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

Aligning Logical and Psychological Perspectives on Diagrammatic Reasoning

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

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

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

1. Introduction

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

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

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

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

1.1. *Availability of constraints*

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

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

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

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

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

1.2. *Outline*

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

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

2. What Are Diagrams?

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

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

2.1. *Properties of diagrammatic representations*

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

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

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

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

Our basic observations are that:

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

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

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

Sentential languages are typically constructed using a vocabulary of symbols:

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

Along with some rules of combination:

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

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

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

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

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

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

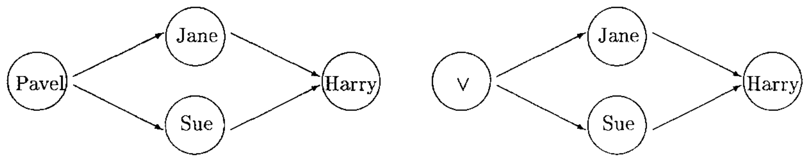

Figure 1. Direct interpretation vs. abstract syntax.

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

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

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

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

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

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

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

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

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

3. Desiderata on a Theory of Efficacy

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

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

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

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

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

4. A View from Logic

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

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

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

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

5. A View from Psychology

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

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

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

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

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

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

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

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

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

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

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

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

6. Cases Studies in Diagrammatic Reasoning

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

6.1. *Efficacy and inefficacy in psychological studies*

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

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

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

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

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

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

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

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

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

6.2. *A case study in inefficacy*

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

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

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

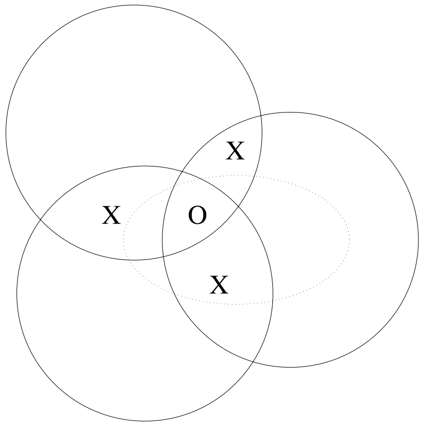

Figure 2. The Helly constraint for convex regions in two dimensions.

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

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

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

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

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

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

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

7. The Theoretical Framework

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

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

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

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

ii) these constraints are available to the agent;

iii) inference with the representations is tractable.

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

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

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

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

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

8. A Sample Analysis

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

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

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

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

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

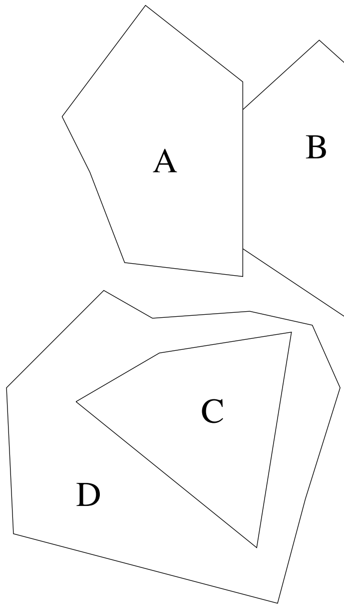

Figure 3. "Some A are B, no B are C, no A are C, no A are D, no B are D, all C are D".

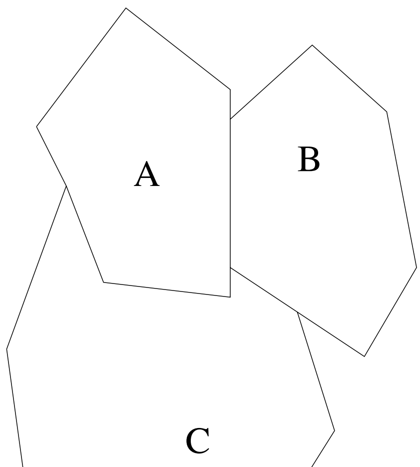

Figure 4. "Some A are B, some B are C, some A are C".

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

8.2. *The analysis*

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

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

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

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

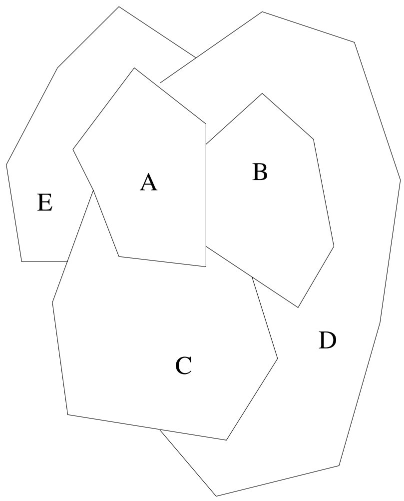

Figure 5. An attempt to realize situation S.

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

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

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

9. Extending the Logical Approach

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

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

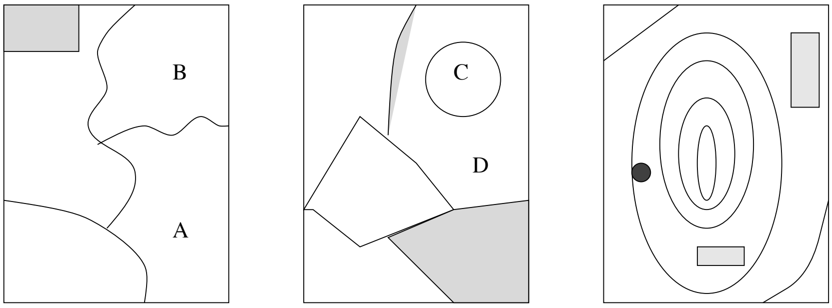

Figure 6. Some simple maps.

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

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

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

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

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

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

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

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

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

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

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

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

10. Extending the Psychological Theory

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

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

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

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

11. Conclusion

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

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

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

Acknowledgements

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

Notes

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

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

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

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

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Stenning, K., Cox, R. & Oberlander, J. (1995). The effect of Graphical and Sentential Logic Teaching on Spontaneous External Representation. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* **2**(4): 56–75.

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

Stenning, K. & Oberlander, J. (1995). A Cognitive Theory of Graphical and Linguistic Reasoning: Logic and Implementation. *Cognitive Science* **19**(1): 97–140.

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

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

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

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

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

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

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