From: AAAI Technical Report SS-92-02. Compilation copyright © 1992, AAAI (www.aaai.org). All rights reserved.

Diagrammatic Semantics for Spatial Prepositions

Thomas R. Ioerger*
Department of Computer Science, University of Illinois
405 N. Mathews Ave., Urbana, IL 61801

Abstract

Attempts to use images as mental models of natural language sentences with spatial prepositions have been hindered by differences in level of detail between propositional and diagrammtic representations. Specifically, when propositional knowledge is modeled with an image, the level of detail of the diagrammatic representation often requires some details to be assumed. Subsequently, it becomes unclear what details in an image are necessary versus arbitrary; this is the Indeterminacy Problem. Previously, a computational model of imagery, ISR, was introduced that can avoid the Indeterminacy Problem by dynamically manipulating images, rather than treating them as static models. This paper reports our attempts to apply ISR in a fairly realistic domain in which spatial prepositions are used to relate the locations of objects in a room. The major difficulty encountered was in modeling the constraints of gravity, and the analysis of this problem exposes a subtle but crucial interaction among various components of ISR. We describe a heuristic solution, and suggest, by similarity to planning, that natural language semantic processors require such heuristics that trade completeness for speed.

1 Introduction

The use of images as mental models for representing the semantics of natural language sentences with spatial prepositions has been suggested as more psychologically valid than other representations such as propositions (Johnson-Laird 1983). However, a difficulty encountered in attempts to implement such reasoning systems has been the difference in level of detail required by diagrammatic and propositonal representations (Waltz & Boggess 1979). One of the classical "imagery-debate" arguments against image-based representations is that they are often over-specific (Pylyshyn 1973). Knowledge expressed in propositional format can determine part of the state of the world while conveniently leaving other parts undetermined; whereas, an image determines everything about a particular state of the world. Thus propositional knowledge can often be represented by many images, each of which must make assumptions about details that are independent of the propositions. If such an image were examined (as an image-based reasoner might do), it would not be clear what details were necessitated by the propositional knowledge and what details were based on assumptions. How then can an image be used to represent and reason about propositional knowledge without being over-specific? We call this the Indeterminacy Problem.

The Indeterminacy Problem stems from an unnecessarily naive assumption about how images could be used for reasoning. Those who would claim that images cannot be used to implement a reasoner due to the Indeterminacy Problem apparently assume that only a single, static image is used. However, images can be manipulated dynamically; positions, shapes, etc. can be adjusted to create new images out of old ones. Previously, we introduced a sophisticated computational model of imagery, ISR (Indeterminacy in Spatial Reasoning), which exploits this dynamic nature of images and which, at least in some domains, can permit an image-based reasoner to avoid the Indeterminacy Problem (Ioerger 1991).

In this paper, we will review the ISR model and its application to reasoning with images in an abstract spatial domain. Then we will report our most recent attempt to apply ISR in a more realistic domain which involves reasoning about the spatial relationships of objects in a room. Although the program was moderately successful, modeling the effects of gravity was complicated. The difficulties point out a subtle but crucial interaction among various components of ISR. We conclude by discussing the similarity to planning and the need for heuristics that trade completeness for speed.

2 The ISR Model of Imagery

ISR is a computational model of how images can be used to reason about the spatial relationships among objects. The model assumes that some external agent is submitting a sequence of assertions and queries in a propositional language to a spatial reasoner. For example, the agent might attempt to use the reasoner to see if C is to the right of A when A is to the left of B and B is to the left of C by submitting **assert(LEFT(A,B)) assert(LEFT(B,C)) query(RIGHT(C,A))**. The reasoner should accept the assertions in sequence, and when a query is received, an answer such as "yes," "no," or "unspecified" should be returned. These responses indicate, respectively, that the query is necessarily implied by the previous assertions, that the query is necessarily false based on the previous assertions, or that the previous assertions do not constrain the truth-value of the query.

An important component of the ISR model is the set of recognition functions which compute the truth-value of a proposition (a predicate with specific arguments) in an image. The value returned is a real number between 0 and 1 inclusive, indicating the degree to which the image satisfies the proposition (0 means false; 1 means true). For example, a common word-sense (the one taken in this paper) of "left" has graded membership; one object can be *directly* to the left of another, *absolutely not* to the left, or *somewhat* to the left, depending in a strictly decreasing manner on the angle between the vector from the reference to the located object and a vector pointing left. Note that, even when the located object is in the half-plane to the right of the reference object, there are some positions that hardly satisfy the proposition but nonetheless do so *more* than other positions. The evaluation of an image with respect to the negation of a proposition is taken to be one minus the evaluation with respect to the proposition itself, and the evaluation of an image against a conjunction of propositions is taken to be the minimum over the independent evaluations by each conjunct.

At all times, a single current image, as well as the set of previous assertions, is maintained. As assertions are made, they are inserted into the set, and the image is adjusted to accomodate the new proposition along with the previously asserted ones "as well as possible." The component of ISR that carries out this update with respect to the enlarged set of propositions is called the Adjustment Procedure. From the current image, the Adjustment Procedure generates a set of adjusted images in which the positions of some objects

*This material is based upon work supported under a National Science Foundation Graduate Fellowship and ONR grant N00014-88K0124. E-mail: ioerger@cs.uiuc.edu

are changed. Exactly which adjustments are made depends on the domain. Then the current image and each adjusted image is evaluated against the set of propositions. If some adjusted image is better than the current image, a best image is selected to replace the current image, and the procedure iterates. The Adjustment Procedure thus hill-climbs until it finds the best image "possible" that satisfies the new assertion along with the all the previous ones. Of course, this procedure can be subject to the various problems often encountered with hill-climbing. It is this issue that we will discuss below in the context of a realistic domain.

When a query is made, it is not sufficient to simply evaluate the current image against the proposition. Because of the Indeterminacy Problem, the specific answer to the query which is derived from the image could be either necessitated by the assertions or merely based on a consistent assumption. Thus the ISR model suggests that we consider both the best adjusted image that could be found if the queried proposition had been asserted, and the best adjusted image that could be found if its negation had been asserted. Suppose that, by asserting the query, the Adjusment Procedure finds an image that has a high evaluation (above some threshold) with respect to the set of previous assertions plus the query, but by asserting the negation of the query, the best image the Adjustment Procedure can find only has a low evaluation (below some possibly different threshold). Then the query is consistent with the previous assertions and its negation is inconsistent with them, so the queried proposition is necessarily true and the response to the query should be **YES**. In the symmetrical case, the query is found to be inconsistent with the previous assertions but its negation is consistent with them, so the response should be **NO**. If, by asserting either the queried proposition or its negation, the Adjustment Procedure can find images which evaluate high with repect to the appropriately enlarged set of propositions, then either is consistent with the previous assertions, so the response should be **UNSPECIFIED**. Thus the ISR model of imagery demonstrates a sophisticated method for reasoning with images that can avoid the Indeterminacy Problem.

3 An Abstract Spatial Domain

The first domain in which the ISR model was used to reason with images was an abstract spatial domain. This domain consisted of a square region (2-dimensional and bounded) containing three volumeless, non-overlapping objects (points), call them A, B, and C. For the reasoning task, it was sufficient to discretize the region into sub-regions by dividing each edge of the region into 10 units of length. Thus any particular arrangement could be represented with a 10×10 array in which each element indicates that the corresponding sub-region is empty or contains one of the objects. In the abstract spatial domain, the propositions that could be asserted or queried were constructed from a predicate (LEFT, RIGHT, ABOVE, or BELOW) and two arguments naming different objects. As an example, LEFT(A,B) means "A is to the left of B."

In order to discuss the performance of an implementation of the ISR model in the abstract spatial domain, it is necessary to specify exactly what "adjustment" means and what the recognition functions are like. In this case, an adjustment of an image is a nearly identical image, but with some object moved a small distance (1 unit) in any direction (as long as the object does not overlap another or go outside the bounds of the region). In the Adjustment Procedure, the objects to be adjusted are those involved in some proposition that is least satisfied in the current image, and they are moved one step in any one of four directions: up, down, left, or right. The recognition function for LEFT first computes the angle (θ) between the vector from the reference object (the second argument) to the located object (the first argument) and the unit vector pointing left, and then returns the linear transformation $(1 - \theta/\pi)$ of this angle from the range 0 to π into the range 1 to 0. The other recognition functions can be computed by rotating an image and calling the function for LEFT with the same arguments.

The implementation of the ISR model as described above worked fairly well in the abstract spatial domain. Prior to receiving any assertions or queries, a specific but arbitrary arrangement was adopted as the initial image. However, the fact that no propositions had been asserted meant that no query would be constrained to be particularly true or false; the positions of the objects could always be adjusted from the initial image independently to accomodate the queried proposition or its negation, so initial responses were always **UNSPECIFIED**. As assertions were received, the current image was incrementally updated and the propositions were added to the set of previous assertions. When a query was received, the recognition functions for the previously-asserted propositions constrained the search for images that were examples or counter-examples of the queried proposition. Because the images were dynamically manipulated, indeterminacy did not cause any problems. As an example, assert(**RIGHT(B,A)**) query(**ABOVE(B,A)**) would elicit the response **UNSPECIFIED**. To extend the example, the same response would be elicited for assert(**RIGHT(B,A)**) assert(**ABOVE(C,B)**) query(**ABOVE(C,A)**) because, the following two images are reasonable examples of the query and its negation:

```
 ----------      ----------
|        C |    | A        |
| A      B |    |        C |
|          |    |        B |
 ----------      ----------
```

The power of the ISR-based reasoning system was apparent in some of the inferences that could be made in this domain. For example, suppose after some assertions, query(**LEFT(A,B)**) elicited **YES**. Then query(**LEFT(B,A)**) would elicit **NO**, which we would expect based on the anti-symmetry of the LEFT relation. Also, query(**RIGHT(B,A)**) would elicit **YES**, which we would expect based on the symmetry of LEFT and RIGHT. Similarly, the transitivity of each relation was evident. Note, however, that all of these inferences were based solely on the recognition functions; the interacting constraints among propositions could be detected by the hill-climbing Adjustment Procedure. No rules that state the relationships among propositions were needed.

4 A More Realistic Domain

In order to challenge the capabilities of ISR, we next applied the model to reasoning with images in a more realistic domain. In particular, the domain consisted of a room with four objects in it: a book, a table, a penguin, and a fly. This domain had two minor differences with the previous abstract spatial domain. First, there was a third dimension. Hence possible adjustments to the position of an object had to include a small distance forward and backward, as well as left, right, up, and down. Second, it was important to represent the different volumes occupied by different objects because relative sizes affected inference. For example, if the fly is on the book, which is on the floor, then the fly cannot be above the table. But if the fly is on (the top of) the table and the book is on the floor (sufficiently close to the table), then the fly could be above the book. Thus each object was given a default length, width, and height. A particular arrangement of objects in the room could be represented with a $10 \times 10 \times 10$ array in which elements corresponding to multiple, contiguous sub-regions might indicate occupancy by the same object. As in the abstract spatial domain, objects were not permitted to overlap.

The assertion/query language in this more realistic domain was similar to that in the abstract spatial domain. The predicates BEFORE and BEHIND were added, and of course the terms for proposition-arguments named particular objects in the room. The recognition functions for LEFT, RIGHT, ABOVE, BELOW, BEFORE, and BEHIND, like those in the abstract spatial domain, were linear transformations of the angle (in 3 dimensions) between the vector from the reference object to the located object and some special vector (for example, the unit vector pointing left). Since the objects were not just points, the vector in an image between two objects was defined by their centers, which could be found easily by an image-scanning procedure.

As in the abstract spatial domain, one arbitrary arrangement of the objects in the room had to be given prior to any assertions or queries. In this particular image, the book, table, and penguin are all on the floor, and the fly is hovering in air in the middle of the room. Initial experiments with an implementation of the image-based reasoner using ISR as described, however, revealed a significant difference between this more realistic domain and the abstract spatial domain. In one particular case, we asserted **LEFT(BOOK,TABLE)**. Since in the initial image the book was on the floor nearly but not directly to the left of the table, some adjustment had to be done. Upon inspection of the updated image, the book was indeed to the left of the table, but both objects were represented as floating in mid-air! Clearly, the Adjustment Procedure had not been programmed to model the constraints gravity puts on the arrangements of real-world objects.

Making the Adjustment Procedure model gravity was nontrivial. When a proposition was asserted or queried, the names used as arguments were looked up in a lexicon to obtain a semantic feature called WEIGHTED. If an object was marked +WEIGHTED, it was considered to be affected by gravity in such a way that, in any reasonable image, the object would have to be supported by at least one point of another object (ignoring balance) or be at the bottom of the room (on the floor). The fly was the only -WEIGHTED object and was allowed to be represented as floating freely in mid-air. The Adjustment Procedure was then programmed to be sensitive to this feature. -WEIGHTED objects could be adjusted as before: up, down, left, right, forward, and back. But a +WEIGHTED object could not be adjusted up, since it would be become unsupported, and it could not be adjusted down, since it would then overlap with the object supporting it.

Two interesting special cases could occur while adjusting the position of a +WEIGHTED object in an image. First, the object could be adjusted laterally off the edge of its support. In this case, a procedure iteratively adjusts the object downward until it finds a new object or the bottom of the room (the floor) on which to rest. The inverse case occurs when a lateral adjustment of a +WEIGHTED object causes it to overlap with another object, implying that they are adjacent. If this adjustment were simply ignored, then objects would not be able to stack up. For example, assume that after receiving **assert(ABOVE(BOOK,TABLE)) assert(ABOVE(PENGUIN,TABLE))**, the current image depicts the book and the penguin both on the top of the table. Then, if **assert(ABOVE(BOOK,PENGUIN))** were received, there would be no way for the book to "jump" up on top of the penguin. Thus, when a +WEIGHTED object is blocked from making a particular lateral adjusment, a procedure must scan upward for some free space–hopefully on the edge of the would-be obstacle–to allow the adjusment.

5 Analysis of Incompleteness

Unfortunately, even when the Adjustment Procedure was reprogrammed for +WEIGHTED objects, it still could not properly handle the sequence **assert(ABOVE(BOOK, TABLE)) assert(ABOVE(PENGUIN,TABLE)) assert(ABOVE(BOOK,PENGUIN))**. The reason an image with the book on top of the penguin (which is on top of the table) could not be found illustrates a subtle but crucial interaction among various components of the ISR model of imagery. Recall that the recognition function for ABOVE is based on the angle between the vector from the center of the reference object to the center of the located object and a vector pointing up. In the present case, since the center of the book is below the center of the penguin (when they are both on the table), the only adjustment that would increase the value of the recognition function for ABOVE would be to move the book *away* from the penguin, since gravity prohibits the book from being adjusted upward. Thus the book can never get close enough for the Adjustment Procedure to consider letting the it "jump" up on top of the penguin.

The example illustrates the interaction between adjustment and the recognition functions. Because the Adjusment Procedure is based on hill-climbing, it is potentially susceptible to such causes of incomplete search as the plateau, ridge, and foothill problems (Winston 1984). However, in forfeiting completeness, tremendous efficiency is gained. The number of possible arrangements that could be used to update the current image after a new assertion is made is on the order of 10^{12} (10 units per dimension, 4 objects, 3 dimensions). However, in this domain, the new assertion, if badly violated, at most requires an object or two to move across the room, taking on the order of 10 steps (multiplied by some small constant to account for crooked paths). Since each step requires the generation and evaluation of only 4 to 6 adjustments, the total number of images considered is only on the order of 100–a significant decrease in complexity!

To see why we should expect the hill-climbing Adjustment Procedure to work at all, let us consider how adjustment and the recognition functions work together in the abstract spatial domain. Suppose a single proposition like **LEFT(A,B)** is asserted, and the recognition function for this proposition applied to the initial image only returns 0.6. By generating and evaluating adjusted images in which A or B is moved a small distance up, down, left, or right, the Adjustment Procedure is essentially exploring the local gradient of the recognition function. Since, at each step in the iteration, the Adjustment Procedure chooses to move an object in the direction that will maximize the increase in the evaluation of the updated image, it is crucial that the adjustment that looks best *locally* can in fact be completed with similar steps to find the *globally* best image.

This works in the abstract spatial domain because each recognition function has the following nice property with respect to the adjustments considered. Notice that the adjustments to the position of one object are guaranteed to rotate it in both directions relative to the other object. Thus the strictly decreasing dependence of the recognition function on the angle between the vector between two objects and some special vector means that an adjustment that increases the evaluation of the current image can always be found, unless the proposition is already maximally satisfied. Thus it is guaranteed that, by choosing an adjustment that looks locally best and continuing to do so, the value of the recognition function will increase without encountering local maxima, plateaux, or ridges until the Adjustment Procedure finds a globally best image.

Thus the Adjustment Procedure seems to be complete when searching for best images satisfying single propositions. Unfortunately, when there are multiple propositions to be considered, the Adjustment Procedure cannot be guaranteed to find the best update. Since the evaluation of an image against a set of propositions is taken to be the minimum independent evaluation, this defines a composite function which might have local maxima, plateaux, or ridges. However, in our experiments in both the abstract spatial domain and the more realistic domain, the Adjustment Procedure did not appear to fail frequently for this particular reason.

Another potential cause of incompleteness is the "obstacle" effect of other objects. It is conceiveable that all adjustments that would lead to more highly evaluated images might be rejected because they would cause objects to overlap. In the abstract spatial domain this was generally not a problem because there were only three objects and the positions of two of them were considered for adjustment in each step. In the more realistic domain, the refinement of the Adjustment Procedure for +WEIGHTED objects allowed most objects to pass over or under any obstacles. However, in general, obstacles can be expected to cause the Adjustment Procedure to be incomplete.

The problem mentioned above in which the book could not "jump" up on the penguin while both were on the table can now be analyzed. Due to the constraints of gravity, the Adjustment Procedure only generates a restricted set of adjustments for the book. Although moving it upward would be best, the book cannot become unsupported. Thus moving it away from the penguin appears to be the best adjustment. The restriction of the adjustments causes the local gradient of the recognition function for ABOVE to be misinterpreted. By tweaking the definition of adjustment, we have ruined the nice interaction it had with the recognition functions.

There are three fairly obvious solutions to this particular problem, each of which has significant disadvantages. First, we could consider changing the recognition function for ABOVE so that, if the located object is below the horizontal plane containing the reference object, then the evaluation would increases as the distance between them decreased. This solution tailors the gradient of the recognition function so that the Adjustment Procedure will be lead to find the best adjusted image, but it is burdensome if not incorrect to change the semantics of the recognition functions for this purpose. A second solution would be to relax the requirement that objects not overlap and the constraints of gravity during adjusment. Thus objects could pass through other objects or "float" through air to acheive a satisfactory arrangement. However, it would seem that, at some point before the Adjustment Procedure is done, the constraints must be enforced once again, and doing so at the last moment could cause arbitrary changes to the arrangement that might severely violate the satisfaction of some propositions.

The third solution is to allow objects to be adjusted from their locations by longer distances than just one unit. In fact, the refinement of the Adjustment Procedure in which +WEIGHTED objects were enabled to "fall off" or "jump up on" other objects is an example of this heuristic. As another example, suppose that the best image the Adjustment Procedure is able to find only satisfies the propositions to a low degree. The Adjustment Procedure might have been stuck at a local maximum, or perhaps the objects it was trying to adjust were trapped by some obstacles. The only way to allow the Adjustment Procedure to do better would be to consider moving some object far enough away from its current position so that it can get out of the local maximum or the trap. The ability to make larger adjustments must be employed judiciously, however. It is primarily the restricted set of adjustments that keeps the complexity of the Adjustment Procedure under control. Also, by making only small adjustments with each step, the hill-climbing Adjustment Procedure can detect interactions among propositions. It is necessary to take these small steps in order to identify the point at which the best adjustment for one proposition intolerably violates the satisfaction of another.

6 Conclusion

One can think of the task of finding, by incremental adjustment, an image that best satisfies a set of propositions as a planning task. In fact, the use of local information to plan each step in search of a globally best adjusted image suggests that the ISR model is essentially a reactive planner (Schoppers 1987). As is an advantage with most reactive systems, the exact effects of each step do not have to be given explicitly; interactions are more or less discovered and handled dynamically. It is also clear from a planning perspective that multiple assertions act as conjunctive subgoals, and certain situations where the Adjustment Procedure is incomplete, especially those in which obstacles block adjustments, can be described as non-linear interactions. Chapman (1987) has shown that such planning problems are NP-hard. Thus we suggest that natural language semantic processors based on mental models require heuristics, like the one implemented in this paper, that trade completeness for speed.

In this paper we have reported our most recent attempt to apply the ISR model of imagery to reasoning in a more realistic domain in which spatial prepositions are used to relate the locations of objects in a room. The constraints of gravity were difficult to incorporate into the ISR-based reasoner. The solution that was implemented made a domain-specific change in the search procedure that caused incompleteness. A few alternative solutions, each with disadvantages, were discussed, and in the light of the similarity to planning, it was observed that heuristics that forfeit completeness (like the solution implemented) are necessary in order to obtain the speed-up offered by diagrammatic approaches to reasoning about spatial prepositions while avoiding the Indeterminacy Problem.

References

Chapman, D (1987) Planning for conjunctive goals. *Artificial Intelligence*, 32:333-377.

Ioerger, TR (1991) Imagery and categories: The Indeterminacy Problem. *Thirteenth Annual Conference of the Cognitive Science Society*, 37-42.

Johnson-Laird, PN (1983) *Mental Models*. Harvard University Press.

Pylyshyn Z (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80:1-24.

Schoppers MJ (1987) Universal plans for reactive robots in unpredictable environments. *Tenth International Conference on Artificial Intelligence*, 1039-1046.

Waltz DL and Boggess L (1979) Visual analog representations for natural language understanding. *Sixth International Joint Conference on Artificial Intelligence*, 926-934.

Winston PH (1984) *Artificial Intelligence*. Addison-Wesley Publishing Company, Menlo Park, CA.