DOES A STORY UNDERSTANDER NEED A POINT OF VIEW?

Robert P. Abelson
Yale University

At the Carbonell Memorial Conference in 974, there was a good deal of informal iscussion of the use by people of analogue imulations in knowledge retrieval or uestion-answering. We asked each other uestions like, "How many traffic lights are here along your usual route from the ailroad station to your house?" Or, "Can a alt shaker be used as a stool?". The ormer type of question usually gives rise o introspective reports of a mental imulation of the traversal of the requested oute, replete with visual imagery. The atter type of question may or may not give ise to a mental simulation. Some people eport knowing propositionally that a salt haker cannot be used as a stool because its ize is insufficient. Others report entally playing through the motor sequence f sitting on a salt shaker, whence they idely discover the negative answer.

People with different cognitive styles an become quite exercised over whether the ropositional or simulational account is the orrect" psychological description of this ype of question-answering. People who xperience difficulty constructing visual or otor images are prone to be strong ropositionalists (e.g., Pylyshyn, 1973). I ad a recent argument on this issue with a ell-known psychologist not given to visual magery. "Can a salt shaker be a stool?", I sked her. "No," she said immediately, Obviously not. It's the wrong size. I new that right away because the features of salt shaker don't match the criticial eatures of a stool. I didn't need any isual imagery".

I then found a subtler example to test er: "Can a shoe be a hammer?". She esitated. Absent-mindedly making a epeated hammering motion up and down with er hand, as though grasping a hypothetical hoe, she answered, "Well, yes." I pounced t her gesture. "Aha! Why were you moving our hand like that?" "Oh, I often gesture ike this," she replied, switching slyly to side-to-side motion. "No, not like <u>that</u> side-to-side) I said, "like <u>this</u>!" (up and own). She reluctantly conceded the point. core one against the propositionalists.

Of course the idea that a knowledge ystem can know by doing is familiar in rtificial intelligence under the rubric of procedural knowledge" (Winograd, 1972; umelhart, Lindsay, & Norman, 1972). It is lso basic tenet of Piaget's psychology of nowledge. Nevertheless, the procedures nvolved in particular mental simulations re not well understood. While in principle t may be possible to separate the concept f simulation from the discussion of on-linguistic codes or images, in practice he two areas seem intertwined. In this aper I will discuss the mental simulation of spatial traversals, as provoked by stories of individuals going from here to there and encountering various events and objects along the way.

<u>The Simulation of Traversals</u>

Our assumption is that when an individual hears a story about a traversal through a spatial territory, he will tend to construct a simulation of this traversal, employing some mixture of linguistic and spatial elements in this construction. This assumption has considerable intuitive appeal, although it may not appeal to the intuitions of everybody, and indeed, need not be true of everybody.

There are at least two interesting consequences of this assumption, both amenable to experimentation. One is that the simulation process will recruit imagined acts and objects not present in the story but necessary for carrying out the simulation. A simple example arises in the following story fragment: "He stood watching the house from outside the white picket fence. Finally, he opened the gate and went into the yard..." Most listeners to these lines report the spontaneous invention of some particular kind of latch used to open the gate. It is either "seen" or represented by a motor sequence, or both. Another type of example is the following. "As Jack approached the intersection, he could make out the sign reading 'Broad Street'. Turning the corner, he quickly spotted the drug store he had been looking for". To the subsequent question whether Jack turned left or right at the corner, most people report having a definite opinion that one or the other was correct, or at any rate was the way they saw it.

It would be interesting to demonstrate and delimit a strong version of this phenomenon, wherein subjects would feel certain that particular details had actually occurred in the traversal story even when they hadn't. Some cognitive psychologists have documented a milder confusion between presented and remembered information, in which conclusions strongly implied by a text are stored as though they were explicit in the text (Bransford, Barclay, & Franks, 1972; Kintsch, 1974). It has also recently been conjectured by Schank and Abelson (1975) for textual contexts replete with cliche -- "situational scripts" such as eating in a restaurant -- that listeners will insert missing obvious details without later realzing that they have done so. The similar pheonomenon conjectured here would be even more striking because the inserted details would be essentially gratuitous: the stories do not imply any particular type of gate latch or direction of turn, etc.

I would like to be able to report to you that we had indeed done an experiment successfully establishing this phenomenon. Then we could debate whether it was something special to simulaton and non-linguistically coded materials (as I would hope it would be), or whether it had

to do more generally with what Minsky (1974) would call "filling in default values" in knowledge frames. Unfortunately I am not able to report this. We tried one experiment last summer but were unable to introduce a long enough time delay to permit subjects to lose the short-term surface cues permitting the discrimination whether a given detail was actually in the story or not. We are going to try again soon.

What I can report to you is an experiment on a second phenomenon. If in theory we accept simulation as a reasonable process by which a traversal can be understood, we must face the ambiguity that there are different vantage points from which a given simulation may be conducted. The most obvious vantage difference is between a simulation of the individual himself performing the traversal, and a simulation of someone watching that individual. This distinction, as we shall see, can have important consequences for memory of the story.

Points of View of the Self and the Observer: An Experiment

A listener simulating a traversal from the point of view of the story character will presumably generate motor images as well as visual images such as might appear to the actor. On the other hand, a simulator from an observer's point of view would presumably not be disposed toward motor or other body sensation images, and his visual images would have a different perspective from that of the actor. For example, very large objects might be in focus for a distant observer, but not well perceived by an actor too close to them. These considerations served as the basis for an experiment designed and run in collaboration with Richard Pinto.

A 68-sentence story about a character leaving a hotel and strolling a block down the street was read in common to subjects given three different instructions. All subjects were told to close their eyes and "imagine, as best and as vividly as you can, along with what you hear". Each sentence was to be rated on a scale of vividness. (this instruction was intended to disguise the memory nature of the study.) Additional instructions were given to constitute two different vantage point conditions. In the Self group, subjects were told to imagine themselves being the main character in the story. In the Balcony group, subjects were told to imagine themselves watching from a fourth-floor hotel balcony. A No Vantage Point group got no specific vantage point instruction.

There were three crucial kinds of details in the story: "far visual" details, items which in realistic perception can best be seen from far away (such as a sign over a bank); "near visual" details, natural to view from close range (such as a wristwatch); and "body sensation" details (such as aggravating a sore arm, or drinking hot coffee). We hypothesized that when faced with a long, meandering story, the listener in the Self condition would tend preferentially to absorb "body sensation" details, and the listener in the Balcony condition would tend preferentially to absorb "far visual" details. The matter is not so simple for "near visual" details. In a pilot study, we found that Balcony subjects sometimes reported "floating down" off the balcony from time to time, as it were, to peer vicariously over the actor's shoulder at details otherwise too small to imagine seeing clearly from a distance. (Such flexible perspective is reasonable in the light of Kosslyn's (1974) experiments on visual image size, showing that it is disadvantageous to retrieve detailed information from small visual images.)

Table 1 gives the mean proportions of the three different types of details correctly recalled by subjects in each experimental condition, on a 21-item cued recall test administered approximately twenty minutes after hearing the story. Our predictions were clearly supported. The Balcony group averaged 17.6% better recall of far visual details than the Self group, and the Self group averaged 15.0% better recall of body sensational details than the Balcony group. There were no differences for near visual details. For each type of detail, the No Vantage Point group did about as badly as whichever of the other two groups had the "wrong" vantage point.

One other dependent measure in the study produced interesting results. Preceding the cued recall test was a free recall task in which all subjects were asked to try to recreate the story as best they could, line for line. Each recalled line was scored for whether it captured the gist of a corresponding original story line. Subjects often left out whole scenes, and reversed the order or events, although they did not often invent details which were not there. The pattern with the Balcony group relatively more correct than the Self group on far visual details as opposed to body sensation details was present in these free recall data, although not as strongly as in cued recall, and not statistically significant. A very significant difference was present, however, in proportion correct on all types of items (including "fillers") for the No Vantage Point group. The total proportions correct were .344 for No Vantage Point, .560 for Self, and .571 for Balcony. The corresponding proportions over critical items on cued recall (Table 1) were .508 for No Vantage Point, .572 for Self, and .572 for Balcony. In other words, the No Vantage Point group exhibits an acute disadvantage in free recall which it does not suffer in cued recall. Some of the story material which is available in memory presents an access problem for the No Vantage Point group.

One further result is worthy of mention. All subjects were given Betts' Test of vividness of visual imagery. The superiority of the Balcony over the Self group on far visual details was sharply enhanced in subsamples scoring high on

general visual imagery (a 32.6% superiority, vs. a mere 2.4% superiority among non-imager subsamples). This supports the reasonable supposition that processing style depends upon the proclivities of the individuals as well as the task orientation given whole groups of subjects.

The overall pattern of these results is I think very difficult to explain from a propositionalist point of view. All subjects heard exactly the same story. They were all told to imagine along with what they heard, so that one cannot argue that some subjects were oriented toward linguistic and others toward non-linguistic codes. All that was different between subject groups was the vantage point from which imagination was to be exercised. The results, it seems to me, support not only the existence of non-linguistic codes, but even more theory-bogglingly, the existence of different forms of non-linguistic codes which depend on the point of view of the listener. Having made this strong statement, I hasten to add the disclaimer that I intend the word "existence" in a very weak sense. To pursue whether non-linguistic codes are fundamentally different mental entities from linguistic codes is to walk into a hopeless metaphysical quagmire. Minsky (1974) has compellingly argued that visual scenes can in principle be described by frames which are essentially conceptual networks. Nevertheless it may be very useful to distinguish vision-based concepts from other concepts, because there is a specialized character to processes which operate on them, such as mental rotation, image magnification, etc. This heuristic argument has recently been put forward by Kosslyn and Pomerantz (1975). In other words, I am arguing that different specialized processing modes are keyed to different vantage points, and have different consequences for what is best remembered. Furthermore, the No Vantage Point group seems to suffer in free recall from the lack of a special processing mode. Perhaps the vantage point provides a set of higher-order nodes in the network representing the story, facilitating access to the lower-order, story details.

What are the Implications for Artificial Intelligence?

On the surface, this experiment may seem barren of implications for artificial intelligence, because smart programs might ideally be able to reduce all sentences in a story to an interconnected meaning representation for which there is no memory loss. Looking deeper, there are two responses to such a complacent view. First of all, even if there were no memory loss, there is still the problem of the non-uniqueness of the meaning representation. From a different point of view, the meaning of the same story can be different. In AI parlance, one might say that different frames are invoked, or that different inferences (or different numbers of inferences) are attached to the story network. The extent to which this psychological truism will prove important in AI applications remains to be seen, but I believe that it is short-sighted to overlook this non-uniqueness problem.

A more radical view is that programs should not be designed to preserve all the details of understood texts, that they should in fact throw away "less interesting" information -- or at least bury it in remote storage so that it doesn't clutter the working memory whenever the story is referenced. From this perspective, the constraints of human memory are seen as an advantage to intelligence, rather than as a deficiency. People's skill in forgetting things, particularly over the long run, might provide a good model for AI systems to strive for.

The trick lies in knowing what to forget, and that is where the vantage point might come in. From a vantage point on a balcony, the kinaesthetic and tactile sensations of the main character are nonessential, and it is natural to spend less effort processing them. (Perhaps there is also a processing cost in switching processing modes.) From the point of view of the main character, items best "understood" by encoding them from long visual perspective are the least natural and most effortful.

What we are saying here is perhaps nothing more nor less than the conventional wisdom that the understanding process ought not to be sensitive to the understander's style and purpose. Yet vantage point is a more specific variable than the vaguer concept of purpose, and more amenable to experimental manipulation. We hope in future experiments to manipulate vantage point in more subtle ways, such as by telling stories with several characters, inducing subjects to identify emotionally with one or another character because of similarity to self. Of course the different characters will have different spatial perspectives in the story, and therefore will experience different body and visual sensations.

REFERENCES

Bransford, J.D., Barclay, J.R., and Franks, J.J., Sentence memory: A constructive vs. interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Kintsch, W., The Representation of Meaning in Memory. Hillsdale, NJ: Erlbaum Associates, 1974.

Kosslyn, S., Effects of imagined object size on response time in mental imagery tasks. Unpublished Ph.D. dissertation, Stanford University, 1974.

Kosslyn, S., and Pomerantz, J., Mental imagery reconsidered: An analysis of Pylyshyn's critique. Mimeographed. Johns Hopkins University, 1975.

Minsky, M., A framework for representing knowledge. Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo No. 306, 1974.

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

Rumelhart, D., Lindsay, P., and Norman, D., A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972.

Schank, R.C., and Abelson, R.P., Scripts, plans, and knowledge. Presented at the 4th International Joint Conference on Artificial Intelligence, Tbilisi, August, 1975.

Winograd, T., Understanding Natural Language. New York: Academic Press, 1972.

Table 1

Mean proportions of correct recalls of story details

	Type of Detail			
	Far visual	Near visual	Body Sensation	(Overall)
Self	.417	.616	.660	(.572)
Balcony	.593	.613	.510	(.572)
No vantage point	.476	.572	.476	(.508)
(p-value, Self vs. Balcony)	(<.05)	(ns)	(<.05)	(ns)