Effects of Discourse Context on Inference Computation during Text Comprehension

by

Yasunori Morishima

and

Walter Kintsch

Department of Psychology
and
Institute of Cognitive Science
University of Colorado, Boulder

RUNNING HEAD: Discourse Context Effects on Inferences

Send correspondence to:
Yasunori Morishima
10251-A Rancho Place
Cupertino, CA 95014
(408) 366-0545 (home)
(650) 578-8067 (work)
(650) 572-7108 (fax)
ymorishima@dyned.com

Abstract

Comprehenders construct a situation model of a text that they are processing. The situation model is a microworld of agents, objects, actions, events, spatial composition, and states associated with what the text explicitly asserts. The current study asks how a situation model influences inference computation during comprehension. The primary question is how strongly a comprehender infers a high-probability instrument (e.g., a pump for inflating a tire) during comprehension given a brief text. The experiments showed that instrument inferences are highly affected by the discourse context; such inferences are not computed unless the context provides information that strongly supports the inferences (e.g., "John grabbed the pump in the garage."). When the context is compatible with such inferences but weak (e.g., "John found the pump in the garage."), comprehenders do not routinely draw the inference. However, comprehenders do infer instruments when they are motivated to elaborate during comprehension. A simulation model was developed based on Kintsch's (1988) Construction-Integration theory. The model successfully displayed behaviors that are qualitatively in agreement with the experimental results and provides an explicit account for the on-line inference computation process.

1. INTRODUCTION

Consider the following sentence pairs:

1.1 Marvin turned on the light in the shed. He inflated the tire.

1.2 Marvin found the broken pump in the shed. He inflated the tire.

1.3 Marvin grabbed the pump in the shed. He inflated the tire.

Would comprehenders infer that Marvin used the pump to inflate the tire on-line, that is, during comprehension? The first sentences in these pairs set up the situation for the actions described by the second sentences. Let us call the first sentences context sentences and the second ones action sentences. While the action sentence in each pair is the same, each context sentence implies a different situation about availability of the instrument "pump". The context sentence for Text 1.1 does not mention "pump" at all. Text 1.2 implies that the pump was not usable since it was broken, thus it does not lead to the expected inference. In other words, it would not make sense to assume that Marvin used the broken pump to inflate the tire. Text 1.3, on the contrary, suggests that he took the pump and intuitively supports the inference. How is the inference process, or specifically the activation of the high-probability instrument, affected by the context sentences? This is the primary question that we address in this article.

Instrument inference has been studied since the early days of inference research. Although early studies (e.g., Johnson, Bransford, & Solomon, 1973; Paris & Lindauer, 1976) claimed that an implicit instrument is encoded in reading a sentence such as *The lawyer cooked dinner* (the instrument is a *stove*), there is considerable evidence today that an instrument is inferred only when certain conditions are met. An explanation for it is that such an inference, called an elaborative inference, only provides additional information rather than establishing local coherence of a text. Inferences that performs the latter function are called bridging inferences. It is generally agreed in the literature that bridging inferences normally accompany comprehension (Corbett, 1984; Corbett & Chang, 1983; Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; Graesser, & Kreuz, 1993; Haviland & Clark, 1974; Kintsch & Vipond, 1978; Lesgold, Roth, & Curtis, 1979; McKoon & Ratcliff, 1980, 1986) whereas elaborative inferences are not routinely computed on-line unless the context of a text is highly constraining or are encoded into the representation only "minimally" (Anderson & Ortony, 1975; Fincher-Kiefer, 1995; Keefe & McDaniel, 1993; Mauner, Tanenhaus, & Carlson, 1995; McKoon & Ratcliff, 1986, 1989; Murray, Klin, & Myers, 1993; O'Brien, Shank, Myers, & Rayner, 1988; Potts, Keenan, & Golding, 1988; Singer & Ferreira, 1983; Whitney, Ritchie, & Crane, 1992).

Several studies show that unless an instrument has been introduced in the text as in Text 1.1, the instrument inference would not be drawn on-line, and the instrument would not be activated (Corbett & Dosher, 1978; Dosher & Corbett, 1982; Singer, 1979, 1980). McKoon and Ratcliff (1981) showed that the instrument is activated at the time of the reading a sentence mentioning the action if the action is highly associated with the instrument. The researchers used texts as the following:

1.4. Bobby got a saw, hammer, screwdriver and square from his toolbox.

1.5a. Then Bobby pounded the boards together with nails.

1.5b. Then Bobby stuck the boards together with glue.

The passages used in the experiments included a combination of Text 1.4 and either Text 1.5a or 1.5b with a few intervening sentences. They found that the reaction time for recognition of the test word "hammer" was shorter when Text 1.5a was presented than when Text 1.5b was presented. On the other hand, when the test word was "mallet" (in this case, "mallet" is included in Text 1.4), there was no significant difference in reaction time between the two versions. Lucas, Tanenhaus, and Carlson (1990), using the lexical decision task, obtained similar results. Namely, instrument inferences are drawn if previous discourse has explicitly mentioned the instrument, but not if sentences with actions implying the use of an instrument are presented out of context. Furthermore, van Meter and Pressley (1994) also found supporting evidence with 10- to 14-year-old children.

To summarize these studies, instrument inference takes place during comprehension under two conditions. The first condition is semantic association between an instrument and an action; namely, an action should have a strong association with a particular instrument (e.g., pounding -- hammer; driving -- car). The other condition has to do with the structure of the text; that is, an instrument should be explicitly mentioned in the text prior to the mention of an action.

However, in these studies, discourse context effects were not systematically investigated. Either the degree of relatedness between the context sentence and the action sentence was not controlled, or context effects were examined only in long-term memory representation rather than on-line processes (see McKoon & Ratcliff, 1981). Thus, the question arises here; are those conditions stated above sufficient for activating the instrument, or does discourse context influence the inference computation?

We argue that the global structure or situation model of text influences on-line computation of such inferences. In the present study, we focus on instrument inferences like the above examples. A situation model is a microworld of agents, objects, actions, events, spatial composition, and states associated with what the text explicitly asserts (Johnson-Laird, 1983; Kintsch, 1988; van Dijk & Kintsch, 1983). The view that a situation model plays an important role during comprehension has been put forward by several theories (Collins, Brown, & Larkins, 1980; Johnsonprobability instruments. Each experiment measured reaction times for instrument priming and reading times for processing critical sentences. Experiment 1 was conducted to establish that the experimental setting employed by the present study was sensitive enough to detect inference, independent of associative priming. In Experiment 2, we manipulated the degree of relatedness between the context sentence and the action sentence to examine its effects on inference computation. In Experiment 3, effects of depth of processing on inference computation was tested in effort to demonstrate evidence for a chain of inferences proposed above. In the second part, we present a computational model based on Kintsch's (1988) Construction-Integration Model. By developing the computational model, we were able to specify more explicitly the memory structure, control mechanisms, representation structure, and processes involved in inference computation.

2. EXPERIMENTS

In the present study, three experiments were conducted. Prior to the experiments, a series of pilot studies were conducted to validate empirically the experimenter's intuitions with regard to the qualities of the texts for the subsequent experiments.

Pilot Studies

Forty-nine sets of experimental texts like the example below were created with action-instrument associations based on Corbett and Dosher (1978), and Lucas et al. (1990). For each of the action sentences, four levels of context sentences were prepared and labeled as follows: Contradictory, Weakly Related, Moderately Related, and Strongly Related. The contradictory context was intended to provide the context that did not lead to the use of the instrument mentioned by the sentence. The three related context sentences were designed to provide the context in which the use of the instrument was plausible, but varied in terms of the degree of supporting the implication. As the labels suggest, the strongly related context intuitively supported the use of the instrument most strongly. The other related contexts, on the other hand, do not readily provide such support, and would require the longer chain of events in order to establish the availability of the instrument for use.

Context Sentences:

(Contradictory)	Marvin found the broken pump in the shed.
(Weakly Related)	Marvin searched for the pump in the shed.
(Moderately Related)	Marvin spotted the pump in the shed.
(Strongly Related)	Marvin grabbed the pump in the shed.

Action Sentence: He inflated the tire.

A series of rating and ranking experiments were conducted on each version of the texts (Myers, Shinjo, & Duffy, 1987). Undergraduates at the University of Colorado participated in the experiments for course credit. They were all native speakers of English. Participants rated each context sentence, using a 7-point rating scale, on how strongly it was related to the action sentence. Similarly, in the ranking experiments, they rank ordered the context sentences.

Twenty-four best texts were selected and analyzed. The mean rating scores and standard deviations for all context levels across texts are shown in Figure 1. An analysis of variance with items as a random factor was performed on the data. The analysis was based on mean rating scores for each text for each context level. There was a significant monotonic trend of the context level [$F(1, 23) = 1788.22$, $p < 0.001$]. Therefore, the results established that the context sentences were well separated and were ordered according to the hypothesized degree of relatedness to the action sentence.

Insert Figure 1 about here.

Experiment 1

The purpose of Experiment 1 was to establish that the present experimental paradigm was suitable for measuring priming effects due to inference, but not to word-based associative priming because, as discussed above, priming effects in inference experiments may be contaminated with the latter type of inference. The experiment employed self-paced reading and a lexical decision task. There were two dependent variables in the experiment. One is the reaction time for the lexical decision, and the other the reading time of the action sentence.

Method

Participants

Fifty-six undergraduates from the University of Colorado at Boulder participated in the experiment for credit in an introductory psychology course. All participants were native speakers of English.

Materials and Design

There were three levels of context in this experiment: Related, Contradictory, and Unrelated. The context was a within-participant independent variable. There were 18 experimental texts, which were randomly chosen from the set of the 24 experimental texts discussed above. Each text was assigned to one of the three context conditions for each participant. A sentence pair was composed of the context sentence followed by the action sentence. Shown below is an example of the text:

2.1a. Marvin grabbed the pump in the shed.

2.1b. Marvin found the broken pump in the shed.

2.2a. He inflated the tire.

2.2b. He flew to Boston.

The sentence pair for the related context consisted of Sentences 2.1a and the action sentence 2.2a. The contradictory condition contained Sentences 2.1b and 2.2a. The sentence pair for the unrelated control context was constructed by pairing Sentence 2.1a and the action sentence from another set (Sentence 2.2b). In this case, care was taken so that the gender/number agreement, and hence referential coherence, was maintained in the resulting sentence pair. The targets for the lexical decision were instrument words mentioned in the context sentence ("pump"). Note that in all conditions, the target word appeared in the context sentence to control for the repetition effect on the lexical decision task. There were 42 filler texts, 30 of which had non-word targets that provided "no" responses for the lexical decision task. For the other 12 sentence pairs, nouns other than the instrument words from the context sentences ("shed") were used as lexical decision targets. The experimental texts were randomly assigned to the conditions for each participant. The order of text presentation was randomized for each participant. The comprehension questions after the lexical decision task asked aspects of the texts other than the use of the instrument ("Did Marvin go into the shed?").

There were two levels (250 msec and 500 msec) of stimulus onset asynchrony (SOA), which refers to the length of delay time between the offset of the action sentence and the onset of the lexical decision. This was a between-participant variable.

There are three possible outcomes of this experiment. First, if the lexical decision is dominated only by the memory trace of the instrument word remaining in the short-term memory, there should not be any difference due to the context condition. Second, if the effect is due to semantic association, both Related and Contradictory contexts should produce a priming effect because their context sentences explicitly mention the instrument, followed by the same action sentence. Finally, if the priming effect is due to the inference about the instrument, only the related context condition should produce the effect.

Apparatus

A Macintosh Quadra 605 was used for the experiment. A program was written in C to control the experiment.

Procedure

Participants were randomly assigned to either of the two SOA conditions. They were tested individually. First the participants were given the instructions on the computer screen. Then, the instructions were further explained by the experimenter. At the beginning of each trial, a string of plus signs with the same length of the context sentence appeared on the computer screen, which was soon replaced by the context sentence. The participants read the context sentence and hit the designated key called the "next" key (the down-arrow key on the Macintosh extended keyboard) to request the action sentence. Then the context sentence was replaced by the action sentence. The participants read the sentences at their own pace. When the participant hit the “next” key after reading the action sentence, a plus sign appeared in the center of the screen for either 250 msec or 500 msec, and it was replaced by the target, to which the participants were instructed to respond by pressing either the "yes" key (the right-arrow key) or the "no" key (the left-arrow key). Then, the participants responded to a yes/no question which was presented on the screen after the lexical decision task. The participants used only an index finger to make a response. They were instructed to place their index finger on the "next" key while reading and move it when they made a response. They were also instructed to make a response as quickly and accurately as possible. Reaction time and accuracy for lexical decision were measured. Reading time of the action sentence and accuracy for the comprehension were also measured. There was a practice run with 10 trials before the test run. In the practice run, the participants were given feedback about their reaction times for the lexical decision task and responses to yes/no questions. In the test run, no feedback was given to the participant. The first 4 trials in the test run were not experimental trials. The participants were encouraged to take a short break after a trial whenever they felt tired. The entire experiment took approximately 30 minutes for each participant.

Results and Discussion

Reaction Times for Lexical Decision on Instruments.

There were 26 participants assigned to the 250 msec SOA condition and 30 participants in the 500 msec SOA condition. Errors occurred either when a participant made an incorrect lexical decision or failed to respond within 2 sec. They were eliminated from the analyses. Reaction times greater or less than 3.0 standard deviations from the participant mean were replaced by this cut-off value. Two participants were eliminated from the analyses due to high error rates. Mean reaction times for all conditions are reported in Table 1. This treatment was performed for the other experiments.

Insert Table 1 about here.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reaction times for each participant or item in each condition. They are reported here as *Fp* and *Fi* respectively. The main effect of SOA was not significant by the participant analysis [$Fp\ (1, 53) = 0.21$, $p > 0.6$], but was significant by the item analysis [$Fi\ (1, 17) = 8.23$, $p < 0.02$]. However, the SOA was not shown to have any influence on the other effects, as no interactions with other variables were significant by either analysis [$Fs < 1.5$].

The mean reaction time for the related condition was significantly faster than those for the contradictory and unrelated conditions by the participant analysis [$Fp\ (1, 53) = 12.66$, $p < 0.001$]. This effect was marginally significant by the item analyses [$Fi\ (1, 17) = 3.72$, $0.07 < p < 0.071$]. Planned pair-wise comparisons of the related context with the unrelated and the contradictory conditions were also performed. The reaction time for the related context was significantly faster than each of these context conditions by the participant analysis [$Fp\ (1, 53) = 9.85$, $p < 0.003$; $Fp\ (1, 53) = 10.16$, $p < 0.003$]. The item analyses showed that this effect was marginally significant with respect to the unrelated context [$Fi\ (1, 17) = 3.72$, $0.08 < p < 0.081$], but it did not reach significance with respect to the contradictory context [$Fi\ (1, 17) = 2.58$, $p > 0.1$]. The contradictory condition did not differ significantly from the unrelated condition by either analysis [$Fp\ (1, 53) = 0.03$, $p > 0.8$, $Fi\ (1, 17) = 0.03$, $p > 0.8$]. Strictly speaking, the priming effect for the related context condition did not reach statistical significance in the item analyses, thus we note that we should be cautious about generalizing the effect over various texts. Yet, considering the fact that the item analysis did not have as much power as the participant analysis and that the effect was close to significance, it does not seem unreasonable to assume that the effect would be found with different texts.

The results showed that there was instrument priming after reading the action sentence if the discourse context provided the information that supported the inference, but there was no such priming if the context implied that the instrument was not available for use. Since the context sentence contained the instrument word in both the related and contradictory conditions, if the priming effect were due to semantic association between the action and the instrument alone, both conditions should have produced the priming effect. Thus, the results suggest that the instrument priming observed in the experiment was due to the instrument inference, not an associative priming.

Reading Times of Action Sentence.

The reading time data for the action sentence obtained from the 54 participants were analyzed. Reading times greater than 5 sec were eliminated from the analyses. In addition, reading times greater than 3 standard deviations from the participant mean were replaced by this cut-off value. This treatment was performed for the other experiments. The mean reading time for each condition is shown in Table 2.

Insert Table 2 about here.

Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reading times for each participant or item in each condition. They are reported here as *Fp* and *Fi* respectively. The main effect of SOA was not significant by the participant analysis [$Fp\ (1, 53) = 0.96, p > 0.3$], but was significant by the item analysis [$Fi\ (1, 17) = 18.3, p < 0.001$]. However, no interactions with other variables were found to be significant by either analysis [$Fs < 0.29$].

The mean reading time for the related condition was significantly faster than those of the contradictory and unrelated conditions by both analyses [$Fp\ (1, 53) = 25.33, p < 0.001$, $Fi\ (1, 17) = 18.30, p < 0.001$]. Planned paired comparisons of the related context with the unrelated and the contradictory conditions were also performed. The reading time for the related context was significantly faster than each of these context conditions by the participant analysis [$Fp\ (1, 53) = 12.82, p < 0.001$; $Fp\ (1, 53) = 22.8, p < 0.001$]. The item analyses also showed that this effect was significant with respect to both contexts [$Fi\ (1, 17) = 6.74, p < 0.02$; $Fi\ (1, 17) = 20.32, p < 0.001$]. The contradictory condition did not differ significantly from the unrelated condition by either analysis [$Fp\ (1, 53) = 0.24, p > 0.6$, $Fi\ (1, 17) = 0.11, p > 0.7$].

The reading time data showed the same pattern of results as the reaction time data. If the event described by the action sentence did not fit the context and thus was not coherent at the discourse level (e.g., pounding the board in the context of setting up the projector in the lab or finding the broken hammer), the participants took a longer time to process the action sentence. This outcome suggests that the participants tried to elaborate on the text to establish global coherence of the text. Note that the referential coherence was maintained for all the contexts by the use of the proper pronoun for the agent in the action sentence. Yet, the unrelated and contradictory conditions showed significantly longer reading times than the related condition. If coreference or argument overlap is the major source of coherence that comprehenders attempt to maintain during comprehension, there should be no difference in reading time among the three conditions.

Experiment 1 successfully established that the present method is sensitive to inference but not to word-based associative priming. The reaction time data and the reading time data provided converging evidence for this conclusion. These results give support to the hypothesis that global structure of text affects on-line processes, in particular that computation of instrument inference depends on discourse context. However, it is not clear how much elaboration comprehenders would make to draw instrument inference because in Experiment 1 the context sentences in the related condition were not only consistent with the instrument inference but also strongly related to the action. The next experiment was conducted to address this issue.

Experiment 2

Experiment 2 tested the hypothesis concerning the construction of an inference path leading to instrument inference. The experiment tested whether an instrument inference would be computed even if the discourse context calls for an elaboration to infer an instrument.

Method

Participants.

The participants were 36 native speakers of English in the Boulder community, who were paid for their participation.

Materials and Procedure.

Experiment 2 was identical with Experiment 1 for the most part. The only difference was the context condition. In this experiment, there were four levels of context: Unrelated, Weakly Related, Moderately Related, and Strongly Related. An example of the text is shown below. The sentence pair for the unrelated context was constructed in the same manner as in Experiment 1. The three related contexts differ in the degree of relatedness to the action sentence as discussed in the pilot study section.

Context Sentences:

(Weak) Marvin searched for the pump in the shed.

(Moderate) Marvin spotted the pump in the shed.

(Strong) Marvin grabbed the pump in the shed.

Action Sentence: He inflated the tire.

There were a total of 24 experimental texts, thus 6 texts for each context. There were 36 filler texts, 30 of which had non-words as targets to provide "no" responses, and the other 6 texts were positive fillers that had words from the context sentence other than instrument words as targets. The SOA of 500 msec was used in this experiment because no interactions of SOA with any other effects were found in Experiment 1.

Results and Discussion

Mean reaction times for the lexical decision task and mean reading times for the action sentences are presented in Table 3.

Insert Table 3 about here.

Reaction Times for Lexical Decision on Instruments.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reaction times for each participant or item in each condition. The mean reaction time for the strongly related condition was significantly faster than those for the other context conditions [$Fp\ (1, 35) = 24.47, p < 0.001, Fi\ (1, 23) = 22.12, p < 0.001$]. Planned pair-wise comparisons of the strong context condition were also performed. The mean reaction time for the strong context was significantly faster than each of the other context conditions by both analyses [against Unrelated, $Fp\ (1, 35) = 19.56, p < 0.001, Fi\ (1, 23) = 26.46, p < 0.001$; against Weak, $Fp\ (1, 35) = 13.47, p < 0.001, Fi\ (1, 23) = 7.33, p < 0.001$; against Moderate, $Fp\ (1, 35) = 12.32, p < 0.002, Fi\ (1, 23) = 5.73, p < 0.003$]. Though the mean reaction time for the unrelated context was approximately 30 msec slower than those for the weak and moderate contexts combined, this difference was not significant by either analysis [Fp (1, 35) = 2.05, $p > 0.16$, Fi (1, 23) = 0.70, $p > 0.41$]. This means that there was no significant instrument priming for either of the weaker contexts. It was also found that the mean reaction times for the three related contexts were, on average, significantly faster than the mean reaction time for the unrelated context [Fp (1, 35) = 6.87, $p < 0.02$, Fi (1, 23) = 4.91, $p < 0.04$]. This result replicated what had been found in the literature (e.g., Lucas, et. al., 1991; McKoon & Ratcliff, 1981). Previous studies did not control for the degree of relatedness between the context and action sentences and obtained the evidence that supported instrument priming. What this analysis suggests is that the instrument priming found in the previous studies may in fact be due to a particular type of texts, namely the texts for the strongly related condition.

The reaction time data indicates that the instrument inference was computed only when the discourse context provides information that strongly supports such an inference. Thus, the result leads to the interpretation that comprehenders do not routinely engage in an extensive elaboration to construct a long inference path leading to the instrument inference. In fact, in light of the fact that only the strongly related condition yielded the instrument priming, comprehenders integrate only a minimum amount of knowledge during comprehension.

Reading Times of Action Sentence.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reading times for each participant or item in each condition. The mean reading times for the three related contexts were, on average, significantly faster than that for the unrelated context [Fp (1, 35) = 38.41, $p < 0.001$, Fi (1, 23) = 9.94, $p < 0.005$]. The reading time for the strongly related context was approximately 64 msec. faster than those for the other related contexts. This difference was marginally significant in the participant analysis [Fp (1, 35) = 3.22, $0.081 < p < 0.082$], but did not reach significance in the item analysis [Fi (1, 23) = 1.85, $p > 0.18$]. Planned pair-wise comparisons of these related contexts with the unrelated context were also performed. Each of these conditions yielded significantly faster reaction times than the unrelated condition by both participant and item analyses [for Weak, Fp (1, 35) = 15.19, $p < 0.001$, Fi (1, 23) = 6.23 $p < 0.03$; for Moderate, Fp (1, 35) = 26.33, $p < 0.001$, Fi (1, 23) = 7.06, $p < 0.02$; for Strong, Fp (1, 35) = 61.24, $p < 0.001$, Fi (1, 23) = 12.26, $p < 0.002$].

As in Experiment 1, the reading time data showed the coherence effect. However, this disadvantage in processing time was found only for the unrelated and contradictory conditions. In Experiment 2, the mean reading times for the weaker contexts were significantly faster than that for the unrelated condition and were not very different from the reading time for the strong context condition. This suggests that whereas the unrelated and contradictory contexts build a situation model with which the subsequent sentence cannot be integrated, the weaker contexts, though not strong enough to generate the inference, are good enough to set up a situation model that is compatible with the subsequent action sentence. The faster reading time for these contexts also suggests that the texts for the weak and moderate conditions were processed in a very similar manner as those for the strong condition.

Why did only the strongly related context yield the instrument priming? One explanation is that the strong context does not require deep processing to generate the instrument inference because information or knowledge necessary for the inference can be retrieved by the text easily. Namely, 'grabbing the pump' also entails 'having it', and this information and the association between 'pump' and 'inflating (a tire)' in the situation model make it easier to draw the inference. On the other hand, the weaker contexts do require more elaborations because such information is not easily retrieved by the text elements. This explanation leads to the hypothesis that if the participant does engage in deeper processing, these weaker contexts should also show the instrument inference. Experiment 3 tested this hypothesis.

Experiment 3

Experiment 2 found instrument priming only for the strongly related context. The reading time data suggested that there was not a significant difference among the three related conditions in terms of processing time of the action sentence. It is hypothesized that the comprehender devotes only a limited amount of cognitive resource for on-line inference processing unless there is a reason for deeper and elaborative processing. If, on the other hand, the comprehender engages in the deeper processing that activates and integrates more general knowledge, instrument inference should be drawn. This experiment tested this hypothesis by having the participant engage in a strategic elaborative reading. It was assumed that if the hypothesis holds and participants indeed engage in deeper processing of the text, the instrument would be inferred even for the weaker context.

Method

Participants.

The participants were 36 native speakers of English in the Boulder community, who were paid for their participation.

Materials, Design, and Procedure.

This experiment was almost identical to Experiment 2 in that it involved reading of a sentence pair followed by a lexical decision task and comprehension test. The present experiment had three levels of context: Contradictory, Moderately

Related, and Strongly Related. An example of the experimental texts is shown below:

Context Sentences:

(Contradictory)	Marvin found the broken pump in the shed.
(Moderate)	Marvin spotted the pump in the shed.
(Strong)	Marvin grabbed the pump in the shed.

Action Sentence: He inflated the tire.

The difference between the present experiment and the other two experiments is that the comprehension questions were about the instrument words mentioned in the context sentences. For instance, the question for the above example was *Did Marvin use the pump to inflate the tire?* The correct answers to the questions were to be positive for all but the contradictory context. The participant was instructed to pay particular attention to each comprehension question and feedback from the computer during the practice session. The rationale for this manipulation is that it is expected that after several trials, the participant would be led to the interpretation that the instrument was used to perform the action as long as the context sentence implied that that was plausible, thus resulting in drawing the instrument inference while reading the text.

There were 18 experimental texts, which were randomly chosen from the set of texts used for Experiment 2. Thus, each context condition had 6 texts for each participant. There were 30 filler texts, 24 of which had non-words as targets to provide "no" responses, and the other 6 texts were positive fillers. There were an equal number of texts that provided "yes" responses and "no" responses to the comprehension question. That is, 6 of the experimental texts assigned to the contradictory condition were for "no" responses while the other 12 texts were for "yes" responses. Thus, 18 filler texts were assigned to the contradictory condition to provide "no" responses to the comprehension question and 12 were assigned to the related context conditions to provide "yes" responses.

Results and Discussion

Table 4 shows mean reaction times for the lexical decision task, mean reading times for the action sentences, and mean response times for the comprehension questions.

Insert Table 4 about here.

Reaction Times for Lexical Decision on Instruments.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reaction times for each participant or item in each condition. The mean reaction times for the strongly related and moderately related conditions were, on average, significantly faster than the reaction time for the contradictory context condition in the participant analysis [$Fp(1, 35) = 8.09, p < 0.008$] and this difference was marginally significant in the item analysis [$F_i(1, 17) = 3.14, 0.09 < p < 0.095$]. The difference between the strongly related context and the moderately related context did not reach significance by either analysis [$Fp(1, 35) = 0.49, p > 0.48, F_i(1, 17) = 0.58, p > 0.45$].

Reading Times of Action Sentence.

Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reading times for each participant or item in each condition. Although the reading time for the contradictory context was slower by approximately 100 msec, this difference did not