

# THE COMPUTER REVOLUTION IN PHILOSOPHY (1978)

Aaron Sloman

[Book contents page](#)

This chapter is also available in [PDF format here](#).

---

## CHAPTER 7

### INTUITION AND ANALOGICAL REASONING

The previous chapter listed varieties of information that must be represented in an intelligent system. Nothing was said about how different types of symbolism could be used for different purposes. This chapter explores some of the issues, relating them to philosophical debates about inference and reasoning.

#### **Note:**

This is a revised version of

A. Sloman, (1971) 'Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence', in *Proceedings 2nd IJCAI (1971)* Reprinted in *Artificial Intelligence*, vol 2, 3-4, pp 209-225, 1971, and in J.M. Nicholas, ed. *Images, Perception, and Knowledge* Dordrecht-Holland: Reidel. 1977

Also available online <http://www.cs.bham.ac.uk/research/cogaff/04.html#analogical>

See [notes at end](#) for related papers written later.

#### **7.1. The problem**

Within philosophy, there has long been a conflict between those who, like Immanuel Kant, claim that there are some modes of reasoning, or inferring, which use 'intuition', 'insight', 'apprehension of relations between structures', etc., and those who argue that the only valid methods of inference are logical, for instance the use of syllogisms and rules of predicate calculus. This dispute is relevant to problems in psychology, concerning non-verbal forms of thinking and remembering (for example, the problem whether there is such a thing as 'iconic' memory).

It is also relevant to problems about the nature of mathematics and science. For instance, many mathematicians adopt a logicist' position and argue that the only acceptable mathematical proofs are those using the formalisms and inference rules of symbolic logicians. They claim that where diagrams, or intuitively grasped models are used, these are merely of 'psychological' interest, since, although they shed light on how people arrive at valid proofs, the *real* proofs do not contain such things. According to this viewpoint, the diagrams in Euclid's *Elements* were strictly irrelevant, and would have been unnecessary had the proofs been properly formulated. (For some counter-arguments, see Mueller, 1969.)

This issue is clearly relevant to teachers of mathematics and science. Teachers who accept the logicist' position will be inclined to discourage the use of diagrams, pictures, analogies, etc., and to encourage the use of logical notations, and proofs which are valid according to the rules of propositional and predicate logic.

Kant's theories were opposed to this logicist position, insofar as he argued that important kinds of mathematical knowledge could be both *a priori* and *synthetic*, that is, non-empirical and non-analytic. I think he had an important insight, though it has not been possible until recently to say very clearly what it was. The issues can be clarified by discussing different kinds of symbolisms, or representations, and their roles in various kinds of reasoning. Some irrelevant metaphysical digressions can be avoided by noting that such reasoning can occur in computers, as well as in human minds.

One interpretation of what Kant was trying to say is that we sometimes, for instance in mathematical thinking, use non-verbal 'analogical' representations, and make inferences by manipulating them, instead of always using logic. His claim is that these non-logical (but not illogical) modes of thinking may be valid sources of knowledge.

This topic is closely related to current problems in artificial intelligence, for it turns out that different forms of representation may differ greatly in their computational properties.

In particular, methods of representation and inference which meet the approval of logicians will not necessarily be the best ones to use in a computer program which is to behave intelligently. Not all workers in A.I. would accept this. For example, McCarthy and Hayes (1969) argued that an intelligent computer program will need to be able to prove by methods of logic that a certain strategy will achieve its goal. They claimed that this would be an essential part of the process of decision making. I doubt whether they still hold the same views (see Hayes, 1974), but the position they once advocated is worth refuting even if they have changed their mind, since it is very close to the views of many philosophers, especially philosophers of science.

## ***7.2. Fregean (applicative) vs analogical representations***

The main point I wish to make in this chapter is that there are many different types of language, or representational system, and many different ways of making inferences by manipulating representations. The forms of inference codified by logicians are relevant only to languages of the type analysed by Gottlob Frege (see Bibliography), in which the basic method of constructing complex symbols is by applying function-signs to argument-signs. Much mathematical and logical notation, and many (though not all) of the constructions of natural languages are Fregean. For instance, a first rough Fregean analysis of 'Mary shot Tom's brother' would be something like:

### **Shot (Mary, the-brother-of (Tom))**

where the predicate 'shot' is treated as a two-place function and 'the brother of' as a one-place function. Pictures, maps, diagrams, models, and many of the representations used in computer programs are not Fregean. Some of them are 'analogical'.

This contrast between Fregean (or 'applicative') and analogical representations will be more precisely defined later. It is often referred to by people who do not know how to characterise it properly. For instance, it is sometimes assumed that analogical representations are continuous and the others discrete, or that analogical representations are essentially non-verbal (that is, that verbal languages do not use them), or that analogical representations are isomorphic with what they represent. These mistakes (which will be exposed later) also go along with a tendency to assume that digital computers

cannot construct or use analogical representations. (See the writings of Pylyshyn.)

Terminology is also often confused. What I have called 'Fregean' or 'applicative' representations are sometimes called 'symbolic', 'linguistic', 'formal', 'propositional', or 'verbal'.

The word 'symbolic' is unsatisfactory, since the ordinary use of 'symbols', 'symbolism' and 'symbolic' is much more general (for example maps can be said to be symbolic, even though they are analogical). I shall use 'representation' and 'symbol' and their derivatives more or less interchangeably as very general terms, and will refer to any system of representation or symbolism as a language, as in 'the language of maps'. I shall use 'Fregean' and 'applicative' interchangeably.

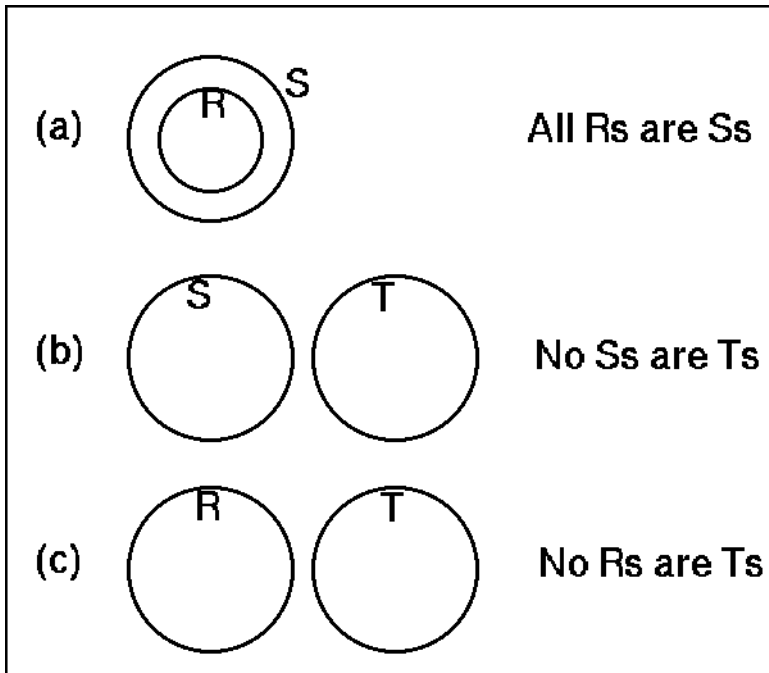
One of the main aims of this chapter is to show that inferences made by manipulating non-Fregean representations may be perfectly valid. I believe this is at least part of what Kant and Intuitionist mathematicians (for example Brouwer) were trying to say.

Before developing the point in detail, I would like to stress that I am not taking sides in the dispute among psychologists who argue over whether people use 'iconic' forms of memory, and reason with images. I believe that contributions from both sides are often riddled with confusions, related to the mistakes referred to above. It is especially important to notice that the points I make about analogical representations are quite neutral on the question whether such representations occur in the mind or not. Even if they occur only on paper (for example in maps and diagrams) the point is that they can still be used in valid reasoning.

Useful discussion of these issues is impossible without careful definitions of some of the main concepts, such as 'valid', 'inference', 'logic', 'verbal', 'analogical', 'Fregean' (or 'applicative'). However, before attempting to be more precise, I shall present a few examples of reasoning with non-Fregean symbolisms.

### ***7.3. Examples of analogical representations and reasoning***

We can reason about set-theoretical relationships using Euler's circles. Suppose we use a circle marked R to represent people in a certain room, a circle marked S to represent students, and a circle marked T to represent taxpayers. Then in figure 1, the three diagrams (a), (b) and (c) all represent possible states of affairs. Geometrical relations in the picture analogically represent relations between sets of people. Whether any one of them represents the way things are in the world is a contingent matter, a matter of fact. It depends, in the case of (a) and (c), on who is in the room at the time in question. This is analogous to the way in which the truth-value of a sentence depends on how things are in the world.



**Figure 1**

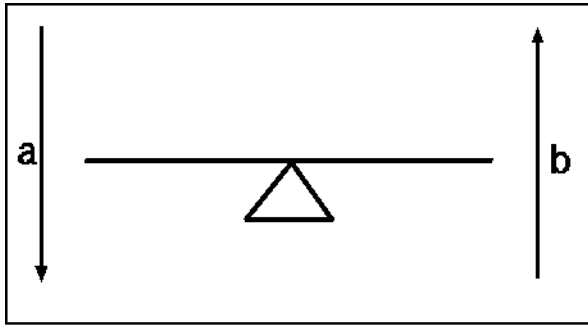
Whether a picture correctly depicts the world is, in each case, a contingent question which can only be answered by examining the world; but we can still discover, without examining the world, that certain *combinations* of correctness and incorrectness are necessarily ruled out. For example, no matter how things are in the world, we can use our understanding of the methods of representation employed in such diagrams to discover that it is impossible for (a) and (b) correctly to represent how things are, while (c) does not, given the stated interpretations of the diagrams. This has to do with the impossibility of creating a diagram containing (a) and (b) simultaneously, without the relation (c). *How* we discover this is not obvious, but *that* we can is.

We are also able to use our understanding of the syntax and semantics of English to tell that the following argument is valid:

All the people in the room are students.  
 No students are taxpayers.  
**Therefore:** No people in the room are taxpayers.

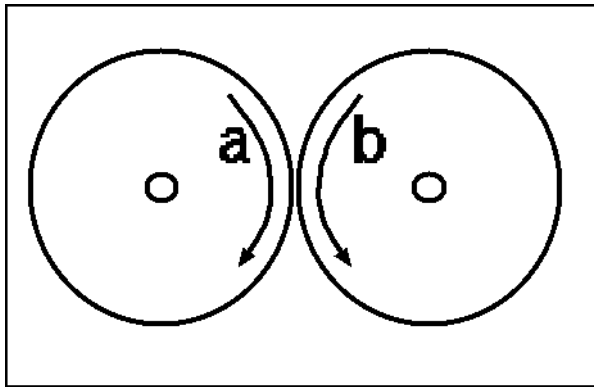
In both the verbal and the diagrammatic representation there are problems about possible ambiguities of reference or meaning. In both cases it is hard for people to explain why the inferences are valid. Nevertheless, we can tell that they are, and the study of such reasoning has occupied great logicians since Aristotle, leading to many logical symbolisms designed to capture the essential form of a variety of inferences.

It is worth remarking that when Euler's circles are used for this kind of reasoning, the three diagrams of figure 1 are normally superimposed in one diagram. This makes it harder to perceive that a method of reasoning from 'premisses' to a 'conclusion' is involved. By contrast, in verbal arguments the premisses and conclusion normally have to be formulated separately. In some of the examples which follow, I shall collapse the different representations involved into one diagram or picture, in the usual way.



**Figure 2**

Here are some more examples. In figure 2 the horizontal straight line is to be interpreted as representing a rigid straight rod, pivoted at the middle on a fixed support. In figure 3 each circle represents a rigid wheel free to rotate about a fixed axle passing through its centre, and contact between circles represents contact *without slipping* between the wheels.



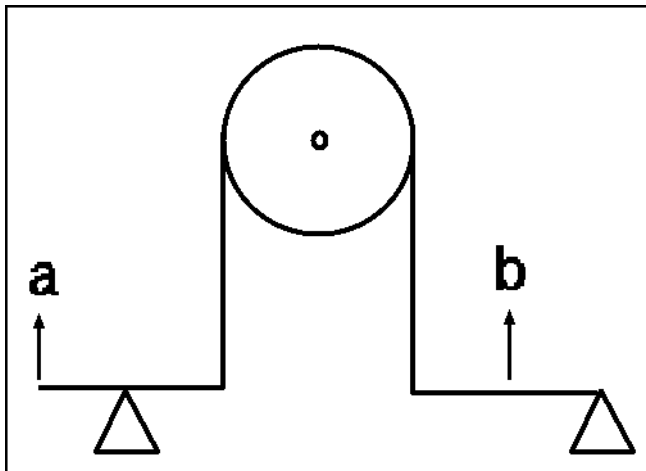
**Figure 3**

In both figure 2 and figure 3 the arrows represent direction of motion, (of what? how can you tell?), so the figures represent changing configurations. However, the arrows labelled (a) are to be interpreted as assumptions, or premisses, and the arrows labelled (b) are to be interpreted as conclusions, inferred from the rest of the picture. In both cases, we can consider a bit of the world depicted by the diagram and ask whether the arrow (a) correctly represents what is happening, and whether arrow (b) correctly represents what is happening. In each case, it is a contingent matter, so empirical investigation is, required to find out whether the representation is correct. (Just as empirical investigation may be used to check the truth of premisses and conclusion in a logical argument.)

However, we can tell non-empirically that it is impossible for arrow (b) to be an *incorrect* representation while arrow (a) and the rest of the diagram represents the situation correctly given the specified interpretations of the arrows, and other features of the pictures. So we can say that the inferences from (a), and the rest of the picture, to (b) is valid, in both figure 2 and figure 3. Both examples could have been replaced by two separate pictures, one containing only arrow (a) and one containing arrow (b), as in figure 1.

Far more complex examples of inferences about mechanical systems, using diagrams could be given. Figure 4 is relatively simple. In figures 4 and 5, horizontal lines again represent rigid levers pivoted at the points indicated by small triangles. The circles represent pulleys free to rotate about their centres,

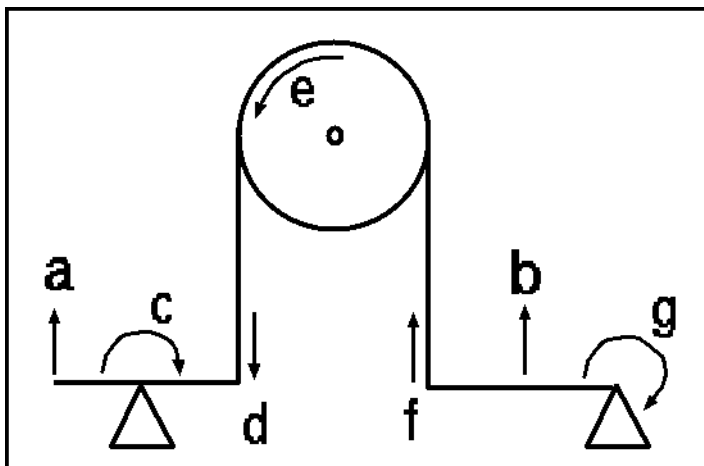
but not free to move up or down or sideways.



**Figure 4**

The vertical lines, apart from arrows, represent inelastic flexible strings, and where two such lines meet a pulley on either side, this represents a string going round the pulley. Where a vertical line meets a horizontal line, this represents a string tied to a lever. As before, the arrows represent motion of the objects depicted by neighbouring picture elements. Once again, we can see that what is represented by the arrow marked (b) can be validly inferred from what is represented by the arrow marked (a) and the rest of the picture.

Where the inference is more complicated, some people may find it harder to discern the validity. In the case of logical or verbal inferences, this difficulty is dealt with by presenting a proof, in which the argument is broken down into a series of smaller, easier arguments. Something similar can be done with an argument using a diagram.



**Figure 5**

For example, figure 5 is just like figure 4, except for additional arrows. The arrows marked (c), (d), (e), (f) and (g) can be taken as representing intermediate conclusions, where each can be validly inferred from the preceding one, and (c) can be inferred from (a), and (b) from (g). Using the transitivity of valid implication, we see that (b) is validly inferrable from (a). Notice that it is not

always immediately obvious what can and what cannot be validly inferred. For instance, if the length of an arrow represents speed of motion, do the inferences remain valid?

It is possible to give a computer program the ability to reason about mechanics problems with the aid of such diagrams. To do so would require us to formulate quite precise specifications of the significant properties and relations in the diagrams, and the rules for interpreting them, so that the computer could use these rules to check the validity of the inferences. Funt (1976) has done this in a program which makes inferences about falling, sliding and rotating objects.

I have experimented with similar programs. Making a program solve problems *intelligently* would involve giving it procedures for searching for significant paths through such diagrams, analogous to the path represented by the arrows (c) to (g), indicating a chain of causal connections relating (a) and (b). Finding relevant paths in complex configurations would require a lot of expertise, of the sort people build up only after a lot of experience. Giving a computer the ability to acquire such expertise from experience would be a major research project given the current state of artificial intelligence. (At the time of writing a group at Edinburgh University, directed by Alan Bundy, is attempting to give a computer the ability to reason about simple mechanical problems described in English.)

I believe that our concept of a causal connection is intimately bound up with our ability to use analogical representations of physical structures and processes. This point is completely missed by those who accept David Hume's analysis of the concept of 'cause', which is, roughly, that 'A causes B' means 'A and B are instances of types of events such that it has always been found that events of the first type are followed by events of the second type'. His analysis explicitly rejects the idea that it makes sense to talk of some kind of 'inner connection' between a cause and its effect. I suspect that we talk of causes where we believe there is a representation of the process which enables the effect to be inferred from the cause using the relations in the representation. The representation need not be anything like a verbal generalisation. However, analysis of the concept 'cause' is not my current task, so I shall not pursue this here.

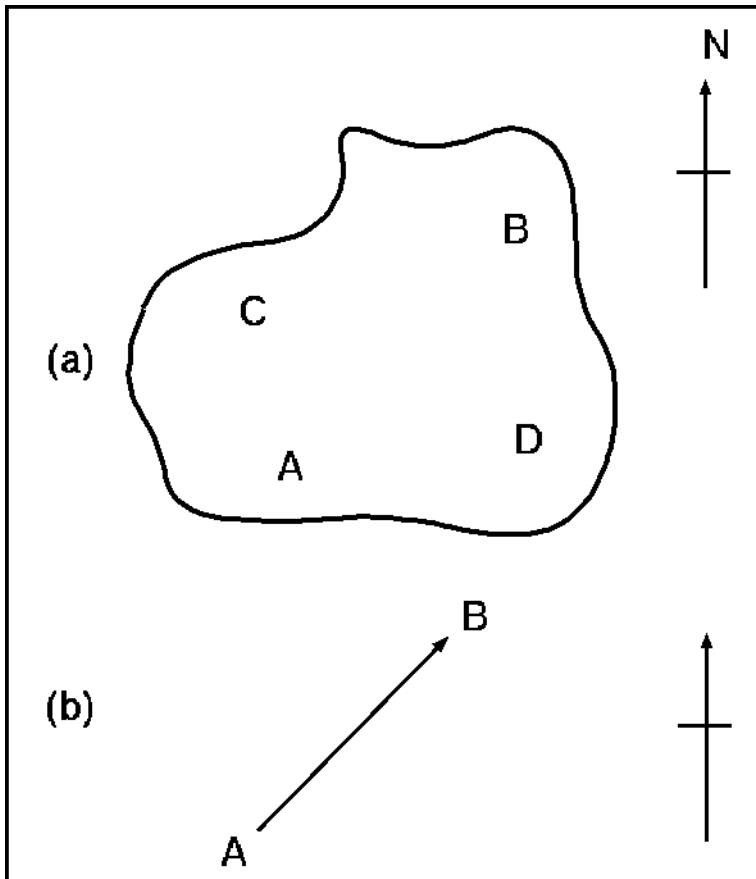
So far my examples of valid reasoning with analogical representations have all used diagrams. It does not matter whether the diagrams are drawn on paper, or on a blackboard, or merely imagined. Neither does it matter whether they are drawn with great precision: detailed pictorial accuracy is not necessary for the validity of examples like figure 4. It is also worth noting that instead of looking at diagrams (real or imagined), we can sometimes do this kind of reasoning while looking at the physical mechanism itself: the mechanism can function as a representation of itself, to be manipulated by attaching real or imaginary arrows, or other labels, to its parts.

So by looking at a configuration of levers, ropes and pulleys, and finding a suitable chain of potential influences in it, we can draw conclusions about the direction of motion of one part if another part is moved.

It is so easy for us to do this sort of thing, for example when we 'see' how a window catch or other simple mechanism works, that we fail to appreciate the great difficulty in explaining exactly how we do it. It requires, among other things, the ability to analyse parts of a complex configuration in such a way as to reveal the 'potential for change' in the configuration. We probably rely on the (unconscious) manipulation of analogical representations, using only procedures which implicitly represent our knowledge of the form of the world. This point is closely bound up with the issues discussed in the chapter on the aims of science, where science was characterised as a study of possibilities and their explanations.

## 7.4. Reasoning about possibilities

This ability to use the scenes we perceive as a representation to be in some sense manipulated in making inferences about possible actions and their effects, is central to our ability to get around in the world. For instance, the ability to select a path across a crowded room is analogous to the ability to use a map to select a route from one place to another. Using the map might be unnecessary if we could get a suitable view of the terrain from a helicopter. We frequently use things as representations of themselves!



**Figure 6**

Figure 6 gives a very simple illustration of the use of a map to make a valid inference. It is instructive in that it also shows a relationship between two representations of different sorts.

In (a) we have a map showing a few towns, marked by dots, with the usual indication of compass points. In (b) we have, not a map, but a representation of the direction (and perhaps distance) between two towns. The arrow represents a vector. Once again we can say that (b) may be validly inferred from (a), though now we have to qualify this by saying that the inference is valid only within certain limits of accuracy.

Many different uses of maps are possible. For instance, from a map showing which crops are grown in different parts of a country, and a map showing the altitude of different parts of the country, we can 'infer' a map showing which regions are both corn-producing and more than 100 feet above sea level.

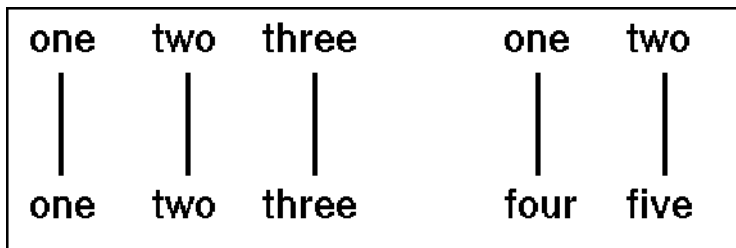
When planning the layout of a room it may be useful to draw diagrams or to make flat movable cardboard cut-outs representing the objects in the room, and to use them to make inferences about the consequences of placing certain objects in certain locations. This has much in common with the use of maps.

This sort of example shows how a representation may be used to reason about what sorts of things are possible. For example, a particular arrangement of the bits of cardboard can be used to show that a certain arrangement of the objects in a room is possible. This is like the use of diagrams in chemistry to show that starting from certain molecules (for example H-H and H-H and O=O), it is possible to derive new molecules by rearranging the parts (giving H-O-H and H-O-H).

### ***7.5. Reasoning about arithmetic and non-geometrical relations***

Reasoning with analogical representations is not restricted to geometrical or mechanical problems. Every child who learns to do arithmetic finds it useful, at times, to answer a question about addition or subtraction by using analogical representations of sets of objects. For example, a child who works out the sum of three and two by counting three fingers on one hand, two fingers of the other, then counting all the fingers previously counted, is reasoning with analogical representations. The same thing can be done with dots, as in figure 7.

An important step in mastering arithmetic and its applications is grasping that number names themselves can be used in place of dots or fingers (that is, 'one two three' followed by 'one two', matches 'one two three four five').



**Figure 7**

The diagram in Figure 7 can be used as a proof that three plus two is Five.

What is the largest possible number of persons who might have been parents of great-grandmothers of yours? What relation to you is your son's daughter's first-cousin? There are various ways you might attempt to answer this sort of question, but one of them involves drawing a fragment of a 'family tree', or possibly several family trees consistent with the problem specification. A family tree diagram is an analogical representation of a bit of the social world. Another example of an analogical representation of a rather abstract set of relationships is a chart indicating which procedures call which others in a computer program. Flow charts give analogical representations of possible processes which can occur when procedures are executed. Both sorts of diagrams can be used for making inferences about what will happen when a program is executed, or when part of a program is altered. A morse code signal is an analogical representation of a sequence of letters.

## ***7.6. Analogical representations in computer vision***

Some people working on computer vision programs have found that it is convenient to use two-dimensional arrays of numbers (representing brightness, for instance) as a representation of a visual image. (See chapter 9 for a simple example.)

Operations on the array, such as examining a set of points which lie on a straight line', or possibly marking such a set of points, make use of the fact that there is a structural relationship between the array and the retinal image. Similarly, when processing of such an image has produced evidence for a collection of lines, forming a network, as in a line drawing of a cube, then it is convenient to build up data-structures in the computer which are linked together so as to form a network of the same structure. A similar network, or possibly even the same one, can then be used to represent the three-dimensional configuration of visible edges of surfaces in the object depicted by the line-image.

Manipulations of these networks (for example attaching labels to nodes or arcs on the network, or growing new networks to represent the 'invisible' part of the object depicted) can be viewed as processes of inference-making and problem solving, with the aid of analogical representations. It may be that something similar happens when people make sense of their visual experiences. (For more on this see the chapter on perception and Clowes, 1971, Waltz, 1975, Winston, 1975, Boden, 1977.)

## ***7.7. In the mind or on paper?***

It should be stressed that most of my examples are concerned with diagrams and other representations which are on paper or some other physical medium. The processes I am talking about do not have to be completely mental, though mental processes will always be involved if the representations are interpreted and used for reasoning. However, in some cases it is possible for the process to be entirely mental, when we merely imagine manipulating a diagram, instead of actually manipulating one. Reasoning of this sort may be just as valid as reasoning done with a real diagram. Unfortunately it is not at all clear what exactly does go on when people do this sort of thing, and introspective reports (for example 'It really is just like seeing a picture') do not really provide a basis for deciding exactly what sorts of representations are actually used. (Pylyshyn, 1973.)

Although we are still very unclear about what goes on in the minds of people, we can understand what goes on in the mind of a computer when it is building arrays or networks of symbols and manipulating them in solving some problem. By exploring such programming techniques we may hope to get a much better understanding of the sorts of theories which could account for human imaginative exercises. Our main lack at present is not data so much as ideas on how to build suitable theories.

The illustrations in the preceding sections should give at least a rough idea of what I mean by saying that sometimes valid reasoning may be done by manipulating analogical representations. Many more examples could be given. It is time now to try to formulate more precise definitions of some of the concepts used.

## ***7.8. What is a valid inference?***

Consider first an inference expressed in sentences in some natural or artificial language. There will be a set of premisses  $P_1, P_2, \dots, P_n$  and a conclusion  $C$ , each of which is a sentence (or is expressed in a sentence). In general, whether a particular sentence says something true or something false, that is, what its truth-value is, depends partly on its form and meaning and partly on how things are in the world. So discovering the truth-value requires the application of verification procedures defined by the sentence and the semantics of the language. So each of  $P_1, P_2, \dots, P_n$  and  $C$  may have its truth-value

determined by 'the world'. In spite of this it may be possible to discover, without examining the world, that is, without applying the usual verification procedures, that there are constraints on the possible combinations of truth-values.

In other words, by *examining* verification procedures, instead of *applying* them, we can discover that certain combinations of truth-values of statements cannot occur, no matter what the world is like. 'London is larger than Liverpool' and 'Liverpool is larger than London' cannot both be true: they are *contraries*. We can discover this by examining the semantics of larger than'. (How is this possible?)

There are many other relationships of truth-values which can be discovered by this kind of non-empirical investigation. For instance, two statements may be incapable of both being false, in which case they are called subcontraries by logicians.

Validity of an inference is a special case of this. Namely, the inference from  $P_1, P_2, \dots, P_n$  to the conclusion  $C$  is valid if and only if relationships between the statements constrain their truth-values so that it is impossible for all the premisses to be true and the conclusion false. So validity of an inference is simply a special case of the general concept of a constraint on possible sets of truth-values, namely the case where the combination

(T, T, .. T: F)

cannot occur. So validity is a semantic notion, concerning meaning, reference, and truth or falsity, not a syntactic notion, as is sometimes supposed by logicians. They are led to this mistake by the fact that it is possible to devise syntactic tests for validity of some inferences, and indeed the search for good syntactic criteria for validity has been going on at least since the time of Aristotle,

It is an important fact about many, or perhaps all, natural languages, that syntactic criteria for some cases of validity can be found. For, by learning to use such criteria, we can avoid more elaborate investigations of the semantics of the statements involved in an inference, when we need to decide whether the inference is valid. The syntactic tests give us short-cuts, but have to be used with caution in connection with natural languages. It is not always noticed that our ability to discern the correctness of these tests depends on a prior grasp of the semantics of key words, like 'all', 'not', 'some', 'if and others, and also a grasp of the semantic role of syntactic constructions using these words. It is still an open question how ordinary people, who have not learnt logic, do grasp the meanings of these words, and how they use their understanding in assessing validity of inferences. (For further discussion see my 'Explaining logical necessity'.)

## 7.9. Generalising the concept of validity

Validity of inferences has been shown to be a special case of the semantic concept of a constraint on possible truth-values of a set of statements, which in turn is a special case of the general concept of a constraint on possible 'denotations' of a set of representations. This provides a basis for giving a general definition of validity.

We have seen from some of the examples of the use of analogical representations, for example, figure 1 and figure 2, that the question whether a particular picture, diagram or other representation correctly represents or 'denotes' a bit of the world is in general an empirical question, which involves using the appropriate interpretation rules to relate the representation and the bit of the world. (Similarly, the truth of what a sentence says is, in general, an empirical question.) We have also seen that it is sometimes possible to discover non-empirically, that is, without examining the world, that if one diagram represents a situation correctly then another must do so too. So we can easily generalise our definition of 'valid' thus:

The inference from representations **R1, R2, . . . Rn** to the representation **Rc** is *valid*, given a specified set of interpretation rules for those representations, if it is impossible for **R1, R2 . . . Rn** all to be interpreted as representing an object or situation correctly (i.e. according to the rules) without **Rc** also representing it correctly.

In this case we can say that **Rc** is jointly *entailed* by the other representations.

This definition copes straightforwardly with cases like figure I, where there are separate representations for premisses and conclusion. The other examples need to be dealt with in the obvious way by treating the single diagram as if it were a compound or two or more diagrams. For example, in figure 2 we can say that there is a 'premiss' which is the diagram with arrow (a) but not arrow (b), and a 'conclusion' which is the diagram with arrow (b) but not arrow (a).

Someone who actually *uses* a picture or diagram to reason with may modify it in the course of his reasoning, and in that case there are really several different diagrams, corresponding to the different stages in the reasoning process.

Explicitly formulating the semantic rules which justify the inference from a set of 'premiss' representations to a 'conclusion' representation, is generally quite hard. We do not normally know what rules we are using to interpret the representations we employ. Many workers in artificial intelligence have found this when attempting to write programs to analyse and interpret pictures or drawings. But the same is also true of the semantic rules of natural languages: it is hard to articulate the rules and still harder to articulate their role in justifying certain forms of inference.

In the case of artificial languages invented by logicians and mathematicians, it is possible to formulate the semantic rules, and to use them to prove the validity of some inferences expressible in the languages. In propositional logic, symbols for conjunction '&', disjunction 'V', and negation '~' are often defined in terms of 'truth-tables', and by using a truth-table analysis one can demonstrate the validity of inferences using these symbols. It is easy to show, for example, that inferences of the following form are valid:

$$\begin{array}{l} \mathbf{P \vee Q} \\ \sim\mathbf{P} \\ \hline \mathbf{so: Q} \end{array}$$

(See for example Copi, *Introduction to Logic*, chapter 8.)

Similarly, in predicate logic the quantifiers ('for all x', 'for some x') may be explicitly defined by specifying certain rules of inference to which they are to conform, like the rule of 'universal instantiation' (see Copi, chapter 10). It is not nearly so easy to formulate semantic rules for words in natural languages. In fact, for some words the task would require much more than the resources of linguistics and philosophy. The semantics of colour words ('red', 'vermilion', etc.) cannot be properly specified without reference to the psychology and physiology of colour vision, for example. The principles by which we interpret pictures, diagrams and visual images may be just as hard to discover and formulate.

If the semantic or interpretative rules for a language or representational system have been articulated, it becomes possible to accompany an inference using that language with a commentary indicating why various steps are valid. A proof with such a commentary may be said to be not only valid, but also *rigorous*. So far relatively few systems are sufficiently well understood for us to be able to formulate proofs or inferences which are rigorous in this sense. Most of the forms of reasoning which we use in

our thinking and communicating are not rigorous.

However, the fact that we cannot give the kind of explanatory commentary which would make our inferences rigorous does not imply that they are not *valid*. They may be perfectly valid in the sense which I have defined. Moreover, we may know that they are valid even if we cannot articulate the reasons.

This is not to suggest that there are some inherently mysterious and inexplicable processes in our thinking. I am only saying that *so far* it has proved too difficult for us.

The use of representations to explain or demonstrate possibilities is not directly covered by the preceding discussion. However, all such cases seem to fit the following schema:

Suppose R is a representation depicting or denoting W, where W is an object, situation or process known to be possible in the world.

And suppose that Tr is a type of transformation of representations which is known (or assumed) to correspond to a really possible transformation Tw of things in the world. (See [chapter 2](#) on the aims of science for discussion of 'really possible'.)

Then, by applying Tr to R, to get a new representation, R', which is interpretable as representing an object, situation, or process W' we demonstrate that W' is possible, if the assumptions stated are true.

This seems to account for the chemical example and the use of bits of cardboard to determine a possible layout of objects in a room.

There are many problems left unsolved by all this. For instance, there are problems about the 'scope' of particular forms of inference. Are they *always* valid, or only in certain conditions? How do we discover the limits of their validity? (See Lakatos, 1976, for some relevant discussion in relation to mathematics, and Toulmin, 1953, for discussions of the use of diagrams in physics.) Does our ability to see the validity of certain inference patterns depend on our using, unconsciously, 'metalanguages' in which we formulate rules and discoveries about the languages and representations we use?

Are children developing such metalanguages at the same time as they develop overt abilities to talk, to draw and interpret pictures, etc.? Questions like these can, or course, be asked about inferences using verbal symbolisms too. (See Fodor, 1976.)

## ***7.10. What are analogical representations?***

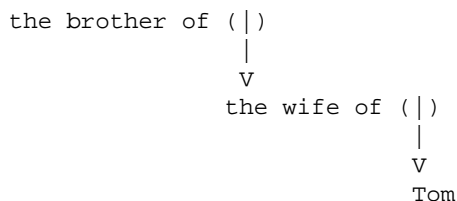
Earlier, I introduced the idea of Fregean or applicative symbolisms, and throughout the chapter have been using the notion of an 'analogical' representation without ever having given it a precise definition. I shall try to explain what I mean by 'analogical' partly by contrasting it with 'Fregean'. I hope thereby to clarify some of the things people have had in mind in talking about 'iconic', 'non-verbal', 'intuitive', 'pictorial' symbols and modes of thinking.

But experience has taught me that readers will project their own presuppositions onto my definitions. So I should like to stress a point which will be repeated later on, namely that there is nothing in the idea of analogical representations which requires them to be continuous (as opposed to discrete). Thus there is nothing to prevent digital computers using analogical representations. A less important source of confusion is the prejudice that analogical representations must be isomorphic with what they represent. This is by no means necessary, and I shall illustrate this with two-dimensional drawings

which represent three-dimensional scenes.

The contrast between Fregean and analogical symbolisms is concerned with the ways in which *complex* symbols work. In both cases complex symbols have parts which are significant, and significant relations between parts. Of course, the parts and relations are not so much determined by the physical nature of the symbol (for instance the ink marks or picture on a piece of paper) as by the way the symbol is analysed and interpreted by users. Only relative to a particular way of using the symbol or representation does it have parts and relations between parts. I shall take this for granted in what follows.

In both Fregean and analogical representations, the interpretation rules are such that what is denoted, or represented, depends not only on the meanings of the parts but also on how they are related. I shall start by saying something about how Fregean symbolisms work. Their essential feature is that all complex symbols are interpreted as representing the application of functions to arguments. Here is a simple example. According to Frege, a phrase like 'the brother of the wife of Tom' should be analysed as having the structure:



The function 'the wife of' is applied to whatever is denoted by 'Tom', producing as value some lady (if Tom is married), and the function 'the brother of' is applied to her, to produce its own value (assuming Tom's wife has exactly one brother). Thus the whole expression denotes whatever happens to be the value of the last function applied.

Frege's analysis of the structures and functions of ordinary language was complex and subtle, and I have presented only a tiny fragment of it. For more details see the translations by Geach and Black, and the items by Furth and Dummett in the Bibliography. I shall not attempt to describe further details here, except to point out that he analysed predicates as functions from objects to truth-values, a notion now taken for granted in many programming languages, and he analysed quantifiers ('all', 'some', 'none', etc.) and sentential connectives ('and', 'or', 'not', etc.) also as functions.

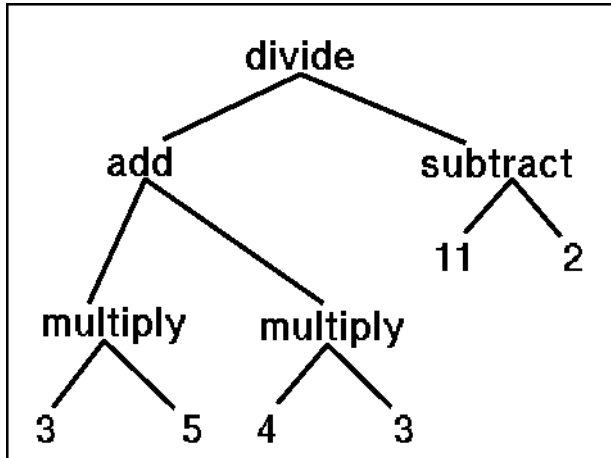
For present purposes it will suffice to notice that although the complex Fregean symbol 'the brother of the wife of Tom' has the word Tom' as a part, the thing it denotes (Tom's brother-in-law) does not have Tom as a part. The structure of a complex Fregean symbol bears no relation to the structure of what it denotes, though it can be interpreted as representing the *structure of a procedure for identifying what is denoted*. In this case, the procedure is first of all to identify whatever is denoted by 'Tom', then use the relation 'wife of' to identify someone else, then use the relation 'brother of' to identify a third object: the final value. (See also my *Tarski Frege, and the liar paradox*(1971). )

We could express this by saying that sometimes the structure of a Fregean symbol represents the structure of a 'route through the world' to the thing denoted. But this will not fit all cases. For instance, in the arithmetical expression:

$$\underline{3 \times 5 + 4 \times 3}$$

$$11 - 2$$

it is not plausible to say that the structure of the whole thing represents a route through the world. However, given certain conventions for grouping, it does represent the structure of a rather elaborate procedure for finding the value denoted. The procedure can also be represented by a tree:



(Notice that in interpreting the expression this way we are using a convention about how expressions involving 'x' and '+' should be 'bracketed'.) The tree-structured procedure is executed by working up to the top of the tree from the bottom. Left-right ordering of components does not signify a temporal ordering in which the sums should be done. In some sense we can say that the sub-expressions, for example, 11, denote aspects of the procedure. But they do not denote *parts* of what is denoted by the whole thing. An arithmetical expression denoting the number three may contain a symbol denoting the number eleven, but that does not imply that the number eleven is in any sense part of the number three.

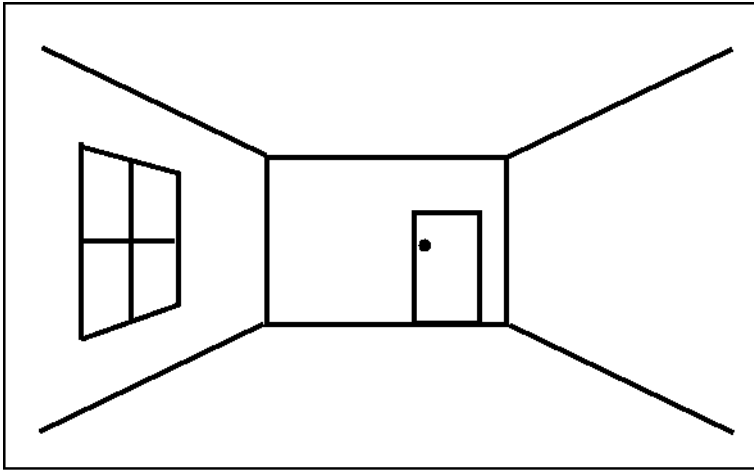
By contrast, analogical representations have parts which denote parts of what they represent. Moreover, some properties of, and relations between, the parts of the representation represent properties of and relations between parts of the thing denoted.

So, unlike a Fregean symbol, an analogical representation has a structure which gives information about the structure of the thing denoted, depicted or represented.

This, then, is my definition of 'analogical'. It is important to note that not ALL the properties and relations in an analogical representation need be significant. For instance, in a diagram the colour of the lines, their thickness, the chemical properties of the paint used, and so on, need not be meaningful. In a map (for instance maps of the London underground railway system) there will often be lines whose precise lengths and orientations do not represent lengths or orientations of things in the world: only topological relations (order and connectivity) are represented. This may be because a map depicting more of the structure of the relevant bit of the world would be less convenient to use. (Why?)

Further, the interpretation rules (semantic rules) need not require that properties and relations within the representation must always represent the same properties and relations of parts of what is represented. The interpretation procedures may be highly context-sensitive. For example, lines of the same length in the scene may be depicted by lines of different lengths in the picture. In figure 8 distances, or lengths, in the picture represent distances in the scene in a complex context-sensitive

way. Further, lines of the same length in the picture may depict different lengths in the scene. Moreover, the relation 'above', in the picture, may represent the relation 'above', or 'further', or 'nearer', or 'further and higher',



**Figure 8**

depending on whether bits of floor, wall, or ceiling are involved. This is connected with the fact that parts of an analogical representation may be highly ambiguous if considered in their own right. Only in the context of other parts is the ambiguity removed. Much work in computer vision is concerned with the problem of enabling global relations to 'resolve local ambiguities. (See bibliography references to Clowes and Waltz, and chapter 9.)

Figure 8 also brings out clearly the fact that although the structure of an analogical representation is related to the structure of what it represents, there is no requirement that the two be *isomorphic*. Indeed, they may have very different structures. In particular, Figure 8 is two dimensional but represents a three-dimensional scene, whose structure is therefore very different from that of the picture.

It should be obvious how to apply my definition of 'analogical' to the sorts of pictures and diagrams used earlier to illustrate inferences with analogical representations. However, it turns out that the precise details of how to interpret relations in a diagram are often surprisingly complicated. Trying to program a computer to do the interpreting is perhaps the best way of discovering the rules. Merely writing down theoretical analyses, you are likely to get the rules wrong. Embodying them in a program helps you to discover that they do not work.

### ***7.11. Are natural languages Fregean?***

Frege was able to apply his function-argument analysis to a wide variety of examples from German, and they transfer easily to the English equivalents. However, not all the complexity of natural language utterances is due to the application of functions to arguments. For example, we often use analogical representations either within sentences or in larger structures, like stories. The order of words, phrases, or sentences often depicts the order of things represented or denoted by the words, etc. Tom, Dick and Harry stood against the wall in that order.' 'He entered the room, saw the body, gasped, and ran out screaming.'

This shows that there is no sharp verbal/analogical or verbal/iconic distinction. A particular symbolism may include both Fregean and analogical resources.

In modern programming languages this is very clear, since there is a great deal of the usual function-application syntax often mixed up with conventions that the order in which instructions occur in a program represents the order in which they are to be executed (and doing them in a different order may produce quite different results). So programming languages, like natural languages, are partly Fregean and partly analogical. This is true even of a logic programming language like Prolog.

But the Fregean/analogical distinction does not exhaust the variety of important kinds of symbolising relations. For example, in a program a symbol may occur which is merely a label' its sole function is to make it easy for other parts of the program to refer to this bit, so that it does not depict either a part of something represented by the whole program nor a thing which is the argument to which a function is applied. Elsewhere in the program may be an instruction to jump to the location specified by this label. The occurrence of such 'jump' instructions can badly upset the correspondence between order of instructions in the program and the time order of events in which the instructions are executed, making programs hard to understand and modify.

The kind of self-referring metalinguistic role of labels in a computer program is clearly something different from the kinds of representation I have called Fregean and analogical.

Natural languages also use self-reference, for instance when the expressions 'the former' and 'the latter' direct attention to order of phrases in a text. They have many other devices which do not fit neatly into these two categories. For example, it is not easy to give a Fregean analysis of adverbial phrases ('He came into the room, singing, leaning heavily on a stick, and dragging the sofa behind him'). So I am not claiming that I have given anything like a complete survey of types of representation. I doubt whether such a thing is possible: for one aspect of human creativity is the invention of new sorts of symbolisms.

One conclusion which may be drawn from all this is that neurophysiologists, psychologists, and popular science journalists who take seriously the idea that one half of the human brain deals with verbal skills and the other half with pictorial and other non-verbal skills are simply showing how naive they are about verbal and non-verbal symbolisms. Presumably, when they learn that besides Fregean and analogical symbolisms there are other sorts, they will have to find a way of dividing the brain into more than two major portions. As for how we deal with combined uses of the two sorts of symbolisms, no doubt it will prove necessary to find a bit of the brain whose function is to integrate the other bits! (Programmers know that there need not be a localised bit of the computer which corresponds to sub-abilities of a complex program.)

## ***7. 12. Comparing Fregean and analogical representations***

Philosophers of science who acknowledge that scientists and mathematicians often use diagrams, models, images, and other non-verbal representations, sometimes claim that this fact is of no philosophical importance. It is a mere empirical fact, of interest to psychologists, but not relevant to philosophical studies of what is 'rational' in scientific methods.

The implication is that the use of non-logical methods of inference, and the choice of analogical representations is an irrational, or at best non-rational, piece of behaviour. Scientists are behaving rationally only when they perform logical deductions from theories and when they use observation and experiment to discover whether certain sentences express truths or falsehoods.

Against this view I shall argue that it is sometimes quite rational to choose to use an analogical rather than a Fregean method of representation. That is, there are often good reasons for the choice, given the purposes for which representations are used, which include storing information for future use, guiding the search for good solutions to problems, enabling new versions of previously encountered situations to be recognized, and so on. I do not claim that analogical representations are *always* best.

If one were designing a robot to be a scientist, or more generally to play the role of a person, it would be advisable, for some purposes, to program the robot to store information in an analogical representation, and to perform inferences by manipulating analogical representations. (See Funt 1977 for a description of a program which solves mechanics problems with the aid of analogical representations.) So it is not merely an empirical fact that people do this too. Of course, neither people nor robots could possibly function with only analogical representations. Any intelligent system will have to use a wide variety of different types of representation and different types of reasoning strategies. But how can we decide which ones to use for which purposes? There are no simple answers.

Fregean systems have the great advantage that the structure (syntax) of the expressive medium does not constrain the variety of configurations which can be described or represented. So the same general rules of formation, denotation and inference can apply to Fregean languages dealing with a very wide range of domains. The formula  $P(a,b,c)$ , or its English variants, like '**a is P to b and c**', can be used for applying a predicate to three arguments no matter what kind of predicate it is, nor what sorts of things are referred to by the argument symbols. The following assertions use the same Fregean structure despite being concerned with quite different domains:

**Between(London, Brighton, Cambridge)**

**Greater-by (three, twelve, nine)**

**Joins(coupling, truck 1, truck 2)**

Contrast the difficulty (or impossibility) of devising a single two-dimensional analogical system adequate for representing chemical, musical, social, and mechanical processes. Fregean systems make it possible to think about very complex states of affairs involving many different kinds of objects and relations at once. For each type of property or relation a new symbol can be introduced as a predicate (that is, a function which, when applied to objects as arguments, yields the result TRUE or the result FALSE). The syntax for making assertions or formulating questions using all these different symbols is the same. There is no need to invent new arrangements of the symbols to cope with a new kind of domain.

The price of this topic-neutrality, or generality, is that it becomes hard to invent procedures for dealing efficiently with specific problems. Very often, searching for the solution to a problem is a matter of searching for a combination of symbols representing something with desired properties. For instance it may be a search for a plan of action which will achieve some goal, or a search for a representation of an arrangement of objects in a room, or a search for a representation of a route between two places which is shorter than alternative routes. For a frequently encountered class of problems it may be advantageous to use a more specialised representation, richer in problem-solving power than a Fregean symbolism.

What makes one representation better than another? To say that it is easier for humans, or that people are more familiar with it is not to give an explanation. An adequate explanation must analyse the structure of the symbolism and show its relationship to the purposes for which it is used, the context of use, and the problems generated by its use. This is often very hard to do, since it is hard to become

conscious of the ways we are using symbols. I shall try, in the rest of this section, to give a brief indication of the sort of analysis that is required.

A method of representation may possess problem-solving power, relative to a domain, for a number of different reasons.

- a) It may have a syntax which makes it impossible to waste time exploring unfruitful combinations of symbols.
- b) It may permit transformations which are significantly related to transformations in what is denoted, so that sets of possibilities can be explored exhaustively and economically.
- c) It may provide a useful 'addressing' structure, so that mutually relevant items of information are located in the representation in such a way that it is easy (using appropriate procedures) to access one of them starting from the other.
- d) It may provide an economic use of space, so that there's lots of room for adding new information or building temporary representations while exploring possible ways of solving a problem. Economy in use of space may also reduce the time taken to search for what is needed.
- e) The representation may make it easy to alter or add to information stored, as new facts are learnt or old information is found to be mistaken or no longer necessary.
- f) The system used may facilitate comparisons of two representations, to find out whether they represent the same thing, and, if not, how exactly they differ.
- g) The representation may facilitate the process of 'debugging', that is tracking down the source of the difficulty when use of the representation leads to errors or disappointments.
- h) The representation may allow similar methods of inference and problem-solving to be used in more than one domain, so that solutions to problems in one domain generate solutions to problems in another domain.

These form just a subset of the problems about adequacy of representations which have had to be faced by people working in artificial intelligence. (See Hayes, 1974, Bobrow, 1975, Minsky, 1975.) The subject is still in its infancy, and criteria for adequacy of representations are only beginning to be formulated. The sorts of issues which arise can be illustrated by the following list of properties of analogical representations which often make them useful:

1. There is often less risk of generating a representation which lacks a denotation. In Fregean systems, as in ordinary language, 'failure of reference' is a commonplace. That is, syntactically well-formed expressions often turn out not to be capable of denoting anything, even though they adequately express procedures for attempting to identify a referent. Examples are 'the largest prime number', 'the polygon with three sides and four corners', 'my bachelor uncle who is an only child'.

In analogical systems it seems that a smaller proportion of well-formed representations can be uninterpretable (inconsistent). This is because the structure of the medium, or the symbolism used, permits only a limited range of configurations. Pictures of impossible objects are harder to come by than Fregean descriptions of impossible objects. This means that searches are less likely to waste time exploring blind alleys.

2. In an analogical representation, small changes in the representation (syntactic changes) are likely to correspond to small changes in what is represented (semantic changes). We are relying on this fact when we use a map to search for a short route between two towns, and start by drawing, or imagining, a straight line joining the two towns, then try to deform the line by relatively small amounts so as to make it fit along roads on the map.

(This is not as simple a process as it sounds.) By contrast, the differences in the forms of words describing objects which differ in shape or size may not be related in magnitude to the differences in the objects. The difference between the words 'two' and 'ten', for example, is in no sense greater than the difference between 'two' and 'three', or 'nineteen' and 'twenty'. 'Circle' and 'square' are not more different in their form than 'rectangle' and 'square'. So substitution of one word for another in a description need not make a symbolic change which is usefully related to the change in meaning. In particular, this means that the notation does not provide an aid to ordering sets of possibilities so that they can be explored systematically.

3. Closely related to the previous point is the fact that constraints in a problem situation (the route cannot go through a wall, a lever cannot bend, the centres of pulleys have a fixed position) may, in an analogical representation, be easily expressed by constraints on the kinds of syntactic transformations which may be applied to the representation. Thus large numbers of possibilities do not have to be generated and then rejected after interpreting them. So 'search spaces' may be more sensibly organised.
4. Often in an analogical representation it is possible to store a great many facts about a single item in a relatively economical way. Each part of a map is related to many other parts, and this represents a similar plethora of relationships in the terrain represented. Using a map we can 'get at' all the relationships involving a particular town through a single 'access point', for example a single dot. If the same collection of relationships were stored in sentences, then for each significant place there would be many sentences referring to it, and this would normally require a large number of repeated occurrences of the name of that place.

Sometimes there are devices for abbreviating sentences repeating a single word, by using 'and' to conjoin phrases, for example, but one could not get rid of all repetitions of place names like this. If the sentences are stored in a list of assertions, then in order to find all the facts concerning any one place it is necessary to search for all the sentences naming it. For some places it is possible to collect together all the sentences concerning them, but since such sentences will generally mention lots of other places too, we cannot collect all the facts about a place under one heading, simultaneously for all places, without an enormous amount of repetition. This problem is avoided in a map.

The same effect as a map can be achieved in a computer data-structure by associating with all objects a set of 'pointers' to all the stored assertions about them, that is, a list of addresses at which assertions are stored in the machine. The facts do not then need to be repeated for all the objects they mention. This sort of technique can lead to the use of structures, within the computer, which include relationships representing relationships in the world. Programmers often make their programs use analogical representations because of the efficiency achieved thereby.

5. Closely related to the previous point is the fact that it is often possible in an analogical representation to represent important changes in the world by relatively simple changes in the representation. For instance, if buttons or other markers on a map represent positions of objects, then moving the buttons represents changes in the world.

From the new configuration the new relationships between objects (which ones are near to which others, which are north of others, etc.) are as easily 'read off' as before the alteration. By contrast, if instead of representing all the initial representations by location on a map, we make a lot of assertions about their relationships, then for each change of position a large number of changes will have to be made in the stored assertions. Of course, this problem can be minimised if we have some way of recording position without doing it in terms of relations to all the other objects, for instance by storing a pair of co-ordinates (latitude and longitude). This also requires good methods for inferring relationships from such stored positional information. Notice incidentally that the use of Cartesian co-ordinates to represent position, and more generally the use of algebraic methods in geometry, involves using sets of numbers as an analogical representation for sets of locations on a line that is, order relations and size relations between numbers represent order relations and distance relations.

### ***7.13. Conclusion***

When an early version of this chapter was published in 1971, many readers thought I was trying to prove that analogical representations are always or intrinsically better than Fregean ones. That would be absurd. I have been trying to show that questions about which should be used can be discussed rationally in the light of the purposes for which they are to be used and the problems and advantages of using them. In some circumstances, analogical representations have advantages.

The problem of deciding on the relative merits of different ways of representing the same information plays a role in the development of science, even if scientists are not consciously thinking about these issues. Similarly a child must be acquiring not only new facts and skills but new ways of representing and organising its knowledge. Very little is currently known about such processes, but the attempt to design machines which learn the sorts of things which people can learn is helping to highlight some of the problems.

The issues are complicated by the fact that one type of representation can provide a medium within which to embed or 'implement' another (see Hayes, 1974). For instance, by using a suitable method of indexing statements in a Fregean language we can get the effect of an analogical representation, as I have already indicated in discussing maps. Another example is the use of two-dimensional arrays to represent two-dimensional images in a computer. There is not really any two-dimensional object accessed by the program, rather a linear chunk of the computer's memory is organised in such a way that with the aid of suitable programs the user can treat it as if it were a two dimensional configuration addressable by a pair of co-ordinates. (Actually the physical memory of the computer is not really linear but it is interpreted as a linear sequence of locations by mechanisms in the computer.)

In [chapter 8](#) on learning about numbers, I give examples of the use of lots of linked pairs of addresses to build up data-structures which in part function as analogical representations, insofar as the order of numbers is represented by the order of symbols representing them. This is another example of one sort of representation being embedded in another.

Computer programs can be given the ability to record and analyse some of their own actions. There will generally be a limit to what a program knows about how it works, however. For instance, programs cannot normally find out whether they are running on a computer made of transistors or some other kind. Similarly, a program may be able to record, and discuss the fact that it is accessing and modifying a two-dimensional array, or moving along a linear list of some kind, without being able to tell how the array or list is actually represented in the computer. So a program could be under the illusion that it is building and manipulating things which are very like two-dimensional pictures on paper, or very like physical rows of objects, not knowing that really it is using scattered fragments of an abstract address space managed by complex storage allocation routines and accessed by procedures

designed to hide the implementation details.

When such a system is asked about its own mental processes, it could well give very misleading accounts of how they work. Phenomenologically, of course, it could not but be accurate. But it would not give accurate *explanations* of its abilities, only *descriptions* of what it does. No doubt people are in a similar position when they try to reflect on their own thinking and reasoning processes. In particular, we see that very little explanatory power can be attached to what people say about how they solve tasks set for them by experimental psychologists interested in imagery.

One moral of all this is that often a discussion of the relative merits of two kinds of representation needs to take account of how the representations are actually constructed and what sorts of procedures for using them are tacitly assumed to be available. (For further discussion, see Hayes, 1974, Sloman, 1975.)

Very many problems have been left unsolved by this discussion. In particular, it is proving quite hard to give computers the ability to perceive and to manipulate pictures and diagrams to the extent that people do. This is an indication of how little we currently understand about how we do this.

---

**[[Notes Added 2001:** It remains very hard to implement working systems with all the features described here, though many partly successful attempts have been made.

See these two books for example (both of which contain papers that are sequels to this chapter):

J. Glasgow, H. Narayanan and Chandrasekaran (Eds), *Diagrammatic Reasoning: Computational and Cognitive Perspectives*, MIT Press, 1995,

M. Anderson, B. Meyer P. Olivier (eds), *Diagrammatic Representation and Reasoning*, Springer-Verlag, 2001.

My own papers in those books are also available online

- [Musings on the roles of logical and non-logical representations in intelligence \(1995\)](#)
- [Diagrams in the mind? \(1998/2001\)](#)

I believe that we cannot hope to understand these issues independently of understanding how human vision works. Likewise, any satisfactory model of human visual capabilities must include the basis for an explanation of how visual reasoning works. [Chapter 9](#) of this book presents some ideas but is still a long way from an adequate theory.

Also relevant are Talks 7 and 8 here:

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/> on visual reasoning and on architectural requirements for biological visual systems, as well as more recent talks in the same directory. **]]**

---

[Book contents page](#)

[Next: Chapter 8.](#)

Last Updated: 4 Jun 2007 (Inserted link to PDF version)

Updated: 26 Feb 2007 (Restored heading for section 7.7)