Effective Diagrammatic Communication: Syntactic, Semantic and Pragmatic Issues

C A Gurr

Human Communication Research Centre
University of Edinburgh
Edinburgh EH8 9LW, UK

email: C.Gurr@ed.ac.uk

Abstract

The study of systems of communication may be divided into three parts: syntax, semantics and pragmatics. Accounts of the embedding of text-based languages in the computational processes of reasoners and communicators are relatively well developed; with accounts available for a spectrum of languages which ranges from the highly formalised and constrained, such as formal logics, to the highly informal and unconstrained natural languages used in everyday conversations. Analogies between diagrams and such textual representations of information are quite revealing about both similarities and differences and can provide a useful starting point for exploring the issues in a theory of diagrammatic communication.

This paper sketches out a theory of diagrammatic communication, based upon recent studies of the syntactic, semantic and pragmatic component issues which such a theory must accommodate. In the context of this theory an exploration is made of the issues involved in answering the question: what makes for an *effective* diagrammatic representation? To achieve this we review a framework in which we may explore properties of representations, and properties of the relation between representations and that which they represent. By relating this framework to our sketched theory of diagrammatic communication, we explore the concept of effectiveness in diagrams. This process of exploration enables us to relate previous studies of the effectiveness of diagrams into a broader, unified framework, which clarifies both the various issues relating to effectiveness and the relations between them.

Theories of diagrammatic reasoning have become more prevalent over the past few years. A collection of both seminal and contemporary papers on diagrammatic reasoning, from cognitive and computational perspectives, is found in Glasgow *et al* [6]. An excellent overview of the different approaches, historical and recent, to this field of research is found in Narayan [19].

This paper investigates the issues concerned in determining what makes for *effective* diagrammatic communication. As yet there does not exist the fully fledged foundation to a theory of diagrammatic communication which such an investigation requires. In Gurr *et al* [11] a careful examination of analogies between the sentential and the graphical cases is shown to be quite revealing about both similarities and differences. Rehearsing some of these analogies provides a decomposition of subproblems for analysing visual representations and diagrammatic reasoning. This decomposition permits us to identify the fundamental issues relating to the effectiveness of visual and diagrammatic representations for communication and reasoning tasks. A particular concern of this paper is the importance of accounting for *pragmatic* issues within a theory of diagrammatic communication. Most previous formal accounts of diagrams have generally ignored cognitive issues concerning human reaction to representations, focusing rather on providing idealised formal semantics for particular notations. To understand what makes diagrams effective, we must consider rather their interpretation by a human, and in this context we see the study of pragmatics as attempting to bridge the gap between formal semantics and this 'real' semantics of diagrams.

The layout of this paper is as follows. Section 1 provides the context, both terminological and philosophical, for the subsequent sections. It presents a discussion of the place of diagrams in the landscape of "visual languages" and reviews the previous studies of Gurr *et al* [10, 11]. It concludes by examining an argument which is central to this paper: that a significant determiner of the effectiveness of a representation is how well it is matched to both what it represents and the task for which it is intended. Section 2 reviews the component issues in a theory of diagrams: syntactic and semantic, as introduced in [10, 11], and also pragmatic. Section 3 presents a framework for examining the correspondence between representations and that which they represent. We discuss, with examples, how syntactic, semantic and pragmatic constraints assist in assuring a 'close match' between representation and represented. Finally, the conclusion reviews the contribution of this paper towards a theory of diagrams and of effectiveness in diagrams, and discusses what is required to complete such a theory.

1 Theories of "visual" languages

The majority of theories of diagrams fall into two broad categories. The first are motivated by the desire to provide a justification for diagrammatic reasoning in formal proofs. Such theories are primarily concerned with providing an account of the correspondence between diagrams and some formal semantics for them, demonstrating such properties as soundness and completeness. The second category of theories are concerned with explaining the impact upon human cognition of graphical representations; seeking to explain what advantages diagrammatic representations hold for the reasoner over other forms of representation.

Theories of the first kind (for example, [18, 26, 27] and the collection in [1]), in common with the formal semantic studies of natural language from which they derive their methods, generally ignore issues concerning the cognitive complexity of inference, and indeed leave out any consideration of the inferential mechanisms that operate over the sentences or diagrams of the languages whose semantics is being specified. Theories of the second kind (for example, [3, 12, 22, 32]) generally lack a fully specified formalism and semantics on which they could base a computational account of how the system of representations is embedded in a user's performance of some task.

An attempt, not to claim any complete theory of diagrams, but rather to sketch a larger structure which accounts for the processes of using diagrams in reasoning or communication, and to clarify the relations between its component problems, is given in [11] based upon a framework introduced in [10]. This is achieved through examination of the analogies (and dis-analogies) between diagrammatic languages and, more classical, textual languages (ranging in complexity from simple algebras, classical logics, more esoteric logics and ultimately to full-blown natural languages). In this paper we both review and extend this previous sketch, so that we may discuss effectiveness in the context of a theory of diagrammatic communication which is subdivided into syntactic, semantic and pragmatic components.

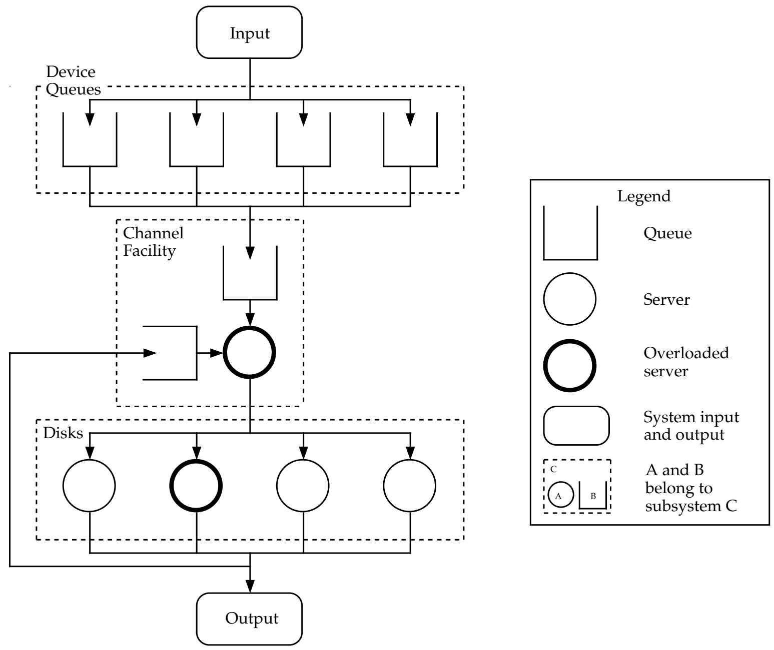

Figure 1: The disk subsystem of a computer

1.1 When is a visual representation a diagram?

A significant motivation for the examination of [10, 11] was the belief that there are fundamental differences between certain aspects of diagrammatic and of textual representations. It was, and is, believed by the authors of [11] that before traditional theories of communication and reasoning, which are based on textual languages, could be convincingly applied to diagrammatic representations, they needed to be re-examined in the light of these differences. Before reviewing the results of this examination, we must clarify what we mean by diagrammatic and textual aspects of representations, and in what sense they differ.

For the sake of argument, a representation may be considered to be a collection of objects and some relations between these objects. The type (visual or otherwise) of a particular representation, and more generally of a language, is determined by the characteristics of the "symbols" used to express these objects and relations. Note that throughout this discussion the term symbol is used in a very broad sense, as it applies to the means of expressing both objects *and* relations in a representation. Thus diagrammatic means of expressing relations – including the use of morphological, geometric, spatial, topological or other such visual features – are also considered visual "symbols" here. Consider, for example, the diagram of Figure 1 (taken from Marks and Reiter's application of linguistic theories of pragmatics to "graphical representations" [16]) which represents a computer's disk subsystem. Here objects (whether physical or conceptual) such as servers or system input/output, are expressed with visual symbols – circles and boxes respectively. The diagram also expresses relations and properties of objects (considered as unary relations) via visual symbols. On the understanding, that is, that not only are features such as the arrows which connect objects considered as symbols, but so also are features such as the use of spatial inclusion within a box drawn with a dotted line (denoting membership of a subsystem).

Consider a textual representation, for example a propositional logic sentence such as "$p \wedge (\neg q \vee r)$". This, if we are to be precise, is also an example of a visual representation.

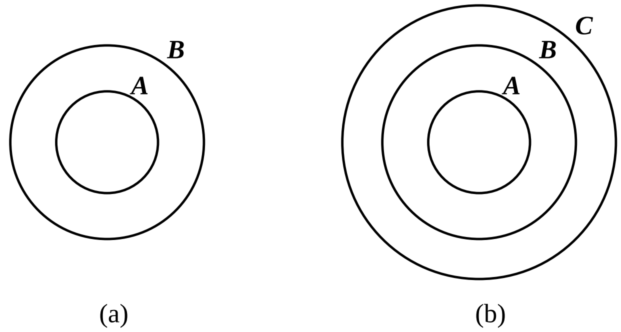

Figure 2: Transitivity in Euler's circles.

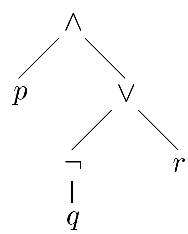

Figure 3: Syntax tree for the sentence $p \wedge (\neg q \vee r)$

After all, the symbols in this sentence are expressed to the reader's visual sense (as ink marks on a page), not to their senses of touch or hearing, as would be the case with braille or speech for example. There is, however, a significant difference between certain of the visual symbols of Figure 1 and those of this propositional sentence. The difference is that certain of the symbols which express relations in Figure 1 exhibit intrinsic properties, and these properties directly correspond to properties in the represented domain.

The language of Euler's circles, used to represent logical syllogisms, gives a classic example of this direct capture of semantic information in a visual symbol. In this language, sets are represented by labelled circles and set inclusion by spatial inclusion. For example, the Euler's circles in the diagram of Figure 2 illustrate the syllogism

i All A are B

ii All B are C

iii (therefore) All A are C

Here the transitive, irreflexive and asymmetric relation of set inclusion is expressed via the similarly transitive, irreflexive and asymmetric visual of proper spatial inclusion in the plane. In textual representations the relationships between symbols are necessarily captured in terms of the concatenation relation, which must then be interpreted by some intermediary abstract syntax that captures the relationship being represented. For example, the relationship between the propositions p and q in the sentence $p \wedge (\neg q \vee r)$ is not captured directly in this sentence, but through an intermediary syntax such as the tree of Figure 3.

Partitioning visual symbols into direct and indirect categories should not be interpreted as suggesting that all visual representations, whether texts or diagrams, are either exclusively direct or indirect. Clearly 'pure' visual languages, containing only direct or indirect symbols, are possible. Examples of the latter would include sequences of English sentences, or conjunctions of algebraic or logical formulae, in completely unformatted text (ignoring line breaks). Such examples, however, are merely extremes in a spectrum of representations which employ both direct and indirect symbols. It is between the extremes of this spectrum, rather than at them, that most visual representations exist. Consider that most texts do contain some formatting, such as page layout, differing fonts or text size. The majority of such means for formatting texts may be seen as a utilisation of directness. Similarly, very few – if any – diagrammatic languages are wholly dependent upon direct symbols. In the diagram of Figure 1, for example, while some symbols (notably spatial inclusion) are directly interpreted, the choice of circles to represent servers carries no direct semantic information. A triangle, octagon or almost any other shape would do just as well to represent servers, were we simply to change the diagram's legend.

Traditional theories of communication and reasoning in textual languages are, naturally, primarily concerned with representations which employ indirect symbols. In [10, 11] the foundations of such theories were re-examined in the light of representations in which direct symbols are employed. The results of this examination are reviewed next. In seeking a fuller understanding of direct aspects of diagrams, of what it is that is peculiar or unique to them, this paper and the earlier studies that are summarised next seek to achieve a fuller understanding of general diagrams, containing both direct and indirect aspects.

1.2 Graphical versus sentential representations

To simplify discussion, the remainder of this paper follows the terminology of [10, 11] and uses the word "graphical" when referring to symbols, representations and languages in which directness is a primary factor. The term "sentential" is used when referring to symbols, representations and languages which employ mainly indirect features. To repeat, this is not to suggest that this division reflects any actual division of classes of visual language in reality. The terms are used simply to clarify whether direct or indirect features of representations are under discussion. In practice, most representations – and certainly most diagrams – will display examples of both. In [10, 11] three major significant differences in the various components of graphical (versus sentential) systems were explored: the directness of diagrams; reasoning in diagrams versus reasoning in texts; and constraints over interpretation.

1.2.1 Directness in diagrams

Certain diagrammatic notations and relations may be directly semantically interpreted. This directness may be exploited by the semantics of diagrams in a systematic way. Such 'systematicity' is not exclusively the preserve of diagrammatic representations, but – with their potential for direct interpretation – diagrams have a head start over sentential representations in the systematicity stakes. This aspect of graphical representations is explored more fully in Section 2.

1.2.2 Diagrammatic reasoning

The second major issue which differentiates diagrammatic from sentential systems relates to theorem provers and reasoning – the dual processes of *constructing* and *recognising* a conclusion are quite different in the diagrammatic and textual cases. In diagrams, the theorem prover is not so easily separated from the 'proof theory'. In fact, the construction process (the combination of two or more diagrams to produce a 'conclusion' diagram) varies substantially between differing graphical systems. In sentential systems, by contrast, there is far more uniformity in the process of constructing inferences. This issue has major implications for our exploration of effectiveness in diagrams. We expand upon this issue, and examine its implications, in Section 1.3.

1.2.3 Constraints over interpretation

The final major difference between graphical and sentential systems, explored in [10, 11], concerns the means by which constraints over interpretation determine the effectiveness of a representation system for the human reader/reasoner. The suitability, learnability and usability of diagrammatic representations (and hence effectiveness) are profoundly influenced by issues of human reaction to representations. It is the constraints over interpretation which determine the suitability of a particular representation for a task of reasoning. For diagrams these constraints fall into three categories: their *origin*; their *point of operation*; and their *availability* to the reader.

More specifically, of these three dimensions the first concerns their origins: whether the constraints are intrinsic to the interpretation of the medium, or alternatively of conventional, external origin. The second dimension of constraints concerns their 'point of operation'. That is, whether constraints apply to, and aid reasoning, over individual representations within a system of representation; or whether they apply to and aid reasoning about the representational system as a whole. The final dimension of constraints concerns their *availability* to the human who reasons with the system. We return to these dimensions of constraints in Section 3, where we explore their effect on the means of assuring that a representation is correctly matched to the task for which it is intended.

1.3 The interaction between representation and task

A primary argument put forward to justify the claim of diagrammatic representation systems being more effective than textual ones is that certain inferences are somehow more immediate, or even are automatic, in diagrams. In such representational systems conclusions appear "for free", as compared with textual systems where a logical inference must be made to produce the conclusion. For example, inferring from the information *all A are B; all B are C* the conclusion *All A are C* is arguably a more straightforward inference in the direct representation system of Figure 2 (Euler's circles) than in the textual case. It can be argued that this is due to the fact that construction of diagram 2(ii) automatically includes the representation of the conclusion *All A are C*, and thus the information appears for free.

This argument is given a formal account by Shimojima [25], where apparent inferences such as that of Figure 2(ii) are termed inferential "free-rides". A free ride is defined as a form of side effect of the manipulation of a diagram. Where a sequence of valid operations is performed which cause some consequence to become manifest in a diagram, where that consequence was not explicitly insisted upon by the operations, a free ride occurs. For the syllogism illustrated by the Euler's circles of Figure 2 for example, the free ride manifests itself when we construct Figure 2(ii). Figure 2(i) is drawn according to the premise *all A are B*, with the circle *A* inside the circle *B*. From the premise *all B are C* we next draw the circle *C* so as to completely surround the circle *B* (Figure 2(ii)). We observe that the circle *C* completely surrounds *A*, and thus conclude that *All A are C*. Note, however, that none of the instructions for constructing Euler's circles which caused us to draw Figure 2(ii) explicitly insists that we should draw circle *C* so as to surround circle *A*. This is a semantically meaningful (and correct) fact which was entailed for "free" by virtue of the construction rules

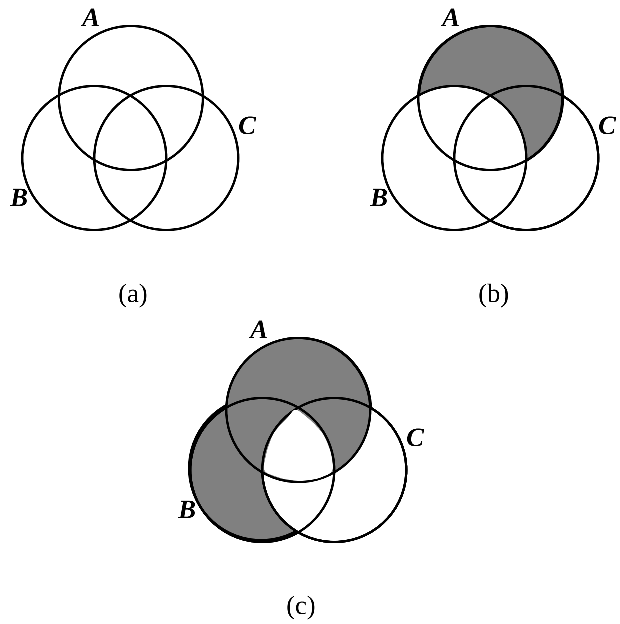

Figure 4: Transitivity in a Venn diagram (shaded areas indicate empty sets).

applied to the premises. This then, is a free ride. However, this issue is more complex than it may appear at first glance.

Consider the Venn diagram of Figure 4, which represents the earlier syllogism of Figure 2 (*all A are B; all B are C* therefore *All A are C*). In this diagram precisely the same graphical relation (proper spatial inclusion) is used to represent the same logical relation of set inclusion. In the same manner as for Figure 2(ii), construction of the final Venn diagram of Figure 4(iii) automatically includes the representation of the conclusion *All A are C*. However, it would be disingenuous to claim that, for a human reasoner, there would be no difference whatsoever between recognising this conclusion in Figure 2(ii) and recognising it in Figure 4(iii). In the light of this example, it would seem more accurate to term such occurrences "cheap rides" rather than free rides; with the addendum that some rides are cheaper than others.

We take the view that reasoning is a two stage process[1]: Firstly, the process of constructing a new diagram which composes premises typically results in a number of inferences being made, all of which are visible in the resulting diagram. Secondly, the desired conclusion must be recognised in the resulting diagram. In diagrammatic languages the diagram resulting from an inference step typically contains numerous logical inferences. For example, both the diagrams of Figure 2(ii) and Figure 4(iii) contain the required conclusion (that *All A are C*), but they also contain a number of other potential conclusions. Recognising the desired conclusion is not automatic and ease of recognition can vary substantially between differing representational systems, as it does in this case.

A significant issue here is the task for which the diagrammatic notation is being used. In the case of solving syllogisms such as the above, the task is a relatively simple one and the Euler's circle diagrams of Figure 2 are significantly more effective than the Venn diagrams of Figure 4. However, this is because the representation is well matched to the task. Venn

[1] Actually, a "three stage process" would be a more accurate statement, but we do not consider here the initial stage of selecting a particular combination of inference rules and premisses. Our argument here is primarily concerned with the performance of a reasoning *step*, rather than issues concerning reasoning *strategy*.

diagrams are more subtle than Euler's circles, being more sophisticated than necessary for this particular task, but capable of benefiting reasoning in other, more complex, tasks for which Euler's circles are insufficient. A major point to note here is that, for a human reasoner, what significantly assists the reasoning process is *recognising* the inherent properties of a particular system of representation. Properties must be recognised for their benefits to be available. Furthermore, as the findings of empirical analyses of the effect of representations upon human reasoning have shown (such as those of Stenning *et al* [28, 30]), humans can display substantial individual variation in their ability to recognise, and thus exploit, the benefits of particular representations.

A brief summary of the argument taken in this paper, the consequences of which are explored in Section 3, is that the effectiveness of a representation is to a significant extent determined by how closely the semantics of the representation resembles that which it represents. One benefit that certain diagrammatic representations offer to support this is the potential to directly capture pertinent aspects of the represented artifact (whether this be a concrete artifact or some abstract concept). To clarify this argument we must explain what is meant here by 'pertinent'.

By 'pertinent' aspects of the represented artifact, we refer to those aspects which are relevant to particular reasoning tasks. We argue that reasoning is strongly influenced by the structure of the representation within which one reasons. Where the structure of a representation matches the primary concepts over which one must reason, reasoning is made easier. Conversely, having the 'wrong' structure in a representation will interfere with reasoning, making it more difficult. This argument is supported by a number of empirical studies of users employing different representations for similar tasks. Notable among these are Stenning and Yule's study [30] of the representations, including numerous diagrammatic representations, used in solving logical syllogisms. The study examined students errors in syllogistic reasoning, and demonstrated that a syllogism's representation was a significant predictor of error. Similarly, a study by Zhang and Norman [32] examined the effect of alternative representations employed in solving the Tower of Hanoi problem. The purpose of this study was to assess the interplay between internal (cognitive) and external (diagrammatic, in these cases) representations. An examination of the results reveals that students were most successful when using those external representations which most accurately reflected the problem task, and were less successful when the representations were either incomplete (being too abstract) or unsupportive of basic reasoning tasks.

The next section explores issues in a sketched theory of diagrams, composed of syntactic, semantic and pragmatic parts. Section 3 returns to the issue of matching representation to task, presenting a framework for assessing the closeness of this match and considering means for ensuring such matching.

2 Syntax, semantics and pragmatics in a theory of diagrams

According to Morris [17], the study of systems of communication can be divided into three parts: syntax, semantics and pragmatics. Syntax is devoted to 'the formal relation of signs to one another'; semantics to 'the relation of signs to the objects to which the signs are applicable' and pragmatics to 'the relation of signs to (human) interpreters'. [2] While the notion of 'signs' can be read very widely, most work in pragmatics has followed in the footsteps of syntax and semantics, and focussed on language use. This section examines syntactic,

[2] See Gazdar and Levinson [4, 15] for useful surveys of the various approaches to the study of pragmatics in text and dialogue.

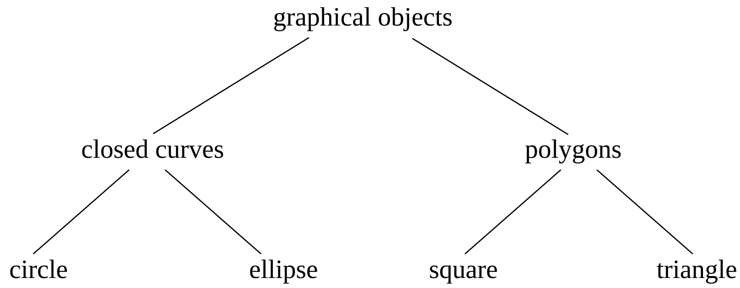

Figure 5: A Hierarchy of graphical Types.

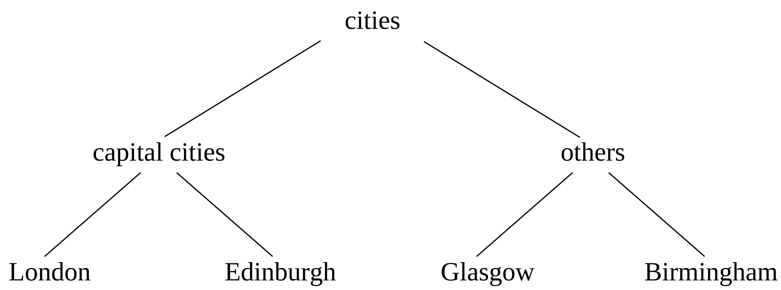

Figure 6: A Hierarchy of City Types.

semantic and pragmatic issues in a theory of diagrams. Firstly, however, it is necessary to consider something even more fundamental than the syntax of diagrams: their vocabularies.

2.1 Diagrammatic vocabularies

For sentential languages the notions of vocabulary, syntax and semantics are readily separable. The syntactic rules which permit construction of sentences may be completely independent of the chosen vocabulary, and may be clearly delineated from a definition of semantics. For example, let P be a propositional logic whose vocabulary consists of the propositions p and q, and the symbols $\wedge$ and $\neg$ representing "and" and "not" respectively. The syntactic and semantic rules for P tell us, respectively, how to construct and interpret formulas using this vocabulary. However, we may substitute the symbols $\{X,\ Y,$ `&`, `~`$\}$ for $\{p,\ q, \wedge, \neg\}$ throughout P to produce a logic which is effectively equivalent to P. Alternatively, we could retain the vocabulary and syntax of P, while altering the semantics to produce a vastly different logic.

In diagrammatic languages the concepts of vocabulary, syntax and semantics do not separate so clearly. For example, a diagrammatic vocabulary may include shapes such as circles, squares, arcs and arrows, all of differing sizes and colours. These objects often fall naturally into a hierarchical typing which will almost certainly constrain the syntax and, furthermore, inform the semantics of the system. Similarly, spatial representing relations, such as inclusion, are part of the vocabulary but clearly constrain the construction of potential diagrams and will likely be mapped to semantic relations with similar logical properties.

A fundamental distinction between sentential vocabularies and common diagrammatic vocabularies is the categorial nature of the latter. Essentially, a naturally ordered type hierarchy holds over many graphical symbols – something that cannot, to any significant extent, be said of textual symbols. This hierarchy may be exploited by the semantics of symbols so as to reflect the depicted domain. For example, Figure 5 shows a hierarchy of graphical symbols in which types are ordered from top to bottom, though not from right to left. A situation which mapped these symbols to cities, such as those of Figure 6, could exploit this type-structure, preserving the ordering of types in each domain across the mapping. The advantage of this constraint is obviously that anything which holds for cities should hold for capital cities, and anything which holds for those should hold for London, say; and the analogous generalisations apply on the graphical side. Naturally there is no formal requirement for diagrammatic vocabularies to exhibit such a type hierarchy, although we shall see in the next section the dangers of failing to take account of a *potential* such hierarchy.

This example demonstrates that, purely as a consequence of choice of symbols for a diagrammatic vocabulary, certain inferences from any given representation will arise, as it were, "for free". *Exploiting the categorial nature of diagrammatic vocabularies in this way requires that the structure inherent in the diagrammatic vocabulary is systematically matched to a relevant structure in the semantics.* We subsequently refer to this systematic mapping of structure in the representing domain (diagrammatic language, in this case) to semantic structures as systematicity. Such systematicity is also often evident in the syntax of diagrammatic languages, where the directness of particular graphical symbols is used to reflect semantic aspects of the represented domain. This further systematicity is explored next.

2.2 Syntax and semantics

A striking feature of many effective diagrams, and a consequence of direct interpretation, is that the spatial relations between their tokens share logical properties with the (not necessarily spatial) relations between denoted objects in the target domain. The classic example of representational efficacy arising from this preservation of constraints is the representation of set inclusion (a transitive, irreflexive, asymmetric relation) by proper spatial inclusion in the plane, as in the Euler's circles of Figure 2. Since proper inclusion is transitive, irreflexive and asymmetric, its efficacy in representing set inclusion is obvious.

This matching of properties between represented and representing (diagrammatic) relations is, again, an example of systematicity. Notice that what we have so far is common to typical discussions of the semantics of diagrams which seek to include cognitive issues. Indeed, account for the systematicity principle may be seen as fundamental to approaches such as Gurr [9], Wang and Lee [31], and also to discussions of analogical reasoning, as in Gentner [5]. Systematicity, in the sense of a structuring of semantics, is certainly not exclusively the preserve of diagrammatic languages. The distinction is that, as diagrammatic vocabularies may possess a hierarchical type structure and diagrammatic relations may be directly semantically interpreted, diagrams, thus, have a 'head start' over sentential representations in the systematicity stakes. However, a significant consequence of this is that the systematicity of diagrammatic languages will vary from language to language.

The systematicity principle is thus a step on the path to achieving effective diagrammatic representations: *where practicable utilise the type hierarchies of diagrammatic vocabularies and, through directness, exploit the intrinsic properties of graphical representing relations to capture semantic information.* However, the potential of the systematicity principle is not without risks. Consider the diagram of Figure 7, a variant of the computer disk subsystem diagram of Figure 1 and likewise taken from Marks and Reiter's application of pragmatic theories to diagrams. The vocabulary of this variant diagram uses instances of the same graphical symbol (a circle) to represent all types of nodes. The node symbols are ordered by size. Following the systematicity principle, this implies (incorrectly) that all nodes in the represented network fall into a single conceptual category, in which they are similarly ordered.

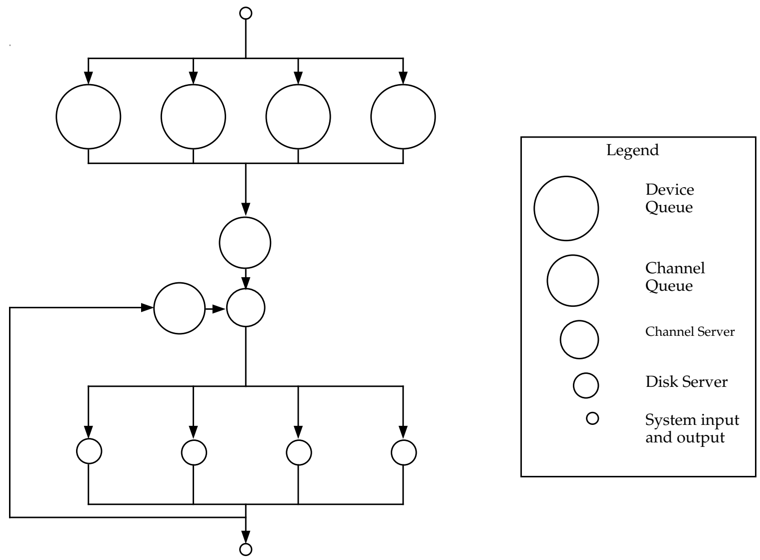

Figure 7: A variant of the disk subsystem diagram of Figure 1

Similarly, Wang and Lee [31] stress the importance of attention to the intrinsic properties of graphical relations, such as the transitivity (and asymmetry) of spatial inclusion: it may be a source of trouble if a transitive relation is used to depict one which is intransitive, or vice versa. In general it is not possible to map all relations exactly, and in different cases different properties are the important ones, so hard-and-fast rules are difficult to devise.

Examples such as the diagram of Figure 7 demonstrate that while exploiting vocabulary and directness in diagrams can contribute to effectiveness, failing to consider such systematicity can result in misleading representations. Note, however, that the diagram of Figure 7 could not be shown to be misleading through purely semantic considerations. The semantics, after all, contains no suggestion whatsoever of an ordering over the nodes of the network. It is, rather, the unwanted – perhaps unconscious – implication that the inherent ordering over the graphical symbols in the representation which carry this suggestion. Examples such as this are thus misleading for pragmatic, rather than semantic reasons.

2.3 Pragmatics for diagrams

Theories of human communication developed initially for linguistic purposes, either for written text or spoken dialogues, may be applied to good effect in the development of theories for communication involving diagrams. Oberlander [20], for example, draws together the findings from several influential studies in graphical communication to argue for parallels between pragmatic phenomena which occur in natural language, and for which there are established theories, and phenomena occurring in graphical representations.

Over time, differing pragmatic traditions have emerged; both Gazdar [4] and Levinson [15] offer useful surveys of the varying schools of thought. Notable among these, work in the Gricean tradition emphasises the view that communication takes place by virtue of a set of assumptions about rational cooperation [7, 8] – people provisionally guess that other people will act in cooperative ways, and these guesses can be confirmed or upset as a conversation unfolds. We examine here the application of this to diagrammatic communication.

A key Gricean idea is that of *implicature*: that a participant in a discourse will both

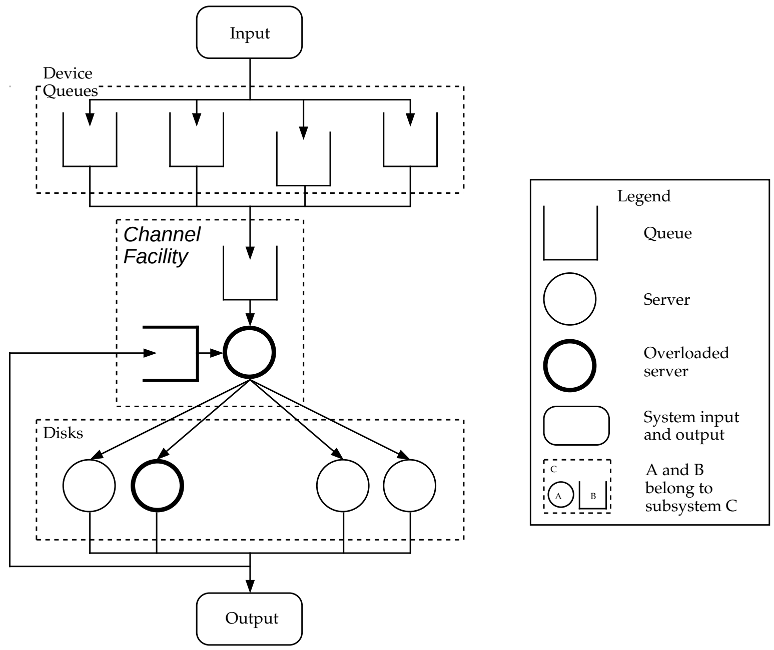

Figure 8: A second variant of Figure 1, also possessing unwanted implicatures

construct and recognise utterances as carrying intended implied information, and avoiding unwanted implications. Consider two concrete examples (taken directly from [15]) of what Grice takes to be implicature, the dialogue in (1) and the alternative texts in (2):

1. (a) A: Can you tell me the time?
   (b) B: Well, the milkman has come.
2. (a) The lone ranger jumped on his horse and rode into the sunset.
   (b) The lone ranger rode into the sunset and jumped on his horse.

(1a) has the implicature that A would like B to tell her the current time. (1b) carries the implicature that B does not know the exact time, but has some information which may permit A to calculate it. (2a)'s implicature is that the jump happened first, followed by the riding. By contrast, (2b)'s implicature is that riding preceded jumping. In both (1) and (2), implicatures go beyond the literal truth conditional meaning. For instance, all that matters for the truth of a complex sentence of the form P *and* Q is that both P and Q be true; the order of mention of the components is irrelevant. An understanding of implicatures, thus, helps to bridge the gap between truth conditions and 'real' meaning.

Grice's account is deceptively simple. His claim is that conversational participants adhere to a Cooperative Principle. This can be phrased as an injunction: make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. A more detailed exploration of this principle, and guidance that it suggests for typical communication, is provided by [7].

Marks and Reiter [16] explore the possibility of applying the Gricean theory of implicature to suggest a reasoned approach to avoiding ambiguities or misleading information in graphical representations. For example, Figures 1, 7 and 8 are presented by Marks and Reiter and display, respectively, a network diagram of a computer's disk subsystem (with no unwanted implicatures) and two further diagrams, each of which carries unwanted implicatures which may be explained by failure to adhere to the cooperative principle. Figure 7, as discussed previously, carries a misleading implicature of ordering among the nodes in the diagram. This is due to the inherent ordering of the graphical symbols used to represent these nodes. Figure 8 is misleading for at least four reasons. First, of the device queues in the upper section of the diagram, the symbol for the second-from-right device queue has been laid out irregularly. The unwanted implication is that this queue must be different to the others somehow, as otherwise it would have been laid out at the same height. Secondly, in the lower section of the diagram the disk-symbols are grouped as two perceptual gestalts. Again the unwanted implication is that this reflects some grouping in the domain, as otherwise they would have been placed regularly. The final two differences which lead to unwanted implicatures both occur in the middle section of the diagram. First, the pen-width used for the lower queue in the channel facility differs from pen-width for all other queues in the diagram. Finally, a different font has been used for the channel facility's text label, implying that it must have different sub-system status.

Note that the diagram of Figure 8, just as that of Figure 1, faithfully portrays all the features of the assumed network model. The problematic implications of the variant diagram are not in fact caused by it being incorrect, but rather by various aspects of it (unspecified in the semantics for the diagram) which are misleading. All of the above problems in the variant diagram arise because there is a general expectation of regularity. Any departures from this are assumed to be meaningful, in accordance with the cooperative principle. The problem with the diagrams of Figures 7 and 8 are that they contain both irregularities where the domain has none and also extra regularities not in the domain.

The examples of Figures 7 and 8 illustrate the potential dangers of failing to take pragmatic aspects of diagrams into account. However, other research has shown that the efficacy of diagrams for communicating information can be increased by actively exploiting such aspects. Studies by Petre and Green [23] of engineers using CAD systems for designing computer circuits, demonstrated that the most significant difference between novices and experts is in the use of layout to capture domain information. In such circuit diagrams the layout of components is not specified as being semantically significant. Nevertheless, experienced designers exploit layout to carry important information by grouping together components which are functionally related. By contrast, certain diagrams produced by novices were considered poor because they either failed to use layout or, in particularly 'awful' examples, were especially confusing through their mis-use of the common layout conventions adopted by the experienced engineers. These conventions, termed 'secondary notations' in [23], are shown in [20] to correspond directly with the graphical pragmatics of [16].

More recent studies, reported in [22], of the users of various other visual languages, notably visual programming languages, have highlighted similar usage of graphical pragmatics. A major conclusion of this collection of studies is that *the correct use of pragmatic features, such as layout in graph-based notations, is a significant contributory factor to the effectiveness of these representations.* The remainder of this paper explores the issues involved in determining how and when the pragmatic (and thus, the 'real' semantic) implicatures of a particular diagrammatic system are correctly matched to the domain and task in question.

3 Matching representation to domain and task

We introduce next a framework for examining the relation, in terms of matching of properties, between diagrams and that which they represent. With respect to this framework we

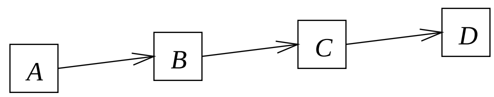

Figure 9: Diagram representing integer ordering.

explore the relationship between directness, systematicity and the pragmatics of reasoning in diagrammatic representational systems, and how they contribute to the effectiveness of a diagrammatic system of representation. The framework arose from a study in Gurr [9] of what it meant for a diagram to be an *isomorphism* of that which it represents. This section first reviews this framework, which typically operates at the level of the relation between individual diagrams and what they represent. We shall see that the intrinsic systematicity of diagrams may often contribute to isomorphism at this level. However, this benefit is only partially carried over at the level of systems of representation. It is in compensating for this shortfall, we shall see, that pragmatic constraints over interpretations play a leading role.

3.1 The relation between representing and represented worlds

One class of terminology which has been applied to diagrams is that they are homomorphisms or isomorphisms of the world which they represent. These terms have precise meanings in algebra, and imply the potential for variation in the strength of match between representation and represented. We discuss here a framework, first introduced by Gurr in [9], which arises from an examination of interpreting these terms more precisely for diagrams.

To say a *homomorphism* exists between two things of the same kind is to say that they possess the same structure. In mathematics, to be precise, a homomorphism is a mapping between the elements of two algebras $\mathcal{A}$ and $\mathcal{B}$, such that the structure of relations over elements of $\mathcal{A}$ is preserved by the relations which hold over their corresponding elements in $\mathcal{B}$. Note, however, that if there exists a homomorphism from $\mathcal{A}$ to $\mathcal{B}$, this does not necessarily imply that the two are identical. It may be, for example, that there are elements of $\mathcal{B}$ which are not mapped to by any element of $\mathcal{A}$. Alternatively, it may be that there are individual elements of $\mathcal{B}$ which are mapped to by more than one element of $\mathcal{A}$. If there exists a homomorphic mapping between algebras $\mathcal{A}$ and $\mathcal{B}$ and *every* element of $\mathcal{B}$ is mapped to by a *unique* element of $\mathcal{A}$ then the two are, effectively, identical and are said to be *isomorphic*.

For example, consider the diagram in Figure 9 (taken from [9]) as a representation of four integers (for example, 1–4) and their ordering, where integers are represented by squares labelled with capital letters and the less-than relation is represented by the transitive closure of *is-arrow-connected*. Here we may use the algebras $(\{1,2,3,4\},\{<\})$ and $(\{A,B,C,D\},\{\text{transitive closure of } \textit{is-arrow-connected}\})$ to describe the represented, and representing (the diagram), worlds respectively. One possible representational mapping between these worlds maps 1–4 to A–D, respectively, and $<$ to the transitive closure of *is-arrow-connected*. Alternatively, the mapping may run in the opposite direction. It may be readily seen that this mapping, in either direction, is isomorphic. Note however, as the subsequent examples shall illustrate, that while this particular diagram is an isomorphism, the system of representation it illustrates (integers represented by labelled squares; less-than represented by the transitive closure of is-arrow-connected) is not isomorphic in all cases. This is an important point, to which we shall return in Section 3.2.

Examining the consequences of varying degrees of strength of matching between properties of relations in the represented and representing worlds amounts to an examination of the

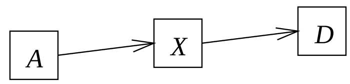

Figure 10: Non-lucid diagram.

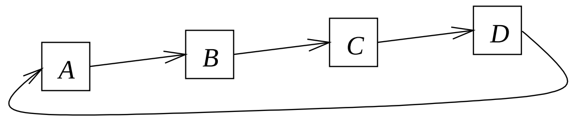

Figure 11: Unsound diagram.

correlation between representing and represented worlds. In extreme cases this correlation may amount to an isomorphic mapping, yet there are numerous ways in which it may fail to be an isomorphism. We do assert, however, that for most purposes non-homomorphic representational systems are likely to be unreasonably intractable, as there would be no guarantee of any connection between the representing and represented objects and relations. Consequently, at the least we should assume that the correlation be an homomorphism. [9] defines four properties: lucidity, soundness, laconicity and completeness which are all required to hold for a homomorphic correlation to be an isomorphism. We paraphrase the definitions of these properties here.

N.B. *In the following examples of Figures 10–13 (adapted from [9]) the world being represented is assumed to be identical to that of Figure 9; that is, the algebra* $(\{1,2,3,4\},\{<\})$. *Similarly, the squares labelled A–D, whenever they appear in these figures, are taken to represent the integers* 1–4 *respectively.*

Definition 1: A mapping from world to representation is injective iff every object (respectively, relation) in the representation represents at most one—although perhaps no—object (respectively, relation) in the represented world. We call such representations *lucid*. A non-lucid representation is one in which some object or relation represents more than one object or relation in the represented world. The diagram in Figure 10, if we assert that the square labelled X represents both of the integers 2 and 3, is an example of an non-lucid diagram.

Definition 2: A mapping from world to representation is surjective iff every object (respectively, relation) in the representation corresponds to at least one object (respectively, relation) in the represented world. We call such representations *sound*. An unsound representation is one in which some object or relation has no correspondence with an object or relation in the represented world. The diagram in Figure 11 is unsound, as the arc connecting the squares labelled D and A represents no corresponding relation in the object world.

Definition 3: A mapping from representation to world is injective iff every object (respectively, relation) in the represented world is represented by at most one—although perhaps no—object (respectively, relation) in the representation. We call such representations *laconic*. A non-laconic representation is one in which some object or relation in the represented world is represented more than once. The diagram of Figure 12, if we assert that the single integer 3 is represented by *two* squares labelled $C1$ and $C2$, is an example of a non-laconic diagram.

Definition 4: A mapping from representation to world is surjective iff every object (respec-

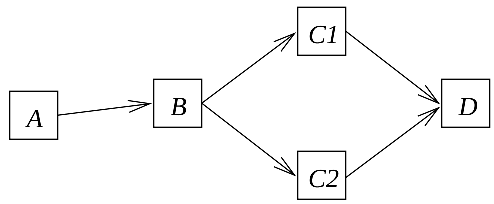

Figure 12: Non-laconic diagram.

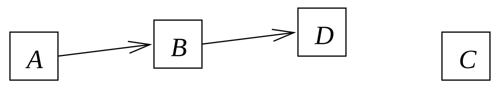

Figure 13: Incomplete diagram.

tively, relation) in the represented world is represented by at least one object (respectively, relation) in the representation. We call such representations *complete*. An incomplete representation is one in which some object or relation in the represented world is not represented. The diagram of Figure 13 is incomplete, as the relations between the integer represented by the square labelled C and the other integers are not represented.

3.2 Enforcing isomorphism of diagrams

Following the discussion of Section 1.3, we argue that the stronger the match between diagram and represented world, the easier the diagram is to reason with and from. Easiest, in fact, if they are isomorphisms. Let us assume that the correlation between a diagram and that which it represents is an isomorphism. The implication of this for the human agent who interprets the diagram is that their interpretation correlates precisely and uniquely with the world being represented. By contrast, where the correlation is not an isomorphism then there may potentially be a number of different target worlds which would match the interpretation. A major issue which must be addressed by a unified theory of diagrammatic reasoning is just *how* this correlation may be guaranteed to be an isomorphism. Note that to assess whether a diagram or system of diagrams is an isomorphism, one must determine what it is that the diagram or system is intended to be isomorphic to. Typically, this requires an understanding not only of the domain which the diagram represents, but also of the task or purpose for which the representation is to be put.

The remainder of this section examines the various means through which representations may be constrained or enhanced so as to ensure lucidity, soundness, laconicity and completeness. Two dimensions of constraints on, or enhancement to, interpretation are examined: origin (whether intrinsic to the representation or external) and point of operation (whether at the level of individual diagrams, or of a diagrammatic system of representation).

In the previous sections we have seen how diagrammatic notations, being directly interpreted, have potentially more *intrinsic* properties than sentential ones. Diagrams, it has been noted, have a head start over sentential representations in the systematicity stakes. Such intrinsic systematicity is thus the initial means through which isomorphism may be achieved in diagrams. *However, while this intrinsic systematicity may often contribute to isomorphism at the level of individual diagrams, this benefit is only partially carried over at the level of systems of representation.* When intrinsic constraints are insufficient to enforce isomorphism

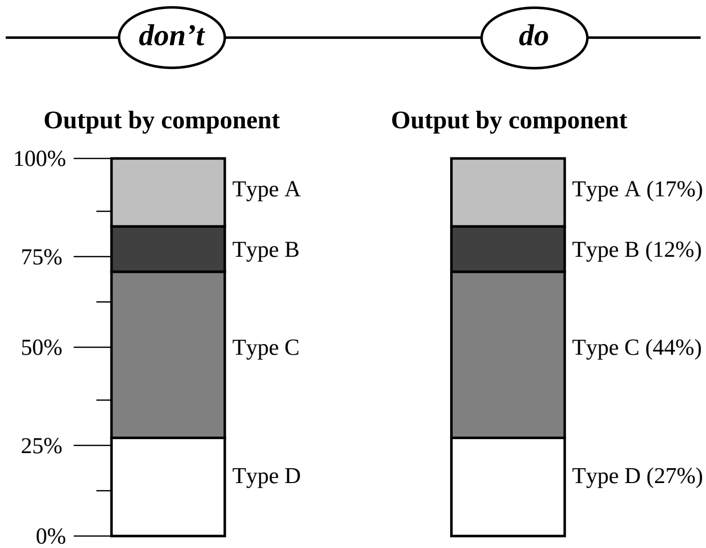

Figure 14: Non-lucid Graph: In the ***don't*** graph both relative and absolute values are represented by area. In the ***do*** graph absolute values are (also) represented separately.

it is typical, as we shall illustrate, for this deficiency to be compensated for by external constraints. Often it is pragmatic assumptions concerning the 'real' semantics of a diagram which constrain or enhance its interpretation as required. These pragmatic constraints, naturally, typically operate at the level of systems of diagrammatic representation.

3.3 Lucid diagrams

For a diagram to be non-lucid suggests that some single object or relation in it represents more than one object or relation in the represented world. This would be potentially highly confusing, and is a clear violation of the Gricean cooperative principle which, in its detailed form, requires that contributions be neither ambiguous nor obscure. Furthermore, it is in fact harder to conceive of a violation of the property of lucidity than of the other three properties examined here. This can be seen by the fact that while previous studies of diagrammatic communication have identified significant properties of diagrams which can be directly related to soundness, laconicity and completeness; no such correlation to lucidity exists.

As an example of the peculiar nature of lucidity, consider Kosslyn's manual for graph design [13], a set of design "do's" and "dont's" presented from the perspective of principles of perception and cognition. In this manual, of 74 examples of 'dont's' (generally supported by corresponding 'do's') in graph design, all of which may be described in some way as examples of non-isomorphic (and on occasion even non-homomorphic) representations, it is interesting to note that only 4 of these examples can be interpreted as, at least in part, non-lucid representations. Figure 14 is an example of a non-lucid graph, based on that of [13] p. 29. Here, in the left-hand ***don't*** graph, the relation *relative value* and the property (unary relation) of *absolute value* are both captured by the same representing relation *area-of-graph*. By contrast, in the right-hand ***do*** graph the scale has been removed and the labels supplemented with numerical values, thus providing a means of representing absolute value which is distinct from the representation of relative value.

Lucidity may thus be viewed as a property of representational systems so eminently

Figure 15: Diagram with intrinsic representation of integer ordering.

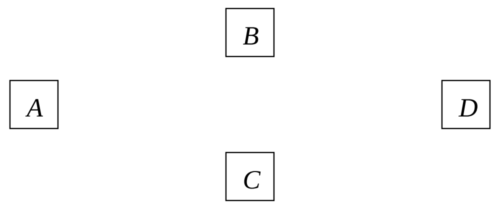

Figure 16: Non-isomorphic diagram of integer ordering.

desirable that it often assumed (as the cooperative principle implies) to be a given. A system of representation should be a priori *designed* so as to (preferably clearly) automatically distinguish between distinct objects and relations in any given representation. Choosing non-lucidity as a feature of a representational system is not a decision that should be taken lightly, and care must be taken to make such a feature apparent to any user.

3.4 Sound diagrams

To be unsound, by our definition, a representational system must be flexible enough to permit the construction of a diagram which does not represent any possible world. That is, a representational system which permitted construction of valid, well-formed and yet inconsistent representations. Our example system for representing the less-than relation among sets of integers is one such system, as the diagram in Figure 11 indicated.

Consider, however, a second representation system for integer ordering, illustrated by the diagram in Figure 15. In this diagram the less-than relation is represented by left-of. The relation left-of is isomorphic in properties to the less-than relation, being irreflexive, anti-symmetric and transitive. We follow Palmer [21] in referring to such isomorphisms as *intrinsic* representations, although they have also been termed *higher-order* homomorphisms by Barwise and Etchemendy [2] and *second-order isomorphisms* by Shepard [24].

Note that representing a relation such as less-than by an isomorphic relation such as left-of does not by itself guarantee that every representation which can be so constructed is an isomorphism. For example, while the diagram of Figure 15 is an isomorphism, not every diagram constructed from this representational system need be so. Consider the diagram of Figure 16, constructed by the same representational system. In this diagram neither of the squares labelled B and C is left of the other. Thus neither integer represented by these two squares is less than the other. Assuming that this representational system illustrated here is homomorphic then the only possible interpretation is that the squares labelled B and C represent the same integer. Alternatively, the diagram could be seen as incomplete, and thus non-isomorphic, if vertical layout (not specified in semantics, but potentially a 'secondary notation') captured *partial* information: that the ordering of B and C was unknown. This example illustrates that the representational system used here is not necessarily laconic, as diagrams can be drawn in which one object in the world is represented by more than one object in the diagram. Thus while certain diagrams constructed using this representational system may be isomorphisms, the system is not isomorphic in all cases.

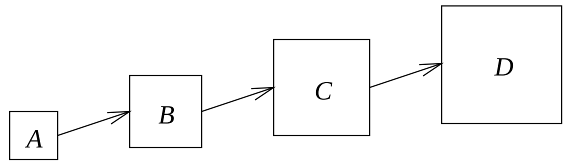

Figure 17: Diagram with redundancy

Stenning *et al* distinguish between those representational systems where such constraints are inherent with those where such constraints must be explicitly imposed, terming the former *self-consistent* systems. For example, the above representation system is clearly not self-consistent. By contrast the representation system illustrated by Figure 15 and 16 is self-consistent. In general, however, it is doubtful that, for anything other than a relatively simple system, such constraints will be totally intrinsic.

Where a representational system such as that illustrated by Figure 15 may produce non-isomorphic diagrams such as in Figure 16, this potential for non-isomorphism must be removed by external constraints. In this case, where the representation system is (in a sense) too expressive, external constraints over the representation system could specify that such diagrams are invalid, or "bad" diagrams. Such external constraints may typically be explained as being motivated by pragmatic considerations; specifically the avoidance of false, or true yet superfluous, information.

3.5 Laconic diagrams

To state that the mapping of objects in a diagram to corresponding objects in the represented world is laconic, is to imply that distinct objects in the diagram refer to distinct objects in the represented world. This property, typically referred to as *icon identity*, very often holds of diagrams. Indeed, in the seminal work of Larkin and Simon [14] it is taken as one of the major defining characteristics which differentiates diagrammatic from textual representations. Diagrams are typically token-referential systems, whereas texts are typically token referential. In token referential systems different tokens of the same type will refer to distinct objects whereas in type referential systems different tokens of the same name-type refer to the same object(following Stenning *et al* [29]). For example, most textual representations are type referential; so that, in a sentence, multiple occurrences of the name-token "Edinburgh" refer equally to the same object (the capital of Scotland).

For the mapping from relations in the diagram to relations in the world not to be injective there must be some relation in the world which is represented more than once in the diagram. Thus some aspect of the diagram would be redundant. Marks and Reiter [16] claim that such redundancy is necessary in certain cases to enforce the correct interpretation of certain relations in the diagram as meaningful. This is offered as a potential solution to the problems of ensuring soundness of relations-sets which were caused by the multitude of potential relations in diagrams that were not a part of the designer's intent. Figure 17 illustrates such an approach where the less-than relation among integers is represented both by the transitive closure of is-arrow-connected and by is-smaller-than. Naturally such a deliberate redundancy must be used with caution to avoid over-emphasising some relation and causing the reader to infer that some further relation must exist (i.e. too strong an emphasis may cause exactly that problem which it was intended to cure).

Examining the example (not presented here) cited by Marks and Reiter, in support of their claim that such redundancy is necessary, it can be seen that rather than using redundancy to emphasise the relevant graphical relations they are in fact effectively using normalisation of irrelevant graphical relations to suppress them. This suppression of irrelevant relations is a particular issue of diagrams precisely because diagrams are typically *overly-specific* in their representation of information. More information than is strictly necessary for the task at hand is typically supplied. For example, where the representation of a triangle is called for, a sentential representation can merely introduce us to "a triangle", and not suggest any further information regarding its properties. A diagram in which a triangle is drawn, however, necessarily will specify whether this triangle is equilateral, right-angled, or otherwise. According to Grice's cooperative principle, apparent relations will be interpreted as meaningful (such as the apparent ordering between symbols in Marks and Reiter's example of Figure 7), as will apparent irregularities (such as those in Marks and Reiter's example of Figure 8). It is by putting extra emphasis on only those relations which are relevant to the task at hand, and by normalising the irrelevant relations, we argue, that a diagrammatic system avoids the possibility of erroneous interpretations of relevance.

3.6 Complete diagrams

Ensuring completeness in the representation of objects amounts to ensuring that every (relevant) object in the represented world appears in the diagram. In certain cases this may be a property which is intrinsic to a given diagrammatic system, although it is difficult to see how this might be true in general. What is certainly true, however, is that in accordance with the detailed account of Grice's cooperative principle (specifically, that all necessary information is included) most diagram designers should attempt to ensure completeness of objects, as most readers would typically assume this to be true.

For relations in the diagram to be incomplete, there must be relations in the world which are not represented in the diagram. Without doubt this is often the case. However, when assessing isomorphism with respect to a given task – and domain – completeness becomes a matter of assessing whether every relation deemed relevant to the task is included in the diagram. Thus in our example system which represents the less-than relation over sets of integers we do not represent properties such as is-even and is-odd, nor do we represent relations such as is-a-multiple-of. These properties (unary relations) and relations exist in the world but are not part of the intent of the diagram and thus not represented.

The decision of which relations in the world are relevant is one made by the designer of the notation. In practice, however, relations not considered relevant by the designer may be important to users of a representation system. Uses of semantically unspecified diagrammatic features are, in this case, an example of objects or (more commonly) relations not deemed relevant by the designer(s) of a representational system but which are deemed relevant by the users of that system. For example, the spatial grouping of related objects exhibited by expert users of CAD-E systems [23], or of visual programming environments [22].

In our node-and-link diagrams for integers we also have not made use of spatial layout and thus a diagram such as that of Figure 18, while sound, may contain additional information (both by virtue of spatial grouping and symmetry of groups). For example, we choose an interpretation of this diagram where the squares labelled A, B, C and D represent the integers 5, 6, 130 and 131 respectively. While the arrows alone cannot help to represent the fact that these four integers fall naturally into two groups of "close" integer-pairs, this information has been represented in Figure 18 by exploiting a spatial relation (length of arrow) as carrying a pragmatic implicature. In the case of established representation systems such pragmatic features form part of the domain knowledge exploited by experienced users.

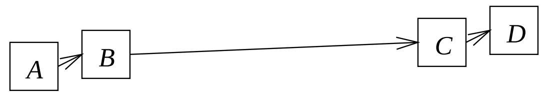

Figure 18: Diagram with meaningful spatial grouping

4 Summary and conclusions

This paper has investigated the issues concerned in determining what makes for effective diagrammatic communication. Given that, as yet, there does not exist the fully fledged foundation to a theory of diagrammatic communication which such an investigation requires, the first step has been to sketch out the component issues of such a theory. Secondly, a framework in which to explore properties of representations, and properties of the relation between representations and that which they represent, has been reviewed. The aims of this paper have been achieved, we hope, by relating the issues which determine the effectiveness of representations, to this framework for exploration, in the context of the sketched theory.

As has been suggested, the study of systems of communication may be divided into three parts: syntax, semantics and pragmatics. An analogy of the resulting tri-partite study of textual (written) communication provides a useful starting point for sketching out the issues in a theory of diagrammatic communication. Examining the similarities and differences which this analogy highlights has permitted us to lay out the issues which make up a theory of diagrammatic communication into similar syntactic, semantic and pragmatic components. Previous exploration of these components has suggested that the effectiveness of a system of representation is determined by the constraints which govern its interpretation. Specifically, there are three dimensions of constraints which are: their origin (whether intrinsic to the system or externally imposed); their point of operation (whether at the level of specific representations or at a meta-systemic level); and their availability to the human reader/reasoner.

We have argued that a significant determinant of effectiveness in representational systems is the degree of closeness of match of structure and properties in a representation to that which it represents. We have furthermore related a framework which permits exploration of this closeness of match to the concepts of constraints over interpretation, in the context of our diagrammatic theory. Examining the implications of such an exploration in the context of a theory of diagrammatic communication has enabled us to link together a number of previous studies and show how they pertain to the assessment of effectiveness. In particular, our exploration of the properties of lucidity, soundness, laconicity and completeness has related studies of the effectiveness of diagrams when viewed as analogical representations (intrinsic systematicity, in our framework); and studies from the perspective of perceptual and cognitive properties of diagrams, such as icon-identity and the expert usage of pragmatic features.

Finally, we note that we have not, at any great length, discussed the third constraint upon interpretation which determines effectiveness: that of the availability of properties of diagrammatic representations to the human reader/reasoner. Availability is clearly significantly determined by the constraints of origin and point of operation, and thus its impact upon effectiveness is to a certain extent indirectly determined by the impact of these previous two constraints. However, availability is also substantially determined by perceptual and other cognitive issues, thus the story related here is as yet incomplete. Work remains to be done in both fleshing out the details of the theory of diagrammatic communication sketched here; and in exploring what impact the availability of properties of diagrams has in completing our account of the issues which pertain to effectiveness in diagrammatic representations.

5 Acknowledgements

The development of the ideas presented here has been greatly assisted by numerous discussions with colleagues in the Human Communication Research Centre; in particular the contributions of John Lee, Jon Oberlander and Keith Stenning have proved invaluable. The author also owes a substantial debt of gratitude, and offers sincere thanks, to the anonymous reviewers of an earlier draft of this paper. Their detailed and insightful comments highlighted the need for clarification of a number of fundamental issues, and contributed greatly to the structuring of this revised presentation. This work was supported by EPSRC Grant # GR/L37953 "Understanding Software Architectures".

References

[1] G. Allwein and J. Barwise. *Logical Reasoning with Diagrams.* Oxford University Press, New York, 1996.

[2] J. Barwise and J. Etchemendy. Heterogeneous logic. In [6], pages 211–234. 1995.

[3] K. J. Campbell, K. F. Collis, and J. M. Watson. Visual processing during mathematical problem solving. *Educational Studies in Mathematics*, 28:177–194, 1995.

[4] G. Gazdar. *Pragmatics: Implicature, presupposition and logical form.* New York: Academic press, 1979.

[5] D. Gentner. Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, 7:155–170, 1983.

[6] J. Glasgow, N. H. Narayan, and B. Chandrasekaran. *Diagrammatic Reasoning: Cognitive and Computational Perspectives.* MIT Press, 1995.

[7] H. Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics: Vol 3, Speech Acts*, pages 43–58. New York: Academic Press, 1975.

[8] H. Grice. Further notes on logic and conversation. In P. Cole, editor, *Syntax and semantics 9: Pragmatics*, pages 113–128. New York: Academic Press, 1978.

[9] C. Gurr. On the isomorphism, or lack of it, of representations. In K. Marriot and B. Meyer, editors, *Theory of Visual Languages*, chapter 10. Springer Verlag, 1998.

[10] C. Gurr. Theories of visual and diagrammatic reasoning: Foundational issues. In G. Allwein, K. Marriot, and B. Meyer, editors, *AAAI Fall Symposium: Formalizing Reasoning with Visual and Diagrammatic Representations.* AAAI Press, 1998.

[11] C. Gurr, J. Lee, and K. Stenning. Theories of diagrammatic reasoning: distinguishing component problems. *Mind and Machines*, 8(4):533–557, 1998.

[12] M. Hegarty. Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5):1084–1102, 1992.

[13] S. M. Kosslyn. *Elements of graph design.* W. H. Freeman and Co., New York, 1994.

[14] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11:65–99, 1987.

[15] S. C. Levinson. *Pragmatics*. Cambridge University Press, 1983.

[16] J. Marks and E. Reiter. Avoiding unwanted conversational implicature in text and graphics. In *Proceedings of AAAI-90*, pages 450–456, 1990.

[17] C. W. Morris. Foundations of a theory of signs. In O. Neurath, R. Carnap, and C. Morris, editors, *International encyclopedia of unified science*, pages 77–138. Chicago University Press, 1938.

[18] K. Myers and K. Konolige. Reasoning with analogical representations. In [6], pages 273–302. 1995.

[19] N. H. Narayan. Diagrammatic communication: A taxonomic overview. In B. Kokinov, editor, *Perspectives on Cognitive Science*, volume 3, pages 91–122. New Bulgarian University Press, Sofia, Bulgaria, 1997.

[20] J. Oberlander. Grice for graphics: pragmatic implicature in network diagrams. *Information design journal*, 8(2):163–179, 1996.

[21] S. E. Palmer. Fundamental aspects of cognitive representation. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorisation*, pages 259–303. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

[22] M. Petre. Why looking isn't always seeing: Readership skills and graphical programming. *Communications of the ACM*, 38(6):33–45, June 1995.

[23] M. Petre and T. R. G. Green. Requirements of graphical notations for professional users: electronics CAD systems as a case study. *Le Travail Humain*, 55:47–70, 1992.

[24] R. N. Shepard and S. Chipman. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1:1–17, 1970.

[25] A. Shimojima. Operational constraints in diagrammatic reasoning. In [1], pages 27–48. 1996.

[26] S-J. Shin. *The Logical Status of Diagrams*. Cambridge University Press, 1995.

[27] J. F. Sowa. Relating diagrams to logic. In *Conceptual Graphs for Knowledge Representation: Proceedings of 1st International Conference on Conceptual Structures*. Springer-Verlag LNAI 699, 1993.

[28] K. Stenning, R. Cox, and J. Oberlander. Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, 10, 1995.

[29] K. Stenning and R. Inder. Applying semantic concepts to analysing media and modalities. In [6], pages 303–338. 1995.

[30] K. Stenning and P. Yule. Image and language in human reasoning: a syllogistic illustration. *Cognitive Psychology*, 34(2):109–159, 1997.

[31] D. Wang and J. Lee. Visual reasoning: Its formal semantics and applications. *Journal of Visual Languages and Computing*, 4:327–356, 1993.

[32] J. Zhang and D. Norman. Representations in distributed cognitive tasks. *Cognitive Science*, 18:87–122, 1994.