A cognitive theory of graphical and linguistic reasoning: logic and implementation

Keith Stenning   Jon Oberlander

Human Communication Research Centre

University of Edinburgh

2 Buccleuch Place

Edinburgh EH8 9LW Scotland

Abstract

We discuss external and internal graphical and linguistic representational systems. We argue that a cognitive theory of peoples' reasoning performance must account for (a) the logical *equivalence* of inferences expressed in graphical and linguistic form; and (b) the implementational *differences* that affect facility of inference. Our theory proposes that graphical representations limit abstraction and thereby aid processibility. We discuss the ideas of *specificity* and *abstraction*, and their cognitive relevance. Empirical support comes from tasks (i) involving and (ii) not involving the manipulation of external graphics. For (i), we take Euler's Circles, provide a novel computational reconstruction, show how it captures abstractions, and contrast it with earlier construals, and with Mental Models' representations. We demonstrate equivalence of the graphical Euler system, and the non-graphical Mental Models system. For (ii), we discuss text comprehension, and the mental performance of syllogisms. By positing an internal system with the same specificity as Euler's Circles we cover the Mental Models data, and generate new empirical predictions. Finally, we consider how the architecture of working memory explains why such specific representations are relatively easy to store.

Acknowledgements The support of the Economic and Social Research Council for the Human Communication Research Centre is gratefully acknowledged. Our initial thinking on the topics discussed here was greatly influenced by the work of John Etchemendy and Jon Barwise. Our research progressed through helpful discussions with colleagues in the Inference Working Group of the HCRC, and it has benefitted from the suggestions of audiences at meetings in Kinloch Rannoch, Edinburgh, and Stanford, and from the constructive advice of our reviewers. Continuation of the work has been supported by three grants: 'SIGNAL', Special Project Grant G9018050, from the Joint Councils Initiative in Cognitive Science and HCI; Collaborative Research Grant 910954, from NATO; and 'GRACE', Basic Research Action P6296, from the CEC ESPRIT Programme.

1 Introduction

Humans can use a variety of external representational systems to perform the same task. The same reasoning task can be performed with *linguistic* representations, such as logical formulae, or with *graphical* representations, such as diagrams. Different representational systems can give rise to different performance characteristics. Humans also use internal representations which may intuitively be differentiated as linguistic or imagistic, and which exhibit different processing characteristics. There is a long history of controversy about how these internal representations can be distinguished (Galton 1883, Pylyshyn 1973, Kosslyn et al. 1979), or indeed whether this is possible even in principle (see Anderson 1978).

In this paper, we argue that a cognitive theory of peoples' reasoning performance is required which can account for two things. First, the fundamental *equivalence* of inferences expressed in graphical and linguistic form; and secondly, the *differences* in facility of inference in the two modes and in heterogeneous combinations. We thus contrast the *logic* of a task with its *implementation*.[1]

This general argument will be advanced with respect to a particular theory of cognitive implementation. The kernel of the theory is that graphical representations such as diagrams limit abstraction and thereby aid processibility. We term this property of graphical systems of representation *specificity*—the demand by a system of representation that information in some class be specified in any interpretable representation. We thus identify specificity as the feature distinguishing graphical and linguistic representations, rather than low level visual properties of graphics. We take specificity to be a general, logically-characterisable property of representational systems, which has direct ramifications for processing efficiency. Our account thus has two virtues. It allows computational specification of the processing differences between differing systems. But also, by detaching the distinctions from low level differences to do with media, it reveals features of natural language discourse which resemble

[1]This parallels Larkin and Simon (1987), who emphasise the distinction between *informational* equivalence and *computational* equivalence. Like them, we depart somewhat from Marr's (1982) terminology. For Marr, the computational level characterises a process in terms of abstract mathematical functions; implementation is a matter of the hardware level. However, relative to a logic, computing with the logic is an implementational issue. We understand computational issues to be less, rather than more, abstract than logical issues, and therefore adopt the latter way of speaking, in which to compute with a logic, we must implement it.

graphical limitations on abstraction. These features will play a similarly important part in maintaining the processibility of natural language discourse.

The paper is structured as follows. In Section 2, we sketch our main working hypotheses. In Section 3, we introduce some of the ideas which underpin our account. The trade-off in processing, between between expressiveness and efficiency, applies to any computational system, human or artificial. We therefore go on to consider two domains which provide suitable empirical tests for the theory. The first is examined in Section 4, where we reconstruct a traditional system for externally supported graphical reasoning, Euler's Circles (ECs), and compare it with another notation for solving syllogisms, Johnson-Laird's (1983) 'Mental Models' system (MMs). In Section 5, we turn to the second domain, and consider the relation between external graphical representations, and internal cognitive structures. There, we discuss how our theory bears on text comprehension, and on the mental performance of syllogism tasks.

2 A general cognitive theory of graphical representations

We can sketch the main points which characterise our working theory in the following way:

1. Graphical representations are one sort of representation which exhibit 'specificity'—they compel specification of classes of information, in contrast to systems that allow arbitrary abstractions.

2. Actual graphical systems permit the expression of some, but not all, abstractions.

3. Together, this means that such representations are relatively easy to process.

4. This specificity helps explain why graphical techniques, such as Euler's Circles, for teaching abstract reasoning are so widespread, and presumably effective.

5. The internal working memory representations we use in some reasoning tasks share with graphical representations this property of specificity.

6. Natural language discourse conventions stay closer to graphics in respect of specificity than do fully abstractive logical languages, in order to preserve processibility.

It is worth observing that our theory is intended to avoid emphasis on the particularly visual properties of graphics. We instead emphasise some *general* logical properties of representations, which have computational ramifications. It is easy to imagine a blind reasoner using embossed Euler's Circles to solve syllogisms.[2] In this paper, we do not discuss point (6), the role of natural language discourse conventions, in any detail; some preliminary remarks are made in Stenning and Oberlander (1991:613–615). Point (4), which relates to processing with external graphical representations, is dealt with in Section 4, and point (5), which relates to processing with internal graphical representations, is dealt with in Section 5. Section 3 lays out some groundwork, by making more precise the idea of specificity and the related notions which underpin points (1) to (3).

3 Specificity and limited abstraction

The first part of our working theory raises a number of questions. First: what does specificity in a representational system actually mean? Secondly, what does it mean to be able to express some, but not all, abstractions? Thirdly, how does this limited expressiveness purchase ease of processing? In answering these questions, we attempt to define specificity more precisely, and therefore make use of some further new terms. In particular, we introduce three types of representational systems, organised by their increasing expressiveness. These are *mimimal abstraction*, *limited abstraction*, and *unlimited abstraction* representational systems. We illustrate them with two simple cases, and then in Section 3.4 indicate their computational significance by comparing them with Levesque's (1988) vivid systems. Such a tripartite hierarchy obviously evokes the Chomsky language hierarchy (cf. Aho and Ullman 1972); we address this parallel, and the cognitive relevance of our proposal, in Section 3.5.

3.1 Minimal abstraction representational systems

The simplest characterisation of specificity can be given in semantic terms. Imagine a represent*ing* world and a represent*ed* world. The former reflects at least some aspects of the latter.

[2]Tactile Venn diagrams have been used with good effect in teaching blind students elementary logic (Goldstein and Moore personal communication).

To characterise a representational system, we must state (i) the represented world; (ii) the representing world; (iii) what aspects of (i) are being modelled; (iv) what aspects of (ii) are doing the modelling; and (v) the correspondences between the two worlds (Palmer 1978:262). Let us require a characterisation ideally to provide an extra component: (vi) a *key*: that part of the mapping from representation to world which has to be made explicit to users of the representation because they do not carry it as part of their general knowledge. A system will then have a set of possible representations, constructible out of basic elements, each of which represents some world as being some way. Rearranging the elements in a particular representation may cause it to correspond to a different possible world.

Now, when a system of representation is a language, either natural or logical, it is relatively straightforward to give a model-theoretic semantics for the system, and for its possible representations. An interpretation function will map representational elements into model elements; differing choices of domain for the model would lead to differing interpretation functions. For example, we could choose to model temporal expressions in natural language using a timeline with a domain of integers, or reals, or whatever. Now, suppose we fix both the domain and the interpretation function; then there is particular question we may ask: how *many* models correspond to a representation? Under the intended interpretation for the language, how many ways are there of making a sentence true?

By contrast, consider a system of representation which is—at least superficially—not like a natural language. Take a graphical system in which a well-formed representation is a fixed arrangement of squares containing a set of solid black circles. The intended interpretation for this system tells us three things. The squares denote the set of offices in my building; the black circles denote researchers; and the relation of spatial containment denotes the relation of working in an office. Just as with a language, we can ask: how *many* models correspond to a particular representation in this system? Under the intended interpretation, how many ways are there of making a graphical representation true?

The basic semantic point here is just this: a *minimal abstraction representational system* (MARS) is one in which there is exactly one model for each representation in the system, under the intended interpretation. We can put this another way via the notion of a relation dimension (cf. Palmer 1978:268). A dimension is a set of mutually exclusive relations, only

Figure 1: Insert the relevant figure around here.

one of which holds for each object or set for which the relation is defined. For example, colour is a unary dimension whose values are properties such as redness; interobject distance is a binary dimension whose values are distances. The current point is thus: take a represented world, choose which relation dimensions the representing world is to capture; a MARS is then one which, for every chosen dimension, must have a single value for every object in the domain.

This semantic characterisation has a syntactic reflex; a representational system which is minimally abstract will embody certain restrictions on its possible representations, which ensure that each representation corresponds to exactly one intended model. The particular manifestation of this syntactic reflex will depend a good deal on the overall form of the representational system. To illustrate this, consider in turn two trivial MARSs: a graphical system of two-dimensional tables, and a linguistic system of restricted predicate calculus.

Two dimensional tables. Consider a system of tabular representation. In the represented world W, there are four objects and five unary property dimensions; each dimension has just two values, which means that an object either has the property, or it doesn't. For the tabular representational system to be minimally abstract, the representing world must always represent each of the objects and dimensions, and must assign each object exactly one value on each dimension. Figure 1 provides a representation which can be interpreted as an element of a minimally abstract tabular system. So: where is the syntactic reflex of the semantic constraint? Our representation contains symbols for objects, properties, 1s and 0s; what is specific about it? The answer is that a well-formed tabular representation has no cells which are not occupied by a 1 or a 0. There are no empty cells (occupied by *Blanks*), and there are no crowded cells (occupied by more than one symbol).

Restricted predicate calculus. A related but rather different syntactic reflex arises when we consider a predicate calculus representation of the same world W. To represent W, we can stipulate that we have a representational language with the following properties.

Figure 2: Insert the relevant figure around here.

We take a first order predicate logic with identity, but without quantifiers and with only negation and conjunction as connectives. We make the unique names assumption, and insist that only one constant denote each element in the domain. Call this language L_0; here, it contains four constants and five predicates. We can say that a sentence of L_0 in conjunctive normal form is *comprehensive* when it contains the minimum number of clauses—here, 20—required to exhaust the combinatorial possibilities of predicate and constant symbols. Figure 2 provides a representation which can be interpreted as an element of a minimally abstract linguistic system based on L_0. In L_0, every sentence corresponds to a single interpretation of its constants and predicates; one and only one sentence is true in any interpretation. The system is minimally abstract because each sentence of the language corresponds to exactly one model. The syntactic reflex here is that well-formed representations have to be comprehensive sentences of L_0. They must have exactle 20 conjuncts, and each combination of predicate and constant symbol must appear once.

We can see that the restriction of a system to minimal abstraction is quite a radical one. For the restricted predicate language, we used only a finite vocabulary in fixed length sentences; we did not use quantifiers and variables, and we did not disjoin comprehensive sentences. Usual logical languages obviously offer these facilities; less obviously, actual uses of tables are rarely as restricted as that exemplified in Figure 1. Let us now therefore turn to a less restrictive class of representational systems, which are related to MARSs.

3.2 Limited abstraction representational systems

Each representation in a MARS under interpretation could represent only one model, only one way for the world to be. Yet real graphical systems surely do not labour under this constraint. Consider again the office-allocation diagram mentioned above. Suppose I wanted to represent the fact that all the offices have two persons in them, apart from one, which has either two or three persons in it. There are two general strategies for enriching diagrams that could be applied here.[3] First I could create *multiple* diagrams: that is, I could produce two alternate diagrams representing the alternatives and place them side-by-side in a complex diagram. Notice that each of the representations gives exactly one value for each object (office) on the relevant dimension (number of occupants). But a representational system which allows multiple diagrams has enriched its expressive power, albeit in a rather simple way. For now, we would say that the complex diagram actually represents two ways the world could be; the single complex diagram represents two models.

We will say that such a system is one type of *limited abstraction representational system* (LARS). This particular type of LARS is such that a complex diagram *abstracts* over several models; the number of its multiple subdiagrams corresponds to the number of models; each subdiagram corresponds to one model. For each type of system, there will be a syntactic reflex for this semantic property. With our tabular system, the reflex will be that we allow juxtaposition of multiple tables, one for each element of the disjunction. With the predicate system, the reflex will be that we allow well-formed formulae to be those which consist of one or more disjuncts, each of which is a comprehensive sentence of the old system.

But we need not adopt multiple diagrams to solve the office-representation problem. A second strategy would be to augment diagrams with *new symbols*. We could introduce a new type of white circle into the squares-and-black-circles representation; one which stands for a worker who might or might not be there. With the new symbol, we can collapse the two diagrams of the multiple method into one. This would contain a set of squares, all but one of which contain only two black circles; the final square containing two black and one white circle.

A system which introduces this type of symbol is another type of LARS. Here, a single diagram corresponds to several models, the number depending on the precise interpretation of the new symbol. In terms of relation dimensions, a LARS of either kind is a system which permits some object to take more than one value on some dimension. The semantic move is, of course, reflected in the syntax of the LARS. With our tabular system, consider introducing a *Blank*, defined as *Either* 1 *or* 0. Now, we can abstract over worlds. Each blank appearing in the diagram doubles the number of cases. In our predicate system, we can introduce a $\diamond$,

[3]There is actually a third strategy, which we turn to when we discuss unlimited abstraction, below.

Figure 3: Insert the relevant figure around here.

so that, for instance $\Diamond Pa \equiv (Pa \vee \neg Pa)$. $\Diamond$ permits disjuncts of this form to be conjuncts in the old comprehensive sentences of our system. Equivalently, we could permit 'partial' sentences, which simply omit such internal disjunctive clauses. Figure 3 illustrates both options. Symbols like these do not permit the expression of dependencies between values in cells of a table (or polarities of clauses in a sentence). Semantically, abstraction is only permitted over models which differ with regard to one object's value on exactly one dimension. So abstraction really is limited, in that little flexibility is allowed in picking out regions of the space of possible models. Using a new symbol to capture abstractions, the number of models abstracted over is exponential in the number of occurrences of the symbol.

Thus, the semantic power introduced by this type of new symbol falls short of that afforded by genuinely 'linguistic' symbols, in the following sense. Only the latter, occurring in a representation, permit the expression of arbitrary dependencies between entities in the represented world. Introducing expressions for arbitrary dependencies corresponds to the third general strategy mentioned in Footnote 3. In the tabular case, for example, we could express the idea that one object's value on a dimension depends on another object's value on another dimension by inserting an equational expression into the appropriate cell of the table. Alternatively, we could place more complex information in the key which is part of the representational system. Our new symbols would be defined here, in terms of the dimensional values to which they correspond. And so too could arbitrary dependencies; these would differ from other parts of the key because they would refer to specific parts of the representation to which they were adjoined. Compare a key entry for a table which stated "Blank anywhere: 1 or 0 in that location" with another entry which said "Blank in column P row b: 1 if $Qc = Sd$, 0 otherwise". Let's call statements of the former type *key terminology*, and of the latter type *key assertions*, loosely following the distinction introduced between terminological and assertional knowledge (cf. Brachman, Fikes and Levesque 1983).

3.3 Unlimited abstraction representational systems

Finally, let us say that a system is an *unlimited abstraction representational system* (UARS) if it expresses dependencies either inside a representation, with equations or whatever, or outside the representation, via key assertions. In itself, this choice of terms is purely stipulative; however, as should emerge below, the processing differences between LARS and UARS are likely to be accounted for in terms of the expressivness of representation and key combined, rather than in terms of representation alone. Hence we choose to say that a LARS is a system which keeps its representations simple, and keeps assertions out of its keys.

The kinds of LARS that are of interest to us, then, are ones which achieve abstraction by using multiple diagrams and key terminology (new symbols). What is limited about multiple diagrams is that we need n diagrams to represent n models. What is limited about diagrams with new symbols is that, for m occurrences of symbol α in a single diagram, we cannot help but represent a number of models exponential in m. We contend that that normal graphical systems are LARSs; much of their usefulness, and their limitations, arise from this property.

3.4 The computational significance of limited abstraction

Our working cognitive theory of graphical representations distinguishes MARSs, LARSs and UARSs on semantic and hence syntactic grounds. But the actual reasons for picking out these classes of representational systems lie in the computational properties which flow from the semantic properties. We would predict that a LARS would be more computationally effective than a UARS, and that this effectiveness would be of use both to human and artificial information processors. To investigate the validity of such a prediction, consider Levesque's (1988) findings. Levesque approached the problem of inferential tractability from a rather different direction. He observed that various well-known metalogical results prevent even first order predicate logic from providing a computationally tractable reasoning system. He then asked: what modifications to or deviations from classical logic "will be necessary to ensure the tractability of reasoning"? His claim is that these deviations are *exactly the same* deviations as are necessary to make logic more psychologically realistic.

Levesque's basic suggestion is that if reasoning tasks are arranged so as to minimise the number of cases to be considered, they can be kept tractable. Requiring a KB to be *vivid* is one way to help minimise cases. A KB is vivid if it is in a certain syntactic form. For sentences of first-order predicate calculus, the KB can only contain (i) ground, function-free atomic sentences; (ii) inequalities between all distinct constants (assumption of unique names); (iii) universally quantified sentences over the domain, which for each predicate and constant express the closed world assumptions; and (iv) the axioms of equality. A vivid KB is consistent and complete; and more importantly, it is tractable, via this theorem of Levesque:

> Theorem 1: Suppose KB is vivid and uses m constants. Let $Q_1, \ldots, Q_n$ be quantifiers, let α be quantifier-free. Then determining if KB $\models Q_1 \cdots Q_n \alpha$ has an $O(m^{n+1} \, |\alpha|)$ algorithm.

The worst cases will be exponential in n, but where n is much less than the length of α, and α is much less than the size of KB, things will be much better; and these are plausible assumptions. Whether or not inference with a vivid KB is tractable will still depend on the algorithm used, but in theory large KBs (of the order of 10^9 sentences) remain tractable.

Like a MARS, a vivid KB cannot represent universals or disjunctions. However, Levesque suggests various extensions to vivid-form KBs which retain its computational characteristics, while increasing expressiveness. Universals are represented by the addition of function-free Horn clauses;[4] disjunctions by a switch to semi-Horn form KBs. The latter encode taxonomies which allow some disjunctions to be re-expressed non-disjunctively, using subsuming predicates. Vividness can also be improved via the use of "observer-centered visually salient properties", which will irresistibly be applied in cases such as Berkeley's triangle.

We agree wholeheartedly with the claim that a primary reason for the appeal of visual information lies in "what it cannot leave unsaid about the observed situation (compared to unrestricted linguistic information)" [p387]. Of course, on neither our view nor Levesque's is this property confined to visual representation. His major point is that tractability is best maintained by minimising the number of cases that must be computed over. His preferred method of case minimisation involves syntactic constraints on representational systems. Our major point so far has been that graphical systems are syntactically constrained; and it turns

[4] Of the form: $\forall x_1 \ldots \forall x_n ((p_1 \wedge \ldots p_k) \rightarrow p_{k+1})$ with $n, k \geq 0$ and p_i atomic.

out that these constraints are very similar to those suggested by Levesque. In itself, this is not surprising, since our claim about limited abstraction is effectively a claim that (i) LARSs can help minimise cases; and (ii) their power to do so is somewhat limited. But the consequence of arriving at a type of representational system which resembles Levesque's is that it too should have computationally desirable properties.

Let us explore the correspondences. An element of a MARS will be of a certain syntactic form, the precise restrictions depending on the particular MARS. In the case of the restricted finite predicate language L_0 discussed earlier, a comprehensive sentence can be regarded as a KB of a special type, actually less expressive than a vivid KB. Inference with respect to this KB will indeed correspond to tractable table-lookup. Of course, all such an inference effectively tells us is the polarity of a given conjunct. LARSs have slightly more complex properties. A system based on L_0 permitting disjunctions of comprehensive sentences will require an upper bound of n look-ups for every query, where n is the maximal number of disjuncts required to cover a set of models. If each look-up gives the same answer, that answer will be returned; otherwise, the lack of an answer will be returned. In principle n for L_0 could be very large, but in practice, the multiple representation technique would not be used when n is large.

A system based on L_0 permitting 'partial sentences', or the new defined symbol $\diamond$, will not always allow the polarity of every conjunct to be found on table look-up, since the answer will not be in the table to be found. However, the lack of an answer can be found on look-up, and the interpretation of any new symbol found by look-up can be determined by consulting the key terminology statements, again by look-up. This type of system will have the same general properties as Levesque's semi-Horn form KBs, since the key terminology is equivalent to the definition of subsuming predicates in a taxonomic component.

When key assertions must be consulted, as when we are dealing with a UARS, the complexity of inference will depend largely upon the syntactic complexity permitted in the assertions. If an L_0-based LARS were supplemented with expressions of unrestricted quantified predicate calculus, inferential complexity would degrade accordingly. However, one can envisage UARS which permit only some syntactically limited key assertions, and thereby maintain a desirable level of tractability. For example, we could require key assertions to contain only universal quantifiers; inference in such a setting should then be tractable.

3.5 The cognitive significance of limited abstraction

There is a sense, then, in which MARS, LARS and UARS form of hierarchy, in which expressiveness and tractability are inversely related. This naturally recalls Chomsky's hierarchy of languages, ranging from type 3 languages (finite-state), through types 2 and 1 (context-free and context-sensitive, respectively) to type 0 languages (recursively enumerable sets). Thus, it is natural to raise two further issues, concerning the relation between the proposed hierarchy and Chomsky's; and the cognitive relevance of such hierarchies.

On the first issue, we have little to say. Chomsky was concerned with systems containing linear sequences of symbols, and we have cast our net somewhat more widely. It is thus not obvious what kinds of correspondences hold; what might constitute a context-free LARS? On the second issue, we would concede that it's obvious that most actual graphical systems function as LARSs. Now, Chomsky's hierarchy has perhaps proved to be of limited use to cognitive science. Most natural languages, after all, are at least type 1, and thus we do not easily locate interesting constraints on processing. By contrast, we would maintain that placing graphical systems at the LA point in the RS hierarchy has significant ramifications for processing. For instance, we indicate in Section 4.2.4 that the LARS version of a particular graphical system is superior to the earlier MARS version, which had been justly criticised on the grounds of its combinatorial inefficiency.

We acknowledge that the computational constraints discussed above are—in a sense—relatively weak, for two reasons. First, such characterisations tend to dwell on worst-cases, which may not be a concern for computational agents which exist in forgiving environments. Secondly, actual performance profiles are only partially determined by the complexity constraints. Even if a representation system has a certain complexity, the choice of a particular representation for a given problem has a considerable impact on its solubility.

Nonetheless, if it is accepted that humans have a set of special purpose reasoning mechanisms (rather than a single general purpose mechanism), then we can show that at least one of these mechanisms performs efficiently precisely in virtue of the limited abstraction permitted by the representations it manipulates. To substantiate our schematic theory, we must develop detailed analyses of actual graphical systems and their cognitive impact. There are two broad lines of enquiry which could provide empirical evidence for the development of the theory.

Study of externally implemented graphical systems can reveal whether their expressive power is that of a LARS. But to show that the logical distinctions between MARS, LARS and UARS have cognitive consequences requires study of human performance and cognitive structure. Evidence may come from tasks involving the manipulation of external graphics, but paradoxically, most of the existing work which addresses our evidential needs actually studies *internal* cognitive structures which arise during the performance of tasks involving no external graphics. We will consider both types of evidence here, in Sections 4 and 5 respectively. Our method is to start from studies of external graphical systems and then to ask how such systems are related to internal computational structures in their users. One implication of our emphasis on general logical and computational properties of graphics is that their tractability for humans arises for the same reasons as their tractability for machines.

4 External graphical representation systems

After briefly reviewing some existing work on graphical communication and exploring desirable properties of domains for testing our theory, we go on to take an example traditional system of graphical reasoning, Euler's Circles, and provide a computational reconstruction of how this system is actually applied in logic teaching. This analysis shows that ECs are LARS when properly interpreted, and reveals some novel properties of the logical fragment which they can be used to reason over. We will then go on to examine another notation for solving syllogisms, Johnson-Laird's 'Mental Models' (MMs; Johnson-Laird and Steedman 1978, Mani and Johnson-Laird 1982, Johnson-Laird 1983, Johnson-Laird and Bara 1984, Johnson-Laird and Byrne 1991). This system was developed partly as a response to earlier work by Erickson (1974) which interpreted ECs as MARS and based a cognitive model of subjects' 'mental' syllogistic reasoning (that is, their reasoning *without* external graphical aids) on this MARS interpretation. We show that MMs are a notational variant of ECs under a LARS interpretation. We close our discussion of Euler's Circles (and of external graphics) by examining the structure of the space of 'registration diagrams' employed in Euler's system.

4.1 Positioning the theory for empirical application

There is a considerable history of work on the cognitive impact of different representations of information. Proposals in the philosophical literature related to our approach go back through Peirce (1977) to Bishop Berkeley (1709). Specificity of graphics is a direct result of their exploitation of homomorphisms which Goodman (1968) placed at the centre of his theory of graphical semantics.

Since the theory hinges on the expressiveness of different representation systems, the most fruitful domains for testing the theory are ones in which there is a need to express some limited range of abstractions (information is not fully determinate). At either end of the dimension of abstraction there is no difficulty in choosing between language and graphics. If, for example, we have total information about the spatial arrangements of a set of objects, then a map (or if we need to read numerical distances, perhaps a matrix of distances) has no real competitor as an information presentation. Twyman (1979) provides an insightful study of such options, and in different ways Tufte (1983) and Mackinlay (1986) both explore methods of presenting determinate statistical information. For us, however, the key material for empirical study is provided by domains in which (i) there is enough determinate information to motivate the use of a graphic; but (ii) there is also a perceived need for the expression of some abstractions which would lead to the need for many-termed disjunctions of MARSs.

There is earlier work in such domains. Gelernter (1963) used diagrammatic representations to control search during geometry theorem proving in an early AI system. Funt (1977, 1980) developed the WHISPER system which employed a spatially organised 'retina' of elements for problem solving. Lindsay (1988) developed a system which exploited the specificity of diagrams, again in the domain of mechanics problems. All of these authors are motivated by observations that graphics aid reasoning and that this is for very general computational reasons. Their emphasis differs from the present theory's in that they see graphics as *eliminating* deduction. This leads them to ignore examples in which graphics are *not* efficacious (tasks which require abstractions which graphics cannot express) and to neglect a comparative approach in which the same information is presented in contrasting modalities.

Larkin and Simon (1987) sought to explain why graphical representations aid reasoning, illustrating their approach again in the domain of mechanics problems. Their paper does adopt a comparative method, translating the same problem into both graphical and sentential systems of representation, and they entertain the possibility that graphics might be bad for some reasoning. However, their approach emphasises differences between token representations, rather than differences of expressive power of the systems the tokens are drawn from. Our formulation in terms of information enforcement *requires* the latter perspective.

Work on intelligent multimedia interfaces (cf. Maybury 1993) and on 'visual languages' within the professional practice of diverse groups (e.g. Petre and Green (1992)) raises questions of cognitive effects of modality choice. But existing systems do not assess alternative presentations of the same information, and visual languages are generally based on semantic network notations. For our initial purposes, they are not ideal, since semantic networks are drawn from the most linguistic end of the dimension of graphical representations; they enforce few specificities.

The most amenable domain we have found for the initial application of the theory is also the one with perhaps the longest history of precise self-conscious use of graphical methods in teaching, namely elementary logic diagrams. At least since Euler (1772) was faced with the problem of teaching a German princess syllogistic reasoning, logicians have used graphical teaching methods based on the analogy between set membership and spatial containment. Venn (1894) modified Euler's method into the one which is in widest contemporary use (cf. Sun-Joo Shin 1991 for a metalogical reconstruction of Venn's system). Peirce (1977) and Lewis Carroll (Dodgson 1896) also worked on graphical methods (Carroll's *Symbolic Logic* is a useful if dated sample). Venn's system is slightly more powerful and probably in commoner use today. Nonetheless, we adopt Euler's as our object of study because we believe that its use of graphics is considerably richer than Venn's and because we believe that this has important perceptual/mnemonic consequences for human performance with the system.

Although we know of no empirical studies comparing teaching elementary logic with and without these graphical aids, (or indeed comparing one graphical system with another) their very persistence is evidence of their usefulness. As we mentioned earlier, there is even anecdotal evidence that these systems are of use to the blind in learning logic (Goldstein and Moore, personal communication); and this itself supports our approach by suggesting that their efficacy stems from general *spatial* characteristics of graphics rather than specifically *visual* properties of human perception. For current purposes we will assume that these graphical systems are useful for at least some didactic purpose and seek an explanation of this fact. When it comes to examining the evidence of human performance for structures in *internal* mental processes which are isomorphic to the external systems, then there is an abundance of empirical evidence which we take up in Section 5.2.

4.2 A graphical algorithm for syllogistic reasoning

We begin by giving a rational reconstruction of Euler's method of using circle diagrams to solve syllogisms. This should be sufficient to test the predictions of our theory that such systems will be LARS but not UARS. We are not primarily concerned with historical exegesis and we will add some notation which does not appear in Euler's (1772) published account. We do believe, however, that this addition merely makes explicit what any logic teacher would interpret Euler to have intended, rather than being a novel system. This point gains certain significance in the light of psychologists' subsequent misinterpretations which will be discussed below (Section 4.2.4). In particular, the reconstruction offered here avoids the combinatorial explosion for which Johnson-Laird and his colleagues have justly criticised other Euler's Circles methods (cf. Johnson-Laird 1983:100,125; Johnson-Laird and Byrne 1991:116–118,201).

Our exposition of this graphical algorithm is divided into three parts: representing premisses; unifying premiss diagrams; and formulating conclusions. It is important to remember that this is a 'competence model' for syllogistic reasoning. It requires the usual sorts of augmentation to serve as a performance model, and we turn to this issue in Section 5.2.

4.2.1 The representation of premisses

There are five topological relations between two circles. These are the Gergonne relations, after the nineteenth century mathematician who made the first attempt to 'formalise' Euler's system (cf. Kneale and Kneale 1962, Faris 1955, and Figure 4). Although syllogistic premisses are mostly modelled by more than one of these five Gergonne relation diagrams, each has a

Figure 4: Insert the relevant figure around here.

Figure 5: Insert the relevant figure around here.

characteristic diagram which is the one which represents the maximum number of types of individual consistent with the premiss. We refer to the models of these diagrams as *maximal models*. It is this characteristic diagram which is used in initial premiss representation. Furthermore, by relating a premiss to its characteristic diagram, we see that within each diagram, there is a sub-area which corresponds to a type of individual which is *established* as existing by the premiss, and there are other areas which correspond to types of individual which are merely *consistent* with the premisses, and which may or may not exist. The area(s) known to be non-empty play a special role in the use of the diagrams. We will say that these regions represent *minimal models* of their premisses. We have suggested elsewhere that this can most easily be brought out by a convention of shading such areas. Here we will use the notation of placing an 'x' in the minimal model regions (Stenning 1989, Figure 5, repeated here as Figure 5). Note that the standard interpretation of the syllogism, which we adopt here, assumes that there are no empty sets.[5]

4.2.2 Registering pairs of diagrams

Diagrams representing the two premisses of a syllogism are registered by making the two middle-term (B) circles in the two premiss diagrams correspond. This sometimes leaves several choices of arrangement of the A and C circles consistent with the premisses. The reg-

[5]The x-marking convention fails to pick out the minimal model in just one diagram—that for *Some A are not B*. It cannot easily be extended to do so because disjunctive x-marking is not possible. However, no untoward conclusions arise from the failure to represent directly this non-emptiness.

istration strategy described here always chooses the arrangement with the maximum number of types of individual consistent with the premisses.

It is useful to consider positive syllogisms and negative syllogisms separately.[6] Positive syllogisms that have valid conclusions have them in virtue of necessary intersection between A and C. Negative syllogisms that have valid conclusions have them in virtue of necessary non-identity between A and C.

Figure 6 illustrates this procedure for the syllogisms which have valid conclusions, and Figure 7 for the syllogisms with no valid conclusions. Figure 8 illustrates it for an interesting group of syllogisms which have no conventionally expressible conclusions, but do neverthless have valid conclusions about the relation between A and C in the domain. These diagrams abstract away from as much linguistic structure as possible. Only the middle term circle is significant, and in each case, the A and C assignments can be reversed. This means that the diagrams abstract over combinations of figure and grammar. 21 diagrams capture 64 syllogisms, and just 8 diagrams capture the 27 syllogisms with conventionally expressible conclusions.

This policy of registration to form diagrams representing maximal models evidently relies on being able to identify logical constraints on circles' placement—on being able to identify whether an arrangement is consistent with the premisses. It might be argued that to assume this ability is to assume the ability to reason syllogistically. We reject this argument. The main problem for human reasoners is calculating implications of *combining* premisses and this problem is not solved by merely being able to assess whether a diagram is consistent with each premiss separately. The role that the graphical representations play is facilitating this process of combination.

It remains to define the fate of xs during the unification of premiss diagrams. If a minimal region marked by an x is sub-divided by the third circle during registration, then the x is removed from the diagram. As an aid to the reader, we have marked such expunged xs with os in the diagrams presented here. Only x-marked minimal regions which persist undivided from premiss diagram into registation diagram remain x-marked. We will call such regions *critical* regions. A critical region corresponds to a maximal type[7] of individual which must

[6] Positive syllogisms have two positive premisses. Negative syllogisms have at least one negative premiss.

[7] Maximal types are types defined for all three properties.

Figure 6: Insert the relevant figure around here.

Figure 7: Insert the relevant figure around here.

exist in any model of the two premisses.[8]

4.2.3 Drawing conclusions

Having specified how diagrams are combined, it remains to describe how conclusions are drawn. It is useful to divide the process into an initial decision *whether* there is a valid conclusion, and a subsequent process of formulating conclusions.

There is a close relation between establishing the (necessary) existence of maximal types and having valid conclusions. All premiss pairs which have valid conclusions establish maximal types. Elsewhere we have called this property of the syllogism *case identifiability* (Stenning and Oaksford 1993). If two premisses warrant a conclusion, it is possible to identify the sort of case which exemplifies the conclusion. This property is what allows graphical methods to be applied to the syllogism without resort to disjunctions of diagrams.[9]

The model theoretic reason for this is not hard to see. Take a region which represents a

[8]We are indebted to Peter Yule for improvements in the formulation of this procedure for deciding whether maximal types are established to exist.

[9]This graphical technique cannot, for example, be applied to disjunctive syllogisms, which constitute a fragment which is not case-identifiable.

Figure 8: Insert the relevant figure around here.

type of individual defined in terms of the two properties of its premiss. Suppose this region is not bisected in the registration diagram which represents the maximum number of types consistent with the premisses. If this is the case, then there is a maximal type established in the final diagram. This is because the x-marked region is wholly included (positive syllogisms) or wholly excluded (negative syllogisms) from the third circle. If it *is* bisected, then there are two maximal types corresponding to it in the final diagram, and either one of these or both may exist, but neither sub-type is necessary.

The converse of case-identifiability holds, except for two registration diagrams which reveal that maximal types are established by pairs of premisses which do not have conventionally expressible conclusions (see Figure 8). It is an arbitrary fact about the quantificational resources Aristotle chose that these conclusions cannot be expressed. Interestingly, they are counterexamples to his principle that two negative premisses never have a valid conclusion. Aristotle presumably excluded these inferences because he did not regard them as involving relations between A and C, and this view is a result of having no distinction between the domain of interpretation and the universal domain; that distinction not clarified until the twentieth century.

The process of formulating conclusions operates directly on critical regions which correspond to established maximal types of individual. Existential conclusions correspond to inferences by conjunction elimination from their descriptions. These inferences are of the form: $\exists x(Ax \wedge Bx \wedge \neg Cx)$ conclude $\exists x(Ax \wedge \neg Cx)$.

This algorithm is simpler than any algorithm for making the strongest valid conclusion. However, it is an empirical fact about human performance under standard instructions that the maximal generalisation is usually made. Although particular conclusions are always safer than universal ones, subjects generally make universal conclusions where they are warranted, and sometimes where they are not. Universal conclusions require that the critical region in a registration diagram be circular and labelled by an end term (A or C). If a critical region is circular, then the label of the circle becomes the subject of a valid universal conclusion. If there is no such circular critical region, there is no valid universal conclusion.

This completes our graphical algorithm for solving syllogisms using Euler's Circles. This algorithm is summarised in Figure 9. We will shortly look at the correspondences and differences

Figure 9: Insert the relevant figure around here.

between this and the Mental Model method.

4.2.4 The EC system as LARS

We now return to our general approach to a cognitive theory of graphical representations. We ask how the current system of graphical reasoning achieves the abstractions required to capture syllogistic logic, and we relate these abstractions to MARSs, LARSs and UARSs.

The pivotal shift from a minimal abstraction interpretation of the diagrams to an abstract one is the shift from interpreting regions as corresponding to types that *do* exist to interpreting them as corresponding to types which *may* exist. Figure 4 shows the abstraction over models which is necessary to express premisses graphically; Figure 5 shows how Euler's reinterpretation allows graphical expression of the abstraction. This shift of interpretation, combined with the subsidiary x-marking convention distinguishing necessary from merely consistent types, is the first prerequisite to achieving a one-to-one mapping between diagrams and premisses. This is because it eliminates the disjunctions of diagrams necessary under the primitive interpretation. The interpretation which psychologists have assumed in the past makes the ECs into MARSs (cf. Erickson 1974, and Ford 1985 for a later defence of this analysis). Our interpretation reflects the actual use of the system by making ECs into LARSs.

In this case, turning a MARS into a LARS is achieved by a change of ontology (from types to possible types) *and* adoption of a definite strategy of diagram choice (represent the maximal model). The resulting compression of diagrams is made usable by the x-marking convention. x-marking plays a role in finding critical regions and therefore in deciding whether there are conclusions, and in formulating them.

The strategy of representing maximal models is what allows all reasoning to proceed with respect to a single diagram. It is a peculiar property of the syllogistic fragment that there is a unique maximal model, and a unique minimal model, for every combination of premisses, and that the minimal model captures all valid inferences. The reason for the latter property is that inferences depend only on the existence of maximal individuals, and never on contingencies between the existences of sets of individuals. These logical properties explain why Euler was able to devise a graphical system for the syllogism. If contingency between maximal types were a determinant of valid inferences, a system more powerful than a LARS would be required.

So the Euler's Circle example substantiates the importance of our distinction between MARS and LARS in understanding how real graphical systems are used. What of the distinction between LARS and UARS? Are there abstractions which this graphical system cannot express? Our analysis of LARS and the upper limits on their expressiveness indicated that the distinction between LARS and UARS remains blurred. Increasingly complex interpretative conventions—which we termed key assertions—mean that graphical representations do less and less work. However, it is clear that the EC system we describe cannot express many abstractions about the domain over which it reasons. For example, only about half the models in this domain correspond to models of EC diagrams (cf. Stenning and Oberlander 1994, p. ??? for a fuller discussion). So, the system cannot implement even the three-predicate fragment of monadic predicate calculus. This system, as it stands, is a LARS rather than an UARS. This observation invites the question whether the EC system could be extended to capture the remaining abstractions in its domain. Sun-Joo Shin (1991) has presented a formalisation of Venn diagrams, and has extended the system to implement the relevant fragment of monadic predicate logic. It is our intuition that this system renounces many of the cognitive advantages of the EC system described here. It seems to do so precisely because its system of 'linking of regions'—introduced to capture contingencies between the existence of types of individual—essentially incorporates a semantic network formalism (cf. Schubert 1976). This comparison clearly warrants empirical investigation.

4.3 The equivalence of Euler's Circles and Mental Models (or $EC = M^2$)

Our aim in considering the correspondence between ECs and MMs is to show what is cognitively important about graphical representations. At one level we will claim that ECs and MMs are equivalent—they implement the same family of theorem provers. But at a more detailed level, ECs exploit the expressive *limitations* imposed by graphical systems in a way that MMs cannot.

In the latter case, there is no natural limit on the notation of two-dimensional arrangements of letters and arcs. We believe this comparison makes it clearer that the importance of MMs is *not* that they are *non-logical* (cf. Johnson-Laird 1983:51) or *non-formal* (cf. Johnson-Laird and Byrne 1991:212), nor that they are a 'model theoretic' rather than a 'proof theoretic' method (cf. Johnson-Laird and Byrne 1991:212–213). ECs provide a graphical computer for syllogistic reasoning. So they *are* a graphical proof theory, however transparent they may make the relation between computation and the space of models. Rather, the important questions posed by both ECs and MMs concern how they are implemented in memory.

Comparison with MM methods is revealing for a cognitive theory of graphics applied to syllogistic reasoning, for three reasons. First, MM theory has been responsible for revealing important empirical observations of subjects' reasoning. Secondly, the theory has also made sweeping claims about the nature of mental representation based on those observations. For example, Johnson-Laird (1983:51) maintains that the theory "solves at a stroke the problems of which particular rules of inference are in the mind, how they are mentally represented, and how children acquire them. These questions simply do not arise, because logic is banished from the mind". More recently, Johnson-Laird and Byrne (1991:215) conclude that methods using MMs provide "the mainspring of human reasoning". Lastly, it has been argued that MMs are distinct from graphical methods for the syllogism. For example, Johnson-Laird specifically excepts his own theory from the generalisation that "all current psychological theories of the syllogism turn out to be variations" on Euler's Circles and Venn diagrams (Johnson-Laird 1983:77). MMs are distinguished from graphical methods on the grounds that they do not suffer the combinatorial explosion which "embarrasses the theories based on Euler circles" (Johnson-Laird 1983:100; cf. also Johnson-Laird and Byrne 1991:116–118).

In comparing ECs to MMs it is important to bear in mind that the latter can be considered either as (i) another externally represented reasoning system or as (ii) a theory about an internally implemented cognitive system bearing some relation to the written theory. We here adopt the former stance in order to compare ECs and MMs as external 'paper-and-pencil' aids to reasoning. In Section 5.2, we consider the empirical evidence about correspondence between MMs and internal mental structures and processes.

One further distinction must also be born in mind in making this comparison. The EC system described here is a normative implementation of syllogistic reasoning corresponding to a normative use of mental models notation. Mental models theory has been extended as a performance theory to explain subjects' errors. It is not difficult to see how to do this with ECs (primarily by specifying sub-optimal registration strategies) but we have not yet done this in our exposition.

ECs, MMs and the relevant fragment of the monadic predicate calculus are evidently equivalent at the logical level—they pick out the same consequence relation. The equivalence that concerns us here is at the level of the theorem provers which are implicit in ECs and MMs, and which we made explicit for ECs in the last section. Both systems operate by: representing all and only maximal types of individual which are consistent with the premisses; by identifying which maximal types of individual are established by the premisses; and then generalising from these types.[10] We therefore proceed by laying out the representational correspondences between the systems and then the correspondences between their proof strategies.

The particular mental model system we use here for comparison is that of Johnson-Laird and Steedman (1978). Other variants exist (for example, Johnson-Laird and Bara 1984, Johnson-Laird and Byrne 1991) but the details of those systems could be reconstructed within the EC framework adopted here. The main complexity involved in establishing the equivalence of any of these systems and ECs lies in MM's treatment of extra notational devices—parentheses and negative links in MMs—and in the procedural elements of the strategies for MMs' use. These annotations and procedures serve to allow one diagram to abstract over several possible states of affairs. We begin by examining the central structural correspondences, and then look at the details which differ within each family.

Columns of letters in MMs and minimal sub-regions of EC diagrams both represent types of individual. Monadic predicate calculus would represent these types by conjunctions of atomic predicates (or their negations), each predicated of the same variable. To illustrate with an example syllogism, Figure 10 shows the development of EC and MM representations for the syllogism *Some A are B, All B are C*.

[10]This is not a general feature of theorem provers. For example, it is commonplace in natural deduction based theorem provers to assume the existence of indviduals inconsistent with the premisses and then proceed by *reductio ad absurdum*.

Figure 10: Insert the relevant figure around here.

At every stage of development, x-marked regions of ECs correspond to columns of letters with no parenthesised elements in MMs. In both notations they represent types of individual whose existence is established by the premisses. Non-x-marked regions and columns with some parenthesised element(s) represent individuals which are consistent with the premisses but not established by them. Note that no region in the EC, nor any column or part of a column in the MM, represents either of the types $A \wedge B \wedge \neg C$ nor $\neg A \wedge B \wedge \neg C$. These are the only types inconsistent with the premisses. The only type consistent with the premisses but unrepresented is the wholly negative type $\neg A \wedge \neg B \wedge \neg C$. Neither ECs nor MMs represent this type of individual, since it plays no role in any inference expressible in the syllogism.

MMs include elements in their representations of first premisses which allow the additions of second premisses to yield any maximal type consistent with both premisses. ECs have a simple and consistent policy of representation which is motivated by the underlying model theory at every point. Registration represents all consistent types. x-marking represents types entailed by the single premisses represented. x-marked regions take their significance directly from the model theory. In contrast, many features of mental model notation are quite arbitrary. Only subsets of consistent maximal types get represented in some syllogisms, but this omission of types plays no part in the system of reasoning. An example of MMs' idiosyncrasies is illustrated in Figure 11. It might at first seem that we can interpret the parentheses in MMs as standing for either a predicate or its negation. However, if this were so, the parenthesised 'a' and 'b' in the representation of *Some A are B* could potentially come to represent individuals which are $A \wedge B$. In fact, the parenthesised 'a' is never completed in this way—it can only come to represent either an A that is $\neg B$, or a B that is $\neg A$. So the parentheses within a column are not to be interpreted independently of each other. This fact could be captured by including a negative link between these two parenthesised elements, as we have done in dotted lines.[11] For example, the first option (with regard to the parenthesised 'a') disappears when

[11] Johnson-Laird claims that universality is represented within mental models by repetition of copies of

Figure 11: Insert the relevant figure around here.

the second premiss forces the parenthesised 'b' to represent a B that is C. This omission of the negative link is particularly arbitrary since the corresponding positive link *is* included in the representation of the negative particular premiss. There, however, the 'b' is arbitrarily not parenthesised (we have inserted square parentheses). If our square additions are included in the MM notation, columns with parenthesised elements then correspond to unmarked EC regions, and unparenthesised columns to x-marked ones.

This example is sufficient to show that mental models adopt arbitrary choices of representation strategy, whereas the Euler's Circle system described here has a motivated policy. What is arbitrariness from a logical point of view might be motivated from a psychological point of view. But in fact it turns out that these particular notational features play no role in capturing empirical observations in the data. Whether subjects take representational short cuts in mentally manipulating ECs is an important psychological question; we return to it in Section 5.2 when we discuss how our EC algorithm might be mentally implemented.

In summary, ECs and MMs should be seen as two families of notations which can each be used to formulate a range of theorem provers. What these theorem provers have in common is that they (i) agglomerate the representation of both premisses into one representation which (ii) represents all individuals consistent with the premisses and (iii) distinguishes those which *must* exist from those which *may* exist. They then specify procedures for reading conclusions off from these representations.

Seeing ECs and MMs as notational variants raises the question: Is one notation more con-

individuals (an analogical representation, he claims). But this sits oddly with the observation that the number of individuals of a type that exist is never relevant to consequence in the monadic predicate calculus, a logic without identity. In fact, closer observation of mental models reveals that the function of repetition at the stage of representing the first premiss is always to allow the representation of any type of individual that may be required when the second premiss is added. Universality, in fact, is captured by ensuring, either explicitly or implicitly, the representation of all consistent types.

strained than the other? It is fairly easy to say what constitutes a natural extension of ECs, though even here, it may be hard to state exactly which annotations are natural. When it comes to MMs, it seems to us that there is no inherent reason why systems of letters, with and without parentheses, linked by arcs and distributed in two dimensions should have any of the limitations that graphical representations exhibit. In fact, it is known that partitioned semantic networks are equivalent in expressive power to polyadic predicate calculus (Hendrix 1979), and that they can be extended to capture the full lambda-calulus (Schubert 1976).

So, although the ECs method described here is equivalent to a theorem prover expressed in MMs notation, there is a great difference in the degree of constraint on the theorem provers expressible in the two notations. The graphical nature of ECs determines that all types represented by a registration diagram must be maximal types. That is because every point in a plane is either inside or outside of every circle drawn in that plane. This is a special case of specificity, and it is precisely this property which captures the human tendency to reason over agglomerative representations (cf. Stenning 1991) composed of maximal types of individuals. For the purposes of formulating psychological theories of reasoning, the more limited the expressive power of a system which can fit the data, the more the notation is contributing to an understanding of the phenomena described.

4.4 Animation

ECs are distinctive among graphical systems based on the analogy of spatial containment to set membership in exploiting movements and constraints on movement in reasoning. We chose to describe an EC algorithm rather than a Venn diagram one because of this exploitation of movement which we refer to as *animation*. Static diagrams are related to each other by movements of circles (and changes in size) and ECs thereby exploit a particular concept of continuity. We believe that this is one of the ways in which they facilitate reasoning and it provides another illustration of specificity, this time in the temporal dimension.

Introducing movement into graphical representations introduces temporal specificity. Time is a dimension, and so representations which employ sequences of states to represent a dimension (whether a temporal dimension or not) have to determine a complete ordering of represented states. An animation sequence is specific with regard to temporal relations just as a diagram

Figure 12: Insert the relevant figure around here.

is with regard to spatial relations. Interpretation conventions may be able to cancel some of these temporal specificities and achieve some abstractions representing partial orderings of states, but not just any abstraction can be expressed.

In the case of ECs, spatial and temporal specificities combine in a very elegant way. Each static diagram represents one of the 128 different models of the syllogism. Moving a circle typically leads to *one* type of individual at a time either being added to, or deleted from the model represented before the movement. In some cases a movement might introduce or delete two individuals but such double events can always be turned into pairs of single ones by changing the angle of movement infinitessimally. This means that the 64 models which have EC diagrams are related in a seven dimensional space, with one dimension for each individual type. Movements of circles in the EC diagram correspond to sequences of transitions from corner to *adjacent* corner of this seven-dimensional hypercube. There are no catastrophes—no cases in which a minimal movement brings about a distant model. Figure 12 portrays the space of three-circle diagrams; the restriction to circles in Euler diagrams is an essential element in establishing this continuity.

Representations that are continuous in the sense used here are computationally tractable for quite general reasons. Roughly, from any point in the structure, one can take a minimal path to any other point on the basis of information in its address. What evidence is there—other than this most general sort of computational ground—for believing that human beings have *internal* reasoning mechanisms which exploit this sort of continuity? Hinton (1979, 1980) argues that it is this type of continuity of structural description which underlies our ability to solve 'visualisation' problems. He further argues that this sort of continuity is sufficient to explain 'mental rotation' phenomena which are usually assumed to require representations which are analogue in a much stronger sense.

Although we know of no explicit experimental study of the role of continuity in the manipulation of such structured spaces in working memory, it seems plausible that our transactions with the mechanical world are underpinned by a mechanism exploiting such continuity. The logical constraints provided by syllogistic premisses can be modelled in the EC representations by the mechanical constraints on the movement of discs in a plane. Imagine driving a nail through the critical region(s) of registration diagrams. For positive syllogisms, this nail constrains the A and C circles from sliding apart. For negative syllogisms, the nail constrains the A and C circles from being superimposed (having been adjusted to have equal size).

So the graphical nature of ECs provides not only the spatial specificity which forces the representation of maximal types of individual. It also provides the spatio-temporal specificity which supplies a mechanism for navigating around the space of models in a continuous fashion (notice that this graphical property does not hold of MMs). The efficacy of this mechanism should not be underestimated. Rather than considering an unstructured set of 128 models, the algorithm positions its user at a particular point in this space corresponding to the registration diagram of the syllogism to be solved. It then helps identify every immediate neighbour of that model, and every immediate neigbour of any subsequent construction.

We return to the role of animation in internal representations when we consider how syllogistic reasoning might be implemented in human working memory.

5 From external graphics to internal imagery

We have so far been concerned with characterising the properties of graphical representations in general and, in particular, ECs and the procedures that manipulate and interpret them so that they implement syllogistic logic. Psychological considerations have only entered through comparison with MMs which are explicitly embedded in a psychological theory of verbal reasoning. But our general approach has cognitive pretensions—we aim to explain differences between peoples' facility in reasoning with graphics, with language and with calculi. So even if we were to restrict ourselves to cases in which external graphics are used, we would still require a theory of how the information they carry is represented and processed internally. In fact, we believe that our approach—through the computational nature of external graphical systems—promises a new perspective on the nature of internal representations used when reasoning without explicit external graphics. However, this does not commit us to the view that the relation between internal and external representations is direct. We would expect the internal representations to show similar specificities to successful external aids, but internal implementations might be expected to differ significantly from external ones.

Our theory can approach the task of analysing internal representations at two levels. By showing that the logic of graphical representations is more computationally tractable than more general logics (as we indicated in Section 3.4), we show that processing the information graphical representations convey is easier than processing some more general class of information. This explains why graphical representations are easier to process *for any reasoning system*. Such an 'architecture-free' approach must be supplemented by empirical observations of human information processing which show that the specific representations used in a task are in fact more easily processed than abstractive ones. Only empirical research can support the view that representations with the characteristics described by the formal theory are actually employed. Theories of the architecture of working memory built on such observations should then be able to explain *how* graphics are processed more easily.

Here we review some of the evidence that specific representations are generally more easily processed by human beings. We first discuss results from studies of text comprehension, and then return to the syllogism and examine the relevance of the behavioural data on 'mental' syllogism solution. Fnally, we consider a theory of working memory which can explain why it is easier to hold and manipulate data-structures which are limited in their powers of abstraction.

5.1 Text comprehension and verbal reasoning

There is already extensive experimental evidence that can be construed along the lines of our theory. The text processing and verbal reasoning literatures deal with the presentation of potentially abstract information and adduce evidence that people derive more concrete representations for it. We can review only a few example findings here. They can be classified by the type of information which creates specificity. The first group study the specificity resulting from graphical representations of spatial relations; the second studies specificity resulting from the unique name axioms implicit in graphical representation systems.

A number of studies have compared the processing of two types of texts. One type continuously determine the spatial relations of a described array of objects; the other type leave spatial relations indeterminate for at least some stretch of text (see for example, Mani and Johnson-Laird 1982, Clark 1969a, McGonigle and Chalmers 1986). The following examples from Mani and Johnson-Laird illustrate indeterminacy which remains at the end of the text.

> **Determinate:** The spoon is behind the knife. The knife is to the right of the plate. The fork is to the left of the *plate*.

> **Indeterminate:** The spoon is behind the knife. The knife is to the right of the plate. The fork is to the left of the *knife*.

In the latter case, the last two sentences only determine a partial ordering of knife, plate and fork. Indeterminate texts prove much harder to process in tasks where subjects have to test the whole range of their models, but not where one will suffice. Both Mani and Johnson-Laird's findings and McGonigle and Chalmers' qualifications are interpretable in terms of subjects having a high-capacity durable memory for representations exhibiting specificity. This memory cannot be employed for indeterminate material unless the addition of contingent information is permitted. These experiments are examples of a large literature on verbal reasoning which bears out the problems caused by indeterminacy. Many of these experiments do not involve spatial relations overtly, but do involve transitive reasoning about dimensions which therefore present the same logical situation (for example, Clark 1969b).

There has been less research explicitly investigating the processing of texts in which reference (rather than spatial relations) is indeterminate giving rise to violations of unique name axioms. Bransford and Johnson's (1972) classic experiment demonstrating the incomprehensibility of a 'trick' prose passage is actually an example, though the authors do not discuss it in those terms. Close examination of the passage used reveals a number of sources of confusion. No syntactically indefinite introductions are made and definite phrases' referents are untraceable. Because the definite noun phrases referring back to earlier mentioned elements are chosen to be abstract, there are multiple possible antecedents for many of them. The picture which the authors used to make the passage comprehensible enables the reader to make unique assignments and thus facilitates processing.

> If **the balloons** popped **the sound** wouldn't be able to carry since **everything** would be too far away from **the correct floor**. **A closed window** would also prevent **the sound** from carrying since most buildings tend to be well insulated. Since **the whole operation** depends on a steady flow of electricity, a break in **the middle of the wire** would also cause problems. Of course **the fellow** could shout but the human voice is not loud enough to carry **that far**. An additional problem is that a string could break on **the instrument**. Then there would be no accompaniment to **the message**. It is clear that the best situation would involve less distance. Then there would be fewer potential problems. With face to face contact the least number of things could go wrong. [Our emboldening marks irretrievable references].

It is not the pictorial nature of the context which is essential. The picture can be replaced by a textual preamble clearly introducing the emboldened elements into the domain in ways which permit the resolution of the later abstractions; the text is rendered equally comprehensible.

> **Context:** *A man* with *a guitar* is serenading *a girl* at *a fifth floor window*. He has *a microphone* connected by *a wire* to *a speaker* suspended at the level of her window by *a bunch of hydrogen balloons* ... [Italics mark indefinite introductions].

Both the provisions of context amount to introducing compliance with the unique name constraints we have mentioned in connexion with minimal abstraction representational systems (MARSs), and their relatives.

A more precisely controlled experiment aimed directly at studying temporary indeterminacy of reference is reported in Stenning (1986). Subjects read texts describing domains which they knew to have just two elements. For example:

> There is a small square. There is a black square. There is a small black thing. There is a small white thing. There is a small circle.

The pattern of identities is not determined until the third sentence in this example. When there was indeterminacy, it caused disruption of processing, particularly at the point where indeterminacy was resolved. This was interpreted as being due to the delayed construction of specific representations. These and other results have been interpreted as demonstrating the existence of a 'non-propositional' memory (cf. Johnson-Laird 1983, Garnham 1987).

A complementary observation that has also been invoked as evidence against propositional representations is that readers cannot remember the surface segmentation of propositions in texts. The observation was made in general terms by Bartlett (1932) and by Bransford, Barclay and Franks (1972) but has recently been used by Garnham (1987) to argue against any 'propositional' account. Garnham's examples are of subjects' failures to discriminate whether they saw *The man with the martini is tall* or *The man standing by the window is tall* when they had been told that the man standing by the window is the man with the martini.

If these texts were represented in a MARS, then this is exactly the result one would expect. The distinction between which properties are identifying and which merely attributed is therefore lost in translation. But both these forms become equivalent in a MARS because both result in representations in which there is one man and one martini, and one window by which he is standing. These inferences are not licenced by the logical form of the isolated sentences but by the discourse interpretation conventions. But that is hardly an argument against the resulting representation being sentential.

5.2 Syllogism data

What light can our theory shed on the empirical literature on human syllogistic reasoning carried out without paper-and-pencil support? Since we have shown that ECs are equivalent to MMs, an account based on ECs will inherit explanations of results explained by MMs up to this equivalence. So, for example, the most important predictor of syllogism difficulty in MM theory is the distinction between one-, two- and three-model problems. In the highly procedural MM system, this distinction is defined by the number of loops of constructing and testing involved in the solution of a given problem. In the graphical algorithm, this property can be defined in terms of the number of possible arrangements of the circles in the registration diagram when the maximal-areas constraint is relaxed. Ardin (1991) showed that most of the variance of problem difficulty is actually captured by the distinction between single model problems and those requiring greater numbers of models. In EC terms, this is just the question whether there is *any* choice of registrations of the two premiss diagrams.

But we need to consider some additional notations (such as directional links) which are used in MM theory but have not yet been replicated in our EC notation, and the role these notations play in capturing empirically observed phenemona. We should also ask whether there are novel predictions suggested by Euler's notation; an important test of an alternative notation is its ability to reveal new generalisations in old empirical data or to suggest new data that should be relevant. We therefore turn next to the role of additional notation in MMs in capturing empirical observations, before reviewing new observations arising from the EC notation.

5.2.1 Notation in MMs not duplicated in the EC system

One feature of the MM notation which plays an important role in capturing psychological data is the directionality of the links between predicate letters within an individual. One of the most original empirical observations which Johnson-Laird and Steedman (1978) made is that the formulation of conclusions is strongly affected by the grammatical organisation of the premisses. In MM notation, grammatical status is marked by the directionality of both positive and negative links, and this effect is captured by the procedures for formulation of candidate conclusions. The procedures prefer to read conclusions in the direction of arrows. So, where 'a' is linked to 'c' through 'b' by two arrows pointing in the same direction, there is a preference for the conclusion which aligns with the arrows.

ECs are not usually used with any grammatical notation. This is probably because of their didactic origin—to learn syllogistic logic is to learn that some grammatical differences are logically immaterial; for example, those in positive particular and negative universal premisses. To learn this is a substantial accomplishment because of the natural tendencies noted in the Figural Effect. These tendencies in turn stem from the functions of subject and predicate in natural language. But if ECs are to be part of a descriptive theory of mental process, annotating subject and predicate is just as natural as it is in MMs. Suffixing the label on each circle by its grammatical category and allowing two matching suffixes to influence the generation of candidate conclusions is equivalent to the directionality of arrows in MMs. Once the annotation is added to ECs, it can play the same role in controlling the formation of conclusions and the testing of their generality: ECs require an interface to the language of conclusions just as much as MMs.

5.2.2 Novel empirical analyses and predictions of the EC notation

Our EC notation revealed the case-identifiability of the syllogistic fragment. This novel property in turn revealed the U-conclusions missed by the Aristotelian proof-theoretic apparatus. This logical curiosity in turn suggests an empirical prediction—that subjects should be capable of drawing the U-conclusions by substantially the same mechanisms they employ for the range of other conclusions. Yule (1991) and Yule and Stenning (1992) test this prediction directly. Their subjects describe maximal types of individuals entailed by syllogistic premisses rather than drawing conventional conclusions. Their results demonstrate that subjects can describe the critical individuals whose existence is entailed by pairs of premisses which have no conventional conclusion. Subjects do not even find these maximal types the hardest ones to discover. This result strongly suggests that the task of identifying maximal types established by premisses is naturally adopted by subjects. We have argued elsewhere that this is because this task is more naturally assimilated to subjects' discourse comprehension strategies than the conventional syllogistic task (cf. Stenning and Oaksford in press). This methodology conforms with our principle that the best evidence for the use of 'graphical' representations is to supply information abstract with regard to an image (syllogistic premisses defining individuals in terms of only two properties) and observe subjects' abilities at constructing maximally specified representations (specifications of maximal individuals).

The graphical approach also reveals some novel formulations of the effects of figure. MM theory explains figural effects in terms of a first-in first-out (FIFO) memory. Our graphical algorithm suggests a more general explanation in terms of the logical, rather than sequential, properties of syllogisms. This novel formulation (cf. Yule and Stenning 1992) focusses on the part played by critical regions in determining conclusions. Where there is a valid conclusion, there is a critical region. Since critical regions are defined as x-marked regions from the premisses *not bisected* during registration, it can be determined which premiss 'contributed' the critical region. In some cases both premisses will contribute the critical region.

Subjects can therefore organise their reasoning around identification of critical regions. If subjects begin organising their conclusion in terms of an x-marked region from the first premiss, in some syllogisms they have to shift attention to an x-marked area from the second premiss in order to find the critical region. Yule and Stenning (1992) have shown that its predictions are born out in subjects' orders of descriptions of critical individuals. The sequences observed are incompatible with the FIFO explanation.

Another difference between the EC method we describe and the MM system is that our algorithm is linear and avoids any loops. This is appropriate in a prescriptive method, but we do not propose this as descriptively adequate for the data of naive syllogistic reasoners. The cycles of construction, testing and reformulation of candidate conclusions of Mental Model Theory can be reproduced in the EC framework proposed here.

There is one further argument in the literature about the relation between MMs and ECs which deserves mention. Johnson-Laird, Byrne and Tabossi (1989) present data from subjects performing relational inferences and claim that this domain demonstrates that MM theory is extendable in ways that graphical methods are not. Their argument goes as follows: ECs capture only monadic arguments. Mental models can capture these relational arguments. Therefore, mental models cannot be equivalent to ECs and are more extensive in their coverage (cf. also Johnson-Laird and Byrne 1991:131).

This argument is all the more extraordinary because the mental model diagrams given as analyses for the relational arguments are positional diagrams which only need a minor rearrangment of letters, and the circles drawn in, to make them into recognisable ECs. For example, Johnson-Laird and Byrne (1991:138) discuss the following case. Let p be painters; m be musicians and a be artists. We use the same notation convention, except that we arrange the as in an equivalent but different order:

Premiss 1 None of the painters are in the same place as any of the musicians

Representation $\mid [p][p][p] \mid\mid [m][m][m] \mid$

Premiss 2 All of the musicians are in the same place as some of the artists

Combined Representation $\mid [p][p][p]\ a \mid\mid [a][a][m][m][m] \mid$

By drawing circles round the ps, as and ms, the EC representation is then derived.

This argument is interesting in the current context because it reveals the involvement in these reasoning tasks of what is sometimes called the 'representation selection' problem (cf. Amarel 1969). Johnson-Laird et al. have chosen a fragment of the relational predicate calculus which collapses trivially to a monadic fragment. This explains why both MMs and ECs can encode the reasoning involved. But what neither system explains is how subjects recognise that these representations can be selected to solve these problems. How do subjects recognise that an apparently relational problem can be reduced to a monadic one by the freezing of arguments to relations? This is a quite general weakness of psychological theories of reasoning which generally focus on the processes which *follow* representation selection.

In summary, the EC algorithm inherits the predictions of MM theory, but the graphical nature of the algorithm suggests some generalisations of some of the most important effects. The method makes new predictions of behaviour on novel tasks, some of which have already been confirmed. Our claim that a reconstructed Euler system is preferable to MMs as a competence theory rests on the inherently weak expressive power of these graphical notations. This weakness stems from the specificity of graphical representations predicted by our general theory. This argument from parsimony is now strengthened by insights contributed by Euler's system into old empirical data, and by the prediction of new experimental results.

5.3 Internal implementation of ECs

On the one hand, our approach to graphical representations through logic places them in the general space of representation systems and focusses on their specificity as their critical logical property. On the other hand, this approach raises questions about how limited abstraction logics can be *implemented* in the mind. This separation of logical and implementational questions is the greatest gain for psychology in our approach. Approaches which conflate mental representation with the specification of an abstract system have failed to identify the single greatest problem for implementing deductive reasoning in human memory—the problem of binding several temporary constellations of attributes in working memory. Stenning and Oaksford (1993) present a discussion of the differences in implementational problem and human performance between long term memory based binding and working memory binding. Syllogisms are difficult for human reasoners precisely because they demand the temporary binding of properties into specifications of individuals; grouping of individuals into models; and possibly the consideration of several models. These bindings cannot be achieved on the basis of bindings already implemented in long term memory.

There is, of course, a very extensive literature on the binding of elements of lists together in human working memory. But there are few memory models which cover the holding of bindings in working memory, which would be suitable for implementing the collections of types underlying syllogistic reasoning. Stenning, Shepherd and Levy (1988) adopted a direct approach to analysing the representations resulting from the processing of very simple texts. They identify the attribute binding problem as the central knowledge representation problem which human memory must solve in order to represent texts in which more than one pattern of bindings is possible. They presented texts describing pairs of individuals in terms of four monadic properties (such as people with professions, nationalities, statures and temperaments) and observed the reading times and errors of cued recall. They demonstrate that binding of property to individual is represented in a distributed fashion; Stenning and Levy (1988) went on to show that the bindings could be retrieved from the distributed representation proposed by a soft-constraint satisfaction system.

This proposal for the representation of binding is distinguished from others (such as Anderson's (1983) ACT*'s semantic network) in that the representation contains within itself an active inferential mechanism which resolves inconsistency in retrieving bindings. This mechanism is what explains why this memory can only hold determinate patterns and so why the system can hold only representations from MARSs. This line of investigation was aimed at explaining the text conventions on the introduction of, and anaphoric back reference to, discourse referents. The conventions were described in Stenning (1978) and shown to resemble graphical specificity in Stenning and Oberlander (1991).

An important consequence of this memory architecture is that the representations underlying verbal reasoning have at least two levels. These are: the underlying associative memory from which the constraint satisfier infers 'best fitting' memories; and mechanisms which operate over the output from this retrieval mechanism. Previous theories have all assumed memory for bindings as primitive and explained only the top-level mechanism. The fact that retrieval from the lower level associative memory is by a constraint satisfaction network explains why this memory can only hold minimally abstract representations. Theories which assume that bindings are primitive network links cannot explain why these links cannot represent a more general range of representations. The empirical evidence for a memory for bindings which holds minimally abstract representations is complicated by the fact that there are certainly other systems of working memory which *can* hold abstractive representations. The most obvious example is the articulatory loop which can hold representations of sentences.

The suggested architecture has various implications for the implementation of syllogisms in human working memory. Because the EC technique employs only abstract representations of a constrained sort, it could be adapted to this architecture. The bindings which must be held are bindings within the narrow compass of MARSs. Because natural deduction systems do not agglomerate representations (cf. Stenning 1991), implementing them would require representing bindings outwith MARSs. The architecture also has the consequence that only one agglomerated pattern of bindings can be held at a time (the constraint satisfier can only satisfy one set of constraints at a time). Thus if subjects employ a strategy of considering several binding patterns (as in Mental Models Theory) this architecture explains why they must be considered serially. It is this requirement of serial consideration of models which MM theory uses to explain the different difficulty of one-, two- and three-model syllogisms.

How could such a memory architecture form the core of a mental implementation of Euler's Circles? It certainly does not represent geometrical entities like circles but rather sets of types of individual. The question can usefully be broken down into two parts: which sets of types of individuals are represented? and how does the strategic reasoning process control the changes to these sets required in the process of considering a range of possible models?

On the first question, Stenning and Levy's model fits well the fact that all regions of ECs represent maximal individuals defined on all three properties. It is a fundamental property of the Stenning and Levy model that all individuals represented are maximal. However, the model has problems representing variable numbers of types of individuals. The interested reader is referred to Stenning & Oberlander 1994 for a detailed discussion of ways of circumventing this restriction. On the second question, the resetting of the set of types represented in the constraint satisfaction system can exploit whatever mechanism underlies our ability to predict the mechanical effects on topological relations mentioned in Section 4.4. Such a mechanism is eminently implementable in a distributed connectionist representation of the multi-dimensional space. Distributed representations derive many of their desirable properties exactly from the continuity of the semantics which is imposed on their states. The underlying representation may actually be thoroughly digital, as in Willshaw nets, in which weights are either 0 or 1 (Willshaw 1981). But the error correction and content addressability of these nets derive from the continuity of their semantics—patterns at small Hamming distances count as similar patterns. The constraint satisfaction model of binding is the only available memory model which explains why representations from MARSs (and their close relatives) are easier to represent than arbitrary binding patterns. It thus provides an approach to the relation between external graphics and internal representations.

This review of a small sample of literature is sufficient to exemplify our current argument. Representational systems that embody various degrees of specificity are a medium of representation which would serve to explain the observations that have been taken to motivate 'non-propositional' representations. Furthermore, they provide some analysis of what these representations *are* like. If images *are* like specific representations, then this theory will explain how they can be processed, discharge the homunculi, and explain many of the observations of people processing abstract stimuli such as texts.

6 General discussion

The contrast between logic and implementation is central to our approach to the cognitive implications of media assignment. Having a logic common to particular graphical and linguistic systems can explain similarities between systems implemented in different modalities, as well as differences between them.

The approach through logic allows explanations of general effects of complexity which are common to people and machines. The framework therefore allows for more subtle differentiation between two types of case. The first includes those cases in which humans exhibit particular performance profiles for very general reasons of computational complexity. By contrast, the second includes cases in which humans' performance profiles are to be explained by particular features of their computational architecture. Such issues are certainly not solved merely by using a logical framework; but at least they can be precisely formulated.

Regarding the imagery debate, we would urge a more careful differentiation of 'propositional' representations into a whole variety of sentential systems. These systems will have quite different inferential properties and ranges of possible implementation. We believe the resulting classifications will be much better guides to the psychological analysis of their internal implementation. Computational models of internal implementations in long term and working memory are not paper-and-pencil-drawing implementations. Nor are they representations in acoustic or printed natural languages. Approaching them armed with a characterisation of their central computational properties—including specificities—might be expected to have advantages in the range of implementations which are thereby suggested.

The analysis of Euler's Circles indicates that our distinction between MARS and LARS at least describes a major distinction between interpretations of this graphical case. And this particular case has historically played a part in arguments for and against graphical reasoning as a psychological model. The clarification of the relation between Euler's method and mental models unites two important strands of research. On the one hand, there is a body of empirical data about human reasoning, with an important tradition in logical thought. On the other, there are contemporary practical interests in the role of graphics in information display. The clarification also indicates that graphical algorithms can make a contribution to extending and deepening the empirical understanding of the data.

Kant argued that as mobile active beings, our reasoning processes are influenced by our spatial experience. His theories were based on a physics subsequently to be generalised, but they had a crucial impact on psychological thinkers such as Vygotsky and Piaget, and on their formulations of theories of the development of reasoning. Linguists have explored the 'localist' hypothesis that the child derives its abstract categories by generalising initially spatial concepts (cf. Anderson 1971, Clark and Carpenter 1989). Cognitive linguists have argued along related lines that our most abstract categories are derived from spatial archetypes (cf. Jackendoff 1983, Langacker 1987); they have further claimed that this leads inexorably to a 'non-realist' ontology for natural language semantics (cf. Lakoff 1987). What we evidently share with all these authors is the following intuition. Mechanisms developed for perceiving and reasoning about the spatial world are likely to be used for reasoning about other domains. However, we are agnostic about the ontological direction which development follows; on balance, we prefer to adopt a realist semantics for space, and to demand from psychology an explanation of how logics are implemented in the mind.

Ongoing theoretical work extends our framework to other systems of graphical and linguistic representation. An analysis of the contrast between graphical and linguistic modalities in this framework motivates a treatment of the more expressive 'visual' formalisms (such as semantic networks) as essentially languages, whereas the least expressive languages (such as finite-state generated languages) as iconic modalities which enforce information in a similar way to graphics. This redrawing of an intuitive distinction on theoretical grounds is shown to throw new light on what is known of the cognitive efficacy of so-called 'visual languages' (Stenning, Neilson & Inder (forthcoming)). In a more empirical vein, we have completed a study comparing the cognitive effects of teaching first-order logic using Hyperproof (a graphically enhanced system due to Barwise & Etchemendy 19??) with the effects of using a conventional syntactic method. Hyperproof uses a graphical LARS involving abstraction 'tricks' analogous to those observed here in ECs. The results show strong interactions between students' pre-course problem-solving aptitudes and their post-course reasoning improvements and proof-styles (Cox, Stenning, & Oberlander (1994)). Oberlander, Cox & Stenning 1994 show that these emergent differences in proof-style revolve around use of the abstraction symbols.

7 References

Aho, A. V. and Ullman, J. D. (1972). *The Theory of Parsing, Translation, and Compiling.* Englewood Cliffs, New Jersey: Prentice-Hall.

Amarel, S. (1969). Problem solving and decision making by computer: an overview. In Garvin, P. L. (ed.) *Cognition: A Multiple View*. New York: Spartan Books.

Anderson, J. M. (1971). *The Grammar of Case*: *Towards a Localistic Theory.* Cambridge: Cambridge University Press.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, **85**, 249–277.

Anderson, J. R. (1983). *The Architecture of Cognition.* Cambridge, Ma.: Harvard University Press.

Ardin, C. (1991). Mental representations underlying syllogistic reasoning. PhD Thesis, Human Comunication Research Centre.

Bartlett, F. C. (1932). *Remembering: a study in experimental and social psychology.* Cambridge: Cambridge University Press.

Barwise, J. and Etchemendy, J. (1990). Visual information and valid reasoning. In W. Zimmerman (Ed.) *Visualization in Mathematics*. Washington: Mathematical Association of America.

Berkeley, G. (1709). *An essay towards a new theory of vision.* Dublin: Rhames.

Brachman, R., Fikes, R. and Levesque, H. (1983). KRYPTON: integrating terminology and assertions. In *Proceedings of the 3rd Annual Meeting of the American Association for Artificial Intelligence*, Washington, DC, 1983, pp31–35.

Bransford, J. D. and Johnson, M. (1972). Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, **11**, 717–726.

Bransford, J. D., Barclay, J. R. and Franks, J. J. (1972). Sentence memory: a constructive versus interpretive approach. *Cognitive Psychology*, **3**, 193–209.

Clark, E. V. and Carpenter, K. L. (1989). The Notion of Source in Language Acquisition. *Language*, **65**, 1–30.

Clark, H. (1969a). Influence of language on solving three-term series problems. *Journal of Experimental Psychology*, **82**, 205–215.

Clark, H. H. (1969b). Linguistic processes in deductive reasoning. *Psychological Review*, **76**, 387–404.

Cox, R., Stenning, K. and Oberlander, J. [**1994**] Graphical effects in learning logic: reasoning, representation and individual differences. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society.*

Dodgson, C. (1896). Symbolic Logic. Chapter Book III in *Lewis Carroll's Symbolic Logic: edited by W. W. Bartley.* Hassocks, Sussex: Harvester Press.

Erickson, J. R. (1974). A set analysis theory of behaviour in formal syllogistic reasoning tasks. In Solso, R. (Ed.) *Loyola Symposium on Cognition*, Volume 2. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Euler, L. (1772). *Lettres a une princesse d'Allemagne*, Volume 2: *Sur divers sujets de physique et de philosophie.* Letters 102–108.

Faris, J. A. (1955). The Gergonne relations. *Journal of Symbolic Logic*, **20**, 207–231.

Ford, M. (1985). Review of 'Mental Models'. *Language*, **61**, 897–903.

Funt, B. V. (1977). WHISPER: a problem solving system utilizing diagrams and a parallel processing retina. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1977, pp459–64.

Funt, B. V. (1980). Problem-Solving with diagrammatic representations. *Artificial Intelligence*, **13**, 210–230.

Galton, F. (1883). *Inquiries into human faculty and its development.* London: Macmillan.

Garnham, A. (1987). *Mental models as representations of discourse and text.* Chichester: Ellis Horwood.

Gelernter, H. (1963). Realization of a geometry-theorem proving machine. In Feigenbaum, E. A. and Feldman, J. (eds.) *Computers and Thought.* N. Y.: McGraw-Hill.

Goodman, N. (1968). *Languages of Art.* Indianapolis: Bobbs-Merrill.

Guyote, M. J. and Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology*, **13**, 461–525.

Hendrix, G. (1979). Encoding knowledge in partitioned networks. In Findler, N. V. (Ed.) *Associative Networks*. New York: Academic Press.

Hinton, G. (1979). Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery. *Cognitive Science*, **3**, 231–250.

Hinton, G. (1980). Frames of reference and mental imagery. In Long, J. and Baddeley, A. (Eds.) *Attention and Performance*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Jackendoff, R. (1983). *Semantics and Cognition.* Cambridge, Ma.: MIT Press.

Johnson-Laird, P. N. and Steedman, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology*, **10**, 64–99.

Johnson-Laird, P. N. (1983). *Mental Models.* Cambridge: Cambridge University Press.

Johnson-Laird, P. N. and Bara, B. G. (1984). Syllogistic Inference. *Cognition*, **16**, 1–61.

Johnson-Laird, P. N., Byrne, R. M. J. and Tabossi, P. (1989). Reasoning by model: the case of multiple quantification. *Psychological Review*, **96**, 658–673.

Johnson-Laird, P. N. and Byrne, R. M. J. (1991). *Deduction.* Hove, Sussex: Lawrence Erlbaum Associates.

Kneale, W. and Kneale, M. (1962). *The development of logic.* Oxford: Oxford University Press.

Kosslyn, S. M., Pinker, S., Smith, G. E. and Schwartz, S. P. (1979). On the demystification of mental imagery. *Behavioural and Brain Sciences*, **2**, 535–581.

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind.* Chicago, Illinois: The University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar.* Stanford, Ca.: Stanford University Press.

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

Levesque, H. J. (1988). Logic and the complexity of reasoning. *Journal of Philosophical Logic*, **17**, 355–389.

Lindsay, R. K. (1988). Images and inference. *Cognition*, **29**, 229–250.

Mackinlay, J. D. (1986). Automatic Design of Graphic Presentations. Technical Report No. STAN–CS–86–1138, Department of Computer Science, Stanford University, 1986.

Mani, K. and Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition*, **10**, 81–87.

Marr, D. (1982). *Vision: A Computational Investigation in the Human Representation of Visual Information.* San Francisco: Freeman.

Maybury, M. (1993). (ed.) *Intelligent Multimedia Interfaces*. AAAI Press.

McGonigle, B. and Chalmers, M. (1986). Representations and Strategies During Inference. Chapter 6 in Myers, T. F., Brown, E. K. and McGonigle, B. (Eds.) *Reasoning and Discourse Processes.* London: Academic Press.

Oberlander, J., Cox, R., and Stenning, K. [**1994**] Proof styles in multimodal reasoning. Presented at the *4th International Conference on Information-oriented approaches to Language, Logic and Computation*, Moraga, Ca., June, 1994.

Palmer, S. E. (1978). Fundamental Aspects of Cognitive Representation. Chapter 9 in Rosch, E. and Lloyd, B. B. (Eds.) *Cognition and Categorization.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Peirce, C. S. (1977). *Semiotic and significs: The Correspondence between Charles S. Peirce and Victoria, Lady Welby.* Bloomington: Indiana U P.

Petre, M. and Green, T. R. G. (1992). Requirements of Graphical Notations for Professional Users: Electronics CAD Systems as a Case Study. *Le Travail Humain*, **55**, 47–70.

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

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

Shin, S-J. (1991). A Situation-Theoretic Account of Valid Reasoning with Venn Diagrams. In Barwise, J., Gawron, J. M., Plotkin, G. and Tutiya, S. (Eds.) *Situation Theory and Its Applications*, Volume 2. Chicago: Chicago University Press.

Stenning, K. (1978). Anaphora as an approach to pragmatics. In Halle, M., Bresnan, J. and Miller, G. A. (Eds.) *Linguistic Theory and Psychological Reality.* Cambridge, Ma.: MIT Press.

Stenning, K. (1986). On making models: a study of constructive memory. Chapter 7 in Myers, T., Brown, K. and McGonigle, B. (Eds.) *Reasoning and Discourse Processes.* London: Academic Press.

Stenning, K. (1989). Modelling memory for models. In Esquerro, J. (Ed.) *Proceedings of the International Colloquium for Cognitive Science*, University of San Sebastian, San Sebastian, 1989.

Stenning, K. (1991). Distinguishing conceptual and empirical issues about mental models. In Rogers, Y., Rutherford, A. and Bibby, P. (Eds.) *Models in the Mind.* Academic Press.

Stenning, K. and Levy, J. (1988). Knowledge-rich solutions to the 'binding problem': some human computational mechanisms. *Knowledge Based Systems*, **1**.

Stenning, K., Neilson, I. and Inder, R. [forthcoming] Applying semantic concepts to the media assigment problem in multi-media communication. Research paper, Human Communication Research Centre, University of Edinburgh. To appear as a chapter in J. Glasgow (ed.) ???

Stenning, K. and Oaksford, M. (1993). Rational reasoning and human implementations of logics. In Manktelow, K. I. and Over, D. E. (Eds.) *Rationality*. London: Routledge and Kegan Paul. Pps. 136-176.

Stenning, K. and Oberlander, J. (1991). Reasoning with Words, Pictures and Calculi: computation versus justification. In Barwise, J., Gawron, J. M., Plotkin, G. and Tutiya, S. (Eds.) *Situation Theory and Its Applications*, Volume 2. Chicago: Chicago University Press.

Stenning, K. and Oberlander, J. (1994). Spatial containment and set membership: a case study of analogy at work. In Barnden, J. and Holyoak, K. (Eds.) *Analogical Connections*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. Pps. 446-486.

Stenning, K., Shepherd, M. and Levy, J. (1988). On the construction of representations for individuals from descriptions in text. *Language and Cognitive Processes*, **2(3)**, pps. 129-164.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information.* Cheshire, Connecticut: Graphics Press.

Twyman, M. (1979). A schema for the study of graphical language. In Kolers, P. A., Wrolstad, M. E. and Bouma, H. (eds.) *Processing of Visible Language: Volume 1.* New York: Plenum Press.

Venn, J. (1894). *Symbolic Logic.* London: Macmillan.

Willshaw, D. (1981). Holography, Associative Memory and Inductive Generalisation. Chapter 3 in Hinton, G. E. and Anderson, J. A. (Eds.) *Parallel models of associative memory.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Yule, P. (1991). An experimental investigation of a new theory of syllogistic reasoning based on Euler. Masters Thesis, Centre for Cognitive Science, University of Edinburgh.

Yule, P. and Stenning, K. (1992). The Figural Effect and a Graphical Algorithm for Syllogistic Reasoning. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pp1170–1175, Bloomington, Indiana, August 1992.