

Diagrams as Physical Models*

B. Chandrasekaran

Laboratory for AI Research
Department of Computer Science & Engineering
The Ohio State University
Columbus, OH 43210 USA
chandra@cse.ohio-state.edu

Abstract. We discuss a variety of roles for diagrams in helping with reasoning, focusing in particular on their role as physical models of states of affairs, much like an architectural model of a building or a 3-D molecular model of a chemical compound. We discuss the concept of a physical model for a logical sentence, and the role played by the causal structure of the physical medium in making the given sentence as well as a set of implied sentences true. This role of a diagram is consistent with a widely-held intuition that diagrams exploit the fact that 2-D space is an analog of the domain of discourse. One line of research in diagrammatic reasoning is that diagrams, rather than being models, are formal representations with specialized rules of inference that generate new diagrams. We reconcile these contrasting views by relating the usefulness of diagrammatic systems as formal representations to the fact that their rewrite rules take advantage of the diagrams' model-like character. When the physical model is *prototypical*, it supports the inference of certain other sentences for which it provides a model as well. We also informally discuss a proposal that diagrams and similar physical models help to explicate a certain sense of *relevance* in inference, an intuition that so-called Relevance Logics attempt to capture.

1 Roles of Diagrams in Reasoning

Diagrams give many different types of assistance during problem solving. We identify five roles here: helping extend short term memory, helping organize problem solving by spatial organization of related information, as sentences in a 2-D language with specialized rules of inference, providing a model of the premises so that plausible subtasks may be hypothesized for formal inference, and providing a model

* This paper was prepared through participation in the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009, and by federal flow-through by the Department of Defense under contract FA8652-03-3-0005 (as a subcontract from Wright State University and Wright Brothers Institute). I am indebted to Peter Schroeder-Heister and Gerard Allwein for significant assistance in thinking about these ideas, to Neil Tennant and Stewart Shapiro for useful discussions, and to one of the reviewers who made useful suggestions for improvement.

of the premises from which consequents can be inferred and asserted. The major concern of this paper is in the last role.

First, diagrams extend short term memory, by providing a spatially organized external location in which to note down information. Second, they help *organize* problem solving. Simon and Larkin (1987) use the example of analyzing a pulley system – they show how the diagram of the pulley system helps the problem solver organize the sequence of equations to solve, or variables to assign values to. The problem solver can use his visual perception to locate the pulley that a strip of rope goes over, and thus to choose which tension variable to consider next. Another example is the spatial organization of addends when we add two numbers: we line up the numbers such that the numerals in the ones, tens, etc., locations line up, and the locations and the spatial relations guide applications of the sequence of problem solving actions.

The third role is that of diagrams as two-dimensional syntax-controlled compositions of diagrammatic symbols¹. Specialized rules of inference can be specified that can generate valid diagrammatic sentences. Allwein and Barwise (1996) contains a number of papers pursuing this perspective in productive ways. In this framework, e.g., theorems in set theory can be legitimately proved using an appropriate sequence of Venn or Euler diagrams.

The fourth and fifth roles both treat the diagram as a physical instance, a *model*, of a state of affairs of interest. That is, it depicts a situation that satisfies the premises. But the fourth and fifth roles deal with different ways of using the physical model. In the fourth role, the model suggests hypotheses to pursue in the formal proof. This is exemplified by the way diagrams are used in proving theorems in Euclid. In this kind of use, a diagram is *a* model of the premises. Not everything that is true in the model is necessarily true given the premises, but nevertheless a careful use of the model suggests possibly productive subtasks for theorem proving. For example, the fact that two angles are adjacent, and the theorem involves one of the two adjacent angles might suggest to the theorem prover that perhaps stored theorems involving adjacent angles may be useful in advancing the proof. Lindsay (1998) provides a review of the issues in the use of diagrams in geometry theorem proving. It has been estimated that the use of the diagram in this way reduces the search space by several orders of magnitude. In the traditional use of Venn or Euler Diagrams in proving theorems in Set Theory, the diagrams play a similar role. It is important to emphasize that the information from the model is not *asserted* as conclusion, but only used to find strategies for arriving at the general conclusion.

The fifth role for diagrams is also based on diagrams as physical models that satisfy the premises. The problem solver sees that the representation is also a model

¹ The Stanford Encyclopedia of Philosophy entry on Model Theory (<http://plato.stanford.edu/entries/model-theory/>) says, "...the overwhelming tendency of this work is to see pictures and diagrams as a form of language rather than as a form of structure. For example Eric Hammer and Norman Danner (Allwein and Barwise, 1986) describe a 'model theory of Venn diagrams'; the Venn diagrams themselves are the syntax, and the model theory is a set-theoretical explanation of their meaning." This quote might overstate the case a bit ("overwhelming tendency"), but "diagrams as sentences or formal representations" is a common enough view. We comment later on how the views of diagrams as representations versus models might be reconciled.

for another assertion that is not explicitly part of the premises, and concludes that the assertion follows from the premises. Exactly when to generalize and by how much are issues for which answers differ from one diagrammatic application to another. Such a use of diagrams is much more common in applied rather than formal reasoning. This role of diagrams is my focus in this paper.

Consider two very simple examples. Given a simple addition problem in arithmetic, say to show $1 + 3 = 2 + 2$, suppose one draws four points (or arrange four stones on the ground) as below:

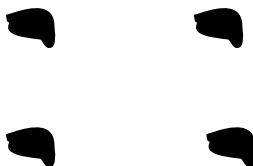


Fig. 1

Under the appropriate mappings, the situation is a model of $1 + 3$. But it is also a model of $2 + 2$. One can demonstrate to a child that $1 + 3 = 2 + 2$ by using the above diagram. Here the generalization issue is trivial: the child could, but typically wouldn't, say, "Maybe this is true when we add 1 star to 3 stars, but is it true when we add 1 slice of pizza to 3 slices of pizza?." Human intuitions about numbers seem sufficiently robust that this issue doesn't arise in a child or an adult. "Individuals that keep their distinct identity" seems to be the background intuition that is operational here, and using that people generalize from star marks on paper or stones on the ground to numbers in general.

Consider another example, one that will find frequent use in the rest of the paper. Given "If A is to the left of B, B is to the left of C, is A to the left of C?," people often draw a diagram as in Fig. 2:



Fig. 2

There is a natural sense in which the physical diagram is a model of the problem situation². The problem solver notices that indeed A is to the left of C, and declares that the inference is true³. Of course, the diagram only represents one specific way in which the points can be located to provide a model. Yet the problem solver makes

² More precisely, it is a model of the conjunct of the given premise with axioms that capture the structure of space in terms of which the predicate Left is defined. For someone for whom the semantics of Left is that of spatially left in ordinary language, the axioms are implicit, and the Figure provides a model for the premise.

³ In fact, the precise role played by the diagram in this inference is more complicated than appears at first, but for the current purposes, this partial account is adequate.

bold to assert that the conclusion is true for all the specific ways in which the points could be located.

Diagrams are just an example of physical models of this type. As mentioned earlier, architectural models and models of molecules constructed out of ping pong balls provide further examples, though their scopes in assisting in human reasoning are not as large as that of diagrams.

Regarding a main concern of logic – accounting for justifiable inferences – this style of reasoning based on a physical model needs to be part of any account of natural reasoning. It puzzles me that more has not been said in logic about the use of such physical models as aids to reasoning, given how prevalent diagrams are in everyday as well as professional reasoning, and the role played by architectural and molecular models in their respective disciplines. So this paper's goal is to raise the profile of physical models in logic. I raise a set of issues for deeper consideration by logicians.

Not everything about a diagram is model-like. It is important to mention that not all diagrams, or all aspects of a given diagram, are models. Actual diagrams have various notations in them, such as letters A and B in Fig. 2, or shadings as in Venn diagrams, that are not model-like. We return to this issue in a later section.

2 Physical Fragments Providing Models for Logical Sentences

A brief remark might be useful on the multiple, sometimes opposite ways, in which the term “model” has been used. In philosophy and practice of science, a description – a set of equations, e.g. – is called a model a domain if the description can be used to predict phenomena in the domain. Thus, Maxwell's Equations model electromagnetic phenomena and physicists speak of the Newtonian model versus the Einsteinian model. In logic, the direction is from description to domain: a domain provides a model of a set of axioms, e.g., arithmetic is a model of Peano's Axioms and plane geometry is a model of Euclidean Axioms. If the axioms are a description, the domain that fits the description is a model of the description. In related usage in logic, a model for a sentence is constructed by assigning truth values to the elements of the Herbrand Universe. In the rest of the paper, we use the term in the sense in which it has been used in logic.

It is useful to start with the standard definition of a model for a sentence in Logic. An *interpretation* for a sentence S consists of:

- A non-empty, possibly infinite, domain D of individuals
- Assignment of specific individuals in D to constant symbols in S
- Assignment to each n -ary function symbol in S of an n -ary function that maps from D^n to D .
- Assignment to each n -ary predicate symbol in S of an n -ary function that maps from D^n to $\{\text{True}, \text{False}\}$.

An interpretation for S is a *model* for it if S evaluates to True under the interpretation.

2.1 Modeling Physical Things

An informal description of how a physical entity may be used to provide an interpretation of a sentence might go as follows. Suppose one finds a physical entity, organizes it into parts, and models the parts in terms of particular sets of physical variables. A specific entity will have specific values for these variables. The causal structure of the physical entity will induce a set of causal constraints on these variables. Suppose, further, one is able to map the individuals in a given sentence S into the “parts” of the entity, and to map the predicates in S into relations between the variables of the parts. If the variable values of the entity are such that they satisfy S under this mapping, the entity could be said to be a physical model for S . What follows is a more formal rendering.

Domain of Individuals. Let Π be a fragment of physical world. Let $\Delta_{\Pi} : \{\pi_1, \pi_2, \dots\}$ be a (possibly infinite) set of entities, each π_i a *part* – a subfragment – of Π . The entities need not be physically disjoint – one entity may be a physical part of another entity; nor is it necessary for Δ_{Π} to exhaust Π , i.e., the totality of physical fragments represented by the elements of Δ_{Π} to be equal to the matter represented by Π .

I will use two concrete examples to illustrate the ideas: Π_1 , the set of points constituting a finite physical horizontal straight line, say drawn on a piece of paper; and Π_2 , a physical object intended to be an architectural “model” of a house.

Examples in Π_1 : The entire finite straight line is Π . Each point in it is a π , thus Π has an infinite number of parts in this model. Another model for the same physical object might subdivide the line into various segments, each providing a π .

Example in Π_2 : The physical entity (the architectural “model”) as a whole is Π , and the physical matter corresponding to various rooms, walls, doors, etc., are the π ’s.

Functions and Predicates. Let $\{\phi_i \mid 0 \leq i \leq k\}$ be a finite set of functions of various arities, such that if ϕ_i is n -ary, it is a function from Δ_{Π}^n to Δ_{Π} . Similarly, let $\{\rho_i \mid 0 \leq i \leq l\}$ be a set of functions of various arities, such that if ρ_i is n -ary, it is a function from Δ_{Π}^n to $\{True, False\}$. The ρ ’s are predicates defined on the physical variables, and thus the values that they take for their various arguments is determined by the causal structure of Π .

Examples in Π_1 : The function, $right_1(\pi_i)$, defined as “the point that is exactly 2 inches to the right; if there is no such point, the right end point” is a unary function. Example of a binary predicate ρ is: $Left(\pi_i, \pi_j)$, with the obvious interpretation.

Examples in Π_2 : Unary function $Entrance\text{-}to(room_i)$, which takes values from the subset of parts of type “door.” Thus, e.g., $Entrance\text{-}to(room_5) = door_6$. Example of ρ : $Bigger\text{-}than(room_i, room_j)$ is a binary function which evaluates to *True* if the area of $room_i$ is larger than that of $room_j$, and *False* otherwise.

Properties and Causal Structure of Π . For the purpose at hand, the physical structure is modeled in terms of a set of variables, selected *attributes* of the physical system. A specific physical instance will have specific values for these variables. Let Θ_i ; $\{\theta_{i1}, \theta_{i2}, \dots, \theta_{ik_i}\}$ be a set of variables in terms of which entity π_i is modeled, and let Θ

$= \bigcup_{i=1}^n \Theta_i$. The *causal structure* of Π , which constrains the values of the variables in

Θ , determines the truth values of the various predicates for various values for their arguments, and thus the truth values of sentences composed out of these predicates.

Thus, part of modeling a physical fragment for the purpose of providing an interpretation for a sentence involves identifying a physical system with the right properties to provide an interpretation, and then setting its parameters – the values of the variables of the part – to those values for which the physical fragments provide a model for the sentence.

Examples in Π_1 : Let part π_i be modeled in terms of a single variable x_i , the x -coordinate of π_i from some origin. $\text{Left}(\pi_i, \pi_j)$ is defined by the values of x_i and x_j . Additionally, the constraints of the physical line result in constraints between predicates: if $\text{Left}(\pi_i, \pi_j)$ and $\text{Left}(\pi_j, \pi_k)$ are both *True*, then $\text{Left}(\pi_i, \pi_k)$ is constrained to be *True*.

Examples in Π_2 : The parts of the house may be modeled in terms of their length, width, height, area, etc. Color and material out of which a part is made may also be in the set of variables. Whether room_5 is larger than room_3 is fully determined by the physical dimensions of Π ; there is no additional freedom to assign *True* or *False*. Further, in a physical architectural model, if room_1 is larger than room_2 which in turn is larger than room_3 , the model will necessarily satisfy the predicate, $\text{larger-than}(\text{room}_1, \text{room}_3)$.

In order to avoid confusion between different usages in science and engineering on one hand and in logic on the other, I use the term *P-model* to refer to a description of a physical entity as in the next definition – the description specifies a *point of view* to look at the physical entity. In order to take a gingerbread house as a possible model of some house, it needs to be viewed as a decomposition of matter into walls, rooms, etc., each as having lengths and heights, rather than as bits of sugar and ginger and flour.

Definition. A *P-model* of a physical fragment Π consists of the following specifications:

- Δ_Π , a set of individuals consisting of parts of Π
- a set $\{\phi_i\}$ of functions of various arities, such that an n -ary function is a mapping from Δ_Π^n to Δ_Π .
- a set of functions $\{\rho_j\}$ of various arities, such that an n -ary function is a mapping from Δ_Π^n to $\{\text{True}, \text{False}\}$.

- a set of variables Θ in terms of which Π and elements of Δ_Π are modeled;
a set Π_{ax} of causal constraints between the variables in Θ .

Remark. There is an infinity of P-models for a given physical entity.

2.2 A Physical Entity Supporting a Logical Model

Let a P-model M_Π of a physical fragment provide an interpretation for a sentence S .

Definition. If a sentence S evaluates to *True* under the interpretation provided by a P-model M_Π of a physical fragment Π , we say that Π provides a *physical model* for S .

Remark. What makes a predicate *True* or *False* in a physical model is that the variables take specific values in the physical fragment, and the causal structure Π_{ax} constrains values between variables.

Examples

Consider the following sentence S :

$$\forall x \forall y \forall z (L(x,y) \ \& \ L(y,z) \rightarrow L(x,z)) \quad (1)$$

Let Π be a physical 1-D spatial line fragment, and let the following be a P-model for Π . M_Π :

Δ_Π : the (infinite) set of points in the line fragment, $\{x_i\}$.

Θ : a single attribute, the co-ordinate of a point x_i with respect to some origin

$\{\phi_i\}$: Null set

$\{\rho_j\}$: a single function, $Less-than(x_i, x_j) = True$, if the co-ordinate of x_i is less than that of x_j ; *False*, otherwise.

(2)

Under the interpretation M_Π , S is *True* in Π . A physical 1-dimensional line fragment is thus a physical model for the sentence (1).

As a more complex example, consider S' : $[(\forall x \forall y \forall z (L(x,y) \ \& \ L(y,z) \rightarrow L(x,z)) \ \&$

$$(\forall x \forall y (L(x,y) \ \& \ L(y,x) \rightarrow Eq(x,y))] \ \& \ L(A,B) \ \& \ L(B,C) \quad (3)$$

Consider the physical diagram in Figure 2 (where the little circles are to be taken as points), with the following M'_Π .

M'_Π : M_Π as defined in (2) plus the following assignments: Constants A , B and C assigned to the points in the 1-dimensional line fragment corresponding to the coordinates as in the Figure. $Eq(x,y)$ assigned to function “Equal(x_i, x_j) = True iff x_i is the same as x_j , and False otherwise.”

(4)

Under the interpretation M'_Π , S' evaluates to *True*, so the diagram in Fig. 2 is a physical model for S' in (3). Readers will recognize (3) as a simple axiomatization of left-ness plus the premises of the problem we stated at the beginning. M'_Π is also a model for the following, S'' :

$$[(\forall x \forall y \forall z (L(x,y) \& L(y,z) \rightarrow L(x,z)) \& (\forall x \forall y (L(x,y) \& L(y,x) \rightarrow \text{Eq}(x,y))) \& L(A,C)] \quad (5)$$

Remark. In applied reasoning, the agent is reasoning in some domain of interest, D , and he is interested in making a model of a sentence, say S . Let D_{ax} be the set of axioms that describe the relevant aspects of the domain of interest. Thus, the agent is looking for a physical model of $D_{\text{ax}} \& S$. Seeing Figure 2 as a model of $\text{Left}(A,B) \& \text{Left}(B,C)$ requires interpreting Left in the spatial meaning of the terms. This interpretation assumes D_{ax} . If instead S were $\text{Goo}(A,B) \& \text{Goo}(B,C)$, one wouldn't see Fig. 2 as its physical model. Successfully making a physical model of S when the agent is reasoning in D involves finding a physical medium such that its causal structure Π_{ax} has the right kind of homomorphism relation with D_{ax} .

There is no requirement that an arbitrary Π have a P-model that provides an interpretation for an arbitrary sentence S . In fact, it is a special situation where a physical model can be constructed so as to provide an interpretation for a sentence. The next section discusses how such physical models are often used.

2.3 Warrant for Generalization

Fig. 2 is a model for S' , but it is just one model. There are infinitely many configurations of points for which the corresponding physical diagrams will provide a model for S' . Nevertheless, we generalize the inference to a class of situations. Fig. 2 also provides a model for “A is farther left of B than B is of C,” but we know that this inference cannot be generalized.

This way of using physical models is quite common in applied reasoning. A chemist, who is considering whether $S_1 \rightarrow S_2$, where S_1 and S_2 are sentences in his domain, might construct a chemical reaction which is a model of S_1 (really a model of his domain axioms and S_1), see if it is also provides a model for S_2 , and, though the specific chemicals in interaction model only instances of S_1 , generalize to the larger class. Of course, a good chemist would know just want sort of model to construct that would bear the generalization.

This style of proof might be called *physical-model-based proof*. The *Model-Based Rule of Inference* may be stated as follows:

Given an inference problem, $S_1 \rightarrow S_2$, where S_2 is not a logical truth, in domain D with domain theory D_{ax} , and given a physical fragment Π such that it provides a P-model M_Π that satisfies $D_{\text{ax}} \& S_1$, if M_Π also satisfies $D_{\text{ax}} \& S_2$, and if M_Π has warrant for generalization with respect to the inference S_2 , conclude $S_1 \rightarrow S_2$ in the general case in D .

One might use the term *prototypical* to describe a model that provides such a warrant for generalization. Asserting logical truths from the model is blocked for reasons related to relevance (see later section on Relevance Logics).

In many cases, the applied reasoner has limited or no access to D_{ax} is an explicit form. However, he has a body of intuitions and practices that help him construct prototypical models for classes of S 's that provide warrant for generalization and help him scope them.

Let S be a sentence in a domain D characterized by axioms D_{ax} . Let $\text{Closure}(D_{ax} \& S)$ be the set of all inferences that are deducible from $D_{ax} \& S$. If Π provides a model for S , it will also provide a model for all elements of $\text{Closure}(D_{ax} \& S)$. However, it will also provide a model for many other inferences that are not in $\text{Closure}(D_{ax} \& S)$. The reasoning agent needs to know how not to make the inferences that are not in $\text{Closure}(D_{ax} \& S)$, even though the model supports it. For example, he needs to know not to infer “ A is farther left of B than B is of C ” from Figure 2.

Different Types of Generalization involving Diagrams. Jamnik (2001) presents a system called Diamond that uses diagrams to prove results such as $1 + 3 + 5 \dots + (2n-1) = n^2$. The proof involves constructing, for a given n , an array of $n \times n$ dots arranged as a square. Because the proof is for a general n , the diagram uses ellipses to indicate the general case. Use of such a diagram involves two kinds of generalization, one of the sort we discuss in this paper, namely, use of n dots or stars (as in Fig. 1) to represent n . The use of ellipses in diagrams involves a very different type of generalization, inductive generalization over n . Jamnik uses an inference rule that she calls the ω -rule that is explicitly intended to help with the latter issue. However, the rule implicitly incorporates generalization over the specific icon (dot or star) used in the array to represent the number n . All the perceptions that the reasoning procedure calls for during reasoning involve treating a star or a dot as a singleton in the unary representation of a number. In the case of numbers, such representations and corresponding generalizations are so deeply rooted in our behaviors – we all know to treat the relevant icon as an integral entity and abstract it as a unit – that we don’t stop a moment to think about them. However, in other uses of models, such as a chemist using chemical mixtures or use of unfamiliar diagrammatic schemes, learning the proper use of the scheme involves learning, implicitly or explicitly, the limits of generalization.

2.4 Diagrams as Models Versus Representations

Some of the well-known work in diagrammatic reasoning can be seen to be based on a diagram being a model, at least in parts. In Hyperproof (Barwise and Etchemendy, 1994), a computer-based system intended to help students learn logic, the left hand side, say, of the screen might contain certain premises and conclusions posed as predicate logic sentences in terms of visual and spatial properties of and spatial relations between objects such as cubes and tetrahedrons of different colors, sizes and locations. The right hand side might show a diagram where an area of space contains cubes and tetrahedrons of different sizes and shapes. The student learns to check if a certain sentence is true or false in the situation represented by the diagram, and learns to make or check inferences by building, or making use of, such diagrams. These parts of the diagram are physical models of the sentences on the left hand side. Similarly, a diagram showing a circle A inside another circle B is a model of the set theoretic assertion, $A \subset B$: the set of points in region A are indeed a subset of points in region B in the physical 2-D space. However, diagrams often have other elements in them whose role is more notational than model-providing, e.g., shading of regions in Venn diagrams to indicate emptiness of the corresponding set. These notational elements often obscure the role of the model parts of the diagrams, and encourage a view of them as simply formal representation, one that just happens to be different in

kind from linear sentential ones. I think that these two views can be reconciled. Here's an outline of how.

First, a model itself is a *composition* of elements, the composition following some syntax. Venn and Peirce's diagrammatic systems for representing set-theoretic assertions can be described (Shin, 1994) as formal languages composed, following a syntax, of regions and various notational elements, and rewrite rules that permit replacement of diagrams satisfying certain conditions with other diagrams.

Suppose, in some sentential system, a rewrite rule supports writing S3, given S1 and S2. Suppose M1 and M2 are physical models for S1 and S2 respectively in a modeling system M, and suppose M3 is a composition of M1 and M2 such that it provides a model for S3. Suppose also that we treat the model system as a formal representation one of whose rewrite rules allows creating M3, given M1 and M2. If generalization from M3 to the assertion contained in S3 has warrant for generalization, we can treat the model system as an alternate formal representation for the assertions in the original sentential system. Thus, the sequence M1, M2 and M3 can be simultaneously viewed as models of S1, S2, and S3, such that inferring S3 has the appropriate warrant for generalization, and as a proof of the content of S3 in an alternate formal representation language M. As I see it, the diagrammatic formal language that Shin develops works because the rewrite rules appropriately embody the generalization that the physical models allow.

Heterogeneous Proofs. When students use Hyperproof to make inferences, their proofs are *heterogeneous* (Barwise and Etchemendy, 1994) – the reasoning involves a sequence of steps some of which are sentential while others are inferences made from the diagrammatic part. As we just pointed out, what the students do with the diagrammatic component can be characterized as model-based inference or, equivalently, inference in a 2-D sentential system whose rewrite rules happen to allow just the kind of model-based inferences that support the needed generalization.

3 Prototypical Models

What makes a prototypical model? The specifics depend on the domain and the predicates of interest, but some general intuitions may be useful. The following ideas might help in the development of a more formal account.

The first idea is *minimality*. Let S be $\text{Left}(A,B) \ \& \ \text{Left}(B,C)$ (we are implicitly in the domain of 1-d space with a directed axis). Just as Fig. 2, Fig. 3 also provides a model for S. However, it provides a model for unrelated things such as $\text{Inside}(D,E)$. Clearly, any inference based on this model, such as $\text{Left}(A,B) \ \& \ \text{Left}(B,C) \rightarrow \text{Inside}(D,E)$, would be a mistake. Fig. 2 is in some sense minimal compared to Fig. 3 for $\text{Left}(A,B) \ \& \ \text{Left}(B,C)$.



Fig. 3

The next idea is that of multiple prototype models. Let S^\wedge be $\text{Left}(A,B)$ & $\text{Left}(A,C)$. While Fig. 2 provides a model for S^\wedge , it doesn't seem prototypical for another reason: it only accounts for a subset of instances. Fig. 4 provides another model.



Fig. 4

Fig. 2 provides a model for $\text{Left}(B,C)$ and Fig. 4 provides a model for $\text{Left}(C,B)$, neither of which follows from S^\wedge , thus neither of these inferences have a warrant for generalization. Applied reasoning in this case requires that two models be set up, each of which allows certain inferences, say $\text{Right}(B,A)$, but not $\text{Left}(B,C)$ or $\text{Left}(C,B)$.

The third idea is a revisit of what we mentioned earlier, that Fig. 2 doesn't support the generalization of "A is farther left of B than B is of C," though the figure provides a model for it. Suppose a new predicate $\text{boogoo}(x,y,z)$ is defined as "x is farther left of y than y is of z." Consider $S^\#$: $\text{Left}(A,B)$ & $\text{Left}(B,C)$ & $\text{Left}(C,D)$, modeled by Fig. 5. Fig. 5 also provides a model for $\text{boogoo}(A,C,D)$, which has a warrant for generalization. Even though $\text{boogoo}(A,B,C)$ is True in the model in Fig. 2, the model does not provide a warrant for generalization.

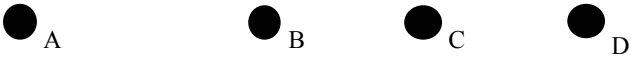


Fig. 5

It can be seen, from a metatheory of $\text{Left}(x,y)$, that it provides an order between x and y. Thus any conjunctions of $\text{Left}(x,y)$'s would specify an order or alternate possible orders. Only information that follows from the order information has warrant for generalization. Such metatheories may be constructed in principle for every domain, but in practice, an agent performing applied reasoning in some domain usually has no access to such metatheories, at least in explicit forms. Even when he has access to the axioms for his domain, such as in the sciences, they are provisional and potentially revisable. So, reasoning in the practical world is aided by models constructed and interpreted with the aid of intuitions based on experience, training and partial theories. Such models play a large role in commonsense reasoning as well, as evidenced by the research of Johnson-Laird (1983) on how people solve syllogistic problems by constructing mental models. These mental models have many points of contact with the models that we describe here. Because people lack access to fully worked-out metatheories, some of the reasoning errors that occur in practical reasoning are due to mistakes in the application of generalization and in the construction of prototypical models.

Everyday reasoning as well as reasoning in professional disciplines is full of implicit and explicit guidelines about how to construct physical models – diagrams in

particular – from which that desired information can be obtained perceptually, and about how to generalize. The ubiquity of such diagrams in human reasoning notwithstanding, it is important to note that discoveries of appropriate diagrams for classes of problems are hard-won. Such discoveries are prized – transmitted culturally for everyday reasoning, and made part of training in professional disciplines, with discoverers often honored with awards.

3.1 Relevance Logics and Prototypical Models

The model-based rule of inference blocks asserting a logical truth based on the physical model. This is because we wish the physical model to *play a role* in the assertion. Why? The intuition is the same as that which drives research on *Relevance Logics*, a summary of which is available in Mares (2006). Here's the main idea.

Relevance logics are a response to what some people take to be paradoxes of traditional implication. The paradox arises in that in some of the inferences authorized by the semantics of traditional implication, the antecedent doesn't seem *relevant* to the consequent. For example, $p \rightarrow (q \rightarrow p)$ and $p \rightarrow (q \rightarrow q)$; q in the first case and p in the second don't seem relevant to the conclusions. People performing applied reasoning, i.e., domain-specific every day argumentation, would object to someone bringing what appear to be irrelevant issues.

Now suppose a reasoning agent performing domain-specific reasoning uses prototypical physical models to assert consequences. That is, given p , he constructs a prototypical model in his domain for W that also models S , and that S has a warrant for generalization. In such cases, his assertion $p \rightarrow S$ will be a relevant inference in W . Technically, any model is also a model for logical truths, so we add a rule forbidding asserting a logical truth as a consequence of any sentence.

Consider as an example:

$$p \rightarrow (q \rightarrow p) \quad (6)$$

In this case p doesn't seem to play a relevant role in making $(q \rightarrow p)$ true. However, suppose there is a domain D in which q is a possible cause for p , and that q would definitely cause p . In that domain, it wouldn't be a surprise to assert that if p is true, then if q is known to be present, q caused p , and consequently that the truth of q would imply the truth of p . If D_{ax} denotes the axioms characterizing D , the following would not fail the test of relevance:

$$(D_{ax} \ \& \ p) \rightarrow (q \rightarrow p) \quad (7)$$

Suppose D describes a physical domain in which one constructs a physical fragment Π that provides a P-model for q . Π , under the same P-model mappings, would also provide a model for p . That is, there is no way to construct a model for q in D without it being a model for p as well. In this case, there is no failure of relevance to assert (6).

The discussion above leads to the suggestion that judgments of relevance in the case of implications arise in applied reasoning in specific domains, whose structures (causal structures if domains are physical) then can be judged to play or not play a role in the antecedent making the consequent true. Further, making such judgments is facilitated in the specific domains by constructing physical models when possible.

The foregoing is a précis of a slightly longer discussion that will appear in a forthcoming paper (Chandrasekaran, to appear). As far as I am aware, current approaches to Relevance Logics don't follow the above approach. My main reason to discuss this is to point to a potentially productive direction of research for the logicians interested in Relevance Logics.

4 Concluding Remarks

Diagrams are often-used reasoning aids in many situations. This paper views diagrams as just the most prominent example of larger class of reasoning aids that provide physical models to premises in some domain. Applied reasoning, where reasoning agents are concerned with inferences in specific domains rather than abstract notions of validity, has not drawn as much attention from logicians as it should. The use of such physical models in applied reasoning raises important issues in logic. I have attempted to formalize the notion of some piece of physical matter providing a model for a sentence. I identified a proof technique called physical model-based inference in which prototypical models in specific domains are constructed that support useful generalizations.

All of this is in accord with a central intuition of Johnson-Laird (1983) that people tend to reason *concretely*, by building small-scale “mental models” of concrete situations that satisfy the premises and generalizing from them, rather than by applying abstract rules of inference. Even though such mental models do not have to be diagrammatic, it is often the case that, when possible, people do construct diagrams of such concrete situations, which then serve as models in the sense that has been elaborated in the paper.

As can be seen from the example of pebbles for adding numbers, whose generalization properties even young children seem to understand, human intuitions on physical models and their prototypicality seem to draw on fairly deep structures in cognition. Developing reasoning skills in various domains includes acquiring or developing intuitions about how to construct prototypical models for specific reasoning situations.

I briefly described how the diagram as model view and the diagram as formal representation view can be reconciled by noting that if the diagrammatic rewrite rule happens to incorporate just those inferences that have the warrant for generalization in the specific model system, then the diagrammatic sequence can be seen as sequence of models or, alternatively, as a sequence of representations that result in the needed inference.

I also related such physical model-based reasoning to issues in Relevance Logics, where the goal is to identify when and how antecedents can be said to have a role in the consequent being true. Physical models, by incorporating the underlying causality of the domain, make it possible, under many conditions, to see whether or not the antecedents play a role in an implication being valid. One of my goals in this paper is to invite the attention of logicians more expert than I to look into what seems to be important lines of investigation.

The paper's account of how the causal constraints of the physical model result in entailments of assumptions also being satisfied by the model raises issues about

mental images of diagrams. The degree to which the substrate of mental images mimics the causality of physical space is controversial. This issue requires a complex and nuanced discussion taking it beyond the scope of the current paper, but pointing out the existence of this issue is relevant to the goals of this paper.

Model-based reasoning of the kind I have discussed is not merely an issue in logic, but in artificial intelligence. AI has focused almost exclusively on what might be called linguistic representations, mirroring the logical form of natural language sentences. However, real reasoning in humans is multi-modal, with perceptual and kinesthetic modalities often contributing to problem solving. Diagrams provide an important window into such multi-modal representations. In (Chandrasekaran, et al, 2004), we describe a diagrammatic representation and reasoning architecture that integrates traditional symbolic reasoning with diagrammatic reasoning. This architecture can be viewed as a kind of generalization of heterogeneous reasoning of Barwise and Etchemendy (1994).

References

- Allwein, G. and Barwise, J. eds. (1996), *Logical Reasoning with Diagrams*, New York, Oxford University Press.
- Barwise, Jon, and John Etchemendy (1994), *Hyperproof*, Stanford: CSLI, and Cambridge: Cambridge University Press.
- Chandrasekaran, B., Unmesh Kurup, Bonny Banerjee, John R. Josephson and Robert Winkler (2004), "An Architecture for Problem Solving with Diagrams," in *Diagrammatic Reasoning and Inference*, Alan Blackwell, Kim Marriott and Atsushi Shimojima, Editors, *Lecture Notes in Artificial Intelligence* 2980, Berlin: Springer-Verlag, pp. 151-165.
- Chandrasekaran, B. (to appear), "Diagrams as Physical Models to Assist in Reasoning," in *Model-Based Reasoning in Science and Engineering*, L. Magnani, ed., King's College Publications, London.
- Jamnik, M. (2001). *Mathematical Reasoning with Diagrams: From Intuition to Automation*. Stanford, CA: CSLI Press.
- Johnson-Laird, P. (1983), *Mental Models: Towards a cognitive science of language, inference, and consciousness*, Cambridge: Cambridge University Press.
- Larkin, J. & Simon, H. (1987), "Why a diagram is (sometimes) worth ten thousand words," *Cognitive Science*, 11:65-99.
- Mares, Edwin "Relevance Logic," *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition), Edward N. Zalta (ed.), forthcoming, URL = <http://plato.stanford.edu/archives/spr2006/entries/logic-relevance/>.
- Lindsay, Robert K. (1998), "Using Diagrams to Understand Geometry," *Computational Intelligence*, 14: 238-272.
- Shin, S., 1994, *The Logical Status of Diagrams*. Cambridge: Cambridge University Press.