop on Intelligence And Multimodality in Multimedia Interfaces. Menlo Park, CA: AAAI Press (in press).

Stenning, K. & Lemon, O. (1997). Diagrams and Human Reasoning: Aligning Logical and Psychological Perspectives. Position paper for 'Thinking With Diagrams II' Conference, University of Portsmouth, January.

Stenning, K. & Oberlander, J. (1995). A Cognitive Theory of Graphical and Linguistic Reasoning: Logic and Implementation, *Cognitive Science* **19**: 97–140.

Tabachneck, H.J.M., Leonardo, A.M. & Simon, H.A. (1994). How Does an Expert Use a Graph? A Model of Visual and Verbal Inferencing in Economics. In Ram, A. & Eiselt, K. (eds), *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, 842–847. Hillsdale, NJ: Lawrence Erlbaum Associates.

Vera, A.H. & Simon, H.A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science* **17**(1): 7–48.

Zhang, J. (1997). The Nature of External Representations in Problem Solving. *Cognitive Science* **21**(2): 179–217.

Zhang, J. & Norman, D. A. (1994). Representations in Distributed Cognition Tasks. *Cognitive Science* **18**(1): 87–122.

Artificial Intelligence Review **15:** 95–114, 2001.

Cognitive Factors in Programming with Diagrams

ALAN F. BLACKWELL[1], KIRSTEN N. WHITLEY[2], JUDITH GOOD[3] and MARIAN PETRE[4]
[1]*Computer Laboratory, University of Cambridge. Cambridge, U.K. (E-mail: alan.blackwell@cl.cam.ac.uk);* [2]*Department of Computer Science, Vanderbilt University, Nashville, Tennessee, U.S.A. (E-mail: whitley@computer.org);* [3]*Human Communication Research Centre, University of Edinburgh (E-mail: judithg@cogsci.ed.ac.uk);* [4] *Centre for Informatics Education Research, Faculty of Mathematics and Computing, Open University, Milton Keynes, U.K. (E-mail: m.petre@open.ac.uk)*

Abstract. Visual programming languages aim to broaden the use of diagrams within the software industry, to the extent that they are integrated into the programming language itself. As a result, they provide an ideal opportunity to study the benefits of diagrams as an external representation during problem solving: not only is programming a challenging problem-solving activity, but the effect of diagram usage can be directly assessed by comparing performance while using a visual programming language to performance with a standard textual language. There have been several misconceptions amongst visual language researchers regarding the role of diagrams in software design, but these are being addressed by empirical studies and by new theories of notation design derived from studies of visual programming. Based on this research, the authors are able to recommend several new directions for research into thinking with diagrams.

Keywords: diagrams, diagrammatic reasoning, visual programming, psychology of programming

1. Introduction

This paper investigates a specific class of diagrammatic notations – Visual Programming Languages (VPLs) – considering them from the perspective of research into psychology of programming. This field has developed a set of research methods specifically aimed at measuring human performance in solving problems with various external notations. As a result, the psychology of programming work dealing with visual programming should offer important insights into the application of diagrams in problem solving contexts.

This paper starts with an overview of VPLs – the reasons for their introduction, and the role that they play within software development. It then gives a brief introduction to research methods that are used in psychology of programming. We review the empirical evidence for benefits resulting from diagram use in programming, then describe new insights into the general properties of notations that have been derived from this empirical work. Finally we propose some of the most pressing topics for ongoing research into diagrammatic properties of VPLs.

2. The Nature of Visual Programming

Computer programming is a challenging intellectual task, involving complexity equivalent to other design and engineering activities. In the term introduced by Reitman (1964) it belongs to the class of "ill-structured" problems that cannot be solved by a strictly defined procedure.

At the same time, programming has two characteristics that make it apparently very appropriate as a research context for studying diagrammatic reasoning. Firstly, many programming activities take place within a technically constrained environment. This makes it possible to observe, record and measure many (but not all) of the actions taken by a programmer. Secondly, diagrams are often associated with computer programming, whether as an external representation employed as a private aid to thought, or as a communication medium between members of a software project team.

Like other engineering design professionals (see Ferguson (1992) for a review), programmers commonly create informal and ephemeral diagrams, both privately and as "talking sketches". As with other engineering disciplines, many of these informal conventions have been formalized in a way that makes them more persistent and suitable for maintenance as a form of design documentation. In programming, furthermore, those that have been formalized are readily amenable to automated translation and immediate testing.

Unlike most diagrams used in other engineering disciplines, software diagrams are not constrained by the need for graphical correspondence to the physical shape of any designed artifact. This has resulted in relatively greater freedom regarding the form of the diagram elements. Also unlike the representations used in other design disciplines, new software diagrams are continually being developed, so it is possible to observe the effect of successive generations of diagram conventions, or even influence the development of future generations of design notation.

VPLs have evolved from diagrammatic notations such as the relatively familiar flowchart. Many of these notations prescribed the behaviour of a program to the same level of detail that a textual programming language would do. When design diagrams started to be drawn directly on the screen of the computer, they could then be translated automatically into equivalent lines of program text. The diagram then becomes part of the program.

There have been at least a hundred proposals for VPLs. Glinert (1990a, 1990b) has edited a two volume collection of landmark papers in this field, dating from 1975. More recent research developments are reported in the proceedings of the now annual IEEE symposia on visual languages, and in the *Journal of Visual Languages and Computing*. Taxonomies of VPLs (as well as of visualization systems) have been proposed by Myers (1990) and by Price, Baecker and Small (1993).

The boundaries dividing VPLs from other classes of diagram are not always clear. We take the view that a programming language must be *executable* – it must have precise semantics defined in terms of expected changes in state of a digital computer. In this paper we do not discuss products such as Microsoft Visual Basic and Visual C++, where the logical behaviour of the program is fully expressed in the form of a textual programming language rather than by diagrams.

3. Psychology of Programming

Research into the cognitive processes involved in computer programming is derived from general cognitive theories of problem solving after Newell and Simon (1972). Some notable early cognitive theories of programming include those of Weinberg (1971) and Shneiderman (1980). Many of the aspects of these models will be familiar to cognitive scientists. They include separate processing of syntactic and semantic information, the collection of expert knowledge into chunks, the structuring of regularly-used information into schemas, and the solution of design problems in terms of previously acquired and frequently modified plans.

An overview of the range of research methods currently used in the psychology of programming can be found in the collection edited by Hoc, Green, Samurçay and Gilmore (1990). This volume also includes a chapter by Gilmore specifically reviewing and advising on empirical research methods that are suited to investigations of programming.

The predominant research technique in psychology of programming adopts the hypothesis testing methods of experimental psychology. Studies generally involve observation of small groups of subjects performing the same constrained, well-defined tasks. Performance might be measured in terms of the time required to complete a simple task, accuracy of response, or classification of statements found in a verbal transcript from the experimental subject. Individual differences between the subjects are not analysed in detail; instead the experimenter compares the differences between groups of subjects, looking for statistically significant variations in performance that were expected to result from the treatment assigned to each group.

Partly as a result of this, psychology of programming, like experimental psychology, has been accused of not addressing the concerns of "real" programmers who deal with messy, ill-defined problems where the main challenges are those of collaboration with customers and colleagues over long periods of time. The focus on statistical analysis also obscures individual differences in strategies used for specific tasks. This is of particular concern where those strategic differences may have far more effect on performance than the experimental hypotheses that are being tested.

There are alternatives to the pure hypothesis-testing approach, including empirical studies that aim to gather qualitative data about the processes involved in programming, as well as longitudinal field studies that chart the course of learning a programming language or observations of larger-scale software development projects over time.

4. Empirical Studies of Diagrammatic Notations Used in Programming

Relatively few empirical studies have focused directly on the use of diagrammatic notations in programming, although recent activity within the visual programming community may reflect growing interest in undertaking more such studies. This section summarizes existing studies, both to show the range of questions investigated and to show common themes in the findings. This summary includes studies on flowcharts, as well as research on VPLs. Interested readers can consult (Whitley 1997a) for a more detailed and extended discussion.

4.1. *Flowcharts*

More studies have focused on flowcharts than on other visual programming constructs. To date, these suggest that flowcharts can outperform text for certain tasks but not for the entire programming process. Scanlan (1989) investigated the comprehensibility of structured flowcharts and textual "pseudo-code" as representations for conditional logic. He asked programming students to view conditional logic and then answer questions about the states required to trigger given actions. Flowcharts had a significant advantage for the time needed to comprehend an algorithm, but also in other variables such as response accuracy. These effects were observed in simple cases as well as in complex ones. Scanlan's results suggest that flowcharts can have a beneficial effect for certain tasks; in particular, they illuminate the control-flow of conditional logic. Similar results have occurred in studies of programming students by Vessey and Weber (1986) and by Cunniff and Taylor (1987).

In contrast, no studies have shown flowcharts having a practical advantage over text across the full range of programming activities. Curtis et al. (1989) studied the flowchart in its historical role as a notation supplemental to textual code (i.e., as a form of design and/or program documentation). They identified two dimensions (symbology and spatial arrangement) capable of categorizing a wide range of notations. The symbology dimension measures the succinctness of a notation and includes three possibilities: unconstrained text (natural language), constrained text and ideograms. The spatial arrangement dimension captures the extent to which a notation's layout highlights the execution paths (in the case of flowcharts, the control flow) of a program; this dimension also has three values: sequential, branching and hierarchical. Combining all possibilities yields nine documentation formats, which were tested in comprehension, coding, debugging and modification tasks. Participants in these experiments were professional Fortran programmers.

Consistent effects of symbology were found in comprehension, coding and debugging. Prose was the most ineffective symbology format for most tasks, whereas constrained language and ideograms were almost equivalent. As for spatial arrangement, only small effects were observed; these occurred in situations in which control flow information was a factor in the task. These results are largely due to the branching arrangement, which improved performance in some tasks where control flow was important. Putting symbology and spatial arrangement together, the constrained/sequential representation typically outperformed the other forms of documentation. In sum, Curtis et al.'s results are consistent with Scanlan's in that the branching arrangement did help in tasks that emphasized control flow; but overall the constrained/sequential representation was equal or better for most tasks.

4.2. *Studies of VPLs in current use*

There are few empirical studies of VPLs that are in current use. A study by Pandey and Burnett (1993) stands out as the strongest controlled study favorable to VPLs. This tested performance on matrix problems using two textual programming languages (Pascal and a modified form of APL) and a diagrammatic subset of the research VPL *Forms/3*. 73% of Forms/3 solutions were completely correct compared to 53% of APL solutions and 40% of Pascal solutions. Their results apply to their language and programming environment as a whole; further study would be required to pinpoint the impact of their particular visual representation.

Aside from this study, the majority of the empirical studies have focused on the few commercially-available VPLs, most notably Pictorius *Prograph*, National Instruments *LabVIEW*, and Hewlett Packard *VEE*. Prograph is the only one of these that is promoted as a general purpose programming language – LabVIEW and VEE are frequently described as measurement and control languages, and characterized as accessible to scientists and engineers having limited programming experience. The visual syntax of LabVIEW and VEE expresses data flow in a way that resembles electronic circuit diagrams (this is particularly explicit in LabVIEW).

Baroth and Hartsough (1995) report on their experience using both LabVIEW and VEE. One case study compared two teams developing the same system in parallel. One team used the textual language C, while the other worked in LabVIEW. At the end of the three-month project, the LabVIEW team had made far more progress. From this study and more than 40 other projects, Baroth and Hartsough enthusiastically report performance benefits for both LabVIEW and VEE. They attribute the productivity gains to increased communication between the customer and developer, which arises, they say, from the visual syntax of the VPLs. Their customers were engineers and scientists comfortable with circuit diagrams, so the circuit-like syntax helped them understand the program.

In contrast, the one controlled laboratory study of LabVIEW seemingly contradicts Baroth and Hartsough's speculations about visual syntax. Green, Petre and Bellamy pitted LabVIEW's two forms of visual conditional logic against two textual notations (Green, Petre and Bellamy 1991; Green and Petre 1992). They started from Gilmore and Green's *match-mismatch* hypothesis, which states that problem-solving performance depends on whether the structure of a problem is matched by the structure of a notation (Green 1977). This was first established using two textual forms of conditional logic: an "if-then-else" notation facilitates answering *forward* questions (i.e., "which action results for given conditions?") while a "do-if" notation facilitates answering *backward* questions (i.e., "which conditions must exist to invoke a specified action?"). LabVIEW happens to provide two notations for expressing conditional logic. Green, Petre and Bellamy proposed that these correspond respectively to a forward and a backward form, and therefore compared them to forward and backward questions using two textual notations. The LabVIEW notations did exhibit the expected match-mismatch effect, but the effect size was less than anticipated. Instead response times were twice as long in comprehension questions using the visual notations.

In subsequent articles, Green and Petre (Petre 1995; Petre and Green 1993) argue that "secondary notation" accounts for a large part of the (un)readability of a visual notation, and that the ability to read a visual notation and its associated secondary notation is dependent upon training. Secondary notation – the use of layout and other informal cues to express structure – is defined below along with Green's other "cognitive dimensions of notations". Green and Petre base their argument on observational studies of novice LabVIEW programmers and expert electronics designers at work. As a group, the experts approached the study questions fairly consistently; they tended to use like strategies and to choose strategies based upon the style of the problem. In contrast, LabVIEW programmers were inconsistent in their strategies, even to the point of changing strategies in mid-task. Green and Petre attribute this to relative inexperience. Also, whereas the electronics experts recognized and took advantage of spatial groupings, the LabVIEW programmers seemed unaware of this secondary notation.

The Green and Petre results have been further explored by Moher et al. (1993) in a study that compared text to petri nets. Moher et al. were interested in whether petri nets show promise as the visual basis for a VPL. They used the same experimental design and comprehension questions employed in the Green, Petre and Bellamy study, but different visual representations. They designed three different petri net notations: one corresponded to the if-then-else statement (a forward form), one to LabVIEW's gates notation and one to a textual do-if statement (both backward forms). The Moher et al. study confirmed the match-mismatch hypothesis for textual notations but not for petri nets. Petri nets were faster for backward than for forward questions, but two of them performed worse than their text counterparts, while the third was not significantly different from text. These results concur with Green, Petre and Bellamy that the match-mismatch hypothesis cannot account for all of the differences seen in this experiment.

4.3. *Algorithm visualization*

A handful of empirical studies have investigated the use of animated pictorial visualization of software algorithms as a teaching tool. Often, such studies fall under the purview of diagrammatic research, for example when the visualization is based on a node and arc graph. Despite popular enthusiasm for the concept of visualization, these studies have found no conclusive evidence to recommend its use. For example, Stasko, Badre and Lewis examined the effects of visualizing a heap algorithm (Stasko, Badre and Lewis 1993). There was no significant difference in comprehension between students who learned the algorithm via a text description and those who also saw a visualization. In view of this failure of empirical validation, Gurka and Citrin (1996) advocate a careful meta-analysis of earlier studies, looking especially for factors that might have produced false negative results.

4.4. *Status of empirical studies*

The existing studies of diagrammatic notations used for programming tasks have contributed examples in which diagrammatic notations resulted in performance benefits. Several have shown visual notations outperforming text in either time or correctness, sometimes in both. Yet many basic assumptions surrounding these studies have not been investigated:

- Diagrams are likely to be better than text for some problems, worse for others. The differences lie in the costs of locating and indexing information, as well as in differences of cognitive processing of symbolic and spatial information – but these are open questions.
- Experts appear to do things differently from novices, but in many ways that are extremely hard to analyse. There are differences in processes of identification, indexing, selection, abstraction, matching strategy to task, finding ways of making complex problems tractable, and so on.
- Every notation makes some information accessible at the expense of obscuring other information. Hence the match-mismatch hypothesis.
- Despite the fact that differences between subjects are not analyzed in detail, individuals do differ.
- When researchers make assumptions about the relationships between programming languages and reasoning, they often don't hold up to empirical scrutiny of how people really program.

This section has raised more questions than it has answered; we don't wish to disguise the fact that many research questions in psychology of programming are both open and difficult. Nevertheless, we consider that these questions are central to the understanding of diagram usage in visual programming – a field where the properties of diagrams are central.

5. Cognitive Dimensions of Notations

The study of the cognitive factors involved in VPLs can be traced to Fitter and Green (1979). Green (with many collaborators, including several of the present authors) remains a central figure in this field. Despite this long history, empirical studies of programmers have had little effect on the design of new programming languages. They have generally addressed quite detailed aspects of programming style or language features, and provide grounds for a critique of specific language features rather than the broader issues of language format. Little has changed since the complaint made by Shneiderman in 1980 that "Computer scientists ... make broad claims for the simplicity, naturalness, or ease-of-use of new computer languages or techniques, but do not take advantage of the opportunity for experimental confirmation" (Shneiderman 1980, p. xiii).

Green (1989, 1991; Green and Petre 1996) has introduced the "cognitive dimensions of notations" framework as discussion tools – descriptions of the artifact-user relationship – intended to raise the level of discourse. (The following description of cognitive dimensions summarizes a more complete treatment in Green and Petre (1996)).

Cognitive dimensions constitute a small vocabulary of terms describing the cognitively-relevant aspects of structure of an information artifact, and show how they can be traded off against each other. Any cognitive artifact can be described in these terms and, although that description will be at a very high level, it will predict some major aspects of user activity. The framework is task-specific, concentrating on processes and activities rather than the finished product. This broad-brush framework supplements the detailed and highly specific analyses typical of contemporary cognitive models in HCI.

5.1. *Partial list of cognitive dimensions*

The framework of cognitive dimensions consists of a small number of terms which have been chosen to be easy for non-specialists to comprehend, yet capture a significant amount of the psychology and HCI of programming. The so-called 'dimensions' are meant to be coherent with each other, like physical dimensions. A partial list of dimensions follows, with thumb-nail descriptions:

Abstraction gradient: An abstraction is a grouping of elements to be treated as one entity, whether just for convenience or to change the conceptual structure. What are the minimum and maximum levels of abstraction? Can fragments be encapsulated?

Closeness of mapping: What 'programming games' need to be learned? Programming requires a mapping between a problem world and a program world. The closer the programming world is to the problem world, the easier the problem-solving ought to be.

Consistency: When some of the language structure has been learned, how much of the rest can be inferred successfully?

Diffuseness: How many symbols or graphic entities are required to express a meaning? Some notations use a lot of symbols or a lot of space to achieve the results that other notations achieve more compactly.

Error-proneness: Does the design of the notation induce 'careless mistakes'? Does it make them hard to find once they have occurred?

Hard mental operations: Are there places where the user needs to resort to fingers or pencil annotation to keep track of what's happening?

Hidden dependencies: A hidden dependency is a relationship between two components such that one of them is dependent on the other, but that the dependency is not fully visible. Is every dependency overtly indicated in both directions?

Premature commitment: Do programmers have to make decisions before they have the information they need?

Progressive evaluation: Can a partially-complete program be executed to obtain feedback on 'How am I doing?' The ability to evaluate their own problem-solving progress is essential for novices and desirable even for experts.

Role-expressiveness: Can the reader see how each component of a program relates to the whole? Role-expressiveness is enhanced by meaningful identifiers, by well-structured modularity, and by the presence of 'beacons' that signify certain code structures.

Secondary notation: Can programmers use layout, choice of naming conventions, grouping of related statements, colour, and other cues to convey extra meaning, above and beyond the 'official' semantics of the language?

Viscosity: How much effort is required to perform a single change? One standard example of viscosity is having to make a global change by hand because the environment contains no global update tools.

Visibility: Is every part of the code simultaneously visible (assuming a large enough display), or is it at least possible to juxtapose any two parts side-by-side at will?

5.2. *Trade-offs among dimensions*

The purpose of the cognitive dimensions framework is to lay out the cognitivist's view of the design space in a coherent manner, and where possible to display some of the cognitive consequences of making a particular bundle of design choices. From the point of view of the designer, there are important trade-off relationships between the cognitive dimensions. Changing the structure of a notation to reduce viscosity, for example, is likely to affect other dimensions (perhaps by introducing hidden dependencies or increasing the abstraction gradient). Far more analysis of trade-off relationships needs to be done. What is important is to bear in mind that because these trade-off relationships do exist, a notable success along one dimension may be reduced by poor performance on another.

The cognitive dimensions framework is by no means a finished entity. Meanwhile, its take-up by other researchers such as Modugno, Green and

Myers (1994), Buckingham-Shum and Hammond (1994) and Yang, Burnett, DeKoven and Zloof (1995) shows the wide applicability of the approach.

6. New Directions

This final section presents several directions for further research into diagrammatic representations. They arise from our current understanding of the cognitive factors involved in use of VPLs, and build on the research described in the body of the paper. These directions were originally proposed as intentionally provocative starting points for discussion at the Thinking with Diagrams Workshop.

6.1. *Basic questions*

The empirical studies that have been conducted have not yielded results in keeping with the enthusiasm of the visual programming community (Blackwell 1996b). Further success depends on the answers to these basic questions. Useful studies would shed light on what kinds of visual representations are beneficial for which tasks and especially on the appropriate class of users for a notation. We have little empirical evidence that any diagrammatic notation is responsible for improved programming performance. Three points are relevant to this issue:

First, any attempt to find such evidence should take care to separate the visual aspects of VPLs from other VPL features. New VPLs include language features orthogonal to the visual/textual dimension. Thus, a study showing a VPL outperforming a textual programming language does not necessarily mean that any visual aspect of the VPL was responsible.

Second, the question of whether a visual notation is appropriate for programming is difficult to answer since programming involves many cognitive tasks. The difficulty lies both in ascertaining which tasks account for a significant proportion of programming effort and whether a given VPL benefits those processes.

Third, the bottom line for much of the programming industry is whether a programming tool produces cost-effective results. To have a successful VPL, the issue is not whether the VPL produces a statistically significant effect, but rather whether the effect size is large enough to be of practical interest.

6.2. *Metaphor and representation*

Much development of visual languages is inspired by the success of the “desktop” metaphor for user interfaces. The reasons for the success of the desktop metaphor are not altogether clear, however. The principles of "direct manipulation" interfaces are agreed – manipulation of abstract entities is simplified if those entities are constrained to obey the rules of the physical world. Other than these fundamental correspondences, the value of the desktop metaphor (and especially extensions to that metaphor) have been widely questioned. Many PC users learn to use the Windows GUI without being aware of the intended metaphor, and researchers such as Halasz and Moran (1982) counsel caution in relying on metaphor.

Nevertheless, VPLs are often conceived with reference to a foundational metaphor. This metaphor generally corresponds to an underlying model of the language. Mayer (1975) demonstrated that an underlying model of the BASIC language – variables represent memory locations – was more easily learned when it was expressed in terms of metaphorical pigeon holes that represented memory locations. Blackwell (1996a) has suggested that VPLs often depend on implicit metaphors. LabVIEW, for example, is based on the data flow paradigm, and its pictorial metaphor depicts wires along which data flows.

Other psychology of programming researchers are working on the evaluation of alternative metaphorical presentations of programs. See, for example, Ploix (1996). The question for diagrammatic reasoning research is whether these metaphors have any effect other than the instructional benefits already observed by Mayer. For example, Cox (1997) has suggested that misinterpretations of common graphical notations such as Euler's circles can be attributed to use of inappropriate metaphors. Blackwell (1998), in experimental investigations, failed to find substantial benefits of instructional or pictorial metaphors in VPL-like diagrams. A more thorough investigation of the metaphors used in VPLs may throw further light on this issue.

6.3. *The role of reusable components*

Many of the commercially available VPLs are promoted as being particularly suitable for end-user programming, a goal that they supposedly achieve by using a visual syntax. However, VPLs such as LabVIEW also provide users with well-stocked libraries of reusable software components. Other examples of languages providing such libraries can be found in Hils' (1992) survey of dataflow-based VPLs.

Supplying these libraries effectively raises the semantics of the language to a higher level than traditional textual programming languages. Thus, the question is raised: How much of these languages' benefits come from semantic level as opposed to their visual aspects? Furthermore, this line of reasoning could lead one to question the importance placed on metaphor. Is the issue of designing a useful VPL really one of communicating computational metaphors or is it an issue of matching the tool to the problem domain?

6.4. *Programming languages for children*

Programming is widely regarded as a skill that would have great educational value when taught to children. Papert's (1980) Logo language was designed for use by children, and has been used successfully for many years. Logo operates in a graphical application domain, but the language itself is purely textual. Di Sessa's Boxer (1986) was an attempt to make Logo more diagrammatic, but Boxer concentrated on issues of context and environment within which text could be organized. Two recent languages are particularly well suited to simple games and simulations. AgentSheets (Repenning and Sumner 1995) and stagecast creator$^{\mathrm{TM}}$ (originally described as "KidSim" in Smith, Cypher and Spohrer 1994) allow children to manipulate graphical elements directly, and define their behaviour within a graphical grid in terms of graphical rewrite rules and condition-action rules.

A number of commercial visual programming products have used the presentational techniques of video games to provide languages that are apparently far more attractive to children. In the mid-1980s the "ChipWits" game allowed the definition of robot behaviour by wiring logical elements together. These robots could then be placed in a maze where they competed with other robots. More recently, "Widget Workshop" provided the basic elements of a data flow language in which fanciful animated devices could be wired together – this program depiction is entertaining in its own right. In Kahn's ToonTalk$^{\mathrm{TM}}$ product (Kahn 1996), three dimensional cartoon characters assemble functional programs as observable actions carried out by robots inside separate houses.

These developments are undeniably entertaining. A critical question is: are they educational? Kahn reports that 8-year old boys using ToonTalk$^{\mathrm{TM}}$ take most pleasure in the fact that, once a function has completed, the house where it was evaluated explodes. This is unsurprising to anyone who has worked with children, but perhaps it reflects a deeper problem with the use of any type of diagrams by children. DeLoache and Marzolf (1992) report the difficulty children have in interpreting symbolic references. They found that more highly salient representations are far less likely to be dereferenced successfully, and that the value of the representation can actually be decreased if it is made more salient by allowing children to play with it.

The use of videogame-like presentation to teach programming to children is certain to become more popular, for the same reasons that VPLs are often immediately appealing to adults. The benefits and dangers of this trend fall within the remit of psychology of programming research, and of diagrammatic reasoning research, but they have not yet (as far as we know) been critically evaluated in these terms. This would seem to be a high priority before this type of product becomes widespread in educational settings.

6.5. *Paradigm and comprehension*

First generation VPLs (in the 1980s) expressed the paradigm of control flow – the order in which the program is executed – as shown in flow charts. In contrast, many current languages (including LabVIEW, VEE and Prograph) are based on the paradigm of data flow. Much empirical work on VPLs has compared relative performance between text and graphics, with little or no consideration given to how the underlying paradigm might affect comprehension. These studies may therefore be confounding the issues of diagram usage and paradigm, at least with respect to program comprehension (program design is considered in the next section). Work on this question is ongoing, and is investigating the relationship between paradigm and representation from various angles in order to assess the relative contributions of each. Control flow and data flow paradigms within VPLs have been studied by Good (1999), while Whitley (1997b) is investigating the data flow paradigm as embodied in both textual and visual notations. Navarro-Prieto (1998) has shown that the data flow organisation of spreadsheets encourages a data-structured mental model of a program.

Initial results therefore show that paradigm does make a difference, at least for novice programmer comprehension activities. Control flow VPLs seem to encourage better understanding of low-level program constructs such as program operations, actions and control flow, while data flow VPLs lead to a higher-level understanding of the program's data flow and functional aspects (Good 1999). This work casts some doubt on the common assumption that visual programming representations are mainly valuable insofar as they depict data flow. Further results which can shed more light on this issue are eagerly awaited.

6.6. *Paradigm and design*

Although paradigm is obviously an important factor in a programming language, studies of program design by experts suggest that design activities are not completely constrained by paradigm. Experienced programmers often plan their solution to a programming problem before they start writing any code. Studies of expert programmers show that they do not solve problems in the target programming language, but rather construct their solution strategies in a personal pseudo-language that is subsequently translated into code. (Petre and Winder 1988) This personal design representation can include pictorial mental images, whether or not the target language is a VPL (Petre and Blackwell 1997).

There is not a necessary correlation between programming languages and solution strategies; on the contrary, strategies volunteered as typical of one paradigm can often be implemented in a language that fits within another. Scholtz and Wiedenbeck (1992) observed that experts learning a new language used strategies already familiar to them, rather than learning how the features of the new language might require a different approach. Furthermore expert programmers do not observe language or paradigm boundaries when constructing solutions but rather borrow useful features across languages or domains. A paradigm might influence strategy, but it may not be the paradigm embodied in the language (Petre 1996). This implies that the cost of reasoning about a given representation might vary, depending on how the programmer's reasoning shifts. Researchers therefore need to consider reasoning paradigm separately from representation paradigm when investigating design and/or coding activities.

6.7. *Multiple representations*

What are the benefits (if any) of using multiple representations in programming? Shneiderman et al. (1977) suggested that if one of the representations is well-known and understood, then additional ones will be redundant. However, this may not necessarily be true, given that:

- different representations will highlight different types of information at the expense of others.
- understanding of a notation is often not complete and/or correct (e.g. when learning a new language, or novel applications of a language).

We know of no empirical work in the field which has addressed the use of multiple representations in program coding. However, Petre et al. (1998) have considered a number of scenarios in which multiple representations might be used in software visualisation, and many of these are applicable to coding: again, empirical confirmation of these hypotheses would be welcome.

6.8. *Scalability*

The scalability question (Burnett et al. 1995) asks whether a software technique suitable for solving small problems can successfully be extended to large-scale ones. This applies to many aspects of computer science. For example, a configuration suitable for a small network may become prohibitively expensive due to traffic bottlenecks as its size increases. Despite common criticisms regarding scalability, Burnett et al. claim that VPLs can successfully be applied to large problems. There are, however, few empirical studies investigating scaling issues that are specific to visual notations, such as problems of perceptual discrimination between large numbers of similar symbols.

On the other hand, there may be no demand for more scalable VPLs: Nardi (1993) has predicted that the most successful VPLs will be the ones aimed at end user programming. Thus, the real impact of these VPLs may stem from enabling more people to get more of their relatively smaller jobs done.

6.9. *Software lifecycle*

New programming languages are often developed with an emphasis on the processes of coding and comprehension. In practice, however, commercial software development projects spend far more effort on activities such as system specification, documentation, testing and maintenance. Whitley and Blackwell (1997) found that commercial users of LabVIEW are concerned with these issues, and that this concern focuses on practical questions that are seldom considered in visual language research – the problem of how visual languages should be printed out, for example, or how source code control systems can construct visual delta files showing the changes between different program versions.

Investigation of the rest of the software lifecycle would be a particularly rewarding area for VPL research. It is also particularly challenging, because it requires long-term observational studies, and is not amenable to investigation through controlled experiments. As our review of empirical studies shows, even small-scale controlled studies have found little conclusive evidence regarding the benefits of VPLs. Large-scale studies of real development projects are unlikely to produce substantially more clearcut results.

7. Conclusions

This paper has described the origins and characteristics of the class of diagram described as visual programming languages, and has reviewed techniques which have been used to study these diagrams within the context of psychology of programming.

A number of researchers have carried out empirical studies using these techniques to directly compare problem solving performance using diagrams and textual notations. These studies have not been conclusive as was hoped, but we do not wish to be overly critical of the intuitive position regarding the value of visual representations. The challenge to researchers in thinking with diagrams is to explain why this intuition might be valid, and to propose the ways in which, if it is valid, it can most effectively be exploited.

The empirical studies completed to date have, however, formed a stepping stone for broader frameworks of notational analysis such as Green's Cognitive Dimensions. These new integrative approaches are moving away from simplistic comparisons of text and graphics, in order to investigate the ways in which tasks can be matched to appropriate representations.

Visual programming is a particularly fertile area for studying the application of diagrams to real-world problem solving contexts. This paper has proposed a number of research questions that are of particular relevance at the time of writing. Any of these would be both challenging and valuable as a starting point for future diagrammatic reasoning research.

References

Baroth, E. & Hartsough, C. (1995). Visual Programming in the Real World. In Burnett, M., Goldberg, A. & Lewis, T. (eds.) *Visual Object-Oriented Programming: Concepts and Environments*, chapter 2, 21–42. Manning Publications Co.

Blackwell, A.F. (1996a). Metaphor or Analogy: How Should We See Programming Abstractions? In Vanneste, P., Bertels, K., De Decker, B. & Jaques, J.-M. (eds.) *Proceedings of the 8th Annual Workshop of the Psychology of Programming Interest Group*, 105–113.

Blackwell, A.F. (1996b). Metacognitive Theories of Visual Programming: What Do We Think We Are Doing? In *Proceedings IEEE Symposium on Visual Languages*, 240–246.

Blackwell, A.F. (1998). *Metaphor in Diagrams*. Unpublished PhD thesis. University of Cambridge.

Buckingham-Shum, S. & Hammond, N. (1994) Argumentation-Based Design Rationale: What Use at What Cost? *International Journal of Human-Computer Studies* **40**: 603–632.

Burnett, M.M., Baker, M.J., Bohus, C., Carlson, P., Yang, S. & van Zee, P. (1995). Scaling Up Visual Programming Languages. *IEEE Computer* **28**(3):45–54.

Cox, R. (1997). Representation Interpretation Versus Representation Construction: An ILE-Based Study Using SwitchERII. In *Proceedings of the 8th World Conference of the Artificial Intelligence in Education Society (AI-ED 97)*. Amsterdam: IOS Press, 'Frontiers in Artificial Intelligence and Applications' series.

Cunniff, N. & Taylor, R.P. (1987). Graphical vs. Textual Representation: An Empirical Study of Novices' Program Comprehension. In *Empirical Studies of Programmers: Second Workshop*, 114–131.

Curtis, B., Sheppard, S.B., Kruesi-Bailey, E., Bailey, J. & Boehm-Davis, D.A. (1989). Experimental Evaluation of Software Documentation Formats. *Journal of Systems and Software* **9**(2): 167–207.

DeLoache, J.S. & Marzolf, D.P. (1992). When a Picture is Not Worth a Thousand Words: Young Children's Understanding of Pictures and Models. *Cognitive Development* **7**: 317–329

diSessa, A. (1986). Notes on the Future of Programming: Breaking the Utility Barrier. In Norman, D.A. & Draper, S.W. (eds.) *User Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum: Hillsdale, NJ.

Ferguson, E.S. (1992). *Engineering and the Mind's Eye*. MIT Press: Cambridge, MA.

Fitter, M. & Green, T.R.G. (1979). When do Diagrams Make Good Computer Languages? *International Journal of Man-Machine Studies* **11**(2): 235–261.

Gilmore, D.J. & Green, T.R.G. (1984). Comprehension and Recall of Miniature Programs. *International Journal of Man-Machine Studies* **21**: 31–48.

Glinert, E.P. (ed.) (1990a). *Visual Programming Environments: Applications and Issues*. IEEE Computer Society Press.

Glinert, E.P. (ed.) (1990b). *Visual Programming Environments: Paradigms and Systems*. IEEE Computer Society Press.

Good, J. (1999). *Programming Paradigms, Information Types and Graphical Representations: Empirical Investigations into the Factors involved in Novice Program Comprehension*. Unpublished PhD thesis. University of Edinburgh, Department of Artificial Intelligence.

Green, T.R.G. (1977). Conditional Programs Statements and Their Comprehensibility to Professional Programmers. *Journal of Occupational Psychology* **50**: 93–109.

Green, T.R.G. (1989) Cognitive Dimensions of Notations. In Sutcliffe, A. & Macaulay, L. (eds.) *People and Computers V*, 443–460. Cambridge University Press: Cambridge.

Green, T.R.G. (1991) Describing Information Artefacts with Cognitive Dimensions and Structure Maps. In Diaper, D. & Hammond, N.V. (eds.) *People and Computers VI*, 297–315. Cambridge University Press: Cambridge.

Green, T.R.G. & Petre, M. (1992). When Visual Programs are Harder to Read than Textual Programs. In *Proceedings of Sixth European Conference on Cognitive Ergonomics (ECCE6)*, 167–180.

Green, T.R.G. & Petre, M. (1996) Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework. *Journal of Visual Languages and Computing* **7**: 131–174.

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 *Empirical Studies of Programmers: Fourth Workshop*, 121–146.

Gurka, J.S. & Citrin, W. (1996). Testing Effectiveness of Algorithm Animation. In *Proceedings of 1996 IEEE Symposium on Visual Languages (VL '96)*, 182–189.

Halasz, F.G. & Moran, T.P. (1982). Analogy Considered Harmful. *Proc. Conf. Human Factors in Computing Systems*: ACM, 383–386.

Hils, D.D. (1992). Visual Languages and Computing Survey: Data Flow Visual Programming Languages. *Journal of Visual Languages and Computing* **3**: 67–101.

Hoc, J.-M., Green, T.R.G., Samurçay, R. & Gilmore, D.J. (1990). *Psychology of Programming*. Academic Press: London.

Kahn, K. (1996). ToonTalk^{TM} – An Animated Programming Environment for Children. *Journal of Visual Languages and Computing* **7**(2): 197–218.

Mayer, R.E. (1975). Different Problem-Solving Competencies Established in Learning Computer Programming With and Without Meaningful Models. *Journal of Educational Psychology* **67**(6): 725–734.

Modugno, F., Green, T.R.G. & Myers, B.A. (1994). Visual Programming in a Visual Domain: A Case Study of Cognitive Dimensions. In: *People and Computers IX*, 91–108. Cambridge University Press: Cambridge.

Moher, T.G., Mak, D.C, Blumenthal, B. & Leventhal, L.M. (1993). Comparing the Comprehensibility of Textual and Graphical Programs: The Case of Petri Nets. In *Empirical Studies of Programmers: Fifth Workshop*, 137–161.

Myers, B.A. (1990). Taxonomies of Visual Programming and Program Visualization. *Journal of Visual Languages and Computing* **1**: 97–123.

Nardi, B.A. (1993). *A Small Matter of Programming: Perspectives on End User Computing*. MIT Press: Cambridge, MA.

Navarro-Prieto, R. (1998). *The Role of Imagery in Program Comprehension: Visual Programming Languages.* Doctoral Thesis, University of Granada. ISBN 84-8497-957-1.

Newell, A. & Simon, H.A. (1972). *Human Problem Solving.* Prentice Hall: Englewood Cliffs, NJ.

Pandey, R.K. & Burnett, M.M. (1993). Is It Easier to Write Matrix Manipulation Programs Visually or Textually? An Empirical Study. In *Proceedings of 1993 IEEE Symposium on Visual Languages (VL '93)*, 344–345.

Papert, S. (1980). *Mindstorms, Children, Computers and Powerful Ideas.* Basic Books: New York.

Petre, M. (1991). Shifts in Reasoning About Software and Hardware Systems: Must Operational Models Underpin Declarative Ones? In: *The Third Workshop of the Psychology of Programming Interest Group* (Hatfield, January).

Petre, M. (1995). Why Looking Isn't Always Seeing: Readership Skills and Graphical Programming. *Communications of the ACM* 38(6): 33–44.

Petre, M. (1996). Programming Paradigms and Culture: Implications of Expert Practice. In: Woodman. M. (ed.) *Programming Language Choice: Practice and Experience*, 29–44. International Thomson Computer Press: London.

Petre, M. & Blackwell, A.F. (1997). A Glimpse of Expert Programmer's Mental Imagery. In Wiedenbeck, S. & Scholtz, J. (eds.) *Proceedings of the 7th Workshop on Empirical Studies of Programmers*, 109–123.

Petre, M., Blackwell, A.F. & Green, T.R.G. (1998). Cognitive Questions in Software Visualisation. In Stasko, J., Domingue, J., Brown, M. & Price, B. (eds.) *Software Visualization: Programming as a Multi-Media Experience*, 453–480. MIT Press: Cambridge, MA.

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.

Petre, M. & Winder, R. (1988). Issues Governing the Suitability of Programming Languages for Programming Tasks. In: *People and Computers IV.* Cambridge University Press.

Ploix, D. (1996). Building Program Metaphors. In Ireland, M. (ed.), *Proceedings of the First Psychology of Programming Interest Group Postgraduate Student Workshop*, 125–129.

Price, B.A., Baecker, R.M. & Small, I.S. (1993). A Principled Taxonomy of Software Visualization. *Journal of Visual Languages and Computing* **4**(3): 211–266.

Reitman, W.R. (1964). Heuristic Design Procedures, Open Constraints and the Structure of Ill-Defined Problems. In Shelly, M.W. & Bryan, G.L. (eds.) *Human Judgements and Optimality.* Wiley and Sons: New York.

Repenning, A. & Sumner, T. (1995). Agentsheets: A Medium for Creating Domain-Oriented Visual Languages. *IEEE Computer* 28(3): 17–25.

Scanlan, D.A. (1989). Structured Flowcharts Outperform Pseudocode: An Experimental Comparison. *IEEE Software* **6**(5): 28–36.

Shneiderman, B. (1980). *Software Psychology: Human Factors in Computer and Information Systems.* Winthrop: Cambridge, MA.

Shneiderman, B., Mayer, R., McKay, D. & Heller, P. (1977) Experimental Investigations of the Utility of Detailed Flowcharts in Programming. *Communications of the ACM* **20**: 373–381.

Smith, D.C., Cypher, A. & Spohrer, J. (1994). KIDSIM: Programming Agents Without a Programming Language. *Communications of the ACM* **37**(7): 55–67.

Stasko, J.T., Badre, A.M. & Lewis, C. (1993). Do Algorithm Animations Assist Learning? An Empirical Study and Analysis. In *Proceedings of INTERCHI '93 Conference on Human Factors in Computing Systems*, 61–66.

Vessey, I. & Weber, R. (1986). Structured Tools and Conditional Logic: An Empirical Investigation. *Communications of the ACM* **29**(1): 48–57.

Weinberg, G.M. (1971). *The Psychology of Computer Programming*. Van Nostrand Reinhold: New York.

Whitley, K. (1997a). Visual Programming Languages and the Empirical Evidence for and Against. *Journal of Visual Languages and Computing* **8**(1): 9–142.

Whitley, K. (1997b). In Search of Visual Programming. In Kann, C. (ed.) *Proceedings of the First Empirical Studies of Programmers Graduate Student Workshop*, October 24, 1997.

Whitley, K.N. & Blackwell, A.F. (1997). Visual Programming: The Outlook from Academia and Industry. In Wiedenbeck, S. & Scholtz, J. (eds.) *Proceedings of the 7th Workshop on Empirical Studies of Programmers*, 180–208.

Yang, S., Burnett, M.M., DeKoven, E. & Zloof, M. (1995) *Representation Design Benchmarks: A Design-Time Aid for VPL Navigable Static Representations*. D.C.S. Tech Rpts 95-60-3. Oregon State University (Corvallis).

Artificial Intelligence Review **15:** 115–134, 2001.

Learning to Think and Communicate with Diagrams: 14 Questions to Consider

PAUL BRNA[1], RICHARD COX[2] and JUDITH GOOD[3]
[1]*Computer Based Learning Unit, University of Leeds, UK;* [2]*School of Cognitive and Computing Sciences, University of Sussex, UK;* [3]*Human Communication Research Center, University of Edinburgh, UK*

Abstract. This paper looks at the particular role which diagrammatic representations, and external representations more generally, play within an educational context. In particular, it considers the way in which the demands on diagrammatic representational systems in educational settings differ with respect to other settings (e.g. professional): in some instances, these demands are increased, while in others, the demands are markedly different.

The paper considers three key issues: the question of whether diagrams make certain tasks easier (and whether this is desirable from an educational point of view), the generalisation and transfer of diagrammatic skills once learnt, and the possible problems associated with simultaneously learning domain knowledge and a novel representational system.

The paper then considers a number of sub-issues, and concludes by highlighting areas of particular interest for future AI research.

Keywords: diagrammatic reasoning, external representations, learning, skill transfer

1. Introduction

This paper is written from the perspective of those interested in the roles that diagrams have, or could have, within the context of learning, and within the broader context of education. It is written for those who wish to apply AI techniques to support learning to think and communicate with diagrams and who also wish to consider the complexities involved. The paper is presented in terms of fourteen questions which require attention.

Viewing "thinking with diagrams" from an educational perspective brings into consideration a number of factors outside those normally considered by persons interested in the formal properties of diagrams. These will be outlined below.

Additionally, however, much of what follows is applicable to the wider issues of "thinking with external representations". The reader can and should consider the issues raised with respect to this broader context.

This relationship is bidirectional: not only can the educational context provide insights which are more widely applicable, but many of the issues found in the literature on the use of external representations are also relevant within an educational context. Thus there is room here for the discussion of issues such as different levels of description, diagrams vs. text, diagram ontologies, linguistics, semiotics, etc.

We have highlighted what we feel are the key issues that are important within educational contexts. However, there are many issues which deserve greater consideration than we are able to give in this paper – for example, the differences between 'animated' and 'non-animated' diagrammatic representational systems,[1] and the ways in which diagrams can complement textual explanations. Neither have we provided a full exploration of the ways in which external representation systems differ from diagrammatic representation systems.[2]

2. The Main Issues

Diagrammatic reasoning in educational contexts (and in more general ones too) can be described as depending on the nature of the task, the semantic properties of the diagram, and the person's prior knowledge, including skills, preferences, experiences, etc. (see, for example (Cox and Brna 1995; Cox 1997; Salomon 1994)).

A major difference between the educational and general contexts is the emphasis in the former on learning, and therefore greater importance is placed on those characteristics which are expected to promote learning.

In both contexts, an individual might typically want to solve a problem. The process of working through from an informal, ill-defined task to a solution has been described by Cox and Brna in terms of the stages of comprehension, selection of external representation system, construction of external representation, and use (including reading-off results) (Cox and Brna 1995), However, it is worth stressing that these different processes interact in very complex ways.

In non-educational (e.g. professional) contexts, it is quite common for groups of colleagues to work on a problem together using diagrams both as a vehicle for problem solving and as a means of performing a variety of communicative acts. Thus the diagram may be both an end product and a means to an end.

However, when we start to consider education contexts in greater details, there is a complication in this view of reasoning with diagrams (and other external representations). While all uses of diagrams can be viewed in discourse theoretic terms, diagrams play a special part in the ongoing discourse between teacher and student.

The difference in status between teacher and students, together with the different expectations present and the ways in which learning is assessed undoubtedly have an effect on the uses to which diagrams are put. The problem here is to outline what is happening. While we will not go into the nature of this discourse in great detail, viewing diagrams as embedded in a discourse may help us understand why the *educational* use of diagrammatic reasoning differs from many other uses.

In considering the educational issues, a number of tensions can be identified. First, there is a tension between 'making things easy' and helping students to learn. This tension derives from the assumption that learning comes through doing rather than from some form of passive observation. If learning to utilise diagrams is to be effective, we can expect that practice in utilising diagrams will be important. For example, if we believe that Venn diagrams are an effective method for solving a certain class of problems, then providing a problem together with a range of possible diagrammatic solutions may make it easier for students to solve the problem but give the student no opportunity to practice diagram construction.

Secondly, there is a tension between situated diagram systems and the desire to achieve a degree of generalisation so that the skills and knowledge gleaned from learning one system can more easily be transferred to a new system.[3] This problem is one that must be considered in the course of any attempt to provide a stand-alone curriculum for education in the use of diagrams. Formally, such a curriculum should be possible as other representational systems have their place in the educational system (grammar, algebra, geometry etc) (Cox 1996).

Thirdly, there is a difficulty that arises in terms of meeting a new diagram system at the same time that new subject-based material is being learned. The tension here is between learning the diagram system versus learning the topic (Brna 1996).

We therefore organise the first part of this discussion along three lines:

1. The tension between making things easy and helping students to learn.
2. The question of how to support the generalisation and transfer of diagrammatic skills.
3. Learning unfamiliar representational systems while learning conceptually new subject matter.

3. Making Things Too Easy?

Question 1: What tasks do diagrams make easier?
Can we consolidate the extensive research on the use of diagrams and other external representations in learning and problem solving to be able to generate useful advice for the instructional designer/teacher?

Question 2: What are the benefits of making tasks easier?
Does making some task easier imply using multiple representations (Barwise 1993)?

One way to facilitate learning is to 'play to the strengths' of the student. Thus areas of difficulty are avoided in some way in order to allow students to exploit their strengths or bypass their weaknesses.

Above, we suggested that learning comes through doing rather than from some form of passive observation. The full range of problem solving skills do need practice to attain both competence in performance and comprehension of the (diagrammatic) representational system itself. However, we also accept that it would be a mistake to assume that observation is inevitably passive: an observer may be highly active in constructing, for example, some interpretation of a diagram without necessarily physically manipulating the diagram in any way.

We take it as given that, whatever the form of internal representations, students are actively processing information obtained from the external representation. It is also reasonable to assume a framework along the lines of Rumelhart and Norman's "Accretion, Tuning and Restructuring" (Rumelhart and Norman 1978), and that sometimes new information can lead to extensive internal restructuring (though the current feeling is that such 'radical' change is quite rare).

Salomon has provided a cognitive analysis of the effect of different media on learning (Salomon 1994).[4] He distinguished between learning external and internal symbolic codes, and between 'stationary' and 'transformational' codes. Simplifying for convenience, stationary is to transformational as problem state is to operator in Newell and Simon's view of problem solving. Stationary and transformational codes are interrelated. For Salomon, a transformational code can be 'short circuited' by providing the stationary code which would have been achieved if the learner had performed the necessary transformations. He argues that short circuiting can save effort but will not lead to the kind of skill internalisation needed for the skill to become a 'mental tool'. This is one of his key concepts for explaining certain effects in the use of media for learning.

Two further valuable concepts are those of activation and supplantation. Activation occurs when the preconditions of a previously learned skill are satisfied, while supplantation is the case where an external transformation has the almost identical effect as the internal transformation that the learner would have applied. Hence supplanting both saves cognitive effort and *models* the transformation explicitly, and more or less accurately. These ideas appear to be fruitful for any analysis that seeks to explain different effects attributable to the external representation systems themselves.

He argues that activation of a cognitive skill is only of benefit if the basic skill is present; short circuiting is useful for saving mental effort but at the cost of failing to provide skill activation. Supplantation is seen as a way in which poor mastery of a skill can be compensated for. However, Salomon does point out that we have no reason to assume that these outline hypotheses work either for all situations or for all students.

As an illustration of supplantation, Salomon (1994) quotes Cronbach and Snow as suggesting that "the high-verbal learner who is weak in visualization might be supplied with extensive diagrams and left to generate his own verbal representations" (Cronbach and Snow 1977, p. 70). (Assuming that it is the process of diagram construction that is being supplanted.)

However, Cox, Stenning and Oberlander (1994) have shown that subjects classed as good diagrammatic reasoners perform better than poor diagrammatic reasoners following a graphically taught logic course (Cox et al. 1994). More precisely, since there was also evidence that syntactic teaching produced better outcomes for non-diagrammatic reasoners in a syntactically taught group, the teaching approach matched to reasoning modality preference produced better learning outcomes.

So it remains unclear whether we should compensate for, or teach to cognitive style differences. Even assuming we could identify the relevant individual preferences/aptitudes, we would expect that there will not be a simple either/or decision to take on this.

Cheng has provided a number of examples of Law Encoding Diagrams (LEDs) which capture some implications of physical theories as constraints on the diagram structure (Cheng 1996). These diagrams make some aspects of the domain easier to manage, making many problems easier to solve than when using standard algebraic techniques. In order to take advantage of LEDs, a new representational system has to be learned, including how to manipulate it. If students can effectively incorporate LEDs into their problem solving repertoire, then there is still a question as to whether they remain a purely personal 'tool for thought' or a way of communicating problem solving solutions. If the former case holds, then perhaps this not a useful way of making the task easier since there is then the problem of how the student maps from the LED to the solution required by teachers, examination boards etc. In other words, LEDs need to become acceptable in the classroom.

Finally, there is the issue of multiple representations. Given that particular representations necessarily highlight some types of information at the expense of others, this property of selective highlighting could potentially be exploited so as to help students focus on the information which is relevant to their current stage in the problem solving process.[5] However, there is some evidence that students cannot always cope effectively with multiple coordinated representations. Ainsworth et al. demonstrate that simply providing coordinated representations does not itself guarantee success (Ainsworth et al. 1996). Indeed, under certain circumstances, it would seem that these representations could 'act against' each other. Schwarz and Dreyfus constructed several measures of information integration from multiple representations (Schwarz and Dreyfus 1993). These are candidates for further work in developing methods for evaluating information integration in systems such as the ones Ainsworth describes.

Cox and Brna have provided a detailed study of the use of SwitchER, a system designed to encourage users to solve problems using multiple serial representation construction (Cox and Brna 1995). Their results indicate that representation selection depends on subjects examining the task requirements carefully. SwitchERII, an intelligent learning environment developed by Cox, features a variety of ways in which the subject is supported in his/her attempt to solve a certain class of word problems (analytical reasoning problems) (Cox 1996). These include an ability to reorder the question's premises, recognition of possible unfamiliarity with a representational system, and recognition of some problem with a specific diagram. However, subjects were often unable to report their awareness of any specific advice given by the system, probably due to the high cognitive load of problem comprehension and external representation construction. Certainly subjects could not avoid noting an intervention since no further progress was possible until they clicked on the appropriate button. There were cases of students who went on to make an error despite a warning about their representation. In this case, SwitchERII made the task easier but the interface did not fully ensure the assistance was recognised. SwitchERII has been used to investigate the relationship between errors made in diagram interpretation and errors made subsequently (by the same subjects) in diagram construction. It was found that errors in representation interpretation do not necessarily predict errors of representation construction – the two tasks differ significantly in terms of the cognitive subsystems involved and in the degree of engagement with the task that they engender (Cox 1996, 1997).

4. Generalising and Transferring Diagrammatic Reasoning Skills?

> **Question 3**: Can we provide a convincing account of how learners gain a high level of competence at operating with relatively unfamiliar external representation systems?
>
> What are the alternative models for generalisation? What is the most convincing explanation?
>
> **Question 4**: How important is translation skill?
>
> Can we accept that one measure of proficiency is the ability to move gracefully between external representation systems? Are there alternative measures?

As mentioned above, learning to use a diagrammatic method for performing a task always occurs in a specific context, and depends to a great extent on the task, the student's prior knowledge, diagrammatic reasoning system's characteristics and the physical and social context. Many of the problems in becoming an effective 'diagrammatic reasoner' can be viewed in terms of how to generalise from experience and transfer skills learned in one context to another.

In seeking to explain transfer, Lowe has argued for the existence of *diagram genres* which enable those familiar with domain dependent, highly specific representations to be highly proficient with superficially very different diagrammatic systems in different domains of use (Lowe 1994a).

There are models of learning to use representations embedded in a computer interface. These are models of exploratory learning often set in the context of learning to use an unfamiliar software application. However, for the most part, this approach has tended to examine skill learning rather than conceptual learning. For example, Rieman, Young and Howes have provided a fine-grain model of exploratory learning (IDXL) which depends on scanning the interface and internal comprehension strategies (Rieman et al. 1996). Kitajima and Polson provide a more abstract comprehension-based model (The LInked model of Comprehension-based Action planning and Instruction taking – LICAI) of exploration of an interface (Kitajima and Polson 1996; Kitajima and Polson, in press) based on Kintsch's construction-integration theory of text comprehension (Kintsch 1988). These models take some steps that may eventually lead to models of exploration-based diagrammatic reasoning in unfamiliar domains as they seek to explain exploration driven by a combination of the situation, prior knowledge and the task. So far, they do not address the problems of using computer applications to learn conceptually difficult material at the same time.

Van der Pal has tried to confront some aspects of learning formal reasoning from a situated action perspective (cf (Brown et al. 1989)) while seeking to respect the need to explain the process of generalisation/decontextualisation (van der Pal 1996). However, it is not just the circumstances surrounding the use of diagrammatic systems which lead to problems connected with generalisation: to an extent, it is connected with the semantic and cognitive properties of the representation systems themselves.

Stenning and Oberlander have argued that a diagrammatic representation system gains some of its cognitive tractability from the limitations on its power of expression (Stenning and Oberlander 1995). Their argument depends in part on the difficulty of representing alternative possibilities using diagrams. This property of diagrams is addressed in the Hyperproof system (Barwise and Etchemendy 1994). This system features a fairly standard natural deduction proof system with a limited set of predicate relationships available. The innovative feature is a chess board and blocks of different shapes, sizes and positions and some mechanisms for allowing for indeterminate representations – abstractions over position, shape and size.[6]

The axioms and goal to be proved are all cast in terms of the blocks in the blocks' world. Provision is made for moving information between the (sentential) proof system and the (graphical) blocks' world (i.e. heterogenous reasoning (Barwise 1993)). The sentential proof is constructed in a window below the blocks world (graphical) window.

One potential source of difficulty for students arises when the proof requires the enumeration of all the possibilities so that each can be tackled one at a time. This 'case split' is difficult to represent in the blocks world since only one world can be shown at a time – so the Hyperproof diagrammatic system 'encourages' students to systematically 'break into cases' and work with a single case (at a time).

Formal logic encourages the user to make the case split explicit prior to handling the individual cases and allows scanning of this structure throughout, while Hyperproof's blocks world only permits one case to be visible at a time in the graphical window. However, Hyperproof's Fitch-style sentential notation goes some way towards giving an overview of the proof structure, and a 'focus slider' is provided to show the blocks world configuration at different steps of the sentential proof.

The issue of generalising and transferring diagrammatic skills has both intra and inter domain extensions. For example, does learning a graphical representation for program design facilitate the learning of, say, a visual programming language? Even further afield, will this same knowledge give students any advantage when they come to learn the diagrammatic representation system used by Hyperproof?

We can regard the transfer of diagrammatic reasoning skills as a complex process of becoming familiar with the current reasoning context, decontextualising what is learned through working on a number of different tasks in a variety of different social situations at different stages of the student's experience, becoming familiar with other representation systems, and then learning to move 'gracefully' between them. This is a complicated process!

In any particular community of practice, facility in translating between commonly-used external representations is a key test of expertise. It can be argued that learning to translate is of vital importance to learning. Kaput, for example, provides an analysis of the ways in which meaning is developed through translation (Kaput 1987).

5. Learning Too Much at the Same Time?

Question 5: How can we partition the cognitive load in a sensible way?

Is it realistic to teach the external representation system prior to teaching the particular domain knowledge of interest? How important is this period of learning? i.e. what effect can we measure or predict will occur in terms of the student's development, as evidenced both by learning the representational system and by learning the new subject-based material?

Question 6: Should this experience be avoided?

If so, how, and if not, why not? Is it better or worse to protect students from this kind of experience? How is student's prior knowledge factored in to provide individualised support? How should this be done (if at all)? What combination of task and semantic characteristics of external representation systems might help to ease the problem?

Students (singly, in pairs and in larger groupings) often learn some domain of knowledge which is represented with the help of a (diagrammatic) external representation system at the same time as also 'learning' the representational system.

In these situations, students are often faced with representations that may be – to them – ambiguous. Therefore they may commit to an interpretation of the representational system (or even some specific representation) which may be inconsistent or flawed in some way in relation to the intended meaning. Thus students have to *interpret* elements of the representation and *identify* (parts of) the representational system (Giere 1988). Things are complicated because students are simultaneously constructing a model of the domain knowledge derived from the domain-based material they are still studying.

For domain specific diagrammatic representation systems, the explicit provision of some of the key components of the domain knowledge may help but Lowe has pointed out how many educators implicitly assume that the representation system itself will help the student to understand these key domain knowledge components (Lowe 1994b). This makes the task much harder for the student.

In the initial stages, the student is struggling with both learning the representation system and learning the domain knowledge. Some may argue that the period of time during which students are 'learning too much' is quite brief compared with the length of time over which the external representation system is used, and hence that it is unlikely to make a difference to the student's development. On the other hand, the period during which students need to actively seek to understand external representations and the external representation system itself may in fact have significant long term effects which we need to take into consideration.

6. Some Further Issues

We can identify some further pervasive issues that have been attracting a fair amount of interest. We consider these to include:

- Sense Making through Diagrammatic Representations
- The Self Explanation Effect and Diagrams
- Diagrams and Educational Discourse
- Sensori-Motor Experience and Diagrammatic Reasoning

6.1. *Sense making through diagrammatic representations*

Question 7: What activities does a learner engage in when 'sense making'?

Can we describe the process of understanding the task in a more detailed way than hitherto? Can we do the same for the process of verifying (partial) solutions?

Question 8: How can diagrams be used to promote 'sense making'?

What techniques do we have at our disposal? What are the best examples from the research literature which demonstrate the benefits of diagrammatic representations as aids to 'sense making'?

For most of us – especially when we are learning material that is conceptually challenging, sense making takes place throughout the problem solving process. From comprehending the task to finding an acceptable solution, we are involved in making sense of the information. This sometimes leads to radical revision of our understanding of the task, of the representational system selected, or even the way in which the representation is constructed.

Lowe has been investigating the ways in which novices make sense of meteorology diagrams for some time (Lowe 1995). He has noted that the student may often select inappropriate low level details upon which to base an interpretation of the domain. Additionally, patterns of search and inference depend to a great extent on prior knowledge (Lowe 1989). He observes that novice meteorologists tend to work from superficial features and rarely find significant relationships – especially when spread across the weather map. This is entirely consistent with research in the area of the novice/expert differences (see, e.g. (Chi et al. 1981)).

Lowe further reports that, in describing weather maps using a card sort methodology, novices use domain-general visuo-spatial descriptions while experienced meteorologists use domain specific terms (Lowe 1996a). His conclusion is that mental model construction can be seriously hindered if the diagrammatic representation is too abstract.

Schnotz and his colleagues have investigated the role of structural analogies in learning to understand and use diagrammatic representations (Schnotz et al. 1993). As a result of examining the course of problem solving in a series of tasks relating to knowledge about time zones, Schnotz has provided an argument that diagrammatic representations have an important role in the construction of an appropriate mental model. He reports that successful and unsuccessful learners accessed and used information derived from the diagrammatic representation in different ways with successful students using the diagrammatic representation to build a schematic mental model and using the supplied text to elaborate the schema. The reason for the worse performance of the ‘unsuccessful’ learners was that they failed to access necessary new information when it was needed, and utilised the graphic information to a much lesser extent.

Hall, Kibler, Wenger and Truxaw have observed that much of a problem solver’s activity is devoted to reaching an understanding of the problem. They collected written protocols from 85 mathematically competent undergraduates as they solved a range of algebra word story 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. With language, learners may re-write or translate a problem in somewhat abstract terms and may even conceal from themselves their less than complete comprehension (Hall et al. 1989). They observed that problem comprehension and solution are complementary processes and that integrating dual representations of a problem is a key aspect of competence.

They write:

> . . . reasoning about the situation 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).

Thus it may be the interplay between problem descriptions in natural language and in diagrammatic form that assists the individual to understand the nature of the task in some deeper way.

Grossen and Carnine provide evidence for the advantages of students to construct their own diagrammatic solutions (Grossen and Carnine 1990). The process of constructing a diagram is a form of sense making that transforms the student's understanding of the task. From a theoretical perspective, the notion of sense making is not well defined. Sense making is a powerful unifying concept but more needs to be done to clarify the ways in which people function when 'making sense'.

6.2. *The self explanation effect and diagrams*

> **Question 9**: Can the self explanation effect be enhanced by the use of multiple representations?
>
> > How does self explanation function with multiple representations? Can self explanation be supported through the use of diagrams – as found, for example, in a computational environment (Kashihara et al. 1996)?
>
> **Question 10**: What relationships are there between task, prior knowledge and self explanation?
>
> > Can Zhang and Norman's (1994) approach be adapted to model this complex interaction? What other promising approaches are there for the elucidation of this interaction?

The self explanation effect, in its original presentation, is associated with good performance on problem solving through the use of example solutions (Chi et al. 1989). It appears that more successful problem solvers have three special characteristics. Relative to the poorer students, the better students tend to: more frequently explain and justify actions to themselves; monitor their comprehension performance more accurately; and refer back to an example for a specific piece of information. In other words, their (metacognitive) self-communication is frequent, rich and accurate.

Cox has described how the use of diagrammatic representations might facilitate this effect via their effect on the nature of the self-discourse (Cox, in press). He suggests that the low expressivity of graphical external representations may assist, to a greater extent than sentential representations, a reflective, self-explaining student by confronting him/her with the need to consider more than one model of the information in question. Cox further argues that there has not been enough work to determine the mechanisms that might lead to the effect, nor on the ways that diagrams play a part in this process.

Wilkin has studied the interaction of self-generated diagrams in relation to the self explanation effect (Wilkin 1994). Poor learners, defined by a median split of subjects, used self-generated diagrammatic representations in trying to understand motion along a curved path. Her results indicated that the adjacency of diagrammatic features would often lead to error, a result quite consistent with Lowe's detailed analysis of the interaction between domain expertise and accuracy in a study of ER usage (Lowe 1996a). The general thrust of the argument is that diagrams may mislead the novice diagram constructor without additional instructional support.

Zhang and Norman have outlined a theoretical framework for analysing the ways in which internal and external representations interact (Zhang and Norman 1994).[7] Their approach provides one starting point for an investigation of the cognitive mechanisms underlying the self explanation effect.

6.3. *Diagrams and educational discourse*

> **Question 11**: In what ways does the teacher-student discourse affect learning new diagrammatic representation systems?
>
> **Question 12**: In what ways does student-student discourse affect the learning of new diagrammatic representational systems?

In the context of education which we are considering, learning is not a fundamentally solitary activity. Even in the physical absence of a teacher, the student's behaviour is still governed by the nature of the teacher/student relationship. Thus we consider there to be an ongoing discourse between the student, his/her peers and the teacher (or teachers).

The nature of the educational use of diagrammatic representational systems is constrained by the kinds of discourse possible. For example, if we wish to provide tools for students to build new diagrammatic reasoning systems then perhaps we should ask "what kind of educational discourse can be realised through this?"

The argument for encouraging learners to participate in the construction of diagrams is often accompanied by a reluctance (parsimony?) to generate new diagrammatic representational systems. We can find many examples where students are expected to use a specific diagrammatic representational system to solve a problem but few examples where the student participates in the construction of a new diagrammatic representational system. The education system may be intrinsically less tolerant of novel (idiosyncratic?) representation systems being generated by students, perhaps in part because here may be a risk that systems containing semantic inconsistencies or errors could reinforce student misconceptions by perpetuating them. Perhaps, from an early stage in the educational process, students should be encouraged to generate their own diagrammatic conventions or to borrow from existing diagrammatic approaches (Lowe 1996b)?

In addition, teachers do not always fully cooperate in the discourse. There is a sense in which they 'hold back' information according to their goals – and these are not always made explicit to the student. Often, teachers require students to offer explanations 'as if' they (the teachers) were ignorant of the concept – both participants are aware of the game. This 'uncooperativeness' can also be seen in the ways in which word problems are posed (Cox 1996).

In terms of student/student discourse, some work has been done to study the ways in which diagrams are, or can be, used to co-construct meaning. For example, Baker and Lund have examined the ways in which such a discourse involving a diagrammatic representation system can be supported through a structured communicative interface (Baker 1996). However, they have not yet provided an analysis of the support needed in terms of the diagrammatic aspects.

Finally, Cox and colleagues report a controlled investigation of 'vicarious' learning out-comes for students who viewed either a) animated diagrams annotated with a previous student's discourse or b) animated diagrams annotated with student-tutor dialogue. Results indicated that both conditions were effective. The results support provide support for the effectiveness of discourse and dialogue-embedded diagrams as useful learning resources (Cox, McKendree, Tobin, Lee and Mayes, in press).

6.4. *Sensori-motor experience and diagrammatic reasoning*

Question 13: How does sensori-motor processing interact with the use of diagrams? Can systems that stress sensori-motor experience (such as Virtual Reality environments) benefit from integrating them with diagrammatic information?

What are the benefits of doing so for education?

Question 14: In what ways does 'presence' interfere with diagrammatic reasoning?

How can we exploit the notion of presence?

The relationship between perception and reasoning has been discussed many times. Recently, it has become apparent that the rise of Virtual Reality technologies offers tantalising possibilities for research into the area of perception, the use of ERs and education. However, there are well reported cases of people effectively learning to perform manual skills but scant evidence that current immersive virtual reality systems provide any significant benefits for conceptual learning (Whitelock et al. 1996; Brna and Aspin, in press). It has been argued that part of the problem is a failure to provide support for more abstract forms of reasoning (Brna 1998). This support might well take advantage of diagrammatic representations.

Little attention has been given in the Virtual Reality (VR) community to the relationship between sensori-motor experience and conceptual learning – the most common provision for symbolic communication is the availability of a whiteboard. The current VR system builders have spent a great deal of effort in rendering objects in 3D with shading and texture. They have also incorporated 3D audio and various forms of haptic feedback in order to provide VR environments with a strong sense of presence (a measure of the fidelity of sensory cues that engender a sense of "physical presence" or "direct experience"). However, the problem of communicating in diagrammatic or textual symbolic form has not been fully confronted. That is, the new communicative possibilities have concentrated on the sensori-motor aspects. There are signs that some researchers want to include various forms of dialogue with pedagogical VR environments (Rickel and Johnson 1997) but there is very little sign that diagrammatic reasoning has been explored within VR environments. No doubt this will change!

Currently, there is a growing interest in modelling how people use diagrammatic representations both for personal uses and for communication. Educational issues include how students work with physical contexts and theoretical concepts to make sense of situations. Brna and Burton have been developing a model of how this process might take place in a collaborative context but have paid little attention to perceptual issues (Brna and Burton 1997). Shrager, for example, has sought to explain how people using different representational systems might align their understanding of the underlying phenomena (Shrager 1989) though he did not really address non-attentional factors. Current models of learning how to use an unfamiliar computer interface, such as Kitajima and Polson's LICAI, are good at explaining how people exploit the cues given but such models do not directly address the interaction between the learner's perception and conceptual aspects of learning.

Another major effect of the introduction of Virtual Reality has been to raise (again) the issue that motivational aspects of educational environments need to be addressed. Virtual Reality has introduced the notion of presence, the sense of 'being there' in some environment. Although this concept has proved to be elusive to measure, either in subjective or objective terms, there is no doubt that VR environments have a complex effect. Whitelock, Brna and Holland (1996) have sought to develop a framework to study conceptual learning in which presence has a place but the relationship, if any, between presence and the use of external representations has not been explored in any detailed way.

So what does the technology of VR bring to the issue of diagrammatic reasoning? If diagrams are defined as intrinsically 2D then VR would seem to be a distraction. If, on the other hand, diagrams can have any dimensionality then what are the benefits of 'entering' a 3D diagram and moving in and around it? If VR provides the possibility for making very compelling presentations of information (e.g. in diagrammatic form) then how can this be exploited educationally? How can the use of diagrams in an educational context take advantage of the possibilities produced by current/future VR technology? For example, what advantages would a 3D AND/OR tree offer over the 'usual' 2D representation?

7. Conclusion

From an AI perspective, there are undoubtedly many possibilities for exploiting existing symbolic and non-symbolic approaches to modelling the processes involved in learning to think and communicate using diagrams. These models might range from detailed, low-level accounts of sensori-motor experience in VR environments to higher-level accounts of the communicational aspects of diagrams in an educational setting.

Several specific modelling approaches have already been discussed in this paper: one has addressed the issue of how to represent the interplay between internal and external representations (Zhang and Norman 1994) while the others seek to account for exploratory learning (Rieman et al. 1996; Kitajima and Polson, in press). However, more work is needed on the issues of how students learn to comprehend and use multiple ERs and how they learn to translate between ERs. Schnotz and his colleagues have made some progress toward explaining how such learning takes place (Schnotz et al. 1993;

Schnotz and Grzondziel 1996) but has not yet developed a computational model. However, it is likely that simple cognitive models of the complex processes involved will appear within the next few years.

Some consideration has also been given to environments that provide support for people to learn and/or use diagrammatic representational systems. For example, just because a system can do some intelligent reasoning with diagrams doesn't mean it should. The learner may need to 'see' the reasoning being modelled by the system. The provision of multiple representations is increasingly popular but some thought needs to be given to the issues of how to a) encourage the transfer of diagrammatic knowledge between representations and b) assist with the integration of information from the diverse sources.

The way representations are linked might need to be dependent on the learner's state of knowledge. Unfortunately, there are still serious problems for presenting possibly large numbers of multiple alternatives – these can be ameliorated to an extent by providing accessible methods for generating new abstractions but they may benefit the stronger students rather than the weaker ones. Some attention needs to be paid to checking the design of diagrammatic systems intended for educational usage.

We have sketched out a number of issues and some key associated questions that need to be considered. We hope that AI researchers will find some promising challenges to accept.

Notes

[1] Mayer and Sims provide a dual coding explanation of the interaction between text and an animation (Mayer and Sims 1994) while Schnotz and Grzondziel give an explanation for their empirical results which argue that static pictures allow for deeper processing than animated pictures (Schnotz and Grzondziel 1996).

[2] In this paper, an external representation *system* is a coherent set of symbols and relationships which are used to represent the information of interest.

[3] However 'near' or 'far' the transfer. That is, whether the transfer is from training problems to new problems of the same overall type or from one domain in another.

[4] His use of the term media corresponds more to Stenning's use of the term modalities (Stenning and Oberlander 1995).

[5] E.g. Brayshaw's MRE (Brayshaw 1993), a program visualisation environment for PARLOG, provides multiple representations at different levels of granularity of the evolution of various program objects over the course of the program's execution.

[6] Individuals differ considerably in the way they use Hyperproof's graphical abstraction devices (Oberlander, Cox, Monaghan, Stenning and Tobin 1996).

[7] They also support the argument that external representations change the nature of the task.

References

Ainsworth, S., Wood, D. & Bibby, P. (1996). Coordinating Multiple Representations in Computer Based Learning Environments. In Brna, P., Paiva, A. and Self, J.A. (eds.) *Proceedings of the European Conference on Artificial Intelligence in Education*, 336–342. Lisbon: Edicões Colibri.

Baker M. & Lund, K. (1996). Flexibly Structuring the Interaction in a CSCL Environment. In Brna, P., Paiva, A. and Self, J.A. (eds.), *Proceedings of the European Conference on Artificial Intelligence in Education*, 401–407. Lisbon: Edicões Colibri.

Barwise, J. (1993). Heterogeneous Reasoning. In Allwein, G. and Barwise, J. (eds.) *Working Papers on Diagrams and Logic*, 211–234. Indiana: Visual Inference Laboratory, Indiana University.

Barwise, J. & Etchemendy, J. (1994). *Hyperproof.* Cambridge: CSLI Publications, Cambridge University Press.

Brayshaw, M. (1993). MRE: A Flexible and Customisable Program Visualisation Architecture. In Diaper, D. et al. (ed.) *People and Computers VIII.* Cambridge: Cambridge University Press.

Brna, P. (1996). Can't See the Words for the Tree: Interpretation and Graphical Representations. In *Proceedings of the Colloquium on Thinking with Diagrams, Digest No: 96/010*, 1/1–1/3. London: IEE.

Brna, P. (1998). Collaborative Virtual Learning Environments for Concept Learning. *International Journal of Continuing Engineering Education and Life-Long Learning* **8**(2). To appear.

Brna, P. & Aspin, R. (In press). Collaboration in a Virtual World: Support for Conceptual Learning? *Education and Information Technology.*

Brna, P. & Burton, M. (1997). The Computer Modelling of Students Collaborating in Learning About Energy. *Journal of Computer Assisted Learning* **13**(3): 193–204.

Brown, J.S., Collins, A. & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher* **17**: 32–41.

Cheng, P. (1996). Thinking, Expertise and Diagrams that Encode Laws. In *Proceedings of the Colloquium on Thinking with Diagrams, Digest No: 96/010*, 10/1–10/3. London: IEE.

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). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science* **5**: 121–152.

Cox, R. (1996). *Analytical Reasoning with Multiple External Representations.* Unpublished Ph.D. thesis, Department of Artificial Intelligence, University of Edinburgh.

Cox, R. (1997). Representation Interpretation versus Representation Construction: An ILE-based Study Using SwitchERII. In *Proceedings of the 8th World Conference of the Artificial Intelligence on Education Society (AI-ED97)*, 434–441. Amsterdam: IOS.

Cox, R. (In press). The Roles of Externalisation in Reasoning with Self-constructed Representations. *Learning and Instruction.*

Cox, R. & Brna, P. (1995). Analytical Reasoning with External Representations: Supporting the Stages of Selection, Construction and Use. *Journal of AI in Education* **6**(2/3): 239–302.

Cox, R., McKendree, J., Tobin, R., Lee, J. & Mayes, T. (In press). Vicarious Learning from Dialogue and Discourse: A Controlled Comparison. *Instructional Science.*

Cox, R., Stenning, K. & Oberlander, J. (1994). Graphical Effects in Learning Logic: Reasoning, Representation and Individual Differences. In Ram, A. and Eiselt, K. (eds.)

Proceedings of the 16th Annual Conference of the Cognitive Science Society, 188–198. Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L.J. & Snow, R.E. (1977). *Aptitudes and Instructional Methods: A Handbook on Research on Interactions*. New York: Irvington.

Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: The University of Chicago Press.

Grossen, G. & Carnine, D. (1990). Diagramming a Logic Strategy: Effects on Difficult Problem Types and Transfer. *Learning Disability Quarterly* **13**: 168–182.

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.

Kaput, J.J. (1987). Towards a Theory of Symbol Use in Mathematics. In Janvier, C. (ed.) *Problems of Representation in the Teaching and Learning of Mathematics*, 159–196. Hillsdale, NJ: Lawrence Erlbaum.

Kashihara, A., Okabe, M., Hirashima, T. & Jun'ichi, T. (1996). A Self-explanation Assistance with Diagram Tailoring. In Brna, P., Paiva, A. and Self, J.A. (eds.) *Proceedings of the European Conference on Artificial Intelligence in Education*, 122–128. Lisbon: Edicões Colibri.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-integration Model. *Psychological Review* **95**: 163–182.

Kitajima, M. & Polson, P.G. (1996). A Comprehension-based Model of Exploration. In *Human Factors in Computing Systems: CHI'96 Conference Proceedings*, 324–331.

Kitajima, M. & Polson, P.G. (in press). A Comprehension-based Model of Exploration. *Human-Computer Interaction*.

Lowe, R.K. (1989). Search Strategies and Inference in the Exploration of Scientific Diagrams. *Educational Psychology* **9**(1): 27–44.

Lowe, R.K. (1994). Diagram Prediction and Higher Order Structures in Mental Representation. *Research in Science Education* **24**: 208–216.

Lowe, R.K. (1994). Selectivity in Diagrams; Reading Beyond the Lines. *Educational Psychology* **14**: 467–491.

Lowe, R.K. (1995). Supporting Conceptual Change in the Interpretation of Meteorological Diagrams. In Aarnoutse C. et al. (eds.) *Abstracts of the 6th Conference of the European Association for Research on Learning and Instruction*, 89–90, Tilburg, The Netherlands: MesoConsult.

Lowe, R.K. (1996a). Background Knowledge and the Construction of a Situational Representation from a Diagram. *European Journal of Psychology of Education* **11**(4): 377–397.

Lowe, R.K. (1996b). Pictorial Information Design for Schools. *Information Design Journal* **8**: 233–243.

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.

Oberlander, J., Cox, R., Monaghan, P., Stenning, K. & Tobin, R. (1996). Individual Differences in Proof Structures Following Multimodal Logic Teaching. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, 201–206. Hillsdale, NJ: Lawrence Erlbaum.

Rickel, J. & Johnson, W.L. (1997). Mixed-initiative Interaction between Pedagogical Agents and Students in Virtual Environments. In *Proceedings of AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction*.

Rieman, J., Young, R.M. & Howes, A. (1996). A Dual-space Model of Iteratively Deepening Exploratory Learning. *International Journal of Human-Computer Studies* **44**: 743–775.

Rumelhart, D.E. & Norman, D.A. (1978). Accretion, Tuning and Restructuring: Three Modes of Learning. In Cotton, J.W. and Klatzky, R.L., (eds.) *Semantic Factors of Cognition*. New York: Lawrence Erlbaum.

Salomon, G. (1994). *Interaction of Media, Cognition and Learning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schnotz, W. & Grzondziel, H. (1996). Effects of Visualizations on the Structure and Applications of Mental Models. Paper presented at the XXVI International Conference of Psychology in Montreal, August 16th–21st, 1996.

Schnotz, W., Picard, E. & Hron, A. (1993). How Do Successful and Unsuccessful Learners Use Text and Graphics? *Learning and Instruction* **20**(3): 181–199.

Schwarz, B. & Dreyfus, T. (1993). Measuring Integration of Information in Multirepresentational Software. *Interactive Learning Environments* **3**(3): 177–198.

Shrager, J. (1989). Reinterpretation and the Perceptual Microstructure of Conceptual Knowledge: Cognition Considered as a Perceptual Skill. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*.

Stenning, K. & Oberlander, J. (1995). A Cognitive Theory of Graphical and Linguistic Reasoning: Logic and Implementation. *Cognitive Science* **19**(1): 97–140.

Van der Pal, J. (1996). *The Balance of Situated Action and Formal Instruction for Learning Conditional Reasoning*. Chicago: The University of Chicago Press.

Whitelock, D., Brna, P. & Holland, S. (1996). What is the Value of Virtual Reality for Conceptual Learning? Towards a Theoretical Framework. In Brna, P., Paiva, A. and Self, J.A. (eds.) *Proceedings of the European Conference on Artificial Intelligence in Education*, 136–141. Lisbon: Edicões Colibri.

Wilkin, B. (1994). The Self-explanation Effect with Self-generated Diagrams. Technical Report No. CSM-9, School of Education, University of California at Berkeley.

Zhang, J. & Norman, D.A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science* **18**(1): 87–122.

Artificial Intelligence Review **15:** 135–149, 2001.

Thinking with Diagrams in Architectural Design

ELLEN YI-LUEN DO and MARK D. GROSS
Design Machine Group College of Architecture and Urban Planning, University of Washington, Seattle

Received March 8 1999; accepted March 15 1999

Abstract. The paper discusses the use of freehand diagrams in architectural design. It examines the roles of diagrams in various contexts: pedagogical books, design studies, designers' introspective accounts and empirical studies of drawing in design. It offers several examples of thinking with diagrams in design and concludes with a discussion of the requirements for computational support for the diagrams in design thinking.

Keywords: diagram, sketch, architectural design, protocol analysis, computational support

1. Introduction

Diagrams are essential representations for thinking, problem solving, and communication in the design disciplines, in particular those concerned with making physical form: mechanical and civil engineering, graphic design, and architecture and physical planning. In architecture, drawings are the primary form of representation; they carry a design from conception to construction. Except for physical models (which can be considered a kind of three-dimensional drawing) all external design representations in architecture are drawings.

Architectural design is ultimately about the configurations, connections, shape, and orientations of physical forms. Even the most abstract design diagrams are early efforts to explore and resolve spatial layout concerns. Architectural diagrams represent not only physical elements, but also forces and flows (e.g., forces of sun and wind and flows of people and materials). Thus arrows, lines, and other symbolic representations of forces and flows appear in architectural diagrams conveying spatial characteristics such as magnitude and direction. In the early phases of designing, architects draw diagrams and sketches to develop, explore, and communicate ideas and solutions. Design drawing, an iterative and interactive act, involves recording ideas, recognizing functions, and finding new forms and adapting them into the design. Thus drawing is not only a vehicle for communication with others; it helps designers see and understand the forms they work with (Edwards 1979).

Thinking with diagrams in architectural design has much in common with thinking with diagrams in other disciplines. The graphical elements and spatial relations of the diagram map to elements and relations in the domain and the spatial representation of the design offers insights and inferences that would be more difficult to see and work with in other representations. In some respects, however, diagrams function differently in architectural design thinking than in other domains. Architectural diagrams employ a full range of graphical indicators: They use topology, shape, size, position, and direction; whereas diagrams in other domains typically employ only one or two of these characteristics. For example, electronic circuit diagrams use only shape and topology to convey the identity and connections of components; the position, direction, and size of the graphic symbols are irrelevant to the meaning of the diagram.

What most distinguishes architectural design diagrams from diagrams in other domains is that the elements and spatial relations correspond to physical elements and spatial relations in the architectural problem. In other domains to make a diagram you must map the problem domain to a set of diagram elements and spatial relations. For example, in an Euler diagram circles represent sets and graphical overlap of the circles represents set intersections. By contrast, an architectural design problem is essentially spatial; and the diagram is simply at one end of a continuum of graphical representations used throughout the design process. The symbols used to represent elements – walls, rooms, building components – in an architectural diagram are not arbitrary, their shapes and sizes derive directly from the physical elements they represent.

In light of the continuum of graphical representations used in architectural design, it is useful to distinguish a diagram from another form of drawing also used in early design: the freehand sketch. We mean by an architectural diagram a drawing that uses geometric elements to abstractly represent phenomena such as sound, light, heat, wind, and rain; building components such as walls, windows, doors and furniture; and characteristics of human perception and behavior such as sight lines, privacy and movement, as well as territorial boundaries of space or rooms.

A diagram is made of symbols and is about concepts. It is abstract and propositional: its elements and spatial relations can be expressed as a set of statements. It explores, explains, demonstrates, or clarifies relationships among parts of a whole or it illustrates how something works (a sequence of events, movement, or a process). Its symbols may represent objects (e.g., a space or a piece of furniture) or concepts (e.g., service area, a buffer zone, accessibility or noise). For example, an arrow indicates the magnitude and direction of a force; a line indicates the ground without specifying material or exact location. A diagram omits detailed scale or realistic pictorial representations; it indicates spatial relationships only approximately using indefinite shapes. For example, a diagram may represent functional spaces in a floor plan as crude 'bubbles', showing only sizes, adjacencies, containment, and connections.

A sketch, in contrast, is about spatial form. It is executed with a finer resolution that indicates attributes of shape. A sketch often comprises repetitive overtraced lines made to explore precise shape, rather than the intentionally abstract shapes of a diagram, and it uses graphic modifiers such as tone and hatching to convey additional information. For example, a plan or elevation sketch may explore the proportions of a building. A perspective sketch provides three-dimensional information about a scene, specifying the shape of physical elements and visual appearance from some location. Although a sketch falls short of precisely determining positions, dimensions, and shapes, it often provides more detailed information than a diagram.

Architects make many other kinds of drawing: softline (freehand) and hardline (drafted) schematic drawings, working drawings, as well as different projections (plans, sections, elevations, elevation oblique, axonometric). We focus on freehand diagrams. In the following section (2) we examine how architects use diagrams in several contexts. We look at books that teach students to design through drawing; we review recent studies on drawing in architecture; we look at architects' introspective accounts of design process. We then review (section 3) several empirical studies of drawing in architectural design. In section 4 we examine three examples of how diagrams support architectural design thinking. We conclude (section 5) with a brief discussion of some desiderata for computational support for thinking with diagrams in design.

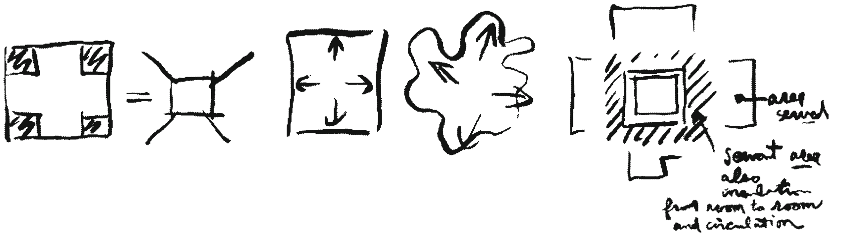
2. Diagrams in Architecture Education and Practice

Many books for architecture students focus on drawing methods and techniques. Lockard's *Design Drawing Experiences* proposes that the ability to "diagram" an architectural context depends on designers' knowledge of issues such as sun, wind, vegetation, traffic, and surroundings. He argues that diagramming can be used to explore variations of design problems and that it allows us to "see, comprehend and respond" to more visual information than we can remember from verbal notes (Lockard 1973). Laseau's *Graphic Thinking* (Laseau 1980) is a guide to making drawings for working out problems and communicating with others. He describes drawing as a means for design development. As an abstraction of an architectural program, the diagram expresses functions, relationships between functions, and the hierarchy of those functions. It is an expression in a graphic language that consists of grammatical rules and vocabulary. Whereas verbal language is sequential, he argues that graphic language is simultaneous: "all symbols and their relationships are considered at the same time."

In professional practice designing often begins with a diagrammatic depiction of the architectural program that is gradually transformed to more complex graphic representations by adding detail. One architect explained designing as a process of transforming and merging diagrams, "trying to take a structural diagram, a functional diagram, and a circulation diagram" and "combine them" (Rowe 1987). Designers often work by making diagrams or transcribing diagrams from their design team colleagues for further development (Graves 1977; Lockard 1977).

Case studies of architects use interviews, observations and works from portfolios to examine the integration of drawing in design practice. Lawson (Lawson 1994) interviewed ten well-known architects to study their methods, concluding that designers "find it hard to think without a pencil in their hand" (p. 141). Fraser and Henmi (Fraser and Henmi 1994) look at how different drawing techniques influence the making of architecture. They describe diagrams as drawings that engage in a "self-conscious reductive process," clarifying a specific interpretation by excluding irrelevant information. They note that architects "symbolize ... intangible factors such as movement, access, sound, view, function, and time ..." (p. 110) in diagrammatic form to represent the abstraction and reduction of information. Herbert (Herbert 1993) examines graphical media in the design processes of six practicing architects. He describes a diagram as an analytic statement, a "composite of graphic marks and written notes." This points out the common practice of mixing text labels and graphics in architectural diagrams. A diagram thus governs and transforms the meanings of verbal statements into a graphic context to help the designer solve problems. Herbert argues that drawings are more than merely a convenient strategy for solving design problems: They are "the designer's principal means of thinking" (p. 1). The designer "must interact with the drawing" (p. 121).

We find diagrams in the sketchbooks of famous architects such as Louis I. Kahn (Brownlee and Long 1991), Le Corbusier (Sekler and Curtis 1978; Guiton 1987), and Peter Eisenman (Eisenman 1987). Kahn in *The Value and Aim in Sketching* (Kahn 1931) mentions that sketches are as important to him as design problems. He argues that "drawing is a mode of representation" and regardless of the medium used, the value of a drawing is in the "purpose" of making. He argues that designers need to interact and work with a sketch, not just "crystallize" thoughts on paper.



(a) (b) (c) (d) (e)

Figure 1. Diagrams with simple geometric shapes for Goldberg House design in Rydal, Pennsylvania (drawn by author after Louis I. Kahn, (Perspecta 1961).

On one hand, Kahn's design diagrams in his proposal for Midtown Philadelphia (Kahn 1953) (Figure 1) show that he has a graphic symbol vocabulary for certain concepts: a plus sign (+) to indicate intersection, an arrow for traffic direction, and a 'greater than' sign (>) for garage. On the other hand, he stresses the importance of shape information in a design. For example, Figure 1(a–e) above shows design diagrams for the Goldberg House in various forms. Kahn uses geometric shapes such as squares, arrows, and lines with textual annotations to explore the spatial arrangement of functional spaces. He explains that Figure 1-a, representing a version of the final design 1-e as four void corners inside a box, is actually (=) a "discovery" from 1-b that the interior (box) can extend out and the diagonals (lines) can be framed (in the endings). Diagrams 1-c and 1-d have the same four directional forces. However, the bounding shape is different (square versus irregular shape) and Kahn has chosen the square shape to develop into the final design 1-e. Figure 1-e shows a configuration of a living room (square) in the middle and buffer zone (hatching indicating "servant area, also insulation from room to room and circulation") and the surrounding rooms ("area served").

Many architects discuss the role of drawing in their design process. For example, Michael Graves explains that the "referential sketch" serves as a "diary" or record of discovery (Graves 1977). It is a "shorthand" notation, a "reference" of an architectural theme recorded to be "used, transformed, ... engaged," elaborated, and combined with other sketches in a later composition.

Christopher Alexander characterizes design as matching program requirements with a corresponding diagram (Alexander 1966). He calls a diagram "any pattern which, by being abstracted from a real situation, conveys the physical influence of certain demands or forces ... " (p. 85). The diagram is the "starting point of synthesis"; the end product is "a tree of diagrams."

The margins of Kevin Lynch's *Image of the City* contain a fascinating collection of tiny diagrams that illustrate ideas about the built environment and cognitive maps described in the text (Lynch 1960). Each diagram comprises only a few lines or symbols, yet the hundred or so diagrams demonstrate the diversity of meaning that a small set of symbols and spatial relations can convey.

These examples all argue that architectural diagramming is necessary for design thinking and the act of making; the shapes drawn influence how architects see and think about design problems. These design diagrams facilitate the designer's reflection, dialogue and self critique and therefore serve the purpose of representing and testing an architect's intent. These anecdotal remarks attest to the widespread use of diagrams in architectural design, and to the belief among designers that diagrams do not merely communicate ideas to others, but that diagrams serve as a primary vehicle for thinking and solving problems. We turn now to look at some of the studies done by design researchers and cognitive scientists that examine what designers do when they draw, and how the visual and verbal modes correlate and interact in design.

3. Empirical Studies of Drawing in Architectural Design

Most empirical studies of design problem solving have been examinations of design protocols. Protocol analysis studies of design problem solving typically collect both verbal and graphical data. Eastman showed that the representations designers use – words and drawings correlate with the problems they find and solve (Eastman 1968). In his study of six subjects performing the task of improving a bathroom layout Eastman documents the design operations used, the objects manipulated and the "control mechanisms" employed.

Akin's (Akin 1986) protocol studies analyzed the "chunking" of architects' design acts while drawing and their shifts in attention. He used the time interval between drawing events to identify how architects group elements in memory. His study revealed several chunks: wall and window segments, steps, furniture elements of similar sizes that have close spatial relations.

Suwa and Tversky videotaped architects sketching to design an art museum (Suwa and Tversky 1996) and asked the architects watching the tape later, to report what they had been thinking (verbal post-design review protocol). Suwa and Tversky studied the relation between concepts as identified by chunks in these protocols and drawing acts. They classified the words mentioned in the verbal protocols into different categories: spaces, things, shapes, views, lights and circulation. They argued that seeing different types of information in sketches drives the refinement of design ideas.

Akin and Lin observed that most protocol studies emphasize recorded verbalizations rather than drawings (Akin and Lin 1995). They discussed symbolic encoding of different modes such as drawing, thinking, examining, and speaking. They asked subjects (1) to reproduce a drawing from a printed transcript, and (2) to predict verbal data from a video tape of the design drawing process with the sound track suppressed. They found that verbal transcripts and drawings were complementary, echoing one other, and that novel design decisions often occurred when the designer was in a "triple mode period": simultaneously drawing, thinking, and examining.

Schön portrays designing as a "reflective conversation with materials." Using protocols of architects sketching to explore possible entrance locations for a library, he argued that design reasoning employs rules (Schön 1988) that derive from previously known types and may be "subjected to test and criticism" by reference to these types. Designers frame a design problem, "set its boundaries, select particular things and relations for attention, and impose on the situation a coherence that guides subsequent moves." Schön uses design sketching protocols to illustrate "reflection-in-action." He argued that designers first 'see' then 'move' design objects. The structure of design is a structure of "seeing-moving-seeing," an alternation of designing (moving) and discovering (seeing). He categorized the kinds of seeing as (1) literal visual apprehension of marks on a page, (2) appreciative judgments of quality, and (3) apprehension of spatial gestalts (Schön 1992; Schön and Wiggins 1992).

Goldschmidt's design protocol studies, like Akin's, examine drawing together with verbalization. She proposes that sketching is a mode of visual thinking and imagery is a conceptual framework for investigation. Sketching, a sequence of design moves and arguments, an "oscillation of arguments" resulting in the gradual transformation of images, is a systematic dialectic between "seeing as" and "seeing that" reasoning modalities. Goldschmidt showed that sketches are not merely representations of images designers already have in mind; rather, the act of sketching is a vehicle for design thinking.

Goel argues that, in sketching, designers employ a different kind of thinking than can be accounted for by the traditional computational theory of mind that is widely held in cognitive science (Goel 1995). Whereas this theory has worked well for building computational models of intelligent behaviors in well defined problem domains such as cryptarithmetic, puzzles, and travel planning, Goel argues that designers work with ill-defined problems, and that sketching exemplifies a kind of representation that can support problem solving in ill-defined domains. He argues that such representations are essentially different from those used in the problem solving domains in which the traditional computational theory of mind has succeeded. By these distinctions, thinking with diagrams (as opposed to sketches) in design would

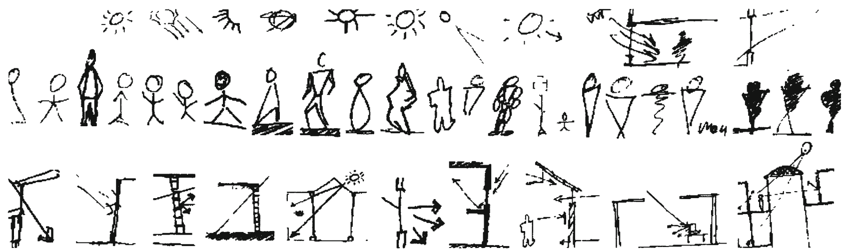


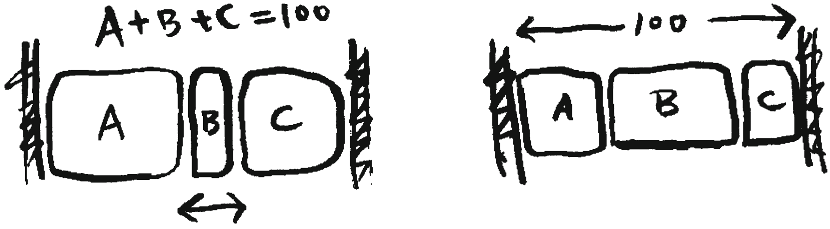
Figure 2. Designers used conventional symbols and configurations for architectural concepts in diagrams.

fall largely (though not necessarily entirely) within the symbol processing domains in which AI has been most successful.

In two studies Do demonstrated that the elements of a graphic vocabulary are associated with specific design concerns (Do 1995). In her first study, which used diagrams and stories from the case based design aid Archie (Zimring, Do, Domeshek and Kolodner 1995), sixty-two architecture students performed four tasks: making diagrams from stories, writing stories from diagrams, pairing diagrams and stories, and commenting on given Archie diagram-story pairs. She found: 1) a small lexicon of symbols were used and arranged in conventional and consistent ways (Figure 2); 2) certain views were used consistently to illustrate certain concerns (e.g., plan views for spatial arrangement and section views for getting light into a building); 3) keywords from stories served as labels in diagrams and vice versa; and 4) most students could read and understand one another's diagrams.

A second study (Do 1997, 1998) verified the conventional use of drawing symbols in design. Architects were given site dimensions and required functions for an office design problem, and they were videotaped designing while an observer took notes. The architects were asked to do a conceptual design and to focus sequentially on four different concerns: spatial layout for zoning different work areas, daylighting, visual access and privacy, and placing a large conference table in the design. The findings from this study of design drawing corroborated the first study in which students were asked to make drawings from stories and vice versa: Architects drew conventional diagrams that correlated with the task at hand.

In sum, the design protocol studies reviewed here, acknowledge that designers use graphic symbols and that the drawing marks they make are linked to verbal protocols and design thinking. These studies discuss several important issues: First, they verify that designers use freehand drawings when thinking about special design concerns; second, they demonstrate that the reasoning is related to design drawings; and third, they suggest that different types of information are embedded in the design drawings. This



(a) (b)

Figure 3. A bubble diagram illustrates dimensions and adjacencies among functions in a floor plan.

indicates that designers do use a lexicon of graphic symbols and configurations when thinking about design and that a design drawing may employ different symbols to represent different types of information.

4. Examples of Using Diagrams in Architectural Design

With this background of anecdotes about the use of diagrams in architectural education and practice, and empirical studies of drawing in design, we turn to some examples of how architects use diagrams in design thinking. We offer three examples of thinking with diagrams in architectural design. The first two examples are similar to the use of diagrams in other domains: the visual representation supports analysis and inference about a problem. The third example appears to be particular to design: the diagram is an abstraction for detail to be filled in later. We argue that the diagram's imprecise quality helps the architect keep in mind its abstract intention.

Figure 3 shows a typical kind of diagram architects sometimes make to consider the layout of functions in a floor plan. Architects use these 'bubble' diagrams to explore relationships among the sizes, adjacencies, and approximate shapes of the spaces needed for various activities. The architect sometimes draws arrows or lines between functions that must communicate, or small tics to indicate an adjacency requirement between two functions, as distinct from pairs of functions that simply happen to be adjacent in the drawing.

A bubble diagram like the ones shown in Figure 3 helps the architect consider possible changes to the design. Each bubble represents the space needed to carry out a function (living, dining, sleeping, etc.) For example, were the architect to enlarge one space, the diagram reveals how the adjacent spaces would need to be correspondingly adjusted to remain adjacent and stay within their own size constraints (e.g., A + B + C = 100). On the other hand, the architect can see when squeezing the diagram would make the dimension

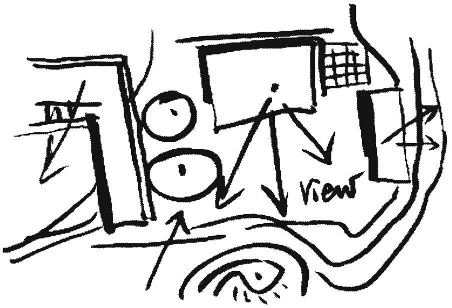


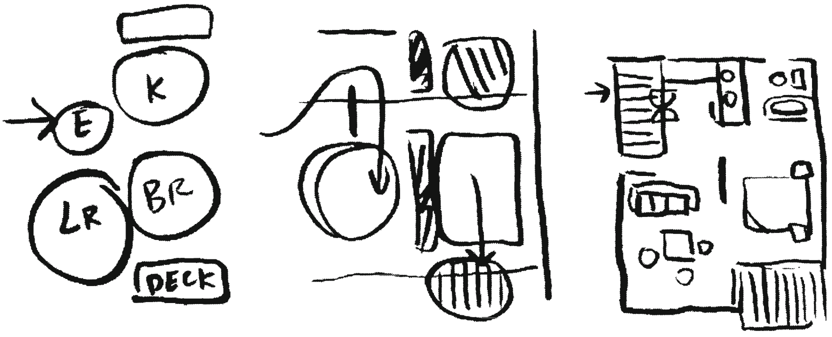
Figure 4. Sight lines to a distant landscape feature display architectural design constraints.

of a space too small for its intended function (when space B is enlarged, space A becomes too small). In short, a bubble diagram helps the architect understand the constraints of a floor plan and the consequences of proposed changes to the design. The diagram makes adjacencies, overlaps, and relative dimensions available by inspection.

Figure 4 shows another use of a diagram in architectural design. Here the architect wants to ensure a view of a landscape feature from a certain place in the building. Sight lines allow the architect to test this predicate. “Is the mountain visible from the living room?” and immediately to see the inference: “If I extend the wing of the house out here, I’ll obstruct the view from the living room”. If the architect retains the sight lines as part of the working design diagram, they can also serve as constraints on the developing design: Don’t put anything in the way.

In these examples, diagrams help the architect keep in mind various constraints that are to apply to the emerging design. The diagram allows the architect to read off visually whether the design in its current state satisfies a certain predicate: an adjacency, a line of sight. In both examples, it would be possible, though tedious, to compute the adjacencies and sight lines geometrically or arithmetically.

The third example, in Figure 5, shows a use of diagrams in architectural thinking that is somewhat different from the first two. Here, the elements of the diagram are abstractions for more specific details that are to be filled in later. The design process is demonstrated here as incremental formalization. For example, the rough diagram in Figure 5-a may later be detailed as shown in Figure 5-b. Alternatively, it may be detailed as in Figure 5-c. The abstract diagram (a) is sufficient to think about design concerns at a certain level; in fact additional detail (b or c) is not especially relevant to decision making and is only likely to distract and perhaps confuse the designer. Thus, an important design skill is matching the level of detail of a diagram to the level of decision making.



(a) (b) (c)

Figure 5. (a) rough diagram; (b) with additional details; (c) with alternative details.

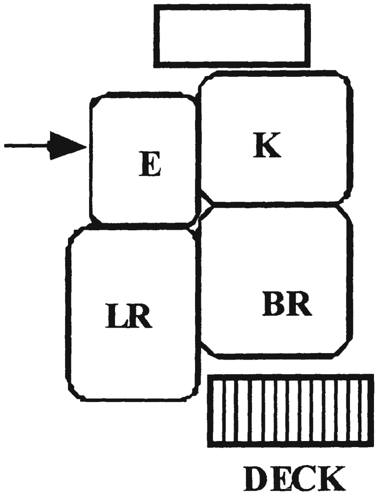


Figure 6. A 'beautified' version of the diagram in Figure 5-a.

The crude character of the freehand diagram in Figure 5-a could be 'beautified' by turning rough and approximate shapes and lines into regular geometric primitives as might be produced with a drawing template or by a structured CAD program (Figure 6). For the purpose of designing, that is, moving from the diagram to a more detailed specification as in 5-a to 5-b or 5-c, this beautified drawing is less useful than the original freehand version.

The rough nature of the freehand diagram serves to remind the architect that it is not a completed work. An inexperienced designer often 'builds the diagram', i.e., proceeds from an initial diagram such as a bubble diagram by simply converting the diagram lines into a floor plan; the seasoned architect elaborates on the diagram, adds detail, and responds to local conditions. The architect will be less tempted to build the diagram if its form is a reminder that additional detail must be worked out, just as a double-spaced typescript invites revision more readily than a tightly formatted page layout of text. The 'beautified' diagram of Figure 6 looks like a floor plan. It is not one. No new information has been added. The architect is likely to keep the geometric shapes and sizes for rooms, the lines for walls, instead of seeing the diagram for what it really is – an abstract plan that must be further refined and specified.

5. Computational Support for Thinking with Diagrams

We come then to the question of computational support. What capabilities are needed in a computer program that would support designers in thinking with diagrams? Based on the above, we believe that the following are desirable components of a system for computational support of thinking with diagrams in architectural design:

- Freehand drawing input, as opposed to structured diagram entry and editing.
- Maintaining spatial relations among elements as the diagram is transformed.
- Recognizing 'emergent' patterns and configurations in a diagram.
- Performing transformations that carry one diagram to another.
- Identifying similarities and differences among diagrams.
- Representing designs at varying levels of abstraction and detail.

These desiderata are represented, to varying degrees, in our prototype system, The Electronic Cocktail Napkin (Gross 1996; Gross and Do 1996). We consider each below.

Freehand drawing input, as opposed to structured diagram entry and editing. Perhaps the strongest argument for freehand drawing is that architects have learned to work that way. They think with a pencil in hand. The kinesthetic act of drawing the diagram seems to help the designer focus on the problems at hand, to consider the relationships among the parts of the design. Although structured diagram entry (choosing elements from a palette or menu) is simpler for the computer program because it avoids the problem of recognition, it also seems to short-circuit the opportunity for the designer to reflect while thinking.

We have observed that the imprecise appearance of a freehand diagram serves to remind the designer that the diagram is not a completed design, and that additional refining and specification must be done to develop the diagram. Therefore diagram beautification and cleanup is not an advantage, at least during the early concept making and manipulation steps of design.

Finally, the designer can draw a diagram faster using a pencil than with a structured editor. Laseau offers time data for making diagrams with a structured drawing program (Laseau 1986): A simple diagram that can be drawn in seconds with a pencil takes several minutes to make with a structured draw program (p. 135).

Recognizing spatial relations among elements and maintaining them as the designer transforms the diagram. The diagrams a designer draws convey elements and spatial relations. The program ought to recognize both the elements and spatial relations in the diagram. It ought to assert the spatial relations on the diagram as constraints, so that as the designer manipulates the diagram further the spatial relations continue to hold.

Recognizing 'emergent' patterns and configurations in a diagram. One feature of diagram representations that seems to aid inference is the ability of the human eye to recognize patterns. These patterns are sometimes called 'emergent' because they are not necessarily drawn intentionally and explicitly, but can be perceived once the diagram is made. A computer system to support thinking with diagrams ought to be able to recognize patterns in the diagrams the designer makes (Edmonds, Moran and Do 1998). The machine ought to be able to recognize not only the patterns the designer draws intentionally, but also emergent patterns that result as unintended side effects.

Performing transformations that carry one diagram to another. A computer system to support diagram reasoning ought to be able to acquire and store graphical transformations of diagrams, which it can then apply on command. A transformation is essentially a graphical production rule, of the sort used in shape grammars (Stiny 1980). The transformation allows the system, upon recognizing the 'left hand side' pattern in a diagram to replace it with the 'right hand side' pattern.

Identifying similarities and differences among diagrams. Thinking with diagrams requires, on occasion, comparing two diagrams and identifying the similarities and differences. For example, the computer system ought to be able to observe that diagram A and diagram B are identical, except that the circles in A are squares in B; that diagram A is a superset of diagram B, containing several additional elements. We can think of several applications of this identification of similarity and difference. One obvious application is to use diagrams as index entries for a database. A similarity measure allows near as well as exact matches to retrieve database items using a diagram index.

Representing designs at varying levels of abstraction and detail. Finally, a computer system that supports diagram reasoning for design should allow the representation of diagrams at various levels of abstraction. As we have observed, diagrams are inherently abstract representations. In architectural design, diagrams are abstract representations of more detailed physical forms. Thus, (and perhaps different from thinking with diagrams in other domains) an architectural design may be represented by diagrams at various levels of abstraction. That is, each element in a diagram at one level of abstraction may be represented by a diagram at a more detailed level down. The computer system should allow for and support these nested levels of abstraction in the diagrams.

Acknowledgments

The authors thank Alan Blackwell for convening the original Thinking with Diagrams workshop, which brought together an interesting collection of people representing a diverse range of disciplines. This research was supported in part by the National Science Foundation under Grant IIS-96-19856 and IIS-00-96138. The views contained in this material are those of the authors and do not necessarily reflect the views of the Natural Science Foundation.

References

Akin, O. (1986). *Psychology of Architectural Design.* Pion: London.

Akin, O. & Lin, C. (1995). Design Protocol Data and Novel Design Decisions. *Design Studies* (April) **16**(2): 211–236.

Alexander, C. (1966). *Notes on the Synthesis of Form.* Harvard University Press: Cambridge.

Brownlee, D.B. & Long, D.G.D. (1991). *Louis I. Kahn: In the Realm of Architecture.* Rizzoli International Publications, Inc.: New York.

Do, E.Y.-L. (1995). What's in a Diagram that a Computer Should Understand. In Tan, M. & Teh, R. (eds.) *CAAD Futures '95: The Global Design Studio, Sixth International Conference on Computer Aided Architectural Design Futures*, 469–482. National University of Singapore: Singapore.

Do, E.Y.-L. (1997). Computability of Design Diagrams – an Empirical Study of Diagram Conventions in Design. In Junge, R. (ed.) *CAAD Futures 97*, 171–176. Kluwer: Munich.

Do, E.Y.-L. (1998). The Right Tool at the Right Time – Investigation of Freehand Drawing as an Interface to Knowledge Based Design Tools. *College of Architecture*, 370. Atlanta, Georgia Institute of Technology.

Eastman, C.M. (1968). On the Analysis of Intuitive Design. In Moore, G.T. (ed.) *Emerging Methods in Environmental Design and Planning*, 21–37. MIT Press: Cambridge.

Edmonds, E., Moran, T. & Do, E. (1998). Interactive Systems for Supporting the Emergence of Concepts and Idea, a CHI 97 Workshop. *SIGCHI Bulletin, a Quarterly Publication of the ACM Special Interest Group on Computer-Human Interaction* **30**(1).

Edwards, B. (1979). *Drawing on the Right Side of the Brain: A Course in Enhancing Creativity and Artistic Confidence*. J.P. Tarcher: Los Angles.

Eisenman, P. (1987). *Peter Eisenman House of Cards*. Oxford University Press: New York.

Fraser, I. & Henmi, R. (1994). *Envisioning Architecture – an Analysis of Drawing*. Van Nostrand Reinhold: New York.

Goel, V. (1995). *Sketches of Thought*. MIT Press: Cambridge MA.

Graves, M. (1977). The Necessity for Drawing: Tangible Speculation. *Architectural Design* **6**(77): 384–394.

Gross, M.D. (1996). The Electronic Cocktail Napkin – Working with Diagrams. *Design Studies* **17**(1): 53–69.

Gross, M.D. & Do, E.Y.-L. (1996). Ambiguous Intentions. *Proceedings, ACM Symposium on User Interface Software and Technology (UIST '96)*, 183–192. Seattle, WA.

Guiton, J. (1987). *The Ideas of Le Corbusier: On Architecture and Urban Planning*. George Braziller: Paris/New York.

Herbert, D.M. (1993). *Architectural Study Drawings*. Van Nostrand Reinhold: New York.

Kahn, L.I. (1931). The Value and Aim in Sketching. *T-Square Club* **1**(6): 19.

Kahn, L.I. (1953). Toward a Plan for Midtown Philadelphia. *The Yale Architectural Journal Perspecta* (2): 10–27.

Laseau, P. (1980). *Graphic Thinking for Architects and Designers*. Van Nostrand Reinhold: New York.

Laseau, P. (1986). *Graphic Problem Solving for Architects and Builders*. Van Nostrand Reinhold: New York.

Lawson, B. (1994). *Design in Mind*. Butterworth: Oxford.

Lockard, W.K. (1973). *Design Drawing Experience*. Pepper Publishing: Tucson.

Lockard, W.K. (1977). *Drawing as a Means to Architecture*. Pepper Publishing: Tucson.

Lynch, K. (1960). *The Image of the City*. MIT Press: Cambridge.

Perspecta (1961). Louis I. Kahn. *The Yale Architectural Journal* **7**: 9–18.

Rowe, P. (1987). *Design Thinking*. MIT Press: Cambridge.

Schön, D.A. (1988). Designing: Rules, Types and Worlds. *Design Studies* (July) **9**(3): 181–190.

Schön, D.A. (1992). Designing as Reflective Conversation with the Materials of a Design Situation. *Knowledge Based Systems* **5**(3).

Schön, D.A. & Wiggins, G. (1992). Kinds of Seeing and Their Functions in Designing. *Design Studies* **13**(2): 135–156.

Sekler, E.F. & Curtis, W. (1978). *Le Corbusier at Work*. Harvard University Press: Cambridge.

Stiny, G. (1980). Introduction to Shape and Shape Grammars. *Environment and Planning B* **7**: 345–351.

Suwa, M. & Tversky, B. (1996). *What Architects See in Their Sketches: A Protocol Analysis*. Workshop Notes: Visual Representation, Reasoning, and Interaction in Design. Artificial Intelligence in Design '96, Stanford University.

Zimring, C., Do, E.Y.-L., Domeshek, E. & Kolodner, J. (1995). Supporting Case-Study Use in Design Education: A Computational Case-Based Design Aid for Architecture. In Mohsen, J. P. (ed.) *Computing in Civil Engineering, A/E/C Systems '95*, 1635–1642. American Society of Civil Engineers. II: Atlanta GA.