TEXTUAL AND NON-TEXTUAL FACTORS IN SITUATION MODEL FORMATION

Charles A. Weaver III

&

Walter Kintsch

Institute of Cognitive Science

University of Colorado

Boulder, Colorado

September, 1985

Technical Report No. 150

ABSTRACT

Using a paradigm established by Perrig and Kintsch (1985), memory for locative and non-locative texts was studied. It was found that if a situation model could easily be formed from the texts, subjects tended to use this model in verifying inferences about the relative locations of objects mentioned in the text. Furthermore, their recalls of the overall text tended somewhat toward a memory reconstruction from this situation model. If, however, the text were changed slightly, so that situation model formation was more difficult, subjects would rely on their textbase memory for making the inferences, and textual factors were more important. Recalls in this case were more reproductive than reconstructive. This was taken as support for the van Dijk and Kintsch (1983) model of discourse comprehension.

INTRODUCTION

The van Dijk and Kintsch (1983) model of text comprehension refined their earlier theory (Kintsch & van Dijk, 1978) by incorporating a third level of representation: the situation model. The three levels of representation in the van Dijk and Kintsch model are surface memory, text-based memory, and the situation model. The surface memory for the text is the verbatim encoding of the text, and is thus totally defined by the text. The text-based memory is derived from the surface structure, but it is concerned with the meaning of the text, and thus is not as rigid. The situation model is the most abstract and variable of the representations. It involves not only the meaning of the text, but also the reader's goals and previous knowledge. Thus, two readers could read an identical passage but comprehend and remember differently. A problem with this new level of representation, though, is how to distinguish it from the text-based memory. With most texts both representations will be verbal.

Perrig and Kintsch (1985), in their study of non-verbal situation model, used a text which described the layout of a small town. The situation model in this case would be the spatial representation of the the town, and could be distinguished from the text-base by its non-verbal nature. One of the results of their manipulations was that Perrig and Kintsch noticed that subjects apparently could recall both components of an inference and still not correctly make that inference. Using their design, however, Perrig and Kintsch could not quantify this observation.

This study will attempt to look more closely at those situations where

both components of an inference are recalled. When will subjects make the inference correctly, and when will they not? Also, construction of the text will allow further study of the differences between the text based memory and the situation model.

EXPERIMENT 1

METHOD

Subjects. Seventy subjects from the University of Colorado subject pool satisfied a course requirement by participating in the study. There were 29 males and 41 females. Two subjects, one of each sex, did not follow instruction and were not used in the data analysis.

Materials. The text used in this experiment were constructed to fit several criteria. First, the locative information was intended to be presented within the context of a complete story, rather than as a series of discreet, unrelated episodes. Second, the text was to have two distinct parts which would be thematically related but were relatively independent. The text used is presented in Appendix 1. The story describes the robbery of a bank in a small town, and the subsequent pursuit and capture of the robbers. The first paragraph deals with the arrival of the men in the town, the actual robbery and the getaway by the robbers. The second paragraph describes the police chief's investigation of the theft and the arrest of the men. Within each paragraph, three locations are mentioned, all along Main Street. The objects are separated from each other by three different measures of distance: a)physical distance, which is the actual distance between the objects in the town. For example, the bank and the car lot are next to each other, so they were given a physical distance of

one unit. The bank and the grocery store were separated by one object (the car lot), so they were given a physical distance of two.; b)surface distance, which is the distance between the objects as they are mentioned in the paragraphs. The bank was mentioned right after the gas station, so it was given a surface distance of one. The grocery store was presented after the bank, so it had a surface distance of two from the gas station (and one from the bank); c)textual distance, which was scored one if the objects were mentioned in the same paragraph and two if they were in different paragraphs.

The fifteen possible locative inference were constructed from the text and presented for verification to the subject. The order of the sentences was random, and there were four versions of the each sentence. For example, if the sentence for verification was "A is north of B", then a subject might have received in in that form, or as "B is south of A", or "A is south of B", or "B is north of A". (Obviously, the truth value of the last two sentences was different form the first two.) Each subject received 7 or 8 of the sentences as true, and the rest false.

Procedure. All of the subject were told that they would read a crime story, and to pay close attention as they would be asked questions about the story later. The story was presented on a computer terminal, one paragraph at a time, for seventy seconds per paragraph. Immediately after the text was presented, the subjects were asked to recall all they could about the text, using exact words if possible. No time limit was placed on the recall, but no subject took more than six minutes. The subjects were told not to refer back to their recalls at any time during the rest of the experiment. Following the

recall, the subjects were told that about fifteen sentences were to be presented to them, and they were to answer true or false to each sentence as quickly as possible by pressing a button marked "true" or "false" on a box in front of them. Their responses and response times were recorded. Following presentation of all 15 sentences, the subjects were told to place the six objects mentioned in the story on a map of the city. The river and the Main Street were already drawn on the maps for the subjects, and they were given the names of the six objects.

RESULTS AND DISCUSSION

RECALL

The free recalls of the subjects were analyzed by comparing them to the proposition list for the text. Performance was about average for this type of task, with the mean subject recalling 38.8% of the propositions. The text was formally analyzed using the COHERENCE program written by Young and Miller (see Young, 1984). The program takes as its input a list of the micropropositions and macropropositions and analyzes the text according to the Kintsch and van Dijk (1978) model. Among other things, the program constructs a coherence graph and computes the number of times each proposition should be carried over to the next processing cycle.

The subjects' recalls showed a levels effect, in which propositions which are higher in the micrograph are recalled with greater frequency. The levels effect is shown in Figure 1. The recalls also showed as cycles effect, in which propositions which are predicted to be held over for additional processing cycles are recalled with

greater frequency. (Three propositions were held over for more than 2 cycles, but for clarity these were combined with those held over for two cycles.) This cycles effect is illustrated in Figure 2.

Each subject's recall was also scored to determine which of the six location phrases were present. The percent of subjects recalling each of the location phrases is shown in Figure 3. The location phrase mentioned first is number 1, the one mentioned second number two, etc. Note that there appears to be a fairly strong primacy effect, but location phrases 3 through 6 were recalled at similar levels, with no trace of a recency effect. Based on which of the six were recalled, it is possible to determine which of the inferences should be correctly verified. The overall performance on those sentences was excellent, with the mean performance being 89.6% correct. This level of performance was significantly higher than if the item was not predicted to be correctly verified (89.6% vs. 62.7%, Chi sq df=1 87.4, $p < .001$).

VERIFICATIONS

Each subject received all fifteen possible sentence verification. Performance here was also fairly good, with a mean success rate of 68.3%, significantly better than chance, Chi sq, df=1, 58.3, $p < .001$. A signal detection analysis was also done for each of the sentences and, as expected, no response biases existed. The mean Beta for all 15 sentences was 0.94. A truly unbiased item would have a Beta of 1.00. Hence, chance performance is 50%. The mean d prime was .958. The arcsin transform was used for all analyses involving proportion correct, but for clarity the actual proportions will be reported. The

hypothesis that it would be easier to verify two item if both the items appeared in the same paragraph was not supported. Within paragraph inferences were verified with 68.9% accuracy, while performence on between paragraph inferences was only slightly lower, 67.8%. This small difference was not statistically reliable, Chi sq, df=1, 1.13, $p > .25$.

An item analysis was done using each of the sentences as items to test the effects of physical, surface and textual distances. The only statistically significant effect was the effect of physical distance. As the physical distances between the two objects increased the sentence was more likely to be correctly verified ($F(1,13) = 9.45$, $p < .01$). Neither surface distance nor textual distance was a significant factor. The effects are illustrated in Figures 4-6.

REACTION TIMES

Reaction times were measured for each subject. The overall mean for the sentence verifications was 4835.1 msec. Reaction times were not different for true and false sentences, but subjects were significantly faster when they responded correctly (4585.2 msec for correct responses vs. 5368.5 msec for incorrect responses, $F(1,1018) = 18.04$, $p < .001$). There was no main effect of physical distance, there was a physical distance by correct/incorrect interaction, $F(4, 1010) = 2.48$, $p < .05$). Figure 7 shows this interaction. As the physical distance increases, subjects responding correctly get slightly faster, while subjects responding incorrectly get slower.The main effect of surface distance was significant, $F(4,1010) = 2.36$, $p = .051$, but there was no interaction. This is shown in Figure 8. Like

physical distance, the main effect of text distance was not significant, but the text distance by correct/incorrect interaction was significant, $F(1,1016) = 5.68$, $p < .05$. This effect is illustrated in Figure 9. Subjects respond more quickly if they were correct, but that effect is much more pronounced if the objects were in the same paragraph.

These reaction time results are somewhat ambiguous. The physical distance interactions indicate that subjects have indeed formed a situation model, but the effects of textual and surface distances indicate that the subjects were also using a textbase. It appears that subjects used both textbase memory and their situation model in responding, but it is not clear under which circumstances which representation was used.

EXPERIMENT 2

The goal of the study was to examine situations in which both components of an inference are recalled. In Experiment 1 there were quite a number of such situations, but in most (nearly 90%) the subjects correctly made the inference. The text used in Experiment 1 was very simple from the standpoint of situation model formation-- all of the locations mentioned were on the same street with only north/south differences. It appears, therefore, that subjects in Experiment 1 were able to form an adequate situation model for the town described, and that their recall of the locations was at least in part reconstructive, as observed in Weaver and Kintsch (1985). Hence it may not be surprising that whenever recall was good, inferences were also good. To test our original hypothesis we need a situation where subjects have no (adequate) situation model and must rely on

their textbase for both recall and inferences. Perhaps if the text was left basically the same, but the locations were arranged so that they would be more difficult to follow (making the situation model formation more difficult) then subjects would recall the locations, but respond incorrectly to inferences involving these locations.

METHOD

Twenty-seven subjects, 15 female and 12 male, from the University of Colorado subject pool satisfied a course requirement by participating in the study. Two subjects (one of each sex) did not follow instructions and were disqualified.

Materials The text used in the second experiment was very similar to that used in the first experiment. The text was changed slightly to make the path taken by the robbers more difficult to follow. The bank was placed off of Main Street several blocks to the East, and the car dealership was placed several blocks to the West. Other than those minor changes, the text was identical.

It was now possible to construct 6 more inference statements in the verifications, incorporating the East/West aspects of the bank and the car dealership. These new sentences were added to the existing 15 sentences (since the bank and the car lot were displaced only East or West, the North/South verifications remained the same). The four possible arrangements of the new sentences were done as before.

Procedure The tasks were presented exactly as before.

RESULTS AND DISCUSSION

RECALL

Free recall was analyzed as before. Performance was again adequate, with subjects recalling an average of 35.1% of the propositions. The new text was also analyzed with the COHERENCE program of Young and Miller (Young, 1984). As before, the subjects' recalls showed a levels effect ($F(4, 100) = 7.80$, $p < .001$) and a cycles effect ($F(2, 102) = 33.1$, $p < .001$). The levels effects of both experiments are shown in Figure 1, and the cycles effect are shown in Figure 2.

The recalls were also scored to determine which of the six location phrases were present. It was hoped that the more difficult situation model from this text would result in subjects recalling both components of the inference who could not correctly make an inference involving the components. This could not be verified, though, since the subjects did not recall the components! In fact, only three subjects recalled more than one of the location phrases. The percent of subjects recalling each of the six location phrases is shown in Figure 3. Notice that the graph of Experiment 2 shows the same shape as the classic serial position curve, with strong primacy and recency tendencies.

VERIFICATIONS

Each subject received all 21 verifications. The goal of this experiment was to make the situation model more complex, and the indications from the verification data indicates that this goal was accomplished too well! Overall performance was very low, with the sentences answered correctly only 52.9% of the time (chance performance is 50%). This difference was not reliably better than chance (Chi sq, df=1, = 1.96, $p > .10$).

An item analysis was done using each of the sentences as item. No effects of physical distance was found, but both surface and text distances effects were found (Fs (4, 16) and (1, 19) = 4.81 and 6.21, respectively, ps < .05). These effects are shown in Figures 4-6.

REACTION TIMES

Reaction times were measured for each subject. The overall mean for the sentence verifications was 4756.1 msecs. There were no differences for correct and incorrect responses, or physical, surface or text distances, nor were there any correct/incorrect X distance interactions. Subjects were significantly faster for 'yes' responses than 'no' responses, though, 5249 msec. vs. 4323 msec., $F(1, 523) = 12.9$, $p < .01$.

GENERAL DISCUSSION

The goal of the study was to determine under which conditions both components of an inference will be recalled, but the inference will not be correctly made, and to study these conditions. In Experiment 1 if the subjects recalled the components they made the inference with nearly 90% accuracy. In the second experiment, subjects did poorly on the verifications, but neither did they recall the components. There were a number of interesting results, however.

Since the text in both experiments was very similar, except that the location phrases in the second text were more difficult to comprehend. The recalls of the non-locative propositions, therefore, should be very closely related. The data support this. The Pearson product-moment correlation coefficient for the non-locative

propositions in both experiment was .835 ($p < .001$) indicating a high degree of consistency. The correlation coefficient for the locative propositions, on the other hand, was .404 ($p > .05$). The situation model, which in this case would be a map of the town, is formed primarily from the locative information. Since the recalls differed primarily in the recalls of the locative information, there should be significant differences in the situation models of the subjects in Experiment 1 and Experiment 2.

The first difference between the two can be found in the recalls of the propositions which make up the locative phrases. These results are shown in Figure 3. Except for the last phrase, which describes the location of the roadblock, the subjects in the first experiment recalled the location phrases at a much higher level. Four of the phrases, numbers 1, 3, 4, and 6 appeared identically in the two different texts, and as can be seen, they were much lower for the second text. Furthermore, when the first phrase was seen in the two texts there had been no differences in any part of the text. Subjects in both experiments should have recalled this phrase identically. This was not the case; 48% of the subjects in Experiment 1 recalled the phrase, while only 30% of the subjects in Experiment 2 recalled the phrase. There must have been some type of retroactive restructuring of the text for these differences to appear.

Additionally, the subjects in Experiment 2 showed a serial position effect, while the subjects in Experiment 1 did not. Serial positions functions are usually associated with textual factors, which again support the contention that textual factors play a larger role in Experiment 2.

The second bit of evidence that the subjects in Experiment 2 found the situation model difficult to construct can be found in the verifications. These subjects, it appears, after a few location phrases had been presented gave up trying to construct the situation model. This would suggest that perhaps when answering the verifications they were relying on their textbase memory for the text rather than their situation models. The subjects in Experiment 1 did have an adequate situation model and thus responded to the verifications from this model and not from the textbase. The subjects in Experiment 1 did show a physical distance effect, which is evidence that they had formed the situation model. They did not show any textual factors, though (surface or textual distance). The subjects in Experiment 2, though, showed ONLY textual effects-both a surface distance and a textual distance effect; they did not show a physical distance effect. Additionally, though the subjects in both experiments showed both levels and cycles effects, those in Experiment 2 showed enhanced effects (see Figures 1 and 2). This is further indication that textual factors were more important for the subjects in Experiment 2.

FIGURE CAPTIONS

Figure 1: Percent Recall versus Level in STM.

Figure 2: Percent Recall versus Number of processing cycles held over.

Figure 3: Percent Recall for each location phrase, in order of text appearance.

Figure 4: Percent correct in verifications versus physical distance.

Figure 5: Percent correct in verifications versus surface distance.

Figure 6: Percent correct in verifications versus text distance.

Figure 7: Reaction time versus physical distance in Experiment 1.

Figure 8: Reaction time versus surface distance in Experiment 1.

Figure 9: Reaction time versus text distance in Experiment 1.

Figure 1

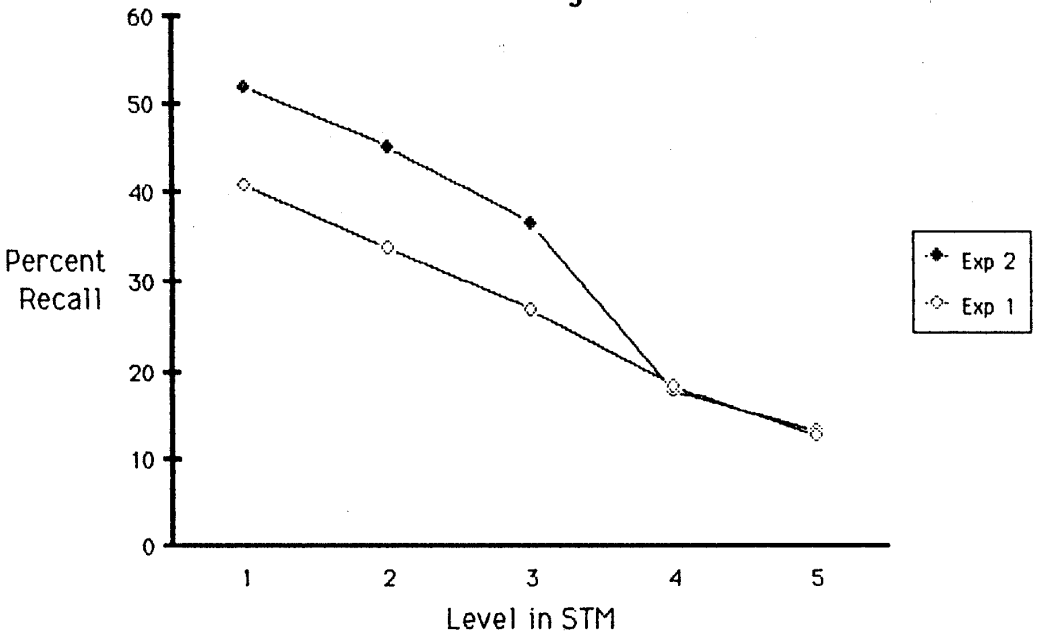


Figure 2

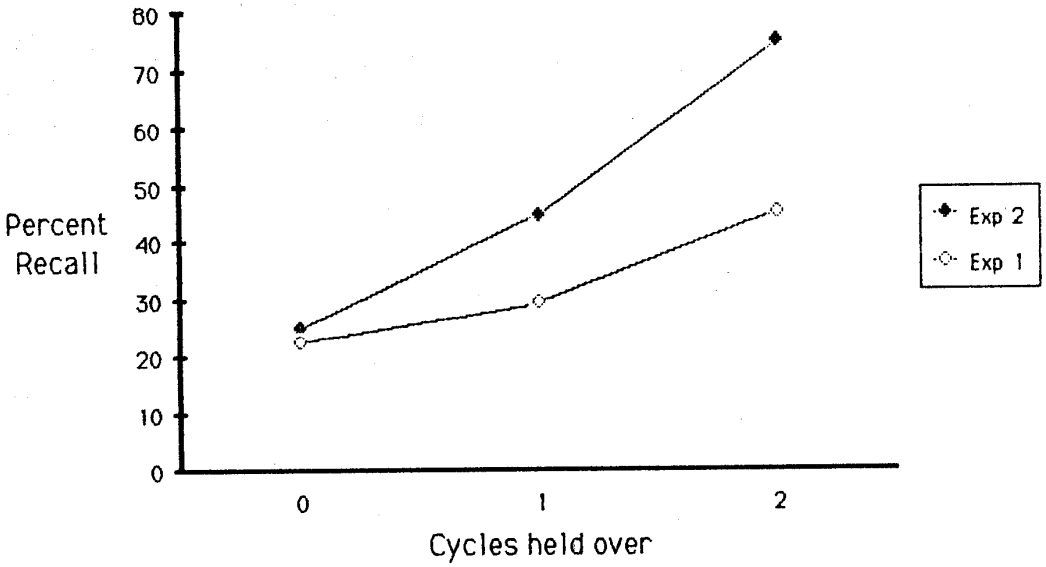

80
70
60
50
40
30
20
10
0
Percent Recall
0
1
2
Cycles held over
Exp 2
Exp 1

Figure 3

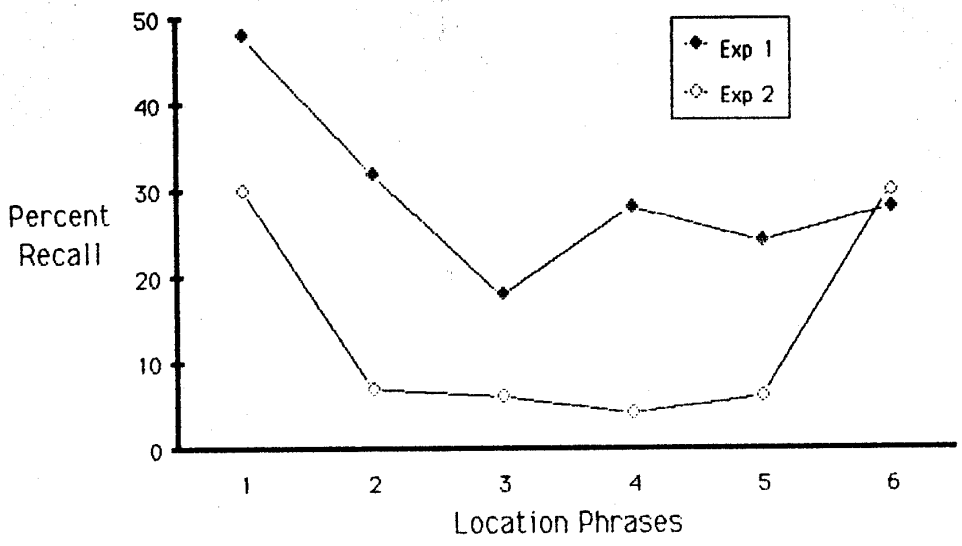

Exp 1
Exp 2
Percent Recall
50
40
30
20
10
0
1
2
3
4
5
6
Location Phrases

Figure 4

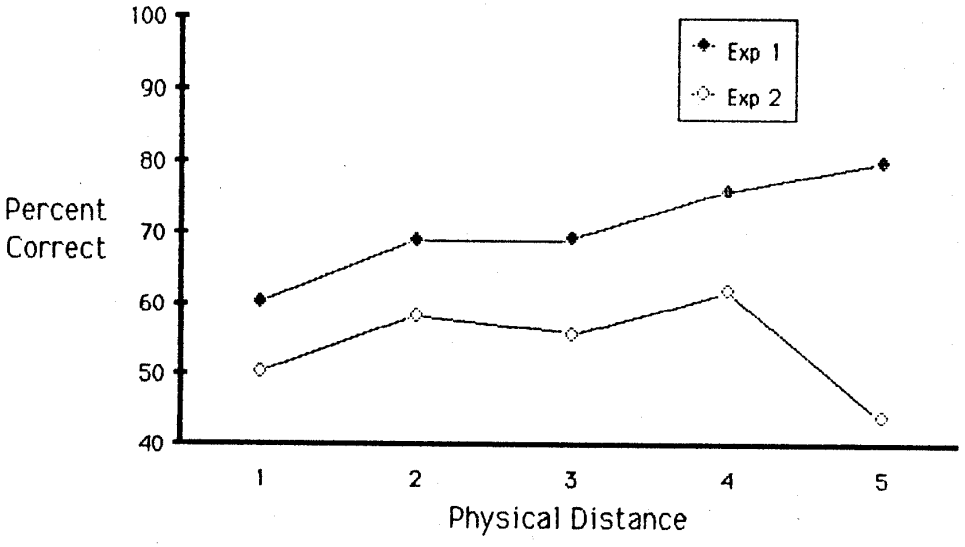

100
90
80
70
60
50
40
Percent Correct
Exp 1
Exp 2
1
2
3
4
5
Physical Distance

Figure 5

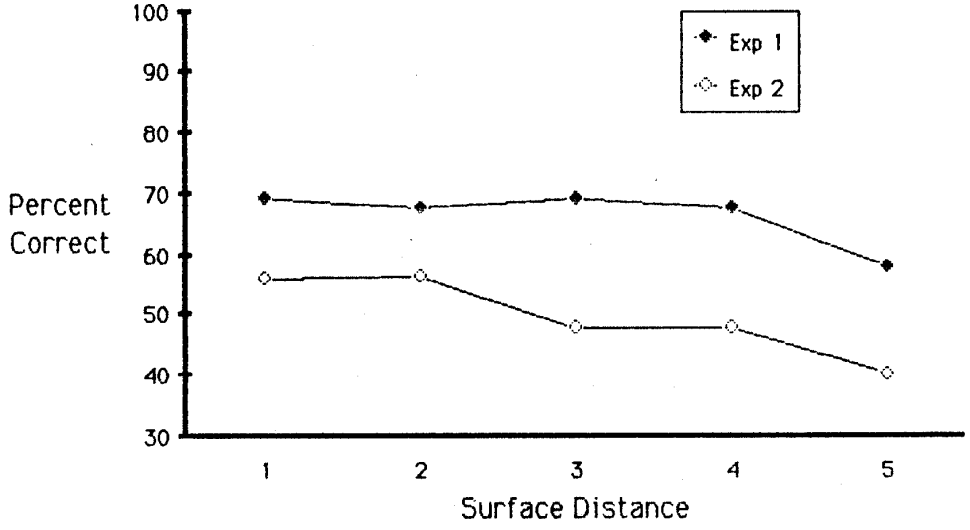

Exp 1
Exp 2
Percent Correct
100
90
80
70
60
50
40
30
1
2
3
4
5
Surface Distance

Figure 6

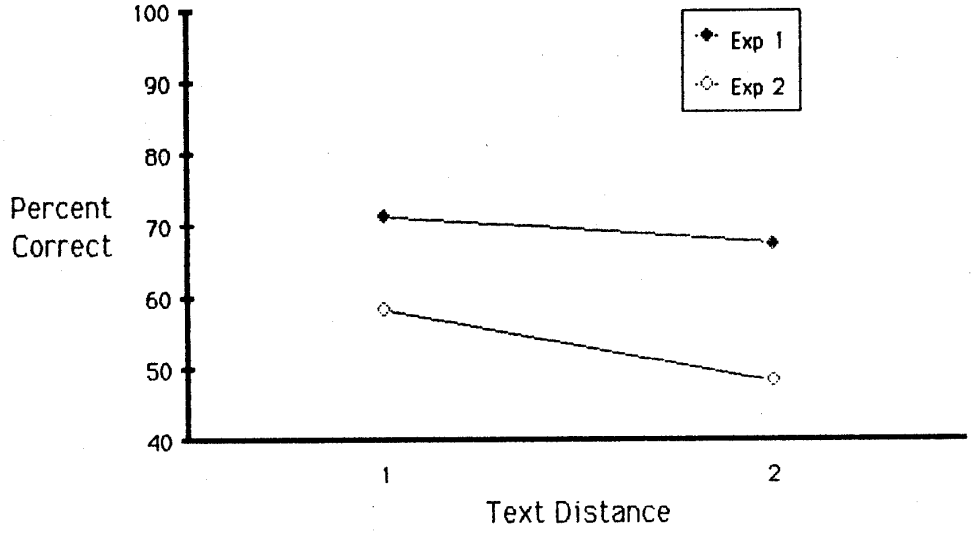

100
90
80
70
60
50
40
Percent Correct
Exp 1
Exp 2
1
2
Text Distance

Figure 7

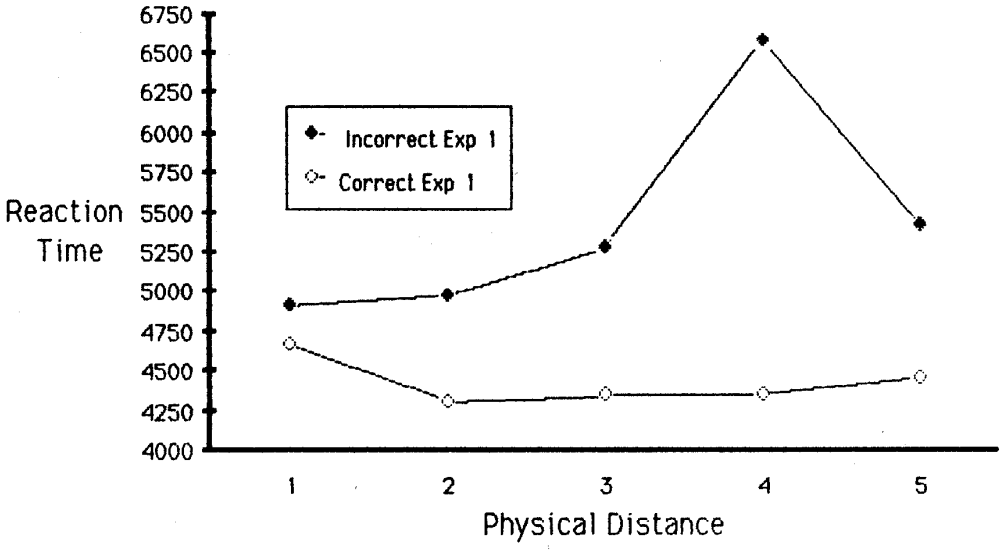

6750
6500
6250
6000
5750
5500
5250
5000
4750
4500
4250
4000
Reaction Time
Incorrect Exp 1
Correct Exp 1
1
2
3
4
5
Physical Distance

Figure 8

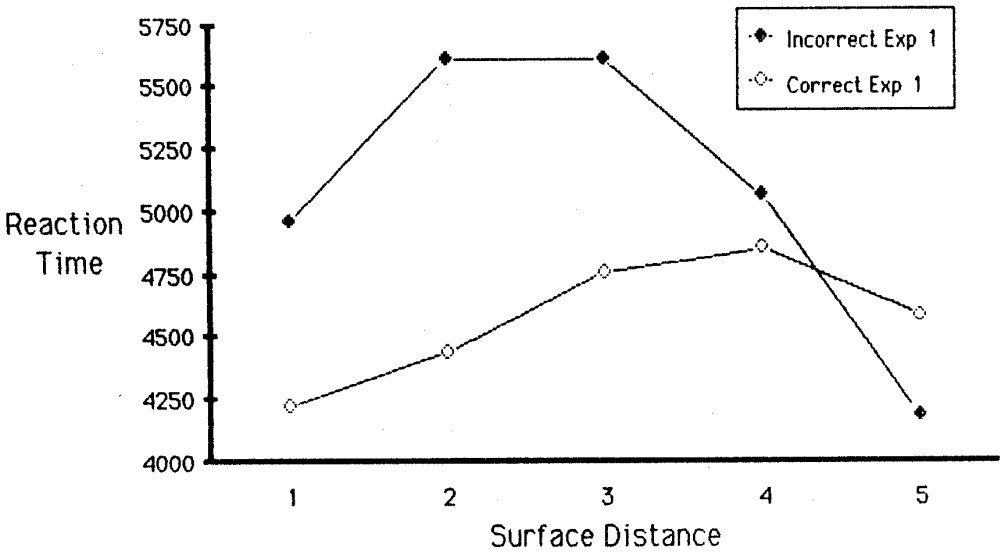
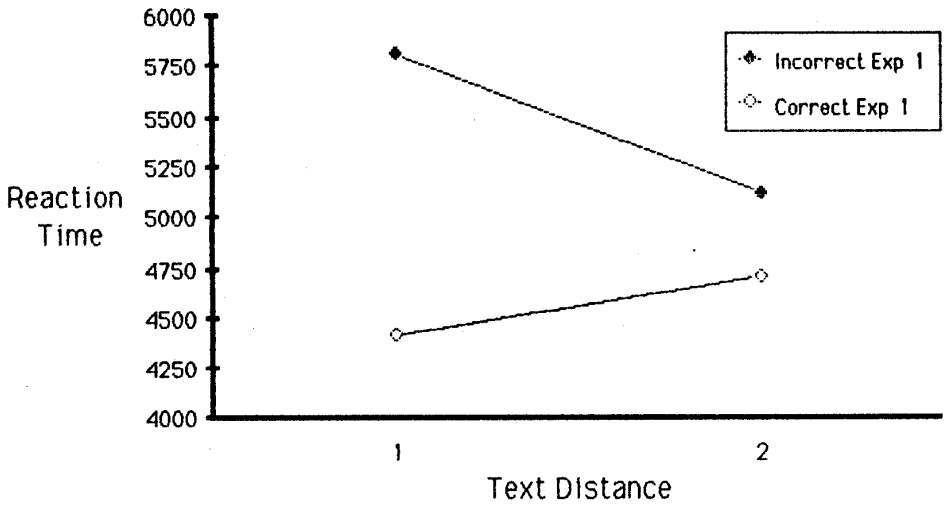

Incorrect Exp 1
Correct Exp 1
Reaction Time
5750
5500
5250
5000
4750
4500
4250
4000
1
2
3
4
5
Surface Distance

Figure 9


6000
5750
5500
5250
5000
4750
4500
4250
4000
Reaction Time
Incorrect Exp 1
Correct Exp 1
1
2
Text Distance

APPENDIX 1

The gas station attendant didn't recognize the car that pulled into his lot. Being 10 blocks north of the river on Main Street, he noticed every car that came into town on Main Street. As he filled their tank, he saw the three men inside the car and thought that it was odd that they would be wearing overcoats in the middle of May. Oh well, he thought, that's out-of-towners for you! After he filled their tank, they screached out of the gas station. The car kept going south on Main, and stopped at the Bank, which was just north of the River that flows east-west through the town. Two of the men got out, carrying a large suitcase, while the third man stayed in the car with the motor running. A few minutes later, the two men came back out of the bank and ran to the car. The men drove quickly away to the south, over the River, and drove for several blocks before abandoning the car in front of the grocery store. An older woman came out of the grocery store just as the men were getting out of their car. Just as she was looking, the suitcase that the biggest man was carrying came open, and money flew everywhere. The woman gasped, looked once again at the men, and rushed back into the store.

Two days later the police chief still had no leads in the case. Since the men no longer had their car, the chief knew that they must still be close by. He had thought that they would hide out in the old warehouse, which was at the Southern edge of Main Street, about 10 blocks south of the river. He and his men checked out the warehouse, though, and found no trace of the robbers. The chief was out of clues, when suddenly he got a phone call. The local car dealer, whose lot was just south of the River, said that he had just sold a brand new car to three men who met the

robbers' description, and furthermore, they paid for the car in new $100 bills. The car dealer said that the men were about to drive away to the north, over the river. That was the break that the chief was waiting for! He immediately got on the police radio and ordered that a roadblock be set up on Main Street, several blocks north of the River. The chief got in his car and rushed to the roadblock. By the time he got there, the robbers had been apprehended.

REFERENCES

van Dijk, T.A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of discourse comprehension and production. Psychological Review , 85, 363-394.

Perrig, W. & Kintsch, W. (1985). Propositional and situational representations of text. Journal of Memory and Language, in press.

Weaver, C.A., III, & Kintsch, W. (1985). Levels of understanding in text comprehension. Manuscript in preparation.

Young, S. R. (1984). A theory and simulation of macrostructures. (Technical Report No. 134). Boulder: University of Colorado, Institute of Cognitive Science.