JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 22, 437–448 (1983)

Inferring Consequences in Story Comprehension

MURRAY SINGER AND FERNANDA FERREIRA

University of Manitoba

The study asked whether readers infer the consequences of events described in stories. Forward and backward inferences were distinguished: only backward inferences contribute to the coherence of a message. The subjects read stories of 9 to 11 sentences and then answered eight questions. The time needed to answer forward inference questions was about the same as for questions paraphrasing the story, but over 0.2 seconds longer than for questions repeating part of the story and for backward inference questions. It was concluded that backward consequence inferences are more reliably drawn during the course of reading than are forward consequence inferences.

In two recent studies, Singer (1979, 1980) asked whether people draw inferences about implied "case-filling" elements like agents, patients, and instruments, during reading. Two general types of inferences were studied. (1) Backward inferences (Thorndyke, 1976; Just & Carpenter, 1978) are ones that specify a connection between the current sentence and an earlier part of the text. For example, the inference that "the pitcher threw the ball" establishes a connection between *the pitcher threw to first base* and *the ball sailed into the field.* Backward inferences enhance the coherence of a message. (2) Forward inferences, in contrast, may be highly probable, but do not contribute to coherence in this fashion.

Singer (1979, 1980) found that readers needed about the same amount of time to verify test sentences expressing the backward implications and corresponding explicitly stated ideas of texts, but more time to verify forward inference test sentences. It was concluded that backward case-filling elements are more reliably "computed" and stored during comprehension than are forward inferences. It was proposed that there are simply too many possible forward inferences that can be derived from a message to permit many of them to be included in the representation of a message (cf. Clark, 1977). Backward inferences, on the other hand, are drawn during reading because, without them, a message would become disjoint.

While many investigators have inspected case relations in comprehension (e.g., Corbett & Dosher, 1978; Carpenter & Just, 1978), there is a growing consensus that there are other relations that are more crucial to the gist of a text. Foremost among these is the relation of cause. Schank and Abelson (1977), for example, have argued that many aspects of the meaning of narrative texts can be represented in terms of "causal chains." A causal chain consists of a sequence of events or actions which function as the causes, enabling conditions, and consequences of one another.

Several investigators have emphasized causal relationships, and even the causal chain construct itself, in recent theoretical formulations concerning text representation and understanding (Glenn, 1978; Graesser, Robertson, & Anderson, 1981; Lehnert, 1977; Trabasso, Secco, & van den Broek, in press). In most of these treatments, it has been asserted that the causal links underlying text meaning are frequently implicit; and that, therefore, causal inferences are essential to ensure comprehension.

The goal of the present study was to deter-

This research was supported by Grant A9800 from the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Linda Wilson for her comments and encouragement. The second author is now at the University of Massachusetts. Requests for reprints should be sent to M. Singer, Department of Psychology, University of Manitoba, Winnipeg R3T 2N2, Canada.

0022-5371/83 $3.00

mine whether causal inferences are drawn during the course of reading. The reason this issue was in doubt is that, as was stated before, human information processing limitations place a severe restriction on the number of inferences that can accompany comprehension. In studying causal inference, it should be possible to avoid certain errors that characterized some earlier studies of inference. In particular, it has been pointed out that neither the false recognition of an implicational test sentence, nor the effectiveness of an implicit recall prompt, nor the reader's ability to answer questions that require inferences from text, prove that the inferences in question were drawn when the text was originally examined (Singer, 1980; Trabasso, 1980).

In this paper, we will take the view that, if an inference is drawn during reading, then the reader's internal representation of a text includes either propositions or propositional arguments that were not directly conveyed by the text. If the sentence *the child stuck the balloon with the pin* permits the inference that the balloon burst, then, using Kintsch's (1974) notation, the representation will include the inferred proposition (BURST, BALLOON). Conversely, the mere spread of activation to a semantic memory node does not constitute an inference (see Corbett & Dosher, 1978, p. 480). For example, the transient activation of the node BROOM that might occur when "sweep" is used in a sentence does not ensure that the reader has inferred that the sweeping was done with a broom.

The complexity of the concept of cause required a refinement of the questions asked in the present study. First, reference was made to the distinction between reasons and causes (Graesser, Robertson, Lovelace, & Swinehart, 1980; Wilks, 1977). Human actions (e.g., *John crushed the lunch*) are said to be caused by the actor's reasons and goals, while physical events (e.g., *the rock crushed the lunch*) are caused by other events. The present study focused upon the causes of physical events, because the study of reasons presupposes a reasonably detailed theory of human motivation (cf. Schank & Abelson, 1977).

A second refinement involved distinguishing between inferring the existence of a causal link, as opposed to inferring an event that participates in a causal relationship. When one encounters a sequence like *the child stuck the balloon with the pin, the balloon burst*, it is likely that one infers the existence of the connection that the first event caused the second. The role of such connections in comprehension has been examined both theoretically (Wilks, 1977) and empirically (Black & Bern, 1981). The present investigation, however, focused on the issue of whether people infer *events* that play a role in causal relations.

Third, it was necessary to distinguish between the two major elements of a causal relationship: namely, the cause and the consequence. The present study examined consequences, for two reasons. First, several investigators have provided evidence that the consequences of actions and events constitute one of the best recalled categories of ideas in a story (Glenn, 1978; Graesser et al., 1981; Nezworski, Stein, & Trabasso, 1982). Second, an examination of relevant materials revealed that while some events have almost certain consequences, fewer have almost certain causes. Consider again the sentences in Sequence (1).

(1) a. The child stuck the balloon with the pin.
    b. The balloon burst.

If one encountered (1a) in isolation, one could be reasonably certain that the balloon burst. Informed of (1b), on the other hand, it seems possible to identify a variety of possible causes. This observation seemed to apply to many of the example sequences that we examined.

To summarize, the goal of this study was to determine whether people infer highly probable consequences. This question needed to be asked because, despite the importance of consequences, information processing limits may be presumed to restrict the amount of inference processing that may accompany reading. Three experiments were conducted. In all three, subjects read brief stories, and were then timed while they read and answered

several questions. In Experiment 1, the main comparison was between forward consequence inferences and paraphrases of explicitly stated consequences. Experiment 2 compared forward inferences with explicitly stated ideas. Finally, Experiment 3 compared forward and backward consequence inferences.

EXPERIMENTS 1 AND 2

The first two experiments were identical except for one important feature of the experimental materials. For this reason, Experiments 1 and 2 will be treated together.

Experiments 1 and 2 were designed to determine whether people infer highly probable consequences during the course of reading, or whether such inferences are delayed until a later time. For example, when one reads *the document was thrown in the fire*, does one immediately and reliably infer that the document burned up? To reiterate, a forward inference is one which might almost certainly be true, but which does not contribute to the coherence of the message.

In Experiments 1 and 2, the subjects read stories that were nine sentences long. Four of the sentences in each story stated or implied specific consequences. After reading the stories, the subjects answered questions about these consequences.

Sentence set (2) illustrates the experimental materials of Experiments 1 and 2. Across the two experiments, all subjects answered questions like (2d) after reading either (2a), (2b), or (2c) in a story. In relation to (2a), question (2d) asks about the explicitly stated consequence that the report was burned. After (2b), (2d) asks about a consequence which is paraphrased by substituting *burn* for its hyponym *incinerate*. Finally, in relation to (2c), (2d) asks about a (forward) implied consequence.

(2) a. The spy burned the report in the fire. (explicit antecedent)
b. The spy incinerated the report in the fire. (synonym antecedent)
c. The spy threw the report into the fire. (inference antecedent)
d. Did the spy burn the report? (question)

Because of the constraints of the procedure, implied consequences were compared only with paraphrased consequences in Experiment 1, and only with explicit consequences in Experiment 2. Consider the possible outcomes of Experiment 1. It would seem that longer answer times for implied than for paraphrased inferences would suggest that the consequence inferences are not reliably drawn during comprehension. Equal answer times, on the other hand, would indicate that consequence inferences are computed and stored during reading. There is, however, an alternate explanation of the outcome of approximately equal answer times. Equal paraphrase and inference answer times might reflect the fact that the consequence inferences are not drawn during reading, but that the time needed to draw the inference at test time is about equal to the time needed to reconcile the question verb (e.g., *burn*) with its synonym (e.g., *incinerate*). This would be consistent with the results of Hayes-Roth and Hayes-Roth (1977), who found that subjects needed 270 milliseconds less to verify sentences identical to ones in antecedent messages than ones which replaced an original word with a synonym.

To permit a clearer interpretation of the results of Experiment 1, Experiment 2 compared implied and explicit consequences. Table 1 shows the possible outcomes of Experiments 1 and 2, and summarizes the conclusions permitted by the four patterns of results. Inference answer times (RTs) approximately equal to both synonym RTs (Experiment 1) and explicit RTs (Experiment 2) (top left-hand cell of Table 1) would strongly suggest that forward consequence inferences are drawn during comprehension. Inference RTs that exceeded both synonym and explicit RTs (bottom right-hand cell) would support the conclusion that consequence inferences are not drawn during comprehension.

If inference RTs exceeded explicit ones but were approximately equal to synonym RTs (top right-hand cell), the outcome would be ambiguous. It could mean that (1) consequence inferences *are* drawn during comprehension, and the only reason that inference RT exceeds

explicit RT is that the explicit questions include one extra word in common with the story; or (2) consequence inferences are *not* drawn during comprehension, and the only reason that inference RT is no longer than synonym RT is that it takes about as much time to reconcile a pair of synonyms (cf. Hayes-Roth & Hayes-Roth, 1977) as to draw a consequence inference. Finally, it was considered highly unlikely that the result pattern shown in the bottom left-hand cell of Table 1 would be obtained.

TABLE 1
CONCLUSIONS SUGGESTED BY THE POSSIBLE RESULT PATTERNS OF EXPERIMENTS 1 AND 2

Experiment 1 outcome	Experiment 2 outcome	
	RT(explicit) = RT(inference)	RT(explicit) < RT(inference)
RT(synonym) = RT(inference)	Consequence inferences drawn during comprehension	Ambiguous[a]
RT(synonym) < RT(inference)	Not considered possible	Consequence inferences not drawn during comprehension

[a]See text.

Method

Materials

Norms. A preliminary norming study was designed to identify a pool of events which have consequences agreed on by most people. Fifty-two naive subjects were asked to "write a brief description of the *most likely* outcome or consequence" of events described in 48 single sentences.

Each of the 48 stimulus sentences had a consequence that the experimenters believed many people would agree with. For the sentence *an egg falls on the floor*, the expected consequence concerned the breaking of the egg. Subjects were given credit for agreeing with this consequence if their answer (1) stated it directly, (2) replaced a crucial word with a synonym or hyponym (e.g., *the egg cracks or smashes*), or (3) expressed the expected consequence plus some additional ideas (e.g., *the egg breaks and makes a mess*). Using these criteria, subjects showed from 0 (*a country goes to war*) to 100% (*a report is thrown in a fire*) agreement with the experimenters' anticipated consequence. Most importantly, there was over 80% agreement for 23 of the 48 phrases and over 90% agreement for 18 of the 48 phrases, providing a reasonable pool from which to construct the experimental materials.

Experimental materials. The materials were four stories written for use in this study. Each story was nine sentences long. Of principal concern in each story were four sentences constructed from ones inspected in the norming procedure, and with which the norming

subjects expressed high degrees of agreement (mean = 92.6%). These experimental sentences occupied positions 2, 4, 6, and 8 in their stories. The odd numbered sentences were fillers.

Table 2 shows one of the stories used in Experiments 1 and 2. In Experiment 1, each experimental sentence could appear in its synonym form (second word in parentheses) or inference form (third word in parentheses). In Experiment 2, each experimental sentence could appear in its explicit or inference form (first or third parenthetical word, respectively).

For each story, eight questions were written. The four questions interrogating the experimental sentences were intended to be answered "yes," while the four that interrogated filler sentences were intended to be answered "no." The eight questions for each story were assigned to a single random order.

In Experiment 1, two tests lists were created. In List 1, two experimental sentences in each story were randomly assigned to the synonym condition and two to the inference condition. The assignments were reversed in List 2. Thus, across the two lists, each experimental sentence occurred once in each condition. The two lists of Experiment 2 were constructed simply by replacing synonym words in the Experiment 1 lists with their explicit counterparts.

The four stories were presented in a single random order to all of the subjects, namely: Trouble at the Factory, Bob the Spy, Carol's First Dinner, and Hunting the Dragon. In each list, the experimental stories were preceded by a practice story of identical form.

Subjects

The subjects in both experiments were male and female students of introductory psychology who participated for course credit. All subjects were native speakers of English. None

TABLE 2
SAMPLE STORY AND QUESTIONS FROM EXPERIMENTS 1 AND 2

Sentence/Question Number	Sentence/Question
S1	Bob the spy read a report by the fire.
S2	A rock (broke/smashed through/flew through)[a] the windowpane.
S3	Bob read a note attached to the rock.
S4	He then (burned/incinerated/threw) his report in(to) the fire.
S5	Next he called the airline.
S6	He (dissolved/melted/put) the coded sugar cube in a glass of water.
S7	Bob left and flew to a tropical resort.
S8	Bob (burned/fried/sat) all the next day in the sun.
S9	Bob knew he was not safe here.
Q1	Did Bob dissolve the sugar cube?
Q2	Did the windowpane break?
Q3	Did Bob call the train station?
Q4	Did the note arrive in an envelope?
Q5	Did Bob burn the report?
Q6	Did Bob fly to a ski resort?
Q7	Did Bob get sunburnt?
Q8	Was Bob reading a newspaper?

[a] Within each set of parentheses, the first word (or phrase) was used to form an explicit antecedent sentence in Experiment 2, the second word was used to form a synonym sentence in Experiment 1, and the third word was used to form inference sentences in both experiments.

of the subjects had participated in the norming study. There were 34 subjects in Experiment 1 and 35 subjects in Experiment 2.

Procedure

The sessions were conducted with groups of one to four individuals. Each subject sat in a separate room. The sentences were displayed on 12-in. video monitor screens, with the subjects seated 40 cm from the screens. The experimental events were controlled by a DEC PDP-8/A computer.

Each subject saw five stories, the first of which constituted practice. For each story, the subjects were asked to read each sentence carefully and then answer the questions appropriately. The sentences in the stories and the questions were displayed one by one. Before each story, a fixation point appeared on the screen for 3 seconds. After a 1-second delay, the nine sentences comprising the story appeared in succession, for a duration of 4 seconds apiece. A 4-second delay followed the story.

Next, the questions were presented. Each question was preceded by a fixation point for 3 seconds, and remained on the screen for 4 seconds. The subjects were asked to respond as quickly as possible without error, and were informed that answers registered after the question disappeared from the screen would be considered incorrect. The interquestion interval was 4 seconds.

The subjects responded "yes" with the index finger of one hand, and "no" with the index finger of the other. The subjects were asked to always keep their fingers resting lightly on the response switches to facilitate responding. Groups of subjects were assigned, on a random basis, to use either their left hand or right hand for the "yes" button. The responses and response times (measured in msecs) were recorded automatically by the computer.

Results

Experiment 1

Error rates of 16.6, 22.1, and 9.0% were measured for the synonym, inference, and filler questions, respectively. Only the response latencies for correct answers to synonym and inference questions were examined. These values are shown in Table 3.

Analyses of variance were applied to the scores in Table 3, treating either subjects (F_1) or sentences (F_2) as the random variable. With subjects random, test status (synonym versus inference) and story were "within" variables. With sentences-random, test status was a "within" variable and story was a "between" variable. An α criterion of 0.05 will be used throughout unless otherwise noted.

No effect reached significance in either analysis, although the subjects-random analysis revealed a main effect of story, $F_1(3,96) = 5.97, MS_e = 373{,}096$. The mean latencies for the stories were Factory, 2292 milliseconds; Spy, 2429 milliseconds; Dinner, 2098 milliseconds; and Dragon, 2354 milliseconds. The subjects-random analysis also revealed one significant interaction involving the list (1 versus 2) variable. The test status effect did not approach significance, $Fs < .5$, with means of 2272 and 2314 milliseconds for the synonym and inference conditions, respectively.

Experiment 2

The error rates were 11.1% for the explicit condition, 20.4% for the inference condition,

TABLE 3
MEAN CORRECT RESPONSE LATENCIES (MSEC) IN EXPERIMENTS 1 AND 2

	Experiment 1		Experiment 2	
Story	Synonym	Inference	Explicit	Inference
Factory	2330	2253	2183	2491
Spy	2370	2488	2182	2513
Dinner	2104	2093	1894	2157
Dragon	2287	2422	2036	2340
Mean	2272	2314	2074	2375

and 10.9% for filler items. Analyses of variance, identical in form to those of Experiment 1, were applied to the mean correct response latencies, shown in Table 3. Test status was significant, $F_1(1,33) = 31.9, MS_e = 199{,}620$, $F_2(1,33) = 28.0$, $MS_e = 49{,}888$, reflecting means of 2074 and 2375 for the explicit and inference conditions, respectively. Story was also significant, although only marginally so with sentences random, $F_1(3,99) = 6.10$, $MS_e = 216{,}282$, $F_2(3,12) = 3.01$, $MS_e = 60{,}850$, $p = .072$. The mean latencies for the stories were 2337, 2346, 2025, and 2187 for Factory, Spy, Dinner, and Dragon, respectively. There was also one significant interaction involving the list variable in the subjects-random analysis.

Comparison of Experiment 1 and Experiment 2

Although Experiments 1 and 2 were conducted separately, it would be desirable to directly compare the response latencies of the two experiments. In particular, it would be important to detect a significant Experiment × Test Status interaction, reflecting the fact that the synonym versus inference difference in Experiment 1 is smaller than the explicit versus inference difference in Experiment 2. Such an analysis was considered to pose difficulties, however, because the inference materials were identical in the two experiments while the "noninference" materials were not. Instead, an alternate analysis was chosen. Difference scores (between synonym and inference or explicit and inference) were computed for each of the test sentences in the two studies. In fact, the difference scores of Experiment 2 significantly exceeded those of Experiment 1, $t(15) = 4.32$.

Correlation coefficients were computed to compare the story latencies of Experiments 1 and 2 with the mean number of words in the four experimental questions for each story. Experiment 1 latencies correlated .35 with mean question length, and Experiment 2 latencies correlated .82 with mean question length. Finally, the correlation between the mean story latencies of the two experiments was .62. These high correlation values suggest that the story effect was due to differences in the mean question length between the stories.

Discussion

Of the four result patterns shown in Table 1, the results of Experiments 1 and 2 clearly conform to the pattern shown in the top right-hand cell. That is, inference latencies were approximately equal to the synonym latencies in Experiment 1, but exceeded the explicit latencies in Experiment 2. As discussed earlier, this outcome is ambiguous with respect to the issue of whether forward consequence inferences are drawn during reading. Suppose that they are. In that case, the mean latencies for all four conditions in the two experiments should have been approximately equal. The fast explicit responses in Experiment 2 are inconsistent with this possibility. However, this inconsistency might be due to the fact that the explicit test questions had one more word in common with their antecedent stories than did any of the other test questions.

Suppose, on the other hand, that forward consequence inferences do not reliably accompany comprehension. Then it might be argued that the inference latencies should have exceeded both synonym and explicit latencies. This alternative is contradicted by the fact that responses to inference questions were not slower than synonym responses. This inconsistency may result from the fact that it may take as much time to reconcile a pair of synonyms (Hayes-Roth & Hayes-Roth, 1977) as to infer a consequence.

There is one aspect of the data that slightly favors the conclusion that forward consequence inferences are not reliably drawn at reading time. It has already been stated that if forward consequence inferences accompany reading, then the speed of responses to the explicit test questions might be attributed to the presence of the extra repeated word. The mechanism that would most likely account for this advantage is the priming of this word. For example, in the question *did Bob burn the report*, the extra repeated word "burn" is primed by "Bob" (Ratcliff & McKoon, 1978). Ratcliff and McKoon have measured priming

effects on the order of 110 milliseconds. This value, however, is less than half of the 246 millisecond difference that exists between the explicit latencies and the mean of the other three conditions. Thus, it might be suggested that it is more likely that the 246 millisecond difference reflects inference time and synonym reconciliation time than an encoding advantage.

Some other aspects of the results should be noted. It had seemed possible, given the complexity of the materials, that error rate would be very high. The mean experimental error rate over Experiments 1 and 2 was about 18%. This was somewhat higher than the rate of 10% measured by Singer (1980, Experiment 3), but not so high as to invalidate the answer time data. Furthermore, the pattern of slightly higher error rates for the longer latency conditions is consistent with that obtained in other studies of this sort, and is inconsistent with the possibility of a speed–accuracy tradeoff.

In summary, the results did not clearly resolve the issue of whether people reliably draw forward consequence inferences during reading. To further inspect this question, a third experiment compared forward and backward consequence inferences. Because backward inferences contribute to coherence, it is likely that they are drawn during reading. Accordingly, equal backward and forward inference latencies would suggest that forward consequence inferences also accompany comprehension; while faster backward than forward inference latencies would indicate that forward consequence inferences are frequently delayed till a later time.

Experiment 3

Method

Materials

The verbal materials were the four stories of Experiments 1 and 2. For each of the four experimental sentences in each story, a new sentence was written which could be inserted into the story. For example, the new sentence *the ashes floated up the chimney* could be inserted immediately after *he quickly threw his report in the fire*. The intended impact of the new sentence was to require a backward inference about the consequence in question, such as, in this case, that the report burned up. It is important to note that the inference-inducing sentences did not include the words expressing the consequence, such as *burn*.

The new lists were constructed from the lists of Experiment 2. Every experimental sentence formerly in the (forward) inference condition was left unchanged. Sentences formerly in the explicit condition were changed in two ways. (1) The "explicit" word (e.g., *burned*) was replaced by the corresponding word from the inference version (e.g., *threw*). (2) The new backward inference inducing sentence was inserted immediately after the former explicit sentence. These changes had two effects: first, every version of each story now included two forward inference sentences and two backward inference sentences; second, all stories were now eleven sentences long, including four experimental, five filler, and two backward inference inducing sentences.

As before, across the two lists, each experimental sentence occurred once in each condition, forward versus backward inference.[1] A few minor revisions of the stories were made to prevent them from sounding awkward. Each list again started with a practice story. One version of the Spy story is shown in Table 4.

Subjects and Procedure

The subjects were 27 men and women from the pool used in the other experiments. They participated either for course credit or a $3.00 payment. The procedure was identical to that of the previous experiments.

Results

Error rates of 6.5, 16.2, and 6.2% were detected for the backward inference, forward inference, and filler questions, respectively. The mean correct response latencies are shown in Table 5. Analysis of variance revealed a

[1] While it would have been desirable to compare explicit, forward inference, and backward inference questions in a single design, the complexity of the verbal materials made this prohibitive.

TABLE 4
ONE VERSION OF THE SPY STORY IN EXPERIMENT 3[a]

Sentence number	Sentence	Sentence function
S1	Bob the spy read a report by the fire.	Filler
S2	A rock flew through the windowpane.	Forward inference
S3	Bob read a note attached to the rock.	Filler
S4	He quickly threw his report in the fire.	Backward inference
S5	The ashes floated up the chimney.	Backward inference inducing
S6	Next he called the airline.	Filler
S7	He placed the coded sugar cube in water.	Backward inference
S8	He poured the clear liquid into the drain.	Backward inference inducing
S9	Bob left and flew to a tropical resort.	Filler
S10	He sat all the next day in the sun.	Forward inference
S11	But Bob knew he was not safe here.	Filler

[a] The questions for this story were identical to the ones shown in Table 2.

main effect of test status, $F_1(1,25) = 23.0$, $MS_e = 86849$, $F_2(1,12) = 19.8$, $MS_e = 19185$. The mean backward inference response latency was 2053 milliseconds, as compared with 2271 milliseconds for the forward inference condition. As in Experiment 2, a marginal story effect was detected, $F_1(3,75) = 9.41$, $MS_e = 192755$, $F_2(3,12) = 2.53$, $MS_e = 98167$, $p = .106$. There was also one significant interaction involving the list variable.

Discussion

The outcome is consistent with the view that people draw backward inferences more reliably than forward consequence inferences during reading. Answer latencies were 218 milliseconds faster for backward inference questions, even though these questions did not have the encoding facilitation advantage of the explicit questions of Experiment 2. The effect was consistent across all four stories examined.

The mean latency for the backward inference questions was 2053 milliseconds, which was actually slightly lower than the mean of 2087 milliseconds for the explicit questions of Experiment 2. Needless to say, comparisons between experiments need to be made very cautiously. However, the questions asked in the two experiments were identical, and the subjects were selected from the same pool. The similarity of the explicit and backward inference means is in close agreement with the results of Singer (1980, Experiment 3), who measured means of 1884 and 1897 milliseconds for explicit and backward inference test items concerning case-filling concepts.

Thus, a comparison of the results of the three experiments indicates that, despite the importance of the cause–consequence relation in prose, people do not reliably draw forward consequence inferences of the sort examined here. Backward consequence inferences, in contrast, appear to be drawn in the course of comprehension.

GENERAL DISCUSSION

In his study of inference, Singer (1980) concluded that readers do not reliably draw

TABLE 5
MEAN CORRECT RESPONSE LATENCIES (MSEC) IN EXPERIMENT 3

Story	Backward inference	Forward inference
Factory	2250	2314
Spy	2173	2469
Dinner	1775	2099
Dragon	2015	2203
Mean	2053	2271

forward case-filling inferences while reading. The present study was considered necessary because physical causes and consequences are more crucial to the gist of a text. The results, however, clearly discourage the conclusion that readers draw forward inferences about event consequences. This outcome might be viewed as conflicting with the fact that people incorrectly report recognizing test sentences that express the implications of antecedent messages (Johnson, Bransford, & Solomon, 1973). Johnson et al., however, acknowledged that their results were consistent with both an inference-during-reading and an inference-at-test interpretation.

It is important to note that many inferences that do not accompany comprehension can, of course, be drawn at a later time. These inferences are typically based upon the reader's knowledge of concepts, situations, and text types. As Singer (1979) and Trabasso and Nicholas (1980) point out, however, it is important not to confuse the ability of the reader to make a correct judgment about an implied idea with the question of whether an inference has accompanied reading.

Because it might be argued that human goals and motives are even more important in a story than event causation, one might be tempted to predict that readers reliably draw forward inferences about goals and motives. Such a possibility, however, seems unlikely to be true, in the light of the present evidence. Downey (Note 1, 1979), in fact, has reported that people need less time to verify explicitly stated motives than forward implied ones, an outcome which is consistent with the results of this study.

Existing evidence thus indicates that forward inference processing during reading is considerably restricted. It is likely that these restrictions stem from human information processing limitations. Accordingly, one of the main goals of inference research should be the identification and inspection of factors that guide the drawing of both forward and backward inferences during comprehension. Above and beyond the coherence requirements inspected here, several factors of this sort have already received some attention. (1) It is quite likely that the *importance* and/or *thematic relatedness* of an idea influences inference processing. Walker and Meyer (1980), for example, found that people were more likely to draw inferences about ideas that were high rather than low in the content structure of a text. This "height" variable might be taken as an index of thematic relevance. Although not focusing on inference, Graesser et al. (1980) reported that ideas that were connected with relatively large numbers of other ideas in a text were recalled more reliably. This variable, "relational density," might be viewed as another measure of the importance of an idea. (2) Hayes-Roth and Thorndyke (1979) have provided evidence that the *distance* between two propositions in a text affects the likelihood that a reader will infer the relation between them (see also Walker & Meyer, 1980). (3) It is reasonable to speculate that people draw more inferences about *interesting* than uninteresting ideas in a text (Kintsch, 1980; Schank, Note 2, 1978). (4) There may be certain types of relations and/or rhetorical devices concerning which people draw forward inferences. As mentioned earlier, it is not inconceivable that readers draw forward inferences about human goals and motives in story comprehension. People also usually understand the indirect speech act conveyed by a statement like *it's cold in here*, uttered when the speaker would like the listener to close a window. The identification of an indirect speech act might be considered a form of forward inference.

One aspect of the results of Experiment 1 merits further comment. It was found that answer times were about equal for paraphrase and forward inference questions. The present interpretation of this outcome is that it takes a moderate amount of time to reconcile the crucial word in the paraphrase question with its synonym or hyponym in the story representation. The position taken is that people do not semantically decompose words in a message unless absolutely required to so do. Accordingly, it is argued that the representation of *the man smashed the vase* is (SMASH,

MAN, VASE) and not (VIOLENTLY (BREAK), MAN, VASE). The strongest case for the nondecomposition view has been made by Hayes-Roth and Hayes-Roth (1977), who presented evidence that the representation of a message is specific to the words appearing in the message. Similarly, Kintsch and Bates (1977) reported that people have surprisingly accurate memories for the precise wording used in a lecture. The nondecomposition view is also an important element of the coherence graph comprehension model of Kintsch and Van Dijk (1978).

Finally, the clarity of the present results is noteworthy, given the complexity of stories and the multiple-question procedure. The present stories were 9 to 11 sentences long, compared with 3 in Singer's (1980) study. Each story was followed by a random sequence of eight questions, whereas Singer (1980) asked only one question after each passage. The length of the present stories raised the possibility of considerable forgetting, and the use of eight questions created the opportunity for interference among the questions. In spite of these possible pitfalls, the magnitudes of the significant effects were very similar to those measured by Singer (1980). Furthermore, the results were highly consistent. In Experiment 3, for example, the backward–forward difference was in the same direction for all four stories, and for 14 of the 16 experimental sentences. What this indicates is that the timed multiple-question procedure promises to function as a very effective technique for assessing the information that readers acquire from text.

REFERENCES

BLACK, J. B., & BERN, H. Causal inference and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, 1981, **20,** 267–275.

CLARK, H. H. Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

CORBETT, A. T., & Dosher, B. A. Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17,** 479–491.

GLENN, C. G. The role of episodic structure of story length in children's recall of simple stories. *Journal of Verbal Learning and Verbal Behavior*. 1978, **17,** 229–247.

GRAESSER, A. C., Robertson, S. P., & Anderson, P. A. Incorporating inferences in narrative representations: A study of how and why. *Cognitive Psychology*, 1981, **13,** 1–26.

GRAESSER, A. C., Robertson, S. P., Lovelace, E. R., & Swinehart, D. M. Answers to why-questions expose the organization of story plot and predict recall of actions. *Journal of Verbal Learning and Verbal Behavior*, 1980, **19,** 110–119.

HAYES-ROTH, B., & HAYES-ROTH, F. The prominence of lexical information in memory representations of meaning. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16,** 119–136.

HAYES-ROTH, B., & THORNDYKE, P. W. Integration of knowledge from text. *Journal of Verbal Learning and Verbal Behavior*, 1978, **18,** 91–108.

JOHNSON, M. K., BRANSFORD J. D., & SOLOMON, S. K. Memory for tacit implications of sentences. *Journal of Experimental Psychology*, 1973, **98,** 203–205.

JUST, M. A., & CARPENTER, P. A. Inferences processes during reading: Reflections from eye fixations. In J. Senders & R. Monty (Eds.), *Eyemovements and the higher psychological functions*. Hillsdale, N.J.: Erlbaum, 1978.

KINTSCH, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

KINTSCH, W. Learning from text, levels of comprehension, or: Why would anyone read a story, anyway. *Poetics*, 1980.

KINTSCH, W., & BATES, E. Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3,** 150–159.

KINTSCH, W., & VAN DIJK, T.A. Toward a model of text comprehension and production. *Psychological Review*, 1978, **85,** 363–394.

LEHNERT, W. Human and computational question answering. *Cognitive Science*, 1977, **1,** 47–73.

NEZWORSKI, T., STEIN, N. L., & TRABASSO, T. Story structure versus content in children's recall. *Journal of Verbal Learning and Verbal Behavior*, 1982, **21,** 196–206.

RATCLIFF, R., & MCKOON, G. Priming in item recognition. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17,** 403–417.

SCHANK, R., & ABELSON, R. *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum, 1977.

SINGER, M. Processes of inference in sentence encoding. *Memory & Cognition*, 1979, **7,** 192–200.

SINGER, M. The role of case-filling inferences in the coherence of brief passages. *Discourse Processes*, 1980, **3,** 185–201.

THORNDYKE, P. W. The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15,** 437–446.

TRABASSO, T. On the making of inferences and their as-

sessment. In J. T. Guthrie (Ed.), *Reading comprehension and education*. Newark, Del.: International Reading Association, 1980.

TRABASSO, T., & NICHOLAS, D. W. Memory and inferences in the comprehension of narratives. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children*. Hillsdale, N.J.: Erlbaum, 1980.

TRABASSO, T., SECCO, T., & VAN DEN BROEK, P. Causal cohesion and story coherence. In H. Mandl, N. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, N.J.: Erlbaum, in press.

WALKER, C. H., AND MEYER, B. J. F. Integrating different types of information in text. *Journal of Verbal Learning and Verbal Behavior*, 1980, **19**, 263–275.

WILKS, Y. What sort of taxonomy of causation do we need for language understanding? *Cognitive Science*, 1977, **1**, 235–264.

REFERENCE NOTES

1. DOWNEY, D. M. *The role of motivational inferences in story comprehension*. Unpublished manuscript, University of Colorado, Department of Communication Disorders and Speech Science, 1979.
2. SCHANK, R. C. *Interestingness: Controlling inferences*. Research report 145, Yale University, Department of Computer Science, 1978.

(Received June 4, 1982)