BAD-MOUTHING FRAMES

Jerry Feldman
University of Rochester
Rochester NY

It appears that many people in the AI/psycholinguistics community are like my old friend (in California) who said: "How can I understand something unless I believe it for a while." This seems to me to indicate the role of "paradigms" such as "frames" in the study of thought. Since I do not, myself, work that way and also do not (despite years of the New York Review) function well as a critic of scientific developments, I will limit myself to three rather concrete sets of remarks. These concern vision, interactions with the world and net models in the context of "frames".

Much of the discussion of Minsky's frames paper is concerned with visual perception. There are a number of conclusions about visual perception and visual memory that result from the frames paradigm which appear to me to be wrong and even wrongheaded. The notion that we store a large number of separate views in purely symbolic form is one such. There is a good deal of evidence that people do use three-dimensional models (e.g. Shepard) and that they regularly integrate several views into a single visual model which seems to be the predominate one (e.g. stereopsis, motion parallax). There are other examples of this sort, some of which are clearly the scars of old battles against perceptions, analog mysticism and the like. None of these are beyond incorporation into a frames paradigm. That this is true results from and illuminates the main difficulty with the frames paradigm as a theory -- it seems to be extensible to include anything at all. Without going as far as Popper ( ); we can ask that a theory be at least conceptually refutable. We propose an anti-frames hypothesis below.

Another serious problem with a frames approach to vision lies in the strong assumption of default values for slots. The assumption is that, when viewing a new scene, we are basically just verifying our frame picture of it. Here a little intellectual history is called for. Early efforts in machine perception (and much perceptual psychology) were concerned with visual processes which operated independent of context. We studied edge detectors, pattern classifiers and algorithms for partitioning general straight-line drawings. This not only proved difficult but offered no promise of extension to typical real-world scenes. There then came a concerted effort to overcome (or circumvent) perception problems by giving programs lots of domain knowledge. This has been carried to the extreme of visual perception without vision, viz. anything black and on a desk is a telephone. The frames paper seems to approach vision in this way. Once again, there is substantial evidence (and overwhelming intuition) that this is not what occurs. For example, people can understand totally unexpected images presented for quite short periods. A recent study by Potter indicates that verbal descriptions (which could give rise to very many images) are almost as efficient as a preview for choosing a target in such a set.

> To return to the questions posed at the beginning, one does not need to know exactly what a thing will look like to detect it in a 1/3-second glimpse. In fact, knowing the exact appearance of a target was little better than knowing only its general meaning, which suggests tht a scene is processed rapidly to an abstract level of meaning before intentional selection occurs (Potter 1975).

Well, even Minsky doesn't claim the frames paper is right -- only that it should help us ask the right questions. In machine perception, it seems to be having a significant impact which I consider negative. Frames and the whole idea of vision as mostly problem solving seem to have subverted work on the understanding of images. It has always been tempting to avoid the hard detailed problems of real images for the higher realms of abstract problem solving. The frames paper has had the effect (it really has) of sanctioning this retreat.

My second comment is an extension of this last point -- the actual effects of the world on a "frames" process is much too weak. This has been a major concern of mine lately. Quoting from a recent paper:

> "The wheelless student problem models the real-life problem of buying a used automobile (in the United States). A typical procedure is first to read newspaper advertisements and bulletin boards to assess the situation generally. Then, at relatively low cost, one can telephone various purveyors of cars and inquire about them. At some point, one must actually go to the effort of seeing and driving certain of these. There are professional diagnostic services which can be employed (at considerable cost) to further test the car. In each of these steps, one must decide when to stop that stage and go on to the next one. One does not, of course, proceed in strict order; there will normally be alternatives at several different levels of investigation.
>
> Notice that the "plan" itself is trivial: read, telephone, look, drive, professionally test and buy. It is the application of this plan to the world situation which is difficult. We believe that much intelligent activity is characterized by complex

applications of simple plans and this belief has led to concentrate on the closely related questions of plan elaboration and execution."

The point to be made here is that the world does not hold still for us. Theories which propose a narrow channel between the mind and the world have considerable value, but we will have to move beyond them to achieve real explanatory or robotic power.

It is the relatively static nature of the frames paradigm which makes me most uncomfortable. I consider the main substantive claim of the frames paradigm to be that knowledge is stored in relatively coherent chunks which change only slowly over time. The opposing hypothesis is the older "net" model which says that the collection of knowledge being brought to bear at a given instant is a rapidly changing function of the situation. Once again, we can blur the distinction, but then I don't know what the frames paradigm is.

Let us consider this question of static versus dynamic clumping of knowledge. We can assume, in both cases, that knowledge can include procedures, contexts and any other clever things that we dream up (Bobrow and Winograd have dreamed up dozens). The question is whether the high bandwidth connections among knowledge primitives can be fairly static. There are important reasons for hoping the frames hypothesis is true for animals or could be employed by machines. The general advantages of partitioning complexity are well known (Simon, Alexander). For any current or projected computers there is a memory hierarchy which strongly favors coherent chunks of knowledge relatively loosely coupled with other chunks. Unfortunately, I see great difficulty with static frames of anywhere near the scale discussed by Minsky.

Try to introspect as you slowly read the following sentence:

"Imagine yourself walking into a room; it is the master bedroom of a quiet Victorian house, in a slum of Bombay, which has just had a fire and been rebuilt in modern style, except for the master bedroom which is only half remodeled having its decorative panelling intact but badly visible because of the thick smoke."

The sentence above causes several shifts and refinements of the image. The question is, of course, where are the frames. It is possible that there are a very large number of room frames embodying all the combinatorial possibilities hinted at above. Alternatively, there could be a single room frame that incorporated all these possibilities. Neither of these alternatives strikes me as plausible. What seems to happen is that we build our model dynamically as we process the sentence. The anti-frame hypothesis here is that the connections which are most important (heavily used, etc.) are not specifiable in advance.

Let us return to the problem of buying a used car. I believe that what happens is roughly as follows. We construct a goal-oriented subsystem making use of our knowledge of cars, buying, our locale, friends who know about cars, etc. This subsystem seems to have great internal coherence and rely only slightly on our total world knowledge about e.g. buying. The used-car-buying subsystem probably gets drastically changed each time it is reinvoked with only a few general principles carried over from its last incarnation. While it is active, the subsystem effects much of our perception of the real world and of our thinking. We notice different things, seek different people, make different mental associations, etc. My mind boggles at trying to model a system of such flexibility, but the frames hypothesis (as I understand it) seems more likely to lead me astray than to help.

REFERENCES

Feldman, J. and Sproull, R., "The Hungry Monkey", Technical Report 3, University of Rochester, 1975.

Minsky, M., "A Framework for Representing Knowledge", MIT AI Lab Memo no. 306, 1974.

Popper, K., The Logic of Scientific Discovery. London: Hutchinson, 1959.

Potter, K., "Meaning in Visual Search", Science, 187, p.965, 1975.

Shepard, R., and Meltzer, J., "Mental Rotation of 3d Objects", Science, 171, p.701, 1971.