UNDERSTANDING SCENE DESCRIPTIONS AS EVENT SIMULATIONS[1]

David L. Waltz
University of Illinois at Urbana-Champaign

The language of scene descriptions[2] must allow a hearer to build structures of schemas similar (to some level of detail) to those the speaker has built via perceptual processes. The understanding process in general requires a hearer to create and run "event simulations" to check the consistency and plausibility of a "picture" constructed from a speaker's description. A speaker must also run similar event simulations on his own descriptions in order to be able to judge when the hearer has been given sufficient information to construct an appropriate "picture", and to be able to respond appropriately to the hearer's questions about or responses to the scene description.

In this paper I explore some simple scene description examples in which a hearer must make judgements involving reasoning about scenes, space, common-sense physics, cause-effect relationships, etc. While I propose some mechanisms for dealing with such scene descriptions, my primary concern at this time is to flesh out our understanding of just what the mechanisms must accomplish: what information will be available to them and what information must be found or generated to account for the inferences we know are actually made.

1. THE PROBLEM AREA

An entity (human or computer) that could be said to fully understand scene descriptions would have to have a broad range of abilities. For example, it would have to be able to make predictions about likely futures; to judge certain scene descriptions to be implausible or impossible; to point to items in a scene, given a description of the scene; and to say whether or not a scene description corresponded to a given scene experienced through other sensory modes.[3] In general, then, the entity would have to have a sensory system that it could use to generate scene representations to be compared with scene representations it had generated on the basis of natural language input.

In this paper I concentrate on 1) the problems of making appropriate predictions and inferences about described scenes, and 2) the problem of judging when scene descriptions are physically implausible or impossible.

I do not consider directly problems that would require a vision system, problems such as deciding whether a linguistic scene description is appropriate for a perceived scene, or generating linguistic scene descriptions from visual input, or learning scene description language through experience.

I also do not consider speech act aspects of scene descriptions in much detail here. I believe that the principles of speech acts transcend topics of language; I am not convinced that the study of scene descriptions would lead to major insights into speech acts that couldn't be as well gained through the study of language in other domains.

I do believe, however, that the study of scene descriptions has a considerable bearing on other areas of language analysis, including syntax, semantics, and pragmatics. For example, consider the following sentences:

(S1) I saw the man on the hill with my own eyes.
(S2) I saw the man on the hill with a telescope.
(S3) I saw the man on the hill with a red ski mask.

The well-known sentence S2 is truly ambiguous, but S1 and S3, while likely to be treated as syntactically similar to S2 by current parsers, are each relatively unambiguous; I would like to be able to explain how a system can choose the appropriate parsings in these cases, as well as how a sequence of sentences can add constraints to a single scene-centered representation, and aid in disambiguation. For example, if given the pair of sentences:

(S2) I saw the man on the hill with a telescope.
(S4) I cleaned the lens to get a better view of him.

a language understanding system should be able to select the appropriate reading of S2.

I would also like to explore mechanisms that would be appropriate for judging that

(S5) My dachshund bit our mailman on the ear.

requires an explanation (dachshunds could not jump high enough to reach a mailman's ear, and there is no way to choose between possible scenarios which would get the dachsund high enough or the mailman low enough for the biting to take place). The mechanisms must also be able to judge that the sentences:

(S6) My doberman bit our mailman on the ear.
(S7) My dachshund bit our gardener on the ear.
(S8) My dachshund bit our mailman on the leg.

do not require explanations.

A few words about the importance of explanation are in order here. If a program could judge correctly which scene descriptions were plausible and which were not, but could not explain why it made the judgements it did, I think I would feel profoundly dissatisfied with and suspicious of the program as a model of language comprehension. A program ought to consider the "right options" and decide among them for the "right reasons",[4] if it is to be taken seriously as a model of cognition.

I will argue that scene descriptions are often most naturally represented by structures which are, at least in part, only awkwardly viewed as propositional; such representations include coordinate systems, trajectories, and event-simulating mechanisms, i.e. procedures which set up models of objects, interactions, and constraints, "set them in motion", and "watch what happens". I suggest that event simulations are supported by mechanisms that model common-sense physics and human behavior

I will also argue that there is no way to put limits on the degree of detail which may have to be considered in constructing event simulations; virtually any feature of an object can in the right circumstances become centrally important.

[1]This work was supported in part by the Office of Naval Research under Contract ONR-N00014-75-C-0612 with the University of Illinois, and was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by ONR under Contract No. N00014-77-C-0378 with Bolt Beranek and Newman Inc.

[2]The term "scene" is intended to cover both static scenes and dynamic scenes (or events) that are bounded in space and time.

[3]In general I believe that many of the event simulation procedures ought to involve kinesthetic and tactile information. I by no means intend the simulations to be only visual, although we have explored the AI aspects of vision far more than those of any other senses.

[4]An explanation need not be in natural language; for example, I probably could be convinced via traces of a program's operation that it had been concerned with the right issues in judging scene plausibility.

2. THE NATURE OF SCENE DESCRIPTIONS

I have found it useful to distinguish between static and dynamic scene descriptions. Static scene descriptions express spatial relations or actions in progress, as in:

(S9) The pencil is on the desk.
(S10) A helicopter is flying overhead.
(S11) My dachshund was biting the mailman.

Sequences of sentences can also be used to specify a single static scene description, a process I will refer to as "detail addition". As an example of detail addition, consider the following sequence of sentences (taken from Waltz & Boggess [1]):

(S12) A goldfish is in a fish bowl.
(S13) The fish bowl is on a stand.
(S14) The stand is on a desk.
(S15) The desk is in a room.

A program written by Boggess [2] is able to build a representation of these sentences by assigning to each object mentioned a size, position, and orientation in a coordinate system, as illustrated in figure 1. I will refer to such representations as "spatial analog models" (in [1] they were called "visual analog models"). Objects in Boggess's program are defined by giving values for their typical values of size, weight, orientation, surfaces capable of supporting other objects, as well as other properties such as "hollow" or "solid", and so on.

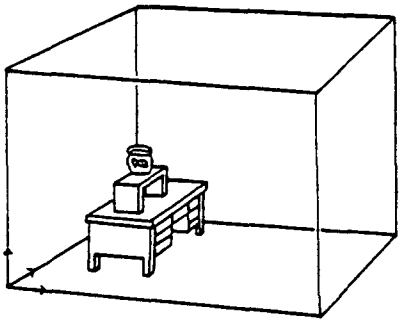

Figure 1 A "visual analog model" of S12-S15.

Dynamic scene descriptions can use detail addition also, but more commonly they use either the mechanisms of "successive refinement" [3] or "temporal addition". "Temporal addition" refers to the process of describing events through a series of time-ordered static scene descriptions, as in:

(S16) Our mailman fell while running from our dachshund.
(S17) The dachshund bit the mailman on the ear.

"Successive refinement" refers to a process where an introductory sentence sets up a more or less prototypical event which is then modified by succeeding sentences, e.g. by listing exceptions to one's ordinary expectations of the prototype, or by providing specific values for optional items in the prototype, or by similar means. The following sentences provide an example of "successive refinement":

(S18) A car hit a boy near our house.
(S19) The car was speeding eastward on Main Street at the time.
(S20) The boy, who was riding a bicycle, was knocked to the ground.

3. THE GOALS OF A SCENE UNDERSTANDING SYSTEM

What should a scene description understanding system to do with a linguistic scene description? Basically 1) verify plausibility, 2) make inferences and predictions, 3) act if action is called for, and 4) remember whatever is important. For the time being, I am only considering 1) and 2) in detail. In order to carry out 1) and 2), I would like my system to turn scene descriptions (static or dynamic) into a time sequence of "expanded spatial analog models", where each expanded spatial analog model represents either 1) a set of spatial relationships (as in S12-S15), or 2) spatial relationships plus models of actions in progress, chosen from a fairly large set of primitive actions (see below), or 3) prototypical actions that can stand for sequences of primitive actions. These prototypical actions would have to be fitted into the current context, and modified according to the dictates of the objects and modifiers that were supplied in the scene description.

The action prototype would have associated selection restrictions for objects; if the objects in the scene description matched the selection restrictions, then there would be no need to expand the prototype into primitives, and the "before" and "after" scenes (similar to pre- and post-conditions) of the action prototype could be used safely.

If the selection restrictions were violated by objects in the scene, or if modifiers were present, or if the context did not match the preconditions, then it would have to be possible to adapt the action prototype "appropriately". It would also have to be possible to reason about the action without actually running the event simulation sequence underlying it in its entirety; sections that would have to be modified, plus before and after models, might be the only portions of the simulation actually run. The rest of the prototype could be treated as a kind of "black box" with known input-output characteristics.

I have not yet found a principled way to enumerate the primitives mentioned above, but I believe that there should be many of them, and that they should not necessarily be non-overlapping; what is most important is that they should have precise representations in spatial analog models, and be capable of being used to generate plausible candidates for succeding spatial analog models. Some examples of primitives I have looked at and expect to include are: break-object-into-parts, mechanically-join-parts, hit, touch, support, translate, fall.

As an example of the expansion of a non-primitive action into primitive actions, consider "bite x y"; its steps are: 1)[set-up] instantiate x[5] as a "biting-thing" -- defaults = mouth, teeth, jaws of an animate entity; 2) instantiate y as "thing-bitten"; 3)[before] x is open and does not touch y and x partially surrounds y (i.e. y is not totally inside x); 4) x is closing on y; 5)[action] x is touching y, preferably in two places on opposite sides of y and x continues to close; 6) x deforms y; 7)[after] x is moving away from y, and no longer touches y.

Finally, lest it should not be clear from the sketchiness of the comments above, I am by no means satisfied yet with these ideas as an explanation of scene description understanding, although I am confident that this research is headed in the right general direction.

4. PLAUSIBILITY JUDGEMENT

The basic argument I am advancing in this paper is this: it is essential in understanding scene descriptions to set up and run event simulations for the scenes; we judge the plausibility (or possibility), meaningfulness, and completeness of a description on the basis of our experience in attempting to set up and run the simulation. By studying cases where we judge descriptions to be implausible we can gain insight into just what is done routinely during the understanding of scene descriptions, since these cases correspond to failures in setting up or running event simulations.

[5] By "instantiate an X" I mean assign X a physical place, posture, orientation, etc. or retrieve a pointer to such an instantiation, if it is a familiar one. Thus "instantiate a baby" would retrieve a pointer, whereas "instantiate a two-headed dog" would probably have to attempt to generate one on the spot. Note that this process may itself fail, i.e. that an entity may not be able to "imagine" such an object.

As the examples below illustrate, sometimes an event simulation simply cannot be set up because information is missing, or several possible "pictures" are equally plausible, or the objects and actions being described cannot be fitted together for a variety of reasons, or the results of running the simulation do not match our knowledge of the world or the following portions of the scene description, and so on. It is also important to emphacize that our ultimate interest is in being able to succeed in setting up and running event simulations; therefore I have for the most part chosen ambiguous examples where at least one event simulation succeeds.

4.1 TRANSLATING AN OLD EXAMPLE INTO NEW MECHANISMS

Consider Bar-Hillel's famous sentence [4]:[6]

(S10) The box is in the pen.

Plausibility judgement is necessary to choose the appropriate reading, i.e. that "pen" = playpen. Minor extensions to Boggess's program could allow it to choose the appropriate referent for pen. Pen1 (the writing implement) would be defined as having a relatively fixed size (subject to being overridden by modifiers, as in "tiny pen" or "twelve inch pen"), but the size of pen2 (the enclosure) would be allowed to vary over a range of values (as would the size of box). The program could attempt to model the sentence by instantiating standard (default-sized) models of box, pen1, and pen2, and attempting to assign the objects to positions in a coordinate system such that the box would be in pen1 or pen2. Pen1 could not take part in such a spatial analog model both because of pen1's rigid size, and the extreme shrinkage that would be required of box (outside box's allowed range) to make it smaller than the pen1, and also because pen1 is not a container (i.e. hollow object). Pen2 and box prototypes could be fitted together without problems, and could thus be chosen as the most appropriate interpretation.

4.2 A SIMPLE EVENT SIMULATION

Extending Boggess's program to deal with most of the other examples given in this paper so far would be harder, although I believe that S1-S4 could be handled without too much difficulty. Let us look at S2 and S4 in more detail:

(S2) I saw the man on the hill with a telescope.
(S4) I cleaned the lens to get a better view of him.

After being told S2, a system would either pick one of the possible interpretations as most plausible, or it might be unable to choose between competing interpretations, and keep them both. When it is told S4, the system must first discover that "the lens" is part of the telescope. Having done this, S4 unambiguously forces the placement of the speaker to be close enough to the telescope to touch it. This is because all common interpretations of clean require the agent to be close to the object. At least two possible interpretations still remain: 1) the speaker is distant from the man on the hill, and is using the telescope to view the man; or 2) the speaker, telescope, and man on the hill are all close together. The phrase "to get a better view of him" refers to the actions of the speaker in viewing the man, and thus makes interpretation 1) much more likely, but 2) is still conceivable. The reasoning necessary to choose 1) as most plausible is rather subtle, involving the idea that telescopes are usually used to look at distant objects.

In any case, the proposed mechanisms should allow a system to discard an interpretatiion of S2 and S4 where the man on the hill had a telescope and was distant from the speaker.

[6] A central figure in the machine translation effort of the late 50's and early 60's, Bar-Hillel cited this sentence in explaining why machine translation was impossible. He subsequently quit the field.

4.3 SIMULATING AN IMPLAUSIBLE EVENT

Let us also look again at S5:

(S5) My dachshund bit our mailman on the ear.

and be more specific about what an event simulation should involve in this rather complex case. The event simulation set up procedures I envision would execute the following steps:

1) instantiate a standard mailman and dachshund in default positions (e.g. both standing on level ground outdoors on a residential street with no special props other than the mailman's uniform and mailbag);

2) analyze the preconditions for "bite" to find that they require the dog's mouth to surround the mailman's ear;

3) see whether the dachshund's mouth can reach the mailman's ear directly (no);

4) see whether the dog can stretch high enough to reach (no; this test would require an articulated model of the dog's skeleton or a prototypical representation of a dog on its hind legs.);

5) see whether a dachshund could jump high enough (no; this step is decidedly non-trivial to implement![7]);

6) see whether the mailman ordinarily gets into any positions where the dog could reach his ear (no);

7) conclude that the mailman could not be bitten as stated unless default sizes or movement ranges are relaxed in some way. Since there is no clearly preferred way to relax the defaults, more information is necessary to make this an "unambiguous" description.

I have quoted "unambiguous" because the sentence S5 is not ambiguous in any ordinary sense, lexically or structurally. What is ambiguous are the conditions and actions which could have led up to S5. Strangely enough, the ordinary actions of mailmen (checked in step 6) seem relevant to the judgement of plausibility in this sentence. As evidence for this analysis, note that the substitution of "gardener" for "mailman" turns (S5) into a sentence that can be simulated without problems. I think that it is significant that such peripheral factors can be influential in judging the plausibility of an event. At the same time, I am aware that the effect in this case is rather weak, that people can accept this sentence without noting any strangeness, so I do not want to draw conclusions that are too strong.

4.4 MAKING INFERENCES ABOUT SCENES

Consider the following passage:

(P1) You are at one end of a vast hall stretching forward out of sight to the west. There are openings to either side. Nearby, a wide stone staircase leads downward. The hall is filled with wisps of white mist swaying to and fro almost as if alive. A cold wind blows up the staircase. There is a passage at the top of the dome behind you. Rough stone steps lead up the dome.

Given this passage (taken from the computer game "Adventure") one can infer that it is possible to move to the west, north, south, or east (up the rough stone steps). Note that this information is buried in the description; in order to infer this information, it would be useful to construct a spatial analog model,

[7] Although one could do it by simply including in the definition of a dog information about how high a dog can jump, e.g. no higher than twice the dog's length. However I consider this something of a "hack", because it ignores some other problems, for example the timing problem a dog would face in biting a small target like a person's ear at the apex of its highest jump. I would prefer a solution that could, if necessary, perform an event simulation for step 5), rather than trust canned data.

with "you" facing west, and the scene features placed appropriately. In playing Adventure, it is also necessary to remember salient features of the scenes described so that one can recognize the same room later, given a passage such as:

(P2) You're in hall of mists. Rough stone steps lead up the dome. There is a threatening little dwarf in the room with you.

Adventure can only accept a very limited class of commands from a player at any given point in the game. It is only possible to play the game because one can make reasonable inferences about what actions are possible at a given point, i.e. take an object, move in some direction, throw a knife, open a door, etc. While I am not quite sure what make of my observations about this example, I think that games such as Adventure are potentially valuable tools for gathering information about the kinds of spatial and other inferences people make about scene descriptions.

4.5 MIRACLES AND WORLD RECORDS

With some sentences there may be no plausible interpretation at all. In many of the examples which follow, it seems unlikely that we actually generate (at least consciously) an event simulation. Rather it seems that we have some shortcuts for recognizing that certain events would have to be termed "miraculous" or difficult to believe.

(S22) My car goes 2000 miles on a tank of gas.
(S23) Mary caught the bullet between her teeth.
(S24) The child fell from the 10th story window to the street below, but wasn't hurt.
(S25) We took the refrigerator home in the trunk of our VW Beetle.
(S26) She had given birth to 25 children by the age of 30.
(S27) The robin picked up the book and flew away with it.
(S28) The child chewed up and swallowed the pair of scissors.

The Guinness Book of World Records is full of examples that defy event simulation. How one is able to judge the plausibility of these (and how we might get a system to do so) remains something of a mystery to me.

The problem of recognizing obviously implausible events rapidly is an important one to consider for dealing with pronouns. Often we choose the appropriate referent for a pronoun because only one of the possible referents could be part of a plausible event if substituted for the pronoun. For example, "it" must refer to "milk", not "baby", in S29:

(S29) I didn't want the baby to get sick from drinking the milk, so I boiled it.

5. THE ROLE OF EVENT SIMULATION IN A FULL THEORY OF LANGUAGE

I suggested in section 3 that a scene description understanding system would have to 1) verify the plausibility of a described scene, 2) make inferences or predictions about the scene, 3) act if action is called for, and 4) remember whatever is important. As pointed out in section 4.5, event simulations may not even be used for all cases of plausibility judgement. Furthermore, scene descriptions constitute only one of many possible topics of language. Nonetheless, I feel that the study of event simulation is extremely important.

5.1 WHY ARE SIMPLE PHYSICAL SCENES WORTH CONSIDERING?

For a number of reasons, methodological as well as theoretical, I believe that it is not only worthwhile, but also important to begin the study of scene descriptions with the world of simple physical objects, events, and physical behaviors with simple goals.

1) Methodologically it is necessary to pick an area of concentration which is restricted in some way. The world of simple physical objects and events is one of the simplest worlds that links language and sensory descriptions.

2) As argued in the work of Piaget [5], it seems likely that we come to comprehend the world by first mastering the sensory/motor world, and then by adapting and building on our schemata from the sensory/motor world to understand progressively more abstract worlds. In the area of language Jackendoff [6] offers parallel arguments. Thus the world of simple physical objects and behaviors has a privileged positions in the development of cognition and language.

3) Few words in English are reserved for describing the abstract world only. Most abstract words also have a physical meaning. In some cases the physical meanings may provide important metaphors for understanding the abstract world, while in other cases the same mechanisms that are used in the interpretation of the physical world may be shared with mechanisms that interpret the abstract world.

4) I would like the representations I develop for linguistic scene descriptions to be compatible with representations I can imagine generating with a vision system. Thus this work does have an indirect bearing on vision research: my representations characterize and put constraints on the types and forms of information I think a vision system ought to be able to supply.

5) Even in the physical domain, we must come to grips with some processes that resemble those involved in the generation and understanding of metaphor: matching, adaptation of schemata, modification of stereotypical items to match actual items, and the interpretation of items from different perspectives.

5.2 SCENE DESCRIPTIONS AND A THEORY OF ACTION

I take it as evident that every scene description, indeed every utterance, is associated with some purpose or goal of a speaker. The speaker's purpose affects the organization and order of the speaker's presentation, the items included and the items omitted, as well as word choice and stress. Any two witnesses of the same event will in general give accounts of it that differ on every level, especially if one or both witnesses were participants or has some special interest in the cause or outcome of the event.

For now I have ignored all these factors of scene description understanding; I have not attempted an account of the deciphering of a speaker's goals or biases from a given scene description. I have instead considered only the propositional content of scene description utterances, in particular the issue of whether or not a given scene description could plausibly correspond to a real scene. Until we can give an account of the judgement of plausibility of description meanings, we cannot even say how we recognize blatant lies; from this perspective, understanding <u>why</u> someone might lie or mislead, i.e. understanding the intended effect of an utterance, is a secondary issue.

There seems to me to be a clear need for a "theory of human action", both for purposes of event simulation and, more importantly, to provide a better overall framework for AI research than we currently have. While no one to my knowledge still accepts as plausible the "big switch" theory of intelligent action [7], most AI work seems to proceed on the "big switch" assumptions that it is valid to study intelligent behavior in isolated domains, and that there is no compelling reason at this point to worry about whether (let alone <u>how</u>) the pieces developed in isolation will ultimately fit together.

5.3 ARE THERE MANY WAYS TO SKIN A CAT?

Spatial analog models are certainly not the only possible representation for scene descriptions, but they are convenient and natural in many ways. Among their advantages are: 1) computational adequacy for

representing the locations and motions of objects; 2) the ability to implicitly represent relationships between objects, and to allow easy derivation of these relationships; 3) ease of interaction with a vision system, and ultimately appropriateness for allowing a mobile entity to navigate and locate objects. The main problem with these representations is that scene descriptions are usually underspecified, so that there is a range of possible locations for each object. It thus becomes risky to trust implicit relationships between objects. Event stereotypes are probably important because they specify compactly all the important relationships between objects.

5.4 RELATED WORK

A number of papers related the the topics treated here have appeared in recent years. Many are listed in [8] which also provides some ideas on the generation of scene descriptions. This work has been pervasively influenced by the ideas of Bill Woods on "procedural semantics", especially as presented in [9]. Representations for large-scale space (paths, maps, etc.) were treated in Kuipers' thesis [10]. Novak [11] wrote a program that generated and used diagrams for understanding physics problems. Simmons [12] wrote programs that understood simple scene descriptions involving several known objects. Inferences about the causes and effects of actions and events have been considered by Schank and Abelson[13] and Rieger[14]. Johnson-Laird[15] has investigated problems in understanding scenes with spatial locative prepositions, as has Herskovits[16]. Recent work by Forbus[17] has developed a very interesting paradigm for qualitative reasoning in physics, built on work by deKleer[18,19], and related to work by Hayes[20,21]. My comments on pronoun resolution are in the same spirit as Hobbs[22], although Hobbs's "predicate interpretation" is quite different from my "analog spatial models". Ideas on the adaptation of prototypes for the representation of 3-D shape were explored in Waltz [23]. A effort toward qualitative mechanics is described in Bundy [24]. Also relevant is the work on mental imagery of Kosslyn & Shwartz[25] and Hinton[26].

I would like to acknowledge especially the helpful comments of Ken Forbus, and also the help I have received from Bill Woods, Candy Sidner, Jeff Gibbons, Rusty Bobrow, David Israel, and Brad Goodman.

6. REFERENCES

[1] Waltz, D.L. and Boggess, L.C. Visual Analog representations for natural language understanding. Proc. of IJCAI-79, Tokyo, Japan, Aug. 1979.

[2] Boggess, L.C. Computational interpretation of English spatial prepositions. Unpublished Ph.D. dissertation, Computer Science Dept., University of Illinois, Urbana, 1978.

[3] Chafe, W.L. The flow of thought and the flow of language. In T.Givon (ed.) Discourse and Syntax. Academic Press, New York, 1979.

[4] Bar-Hillel, Y. Language and Information. Addison-Wesley, New York, 1964.

[5] Piaget, J. Six Psychological Studies. Vintage Books, New York, 1967.

[6] Jackendoff, R. Toward an explanatory semantic representation. Linguistic Inquiry 7, 1, 89-150, 1975.

[7] Minsky, M. and Papert, S. Artificial Intelligence, Project MAC report, 1971.

[8] Waltz, D.L. Generating and understanding scene descriptions. In Joshi, Sag, and Webber (eds.) Elements of Discourse Understanding, Cambridege University Press, to appear. Also Working paper 24, Coordinated Science Lab, Univ. of Illinois, Urbana Feb. 1980.

[9] Woods, W.A. Procedural semantics as a theory of meaning In Joshi, Sag, and Webber (eds.) Elements of Discourse Understanding. Cambridge University Press, to appear.

[10] Kuipers, B.J. Representing knowledge of large-scale space. Tech. Rpt. AI-TR-418, MIT AI Lab, Cambridge, MA, 1977.

[11] Novak, G.S. Computer understanding of physics problems stated in natural language. Tech. Rpt. NL-30, Dept. of Computer Science, University of Texas, Austin, 1976.

[12] Simmons, R.F. The CLOWNS microworld. In Schank and Nash-Webber (eds.) Theoretical Issues in Natural Language Processing, ACL, Arlington, VA, 1975.

[13] Schank, R.C. and Abelson, R. Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[14] Rieger, C. The commonsense algorithm as a basis for computer models of human memory, inference, belief and contextual language comprehension. In Schank and Nash-Webber (eds.) Theoretical Issues in Natural Language Processing, ACL, Arlington, VA, 1975.

[15] Johnson-Laird, P.N. Mental models in cognitive science. Cognitive Science 4, 1, 71-115, Jan.-Mar. 1980.

[16] Herskovitz, A. On the spatial uses of prepositions. In this proceedings.

[17] Forbus, K.D. A study of qualitative and geometric knowledge in reasoning about motion. MS thesis, MIT AI Lab, Cambridge, MA, Feb. 1980.

[18] de Kleer, J. Multiple representations of knowledge in a mechanics problem-solver. Proc. 5th Intl. Joint Conf. on Artificial Intelligence, MIT, Cambridge, MA, 1977, 299-304.

[19] de Kleer, J. The origin and resolution of ambiguities in causal arguments. Proc. IJCAI-79, Tokyo, Japan, 1979, 197-203.

[20] Hayes, P.J. The naive physics manifesto. Unpublished paper, May 1978.

[21] Hayes, P.J. Naive physics 1: Ontology for liquids. Unpublished paper, Aug. 1978.

[22] Hobbs, J.R. Pronoun resolution. Research report, Dept. of Computer Sciences, City College, City University of New York, c.1976.

[23] Waltz, D.L. Relating images, concepts, and words. Proc. of the NSF Workshop on the Representation of 3-D Objects, University of Pennsylvania, Philadelphia, 1979. Also available as Working Paper 23, Coordinated Science Lab, University of Illinois, Urbana, Feb. 1980.

[24] Bundy, A. Will it reach the top? Prediction in the mechanics world. Artificial Intelligence 10, 2, April 1978.

[25] Kosslyn, S.M. & Shwartz, S.P. A simulation of visual imagery. Cognitive Science 1, 3, July 1977.

[26] Hinton, G. Some demonstrations of the effects of structural descriptions in mental imagery. Cognitive Science 3, 3, July-Sept. 1979.