

USE OF THEMATIC INFORMATION TO SPEED SEARCH OF SEMANTIC NETS

Lynne M. Reder
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

John R. Anderson
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

ABSTRACT

A model is proposed for the representation of thematically integratable information in semantic memory and a process for searching through this memory. We assume that related facts about an individual are not directly attached to the individual node, but rather, are indirectly attached via a thematic sub-node. Thematic sub-nodes allow much faster search of memory under certain conditions, but not others. We tested these predictions with human subjects and found support for these notions.

There is little doubt that the time to search a computer data base depends in part on the number of facts stored in it. Similarly, the time for humans to retrieve a specific fact from memory has been shown to depend on the number of other, related facts. A large number of reaction time studies (e.g., Anderson, 1974; Anderson & Bower, 1973; Hayes-Roth, 1977) have found that subjects are slower to recognize a fact the greater the number of facts learned to the same concept. For example, subjects are slower to recognize the fact "The psychologist wanted to go to Japan" if they have also studied "The psychologist thought the research was good" and "Everyone loves Japan" than if no other experimental facts share concepts with the probe. This particular result has been named the *fan effect* because latencies are found to be a function of the number of propositions *fanning* out of a given concept node in a propositional network representation (see Anderson, 1976). (Subjects are also slower to reject unstudied probes as a function of the fan of their concepts.)

There is a paradoxical aspect to the result that the more facts a person knows about a concept the slower he or she is to verify any one of them. It seems intuitive that experts (a) know lots of facts about certain topics, and (b) can answer more questions on those topics and answer them faster than anyone else. However, certain situations can be pointed out where experts seem to experience a type of interference that results from too much similar information. Learning a new text editor in many ways is much easier if one has used a similar editor

before. However, the more editors a person uses, the harder it is to keep the commands straight.

Even though experts may sometimes suffer more interference than non-experts, it also seems that if their information can be *integrated* into various themes, then somehow the information should be less interfering. Consider the following knowledge representation for facts about an individual and process model for searching this knowledge: all facts studied about the individual are attached to that individual's concept node. However, propositions that can be considered to be thematically related are not attached directly to the person node; rather, they are attached to a theme node that is attached to the person node. Figure 1 gives a representation for the following four sentences: *Melvin crossed his skis. Melvin shussed down the slope. Melvin washed his car. Melvin got on the chair lift.* Facts that can be considered as having to do with Melvin-skiing form a sub-node attached to Melvin. Facts that are not so related are attached to Melvin directly. The *fan* out of the Melvin concept, then, is two rather than four.

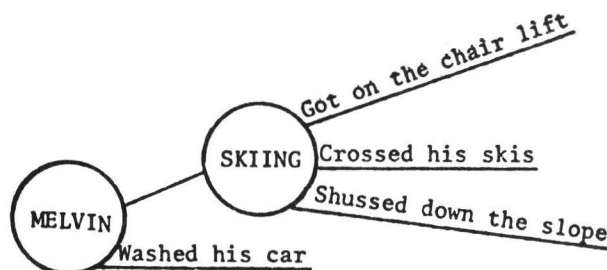


Figure 1: Portion of the Semantic Network encoding facts about Melvin.

This research was supported in part by NSF grants BNS76-00959 and BNS78-17463 to John R. Anderson and by an NIMH post-doctoral fellowship to Lynne M. Reder. We would like to thank T. Iwasawa for assistance in programming and in running subjects.

The search mechanism works as follows: The probe sentence is parsed. Each concept of the probe is activated in memory and activation spreads out along each arc attached to the concept. *Search energy* is assumed to be a constant for each activated concept. How much activation or search energy a particular arc receives is a function of the total number of arcs emanating from the concept (the amount of fan).

In earlier experiments on the fan effect, the facts associated to a person were unrelated. In those situations, the search mechanism predicts a simple effect of number of facts on time to recognize a probe. However, our predictions are more complex for the representation in Figure 1. We would predict that the time to verify *Melvin washed his car* would be the same as if we had learned only two unrelated facts about Melvin. Our predictions about *Melvin crossed his skis* are somewhat more complex.

The fan out of the Melvin concept is two. The fan out of the Melvin-skiing sub-node is three. Whether there is an effect of sub-node fan when attempting to recognize one of several thematically related facts depends on the nature of the foils or false probes. If, for example, the probe to be rejected is *Melvin baked a cake*, search need not continue past the Melvin-skiing sub-node: it is clear to the subject that cake-baking is not part of the skiing theme. If, on the other hand, the probe to be rejected is *Melvin perfected his stem-christies*, search must proceed beyond the sub-node to inspect each proposition attached to it until the correct one is found or all stored propositions are exhausted.

EXPERIMENT

To verify this model* we conducted an experiment where subjects learned facts about various characters. These facts belonged to various themes, e.g., buying a car, jogging, going to a circus. A particular character could have one or two themes associated with him or her and one or three facts associated with a given theme. For example a character with six associated facts would have three facts about each of two themes. A character with four facts would have one fact from one theme and three facts from the other. Each condition in the experiment can be described by two digits, representing the number of facts in each theme associated with a given character. The conditions are (1,0),(1,1),(3,0),(3,1),(3,3). The 0's denote conditions where there was only one theme.

Subjects studied these character-predicate pairings until they knew the facts perfectly (could recall all the facts about each character several times in succession). After this phase was completed, subjects went into the critical phase in which reaction times were collected to answer questions about the material studied. One sentence was displayed on the screen at a time. The subject was told to press the yes or no response key depending upon whether the subject thought it had been studied during the acquisition phase. Test probes were presented in random order. The foils (sentences subjects had not studied) were of two types and were separated into different test blocks. One type was the *thematically related foil* (e.g., *Melvin enjoyed deep powder*), and the other type was the *unrelated foil* (e.g., *Melvin baked a cake*). Both types of foils have the property that their predicates (verb-phrase) were studied with another

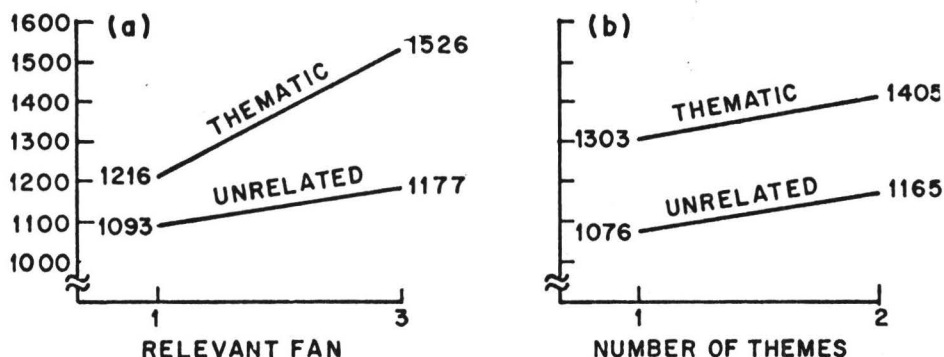


Figure 2: Mean reaction time to judge that a probe had been studied as a function of number of thematically related facts (graph a) and as a function of number of themes (graph b).

*We actually tested two alternative models in this experiment, but space constraints require this simplification.

character during the experiment. This ensured that subjects could not reject a probe simply on the basis of lexical familiarity.

We had two principal predictions for this experiment. First, when the foils are unrelated, there should be a much smaller effect of number of facts studied. This is because subjects should be able to make their judgments just by inspecting the theme subnodes.** They would not have to search the facts attached to these nodes. Our second prediction was that there should be an effect of themes (more themes means more arcs out of the person node), but that it should be small since subjects can quickly reject the wrong theme sub-node, and not be affected by type of foils since the extra sub-node is rejected in the same fashion regardless of foil-type.

RESULTS

Figure 2 displays the data (latencies in milliseconds to decide whether the probe had been studied or not) partitioned two ways: on the left, the data are analyzed as a function of the amount of fan relevant to the probe that is attached to the character in question. For example, the relevant fan for a Melvin skiing probe would be three because the subject studied three Melvin skiing facts. If the probe had concerned Melvin and his car, the relevant fan would have been one. On the right, the data are displayed as a function the number of themes associated with the character in the test probe, which can be one or two. Melvin has two themes, skiing and car washing. (There is no sub-node for car-washing because only one such fact was studied.)

Each graph plots two functions, one labelled unrelated and one labelled thematic. As explained above, the unrelated-thematic factor reflects the type of foils from which the subject had to discriminate the studied probes. The left panel plots the data as a function of the number of facts involving the same theme as the test probe. Subjects take longer to verify that a fact was studied the more other related facts were also studied. However, this fan effect is severely attenuated when the foils are not thematically related. Moreover, as predicted, subjects respond much faster in general when the foils are unrelated. We take this to mean that subjects have a less stringent criterion of saying "yes, this sentence is old" in these blocks. Subjects only search as far as the sub-node before responding when foils are not related to the sub-node. Smith, Adams, and Schnorr (1978) have also found that the fan effect is reduced with unrelated foils. (However, they did not have the other conditions of this experiment viz., testing with thematic foils, and manipulating the number of themes associated with a character.)

**If there is only one theme fact, we assume subjects do not bother to form a subnode. Inspecting one proposition is probably equivalent to inspecting a theme sub-node.

The data on the right shows the effect of number of themes. Here too there is an effect of fan for themes, however this effect is smaller and is not affected by the related/unrelated foil block distinction. The lack of interaction of theme-fan with foil-block type was predicted. Regardless of type of foil, the subject must at least find the relevant sub-node before responding "old". If there is only one theme associated with a character, the fan off the character node will be less. Therefore, search should be faster than with two sub-nodes or a fan of two off the character node.

CONCLUSION

The data support the view put forth in the introduction that facts are not necessarily attached directly to the concepts in their propositions, but rather can be organized into thematically consistent subnodes. It seems to us that, in general, sub-nodes in memory serve several functions. First, sub-nodes make plausibility judgments easy. Typically people are not asked to discriminate thematic foils from related statements; normally one would accept as plausible other related skiing facts about Melvin (see Reder, 1979, for a fuller discussion). Second, by having many sub-nodes in memory, the number of propositions that must be inspected is smaller. If a subject is searching for a fact about Melvin baking a cake, she or he need not search all of the skiing facts to find it--search stops on a wrong path as soon as the sub-node is reached.

References

- Anderson, J.R. *Language, Memory, and Thought*. Hillsdale, N.J.: Lawrence Erlbaum Associates 1976.
- Anderson, J.R. Retrieval of propositional information from long-term memory, *Cognitive Psychology*, 1974, 5, 451-474.
- Anderson, J.R. and Bower, G.H. *Human Associative Memory*. Washington: Winston and Sons, 1973.
- Hayes-Roth, B. Evolution of cognitive structures and processes. *Psychological Review*, 1977, 84, 260-278.
- Reder, L.M. The role of elaborations in memory for prose. *Cognitive Psychology*, 1979, 11, 221-234.
- Smith, E.E., Adams, N., and Schorr, D. Fact retrieval and the paradox of interference. *Cognitive Psychology*, 1978, 10, 438-464.