# The Processing of Referring Expressions within a Semantic Network

John R. Anderson
Yale University

Frege (1892) is credited with emphasizing the distinction between sense and reference. His famous example involved the morning star and the evening star. Despite the fact that they both refer to the same object (i.e., Venus), they have different senses as witnessed by the fact that sentence (1) is not synonymous with sentence (2):

(1) The morning star is the morning star.

(2) The morning star is the evening star.

This philosophical issue has similarities to an issue that is of importance to understanding natural language processing: How do subjects process referring expressions to extract internal representations (a) of their meaning and (b) of their referents in the external world. The example sentence that we will be returning to in this paper is:

(3) The first president of the United States was a bad husband.

It is clear that in understanding this sentence we both process the subject as a description, and identify this as referring to George Washington. This paper will try to explain how this comes about. As I believe that all interesting questions about representation come down to questions about memory, I will approach this question from a human memory perspective.

## Some "self-evident" truths about human memory.

To set up a framework for further discussions, I would like to list some of the facts that I think we know about human memory -- either because of a sophisticated common sense and self-observation or because of a mass of experimental data:

(1) Human memory can be conceived of as a network of associations among concepts.

(2) Some nodes in this network refer to individuals in the external world.

(3) Once information is deposited in memory it cannot be erased.

While there are a number of memory theories that embody these assumptions, I will be using the ACT model (Anderson, 1976) to present the theory and discuss the data in this paper. With this brief statement of the pre-theoretical biases, I would like to turn to an experimental paradigm which captures, in expanded time scale, the processes that I think are going on when we comprehend referring expressions.

## A Mock-up of the Morning Star-Evening Star Example

One of the experiments in this series (see Anderson, 1977; Anderson & Hastie, 1974 for a thorough report) had subjects study a set of facts such as (4) - (8):

(4) The smart Russian is the tall lawyer.

(5) The smart Russian cursed the salesgirl.

(6) The smart Russian rescued the kitten.

(7) The tall lawyer adopted the child.

(8) The tall lawyer caused the accident.

The critical manipulation was whether the identification sentence (4) was learned some time before or some time after sentences (5)-(8). For the identification before condition, part (a) of Figure 1 illustrates, very schematically, the network structure we thought was created. There is a node X set up to represent the individual and attached to that node are the various facts learned about this person. Part (b) of Figure 1 illustrates the network situation in the identification after condition. Because the subject did not learn of the identity between the two individuals until after learning sentences (5)-(8), he was led to create two nodes in memory which turn out to refer to the same individual. It would seem optimal if he could merge nodes X and Y together but this would amount to erasing memory structures, violating principle 3. Rather we assume that the subject encodes a separate proposition to the effect that the two individuals are identical. This is represented in Figure 1b, by the link between X and Y labelled with an '='

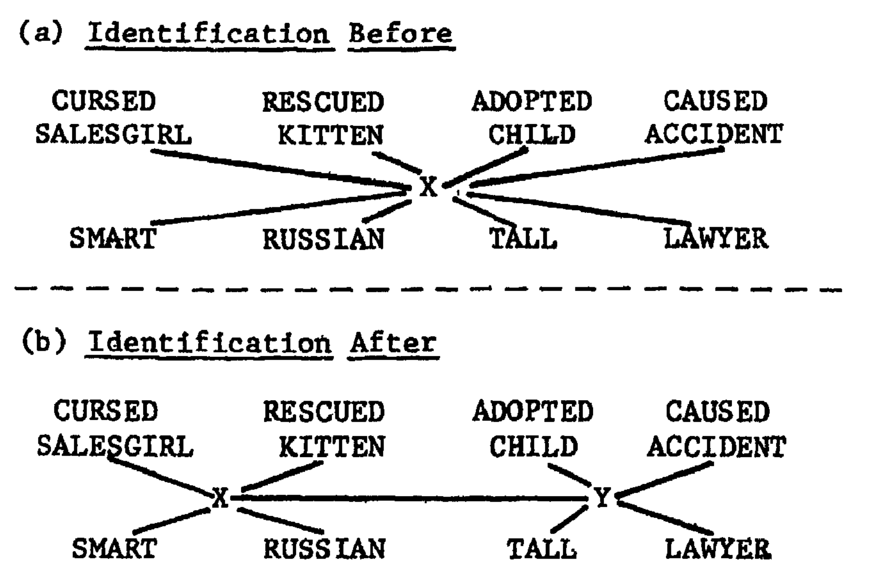



Figure 1. Memory representations at the beginning of the reaction time verification phase.

The memory representations in Parts (a) vs. (b) make different predictions about time to verify statements (9) vs. (10):

(9) The smart Russian cursed the salesgirl.

(10) The smart Russian caused the accident.

Statement (9) is referred to as a direct statement because it is identical to a study statement, while statement (10) is referred to as an inference as it can be inferred from statements (4) and (8).

Table 1 displays subjects' reaction times to verify direct statements and inferences in the identification before and identification after condition. We would expect subjects to show very little advantage for direct statement over inference in a representation like Figure 1a since there is no special connection preserved between the predicates and the referring expressions they

Table 1

Reaction Times (in msec) to verify Statements like 9 and 10

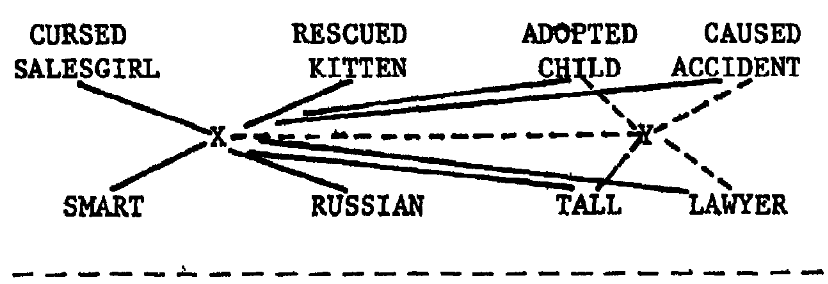
| | Identification Provided | |
|---|---|---|
| | Before | After |
| Direct Statement | 2310 | 1978 |
| Inference | 2388 | 2634 |

were studied with. In fact the verification times are almost identical in the two conditions. In contrast the after condition in Part b of Figure 1 each referring expression is only directly connected to the predicates it was studied with. To verify an inference requires an extra step of activating the path encoding the equality of X and Y. Correspondingly, we find an advantage for direct statements over inference. Finally, note that there are many more links attached to node X in part (a) than to either X or Y in part (b). This means there are more irrelevant paths that can interfere with finding the desired connection. Correspondingly, we find subjects faster to direct statement trues in the after condition.

The data reported in Table 1 come from the first block of reaction time test trials. There were four such blocks of trials. The reaction time data for all four blocks are displayed in Figure 2. Besides illustrating a general speed-up over the course of the experiment, the figure illustrates the differences among the conditions gradually disappear over the course of the experiment. Specifically, the differences between inference and direct statements in the after condition disappears and the differences between identification before and identification after condidisappear.

To account for this across-block trend we propose that the subject begins a process of copying the predicates from one of the nodes in Figure 1b to the other node. That is, one node is chosen to be abandoned and the other to receive all information. Therefore, supposing the subject choses to copy from node Y to X, everytime he encounters a fact attached to X he will attempt to copy it to Y. Figure 3 illustrates our belief about the memory representation by the end of the experiment. Note that the node X has been attach-

Figure 3: Memory representation in the identification after condition after much practice at verfying inference questions.




ed to all the facts learned of Y. Also the connections involving Y are dotted to indicate that they have become weak through disuse. The after

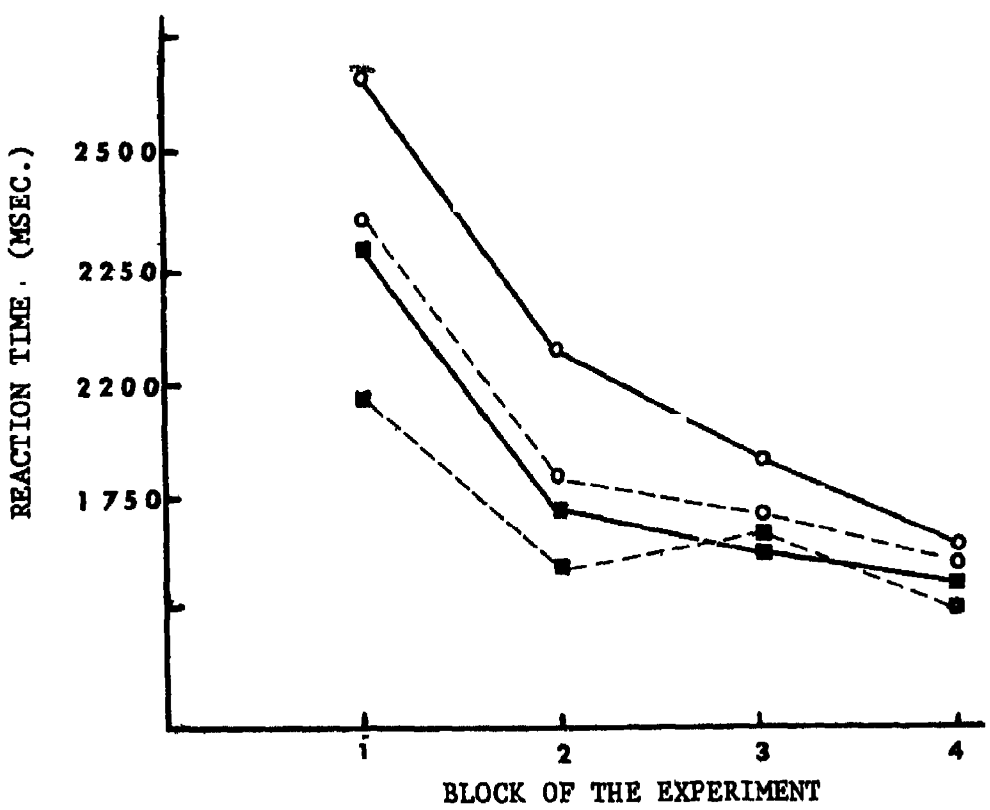



Figure 2: Verification times for various kinds of probes as a function of practice.

representation in Figure 3 has become functionally almost equivalent to the before representation in Figure 1a. Thus there is little difference between inference and direct statement or between the after and before condition.

One might wonder why the subject did not perform this copying when he learned about the identity between the two referring expressions rather than later in the verification phase of the experiment. In the ACT memory model such copying operations cannot be performed unless the data to be copied is active in working memory. At the time of studying the identification statement (4) the predicates needed for copying would not be active in memory. It is only when inferential statements like (10) are encountered in the test that the copying can take place. The referring expression could be copied while learning the identification statement. So the expression tall lawyer might be immediately attached to X. Thus, Figure 1b might be an oversimplification of the state of memory in the identification after condition. But in any case, the inference effect will not go away until the predicates are copied and this will not occur until the reaction time test phase.

Why should we believe this copying explanation rather than any of the multitude of alternative mechanisms that might be offered the explain the data in Figure 2. First, it satisfies the constraint that the subject not be able to erase information from memory and many of the mechanisms would not be. Second, unlike many of the other mechanisms, it assumes an asymetry in the fate of the two individual nodes in Figure 1b. One node is fated to receive all the information and the other node is to be abandoned. It seems reasonable that a subject would choose to preserve that node which had the more information attached and/or had this information attached more strongly. We have been able to demonstrate that subjects do abandon the "weaker" node.

The evidence for this asymetry comes from experiments that use a proper name rather than one of the definite descriptions. That is, the material is the same as in the example except that wherever tall lawyer appears a proper name like James Bartlett would be used. There is evidence (Anderson, 1977) that subjects learn material less well involving the proper name than the definite description. Correspondingly, we would expect subjects to choose to abandon the proper name node and maintain the definite description node. Evidence for this comes from the following analysis: We would propose that, in the initial drilling on the sentence James Bartlett is the Russian, in the identification after condition subjects copy the James Bartlett name to the Russian node. Figure 4 illustrates the memory representation with this asymetry. Note that,

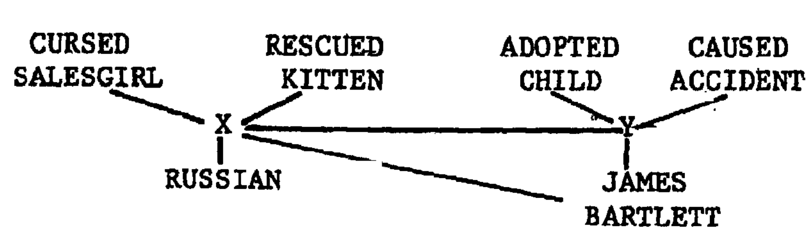



Fig. 4 : Verification times in the identification before and identification after conditions in an experiment that used both proper names and definite descriptions as referring expressions.

according to this representation, subjects should be as fast when verifying an inference predicate of James Bartlett as a direct statement predicate. This is because the proper name is directly attached to both. In contrast, subjects should be much slower for an inference predicate to a definite description because those predicates have not yet directly been attached to node X to which the description is attached. To verify these questions involves the extra retrieval of the proposition that node X equals node Y. Figure 5 presents the data from one of the experiments (Anderson & Hastie, 1974) contrasting definite descriptions and proper names. As predicted there is a large inference effect only for definite descriptions in the after condition.

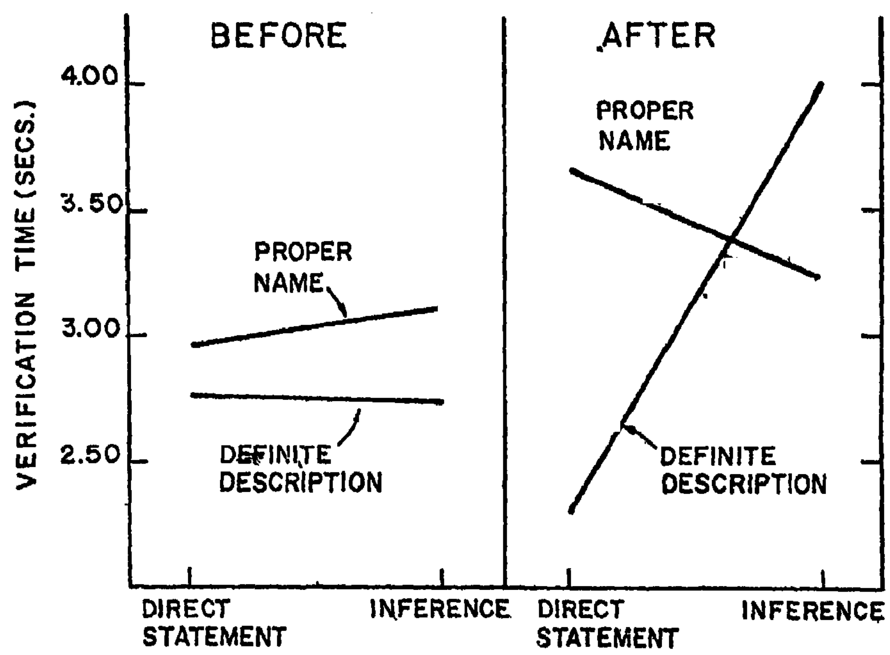



Figure 5.

## Application to Recognition of Referring Expression

The advantage of the paradigm just reviewed is that the sequence of states of memory is sufficiently spread out over time that it is possible to map out the changes in memory. I will be proposing that there is a similar sequence of memory states when subjects process referring expressions as in (3):

(3) The first president of the United States was a bad husband.

However, the processing happens so rapidly it is not as easy to verify each state in the sequence.

Figure 6 illustrates two possible sequences of information processing. Part (a) illustrates the state of memory right after comprehension of

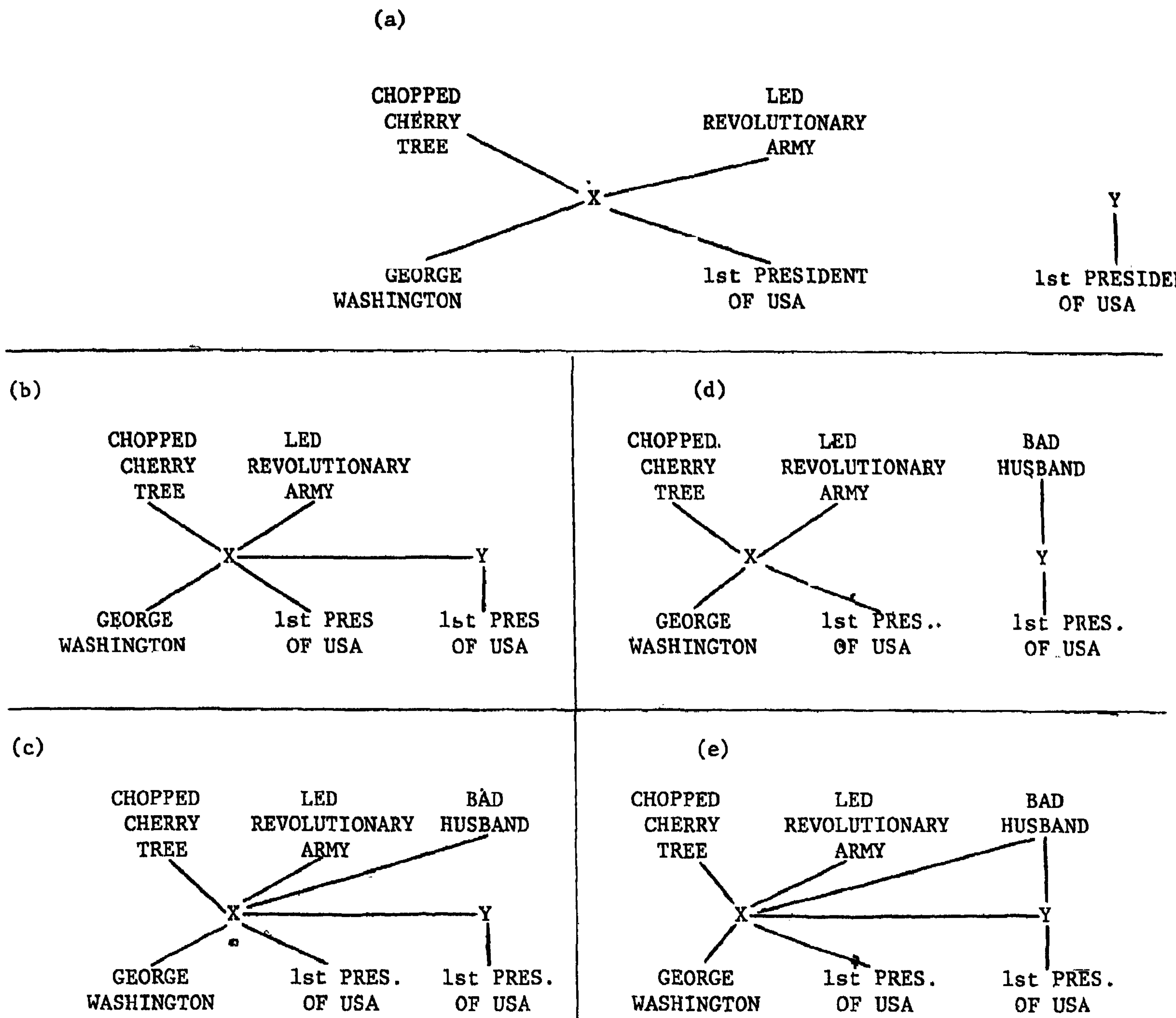



Figure 6. Possible states of memory representation during the processing of sentence (3).

the definite description. A node Y has been created to which there has been attached the "first president of USA" description. A separate node, X, in memory encodes permanent information about George Washington. Part (a) of Figure 6 illustrates a situation analagous to the identification after condition, prior to the identification statement. There are two distinct nodes, unconnected, that refer to the same individual. Introspectively, it seems clear that at least sometimes I comprehend definite descriptions before recognizing their referents. For instance, I understand the description The president of France in 1970 long before I decide that this is George Pompidou. The structure surrounding Y in Part (a) not only provides an embodiment of this pre-identification comprehension, it serves as an encoding of the information that is to guide the search for a referent. The ACT theory would use this representation to build a pattern that could be matched to memory to retrieve the referent. In the case of a description like first president of the USA a direct pattern match should suffice to retrieve the referent. In my case for the President of France in 1970 description, a more complex problem solving strategy had to be evolked.

Once the description of sentence (3) has been comprehended two things can happen: The subject can proceed to recognize the referent of the definite description and he can go on to comprehend the "was bad husband" predicate. Depending on the order of these two events we will wind up with slightly different representations in memory. Part (b) of Figure 5 illustrates the state of memory after recognition of the description and before comprehension of the predicate. As in the after condition (Part b of Figure 1) a link is introduced encoding the identity of X and Y. When the predicate is comprehended a representation of its meaning can be attached directly to X, giving the representation in Part (c) of Figure 6.

Part (d) illustrates the state of memory when the predicate has been comprehended but the definite description has not been identified. In this case the meaning representation of the predicate has been attached to node Y. Part (e) of Figure 6 illustrates the state of memory when the definite description is subsequently recognized. Again a link is introduced indicating the identity between X and Y. The bad husband predicate, which is active in memory, is copied from Y to X. The difference between the final state of the recognize-description-then-comprehend-predicate sequence (Part c) and the comprehend-predicate-then-recognize-description sequence (part e) is that in the latter case the predicate is attached to both nodes. This latter situation is like the situation in the after condition of the previous memory experiments.

What determines which occurs first -- recognition of description or comprehension of predicate? In the ACT model both processes can go on independently. It would simply be a race between two independent processes. Factors such as how quickly the predicate is presented (if spoken) or how quickly the subject turns to the predicate (if printed) will determine the speed of the comprehension success. The speed of recognizing the description will vary with the difficulty of finding its referent. It is clear that neither process waits on the other as witnessed by the sentences:

(11) The first prime minister of Canada was a bad husband.
(12) The first president of the United States pilacked gibs.

In (11) we comprehend the predicate although we never find a referent for the subject. In (12) we find a referent for the subject although we never comprehend the predicate.

Evidence on the Recognition of Referring Expressions.

Right now the contentious reader might be thinking "Yes, that is a possible model for the processing of referring expressions. Yes, it is consistent with the model for your earlier memory experiments. Yes, you presented evidence for that model. But, is there any independent experimental evidence for this model when applied to the real-time recognition of definite descriptions?" Because of its rapid real-time characteristics it is hard to provide particularly direct evidence for this process. But there are some consistent experimental results:

A relevant feature to note about Figure 6c is that it preserves no record that the bad husband predicate was asserted via the first president of USA description. In contrast Figure 6e does preserve this information. Both representations are possible depending on the exact timing of description recognition versus predicate comprehension. To the extent that there is a mixture of these representations we predict both a tendency to make confusions about what referring expression was used (representation 6c) and that subjects will have some residual ability to make this discrimination (representation 6e). An experiment reported by Anderson and Bower (1973) supports this dual prediction. They had subjects study sentences like:

(11) The first president of the United States was a bad husband.
(12) Abraham Lincoln was a good husband.

After studying such sentences subjects were asked to chose among alternatives such as the following:

(13) The first president of the United States was a bad husband.
(14) George Washington was a bad husband.
(15) The first president of the United States was a good husband.
(16) George Washington was a good husband.

These alternatives were presented to the subject randomly ordered but I present them here systematically. Subjects were instructed to indicate the exact sentence that they had studied in which case (13) would be the correct choice. To the extent that subjects false alarm more to (14) over (15) or (16), this is evidence for a representation like Figure 6c where no information is retained about the referring expression used. To the extent that subjects prefer (13) over (14) this is evidence for a representation like Figure 6e. Thus, our predictions in terms of preference is (13) > (14) > (15) = (16). The evidence clearly confirms this prediction with subjects saying that they had seen sentences like (13) 65.2% of the time, like (14) 21.4% of the time, like (15) 7.2% of the time, and like (16) 6.3% of the time. An earlier memory model, HAM (Anderson & Bower, 1973) predicted total confusion in this situation rather than an intermediate level of confusion. In the recognition model for HAM there was no separate memory structure to encode the referring expression. Rather the referent node was directly retrieved from memory without the intermediate step of calculating a representation of the referring expression in memory.

Recently Ortony and Anderson (1977) report a study which replicated and extended this result. They noted that some predicates seemed more appropriate to a proper name and other predicates seemed more appropriate to a definite description. Consider their examples:

(17) The first man on the moon became a national hero.
(18) Neil Armstrong has several children.
(19) The first man on the moon has several children.
(20) Neil Armstrong became a national hero.

Ortony and Anderson point out that the uses in (17) and (18) are somewhat more natural than the uses in (19) and (20). Correspondingly, they found subjects made fewer errors in remembering what the referring expression had been for sentences like (17) and (18) than for sentences like (19) and (20). The error rates were 19.6% versus 30.7%. Note, however, that in both cases subjects identified the original referring expression better than chance (50%).

The Ortony and Anderson result would be expected under the current theory. To the extent that the predicate fits the referring expression subjects might attach it to the new node (e.g., node Y in Figure 5) which has the referring expression attached to it. As Ortony and Anderson noted, the HAM theory had no way to explain this affinity between certain referring expressions and certain predicates. To explain the Ortony and Anderson results in the HAM framework we had to attribute them to a response bias.

In the current ACT theory we can explain this result in terms of the frequency with which subjects chose Part (c) versus Part (e) of Figure 6. The claim is that subjects use representations like Part (e) more frequently when the referring expression is appropriate. This is because it is easier to elaborate on the connection between the referring expression and the predicate.

## Opaque and Transparent References

This analysis of reference has a natural extension to analyzing the difference between opaque and transparent reference. For nstance, contrast:

(21) I am looking for the best lawyer in town.

(22) I am looking for my little old mother.

While both (21) and (22) might be considered ambiguous, the more apparent interpretation of (21) is that I am looking for someone who fits the description "the best lawyer in town" and that I do not have a particular person in mind. In contrast, the more apparent interpretation of (22) is that I do have a particular person in mind. The former is an instance of opaque reference and the latter is an instance of transparent reference. Our discussion has so far focused on transparent reference. To correctly remember an instance of opaque reference it is critical that it not be treated in the same manner as transparent reference. That is, even if the listener knows the reference of "the best lawyer in town", he should not use the node for this reference in representing the meaning of (21). Rather he should create a new node, attach the description to it, and put this node in the representation of (21). Figures 7a and 7b illustrate the different representations appropriate for (21) and (22). In Part (a) there are two dis-

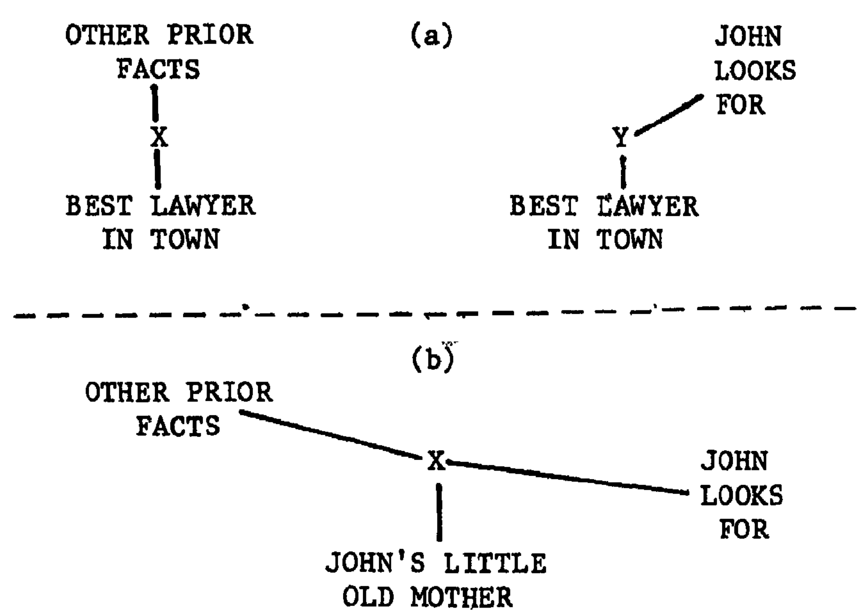



Figure 7: Memory representation for any instance of opaque reference (a) and transparent reference (b).

tinct nodes preserved to represent the best lawyer in town. One node (X) has the prior facts known about the person while the second node (Y) stores information about the opaque reference in sentence (21). There is no such distinction in Part (b) of Figure 7 for the transparent case in (22). All information is attached to the original node X. So, the difference between transparent and opaque reference is whether the new information is copied to an existing node.

## Conclusions

In concluding this paper I would like to return to Frege's morning star-evening star example. The discovery that the morning star was the evening star was an important scientific result. Frege used this fact to make clear the important distinction between sense and reference. The first half of this paper reported experiments where we basically recreated Frege's example and discovered that subjects dealt with that dilemma by the process of copying from one referring node to another. The argument in the second half of the paper was that Frege's examples are not isolated to discoveries of science or to bizarre psychological experiments. Rather, every time we recognize a transparent referring expression we go through a discovery like that of the identity between the morning star and evening star. We create a node to represent the referent of the referring expression and only then discover, with varying difficulty, that this node has the same reference as an established node in memory.

Acknowledgments: Preparation of this manuscript and the research reported was supported by Grant BNS 76-00959 from the National Science Foundation. I would like to thank Paul Kline for his comments on this manuscript.

## References

Anderson, J. R. Language, memory and thought. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Anderson, J. R. Memory for information about individuals. Memory & Cognition, 1977, 5, 430-442.

Anderson, J. R. & Bower, G. H. Human associative memory. Washington, DC: Hemisphere Press, 1973.

Anderson, J. R. & Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. Cognitive Psychology, 1974, 6, 495-514.

Frege, G. Über Sinn und Bedetung. Zeitschrift für Philosophie und philosophiche Kratik, 1892, 100 (all). (Translation available in P. Geach and M. Black (Eds.), Translations from the philosophical writings of Gottlob Frege. Oxford: Basil Blackwell, 1960, 56-78).

Ortony, A. & Anderson, R. C. Definite descriptions and semantic memory. Cognitive Science, 1977, 1, 74-83.