THE EFFECTS OF CONTEXTUAL KNOWLEDGE ON DRAWING $3^{0/}$
INFERENCES FROM CONVERSATIONS

by

TONY MATTHEW DUBITSKY

B.S., State University College of New York at Oswego, 1977

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Psychology

KANSAS STATE UNIVERSITY
Manhattan, Kansas
1980

Approved by:

_____
Major Professor

TABLE OF CONTENTS

ACKNOWLEDGMENTS

LIST OF TABLES

INTRODUCTION

The following review of psychological research on inference-drawing includes a summary of representative research from the prose-processing literature, and some methodological notes on the use of reaction time as a dependent measure. In addition, relevant contributions to the understanding of conversations from the viewpoints of such diverse fields as psycholinguistics, sociolinguistics, artificial intelligence, and ethnomethodology are presented.

Inferences and Implications

Inference-drawing is a central part of natural language comprehension. Language comprehenders often go beyond the information given to infer, integrate, or otherwise fill in elements that are not explicitly stated.

The focus of this research is on implication, a pragmatic relationship between sentences such that one is derived from the other through either laws of logic or heuristic reasoning, the latter of which taps knowledge of the world in addition to knowledge about the linguistic input. By definition, what is implied by or inferred from a sentence is not in its surface structure. Consistent with the dictionary definition, an implication is speaker-based and an inference is hearer-based. X implies W if the speaker says X and the hearer infers W.

Linguistic Considerations

Definitions and examples. Hildyard & Olson (1978) have distinguished among propositional and pragmatic inferences. A propositional inference is a logically valid inference that necessarily follows from the implication and may be based on class inclusion (1a, b), an implicative verb (2a, b), or comparative terms (3a, b), (cf. Paris, 1978). For each of the aforementioned examples from Harris & Monaco (1978), the first sentence in each pair propositionally implies the second sentence, as shown in Table 1.

A pragmatic inference is an inference which is neither propositionally, i.e., necessarily, implied nor directly asserted but is strongly suggested by the interaction of the hearer's world knowledge with the linguistic input. Varieties of pragmatic inferences discussed by Harris & Monaco (1978) are illustrated below. These are instrumental case inferences (4a, b), speech act inferences (5a, b), locative inferences (6a, b), negative continuous adjective inferences (7a, b), temporal sequence inferences (8a, b), causal inferences (10a, b), and context inferences (11a, b). In this last type, the implication is not carried by any single grammatical constituent, but is a function of the interactions of the word meanings in a sentence (cf. Singer, 1977). For each sentence pair (4-11), the first sentence (a) pragmatically implies the second sentence (b), as shown in Table 1.

Harris & Monaco (1978) and Brewer (1977) demonstrated that the conjunction "but" can be used in a "denial-of-expectation" sense as a diagnostic tool for pragmatic, as opposed to propositional, implication. This "but-not" test is performed on a sentence pair consisting of a given sentence (12a), and its predicted implication (12b). A new sentence (12c) is generated by conjoining the previous sentence (12a) by "but" to (12d), the negation of its predicted implication. If this new sentence is semantically acceptable, it passes the test and thus (12b) is pragmatically implied by (12a). If the new sentence would have been contradictory, then the implication would have been propositional.

Psychological Considerations

The encoded meaning of a remembered sentence may not consist entirely of propositions or idea units of only the linguistic information in the sentence. What has been stored may be an amalgamation of some of the propositional content, along with inferences the comprehender has derived from implications of the sentence

Insofar as comprehension occurs through an interaction of the linguistic input and relevant previously stored world knowledge, the language comprehender may have understood from the sentence a concept that was not necessarily true, but was nevertheless consistent with or following from his/her knowledge of the world. Using a variety of dependent measures, studies have found that subjects often remember pragmatic implications of input sentences as well or better than the directly asserted meanings or their propositional implications.

In the paradigmatic sentence-memory experiment, subjects hear a list of unrelated sentences (e.g., 4a-11a in Table 1), and respond to either cued-recall or recognition memory tasks. The latter memory task is, more accurately, a recognition-of-information task in which subjects assign truth values to para- phrases of (a type of propositional inference), pragmatic implications of, contradictions of, and to sentences unrelated to the original input sentences.

Using a recognition-of-information task, Harris (1974) found that subjects remembered pragmatic implications and paraphrases in much the same way as directly stated facts. Brewer & Lichtenstein (1975) and Brewer (1977) replicated these findings in a cued-recall task as well, demonstrating that subjects were more likely to recall the pragmatic implications of sentences rather than the original input sentences. Recalled sentences were scored as pragmatic implications if they passed the but-not test. For example, upon hearing "The safecracker put the match to the fuse" and given the recall cue "safecracker", subjects would often recall the original input sentence as "The safecracker lit the fuse."

The aforementioned results were not obtained, however, in an immediate recognition-of-information task. Harris (1974) found that subjects performing the response task immediately after hearing the stimulus items more accurately discriminated between paraphrases and pragmatic implications, often rating the former as "true" and the latter as "indeterminate" in truth value. However, in

a series of training experiments (cf. Bruno & Harris, in press; Harris, 1977; Harris, Dubitsky & Thompson, 1979), the use of a highly obtrusive and interactive training session in which the propositional-pragmatic distinction was repeatedly emphasized helped subjects to discriminate between these types even at long delays.

<div align="center">Reaction-Time Methodology in Studying Inferences</div>

Reaction-time (RT) methodology has been used to investigate numerous issues about inferential processing, e.g., whether some inferences are computed before others due to a determined need in ongoing comprehension, whether some inferences are computed more quickly than others due to the amount of information that must be either retrieved or reconstructed, and whether inferences are computed at storage or retrieval.

Problems in studying inferences with RT methodology

Variations in materials. The numerous and diverse types of implications used in much research discourages comparison of findings. Although Clark (1973) recommended that "language" should be conceptualized as a random effect, he also noted that it wasn't always necessarily desirable. In some studies, investigators have examined the processing of a small number of specific lexical items in detail, as if they constituted the entire population of interest. Results have been erroneously interpreted as providing support for hypotheses concerning processing differences between general implication types. For example, in testing for differences between pragmatic and propositional implications, Singer (1977) included only "instrument" implications, and Chaffin (1978) used only a limited set of causitive verbs, such as "force", "cause", and "compel".

Experimental tasks. The two major tasks that have been used to examine latencies to respond to inference items are the recognition task (e.g., Kintsch, 1974; Singer, 1977), and the verification task (e.g., Glanzman & Pisoni, 1973; Springston & Clark, 1973; Singer, 1976, 1977; Chaffin, 1978; Keenan, 1978).

In both recognition and verification tasks, during study time the subject is presented with a target sentence or discourse in which some critical information is either pragmatically or propositionally implied. During test time, test statements are presented which are either pragmatic or propositional inferences implied by the antecedent target materials. In the recognition task, these pragmatic inference test statements are derived from target statements while direct assertion test statements are lifted verbatim from the target sentence or discourse. The task is to judge "new" or "old", i.e., whether each test statement in that exact wording had been seen before. In contrast, in the verification task, all test statements are paraphrases of the two types of test statements in the recognition task. In this case the task is to judge whether each test statement is true or false, based on the antecedent target information.

No objective criterion for "correct" task performance. In most other uses of RT methodology in human information processing research, such as the Sperling (1960) task, the Sternberg (1975) memory-scanning task, the Posner (1969) sequential-matching task, and the lexical-decision task (cf. Rubinstein, Garfield, & Milliken, 1970), what constitutes a "correct" response is generally agreed upon and may be readily verified. Recall, however, that the truth value of any pragmatic inference is probabilistic, continuous, relative, and logically indeterminate with respect to its corresponding implication. Whether a pragmatic inference is evaluated as "correct" or "incorrect" depends upon the available context and the criterion the hearer adopts in deciding the classification of true or false.

Inadequacy of binary response choice. In spite of this criterion problem, until recently only "yes-no", "correct-incorrect", "true-false", and "old-new" response choices have been available to subjects evaluating inferences in recognition and verification tasks. For example, Kintsch (1974) found that,

in a true-false verification task, statements in Table 1 like (13), a pragmatic inference from (14), were responded to as "false", although it is logically indeterminate with respect to (14). Singer (1976), using a verification task, cation task, found that a disproportionate number of pragmatic inferences were also responded to as "false". Harris & Monaco (1978) suggest that Singer's (1976) findings might have been more easily interpretable had a third "indeterminate" response category been included, reducing the task demand to answer "false" to less explicitly true items.

Potential task demands in the verification task. In the verification task, unlike the recognition task, it is assumed that normal comprehension processes are being tapped. This assumption may be unwarranted if evidence suggests that subjects are processing materials in non-language-like ways, i.e., as if they were participating in a problem-solving task rather than a language comprehension experiment.

For example, Chaffin (1978) found that subjects who were verifying propositional and pragmatic inferences were using one of two distinct strategies; they either read the entire number of words in both target and test sentences ("read all" strategy) or they read only the verbs in both sentences ("read only verb" strategy). For in this particular study, comprehension of the verbs in target and test materials was apparently sufficient processing to insure adequate task performance. However, an appropriate test of the experimental hypothesis under consideration required that subjects process all of the information in the target and test sentences at least minimally.

Some tentative solutions

Three methods designed to reduce the possibility of such task demands have been reported in the RT inference literature. What follows is a brief discussion of these methods.

Use of strategy as a blocking factor. In a questionnaire administered during pilot work, Chaffin (1978) found that subjects were employing one or the other of the two aforementioned *distinct strategies*. In subsequent data analyses, strategy type was treated as a blocking factor in the analysis of variance, resulting in a gain in statistical power and a reduction in error variance. Use of this technique presupposed that a finite number of systematic strategies can be identified, and that a given strategy is consistently used within a session, which may not be necessarily so.

Use of a variety of stimulus materials. The advantages of this technique are two-fold; in contrast to the studies mentioned in the section on variations in materials, generalizability is increased, and stimulus predictability is decreased. For example, in a series of experiments, Glanzman & Pisoni (1973) examined the verification of sentence pairs related by entailment, grammatical case, consequence of action, and presupposition. In a final experiment subjects verified all sentence-pair types, because earlier work incorporating sentence-pair type as a between-subjects factor suggested that subjects scanned the materials for key words while performing the verification task.

Use of a processing task between study and test. Subjects may be encouraged to process materials deeply if they are told in advance that a task requiring good comprehension of the stimulus materials will be administered between study and test.

Along these lines, Singer (1977) had subjects answer a simple elliptical question after each target stimulus was presented. Although no examples were included in the manuscript, the questions used were probably similar to those used by Bransford & Franks (1971), where an example sentence such as "The rock rolled down the mountain", is followed by the elliptical questions "Did what?" and "Where?"

Processing Differences Between Implication Types

It has been consistently demonstrated that longer latencies are associated with recognition and verification of pragmatic rather than propositional implications (cf. Kintsch, 1974; Singer, 1976, 1977, 1979; Springston & Clark, 1973; Chaffin, 1978).

For example, in a series of three experiments on inferential processes in reading, Kintsch (1974) in the study discussed above, examined latencies to verify a pragmatic inference as either true or false based on information in a two-sentence "paragraph." With respect to the sentence examples in Table 1, longer latencies were associated with (13) when it was preceded by (14) rather than by (15); (15) propositionally implies (13) while (14) pragmatically implies (13).

Consistent results were also obtained by Singer (1976) in studying inferential processing within a verification task. In a similar study involving instrument inferences, Singer (1977) found that verification latencies were ordered from shortest to longest for propositional inferences, pragmatic inferences, and false test statements, respectively.

However, these consistent results were obtained using procedures fraught with some of the methodological difficulties discussed previously. The present research addressed itself to each of these methodological cautions through the inclusion of numerous and diverse types of propositional and pragmatic inferences to minimize predictability of the test statements and allow for greater generalization of findings, and the use of a five-point response scale ranging from "definitely false" to "definitely true" instead of the customary two-point response scale. Furthermore, the realism of the experimental materials and the use of a summarization/title task should also help to suppress the use of problem-solving strategies in task performance.

Prose Research

One of the major goals of the present research was to examine the effects of contextual knowledge on comprehension but using conversations in order to extend the generalizability of results and achieve greater ecological validity. Since the effects of this variable have been observed previously using prose materials, it will be helpful to review some of the findings from this prose-processing literature.

The semantic representation of a linguistic input may vary depending upon the availability of a given context. Contextual knowledge may even accentuate some aspects of meaning and de-emphasize others. This is particularly true of thematic contextual knowledge.

Context-dependent facilitation and inhibition of recall. It has been repeatedly demonstrated that information consistent with the discourse theme is better recalled than theme-irrelevant information. Thus, a preceding context (variously operationalized as a preceding pictorial stimulus, thematic title or phrase, or instructions for the reader to approach the discourse with a certain purpose in mind) may simultaneously inhibit recall of theme-irrelevant information and facilitate recall of theme-relevant information. The contextual knowledge serves as an anchoring point around which new incoming information is interpreted and organized. Unrelated or irrelevant information is minimally interpreted.

In a series of demonstration experiments, Bransford & Johnson (1973) showed the importance of prior knowledge in conjunction with the linguistic input in determining comprehension. Prior knowledge was manipulated experimentally by varying the availability of a preceding context. In the paradigmatic experiment, a vague or ambiguous sentence or discourse is presented to subjects, with or without a preceding nonlinguistic context. Next, a recall test is administered

and comprehensibility ratings are completed. The number of idea-units recalled is scored for each protocol.

For example, in one experiment by Bransford & Johnson (1973), subjects read a passage describing some vague apparatus consisting of balloons, a string, electricity, and a guitar. Only with the addition of a preceding context (a coherent cartoon describing the situation) did it become clear that the passage, as reproduced below, was about a modern-day Rube Goldberg's attempt to woo a fair maiden living in a penthouse apartment.

If the balloons popped, the sound wouldn't be able to carry since everything would be too far away from the correct floor. A closed window would also prevent the sound from carrying, since most buildings tend to be well-insulated. Since the whole operation depends on a steady flow of electricity, a break in the middle of the wire would also cause problems. Of course, the fellow could shout, but the human voice is not loud enough to carry that far. An additional problem is that a string could break on the instrument. Then there would be no accompaniment to the message. It is clear that the best situation would involve less distance. Then there would be fewer potential problems. With face-to-face contact, the least number of things could go wrong.

In this same experiment, the number of idea-units recalled was significantly greater for those subjects shown the picture before hearing the passage than for those subjects shown the picture after the passage or not at all. These latter groups were not statistically significantly different in terms of the number of idea-units recalled. Bransford et al. (1973) used these data to reject the hypothesis that constructive processes (i.e., those involving the generation of inferences) occur at retrieval, favoring rather an encoding explanation.

However, Bransford et al. (1973) themselves found that parallel analyses of comprehensibility ratings replicated the analysis of the number of recalled idea-units across the various context placement conditions, showing that recall covaried with comprehensibility. Thus, the diminished recall in the context-after group may have been due to the subjects' inability to organize the information in passages which were essentially incomprehensible in isolation, without the benefit of a clarifying context. Nevertheless, other theorists (cf. Dooling & Mullet, 1973; Monaco, 1976) have also found support for the notion that constructive processes occur during storage, encoding, or comprehension, and not during retrieval.

Insofar as meaning emerges from interaction of the linguistic input and the extralinguistic knowledge that the language comprehender brings with him or her to the task, any disruption of these two factors may impede comprehension and recall. For example, this may occur if the apparent context is irrelevant to the linguistic input. In another experiment, Bransford et al. (1973) had subjects read a paragraph which was ambiguous with respect to topic; under disambiguating titles the two alternative readings were watching a peace march or going on a space trip.

A critical sentence relevant to one topic but inappropriate to the other was embedded in the paragraph and presented under both alternative titles. Not surprisingly, recall of idea-units from the critical sentence alone was greater in the relevant than in the irrelevant context conditions, since the information in the critical sentence was consistent with the organization of the passage induced by the relevant biasing title.

Schallert (1976) constructed three paragraphs which were ambiguous with respect to topic. For one paragraph, the alternative readings cued by disambiguating titles were "worries of a baseball team manager" and "worries of a

glassware factory manager." Multiple-choice recognition memory tests were administered to subjects after they had read the paragraphs under either one of the preceding contexts. Recognition test items included questions about unambiguous information irrelevant to either context, and information relevant only to one context or the other. Distractors for context-relevant questions included choices relevant to each one of the alternative contexts. It was found that alternative choices consistent with the presented context were chosen more often than those alternatives consistent with the other context. Although Schallert's (1976) use of two-way ambiguous paragraphs is novel, the major focus of the study was the joint effect of processing depth (manipulated through various orienting tasks) and contextual knowledge on prose recall and recognition memory.

Kozminsky (1977) extended and replicated Schallert's (1976) findings in the course of examining the recall of unambiguous narratives containing main and secondary themes. Each narrative was preceded by one of two titles cued to either the main or secondary theme of the passage. Consistent with the results of Schallert (1976), propositions relevant to the theme cued by the title were better recalled than theme-irrelevant propositions.

In a novel manipulation of contextual knowledge, Pichert & Anderson (1977) and Anderson & Pichert (1978) instructed subjects to read a passage about two schoolboys playing "hooky". Embedded in this passage was a description of an interior of a house. Their specific instructions were to read the passage from the perspective of either a real-estate agent or a burglar. Elements in the passage varied in importance or salience depending upon the perspective adopted by the comprehender; that there were valuable possessions in the house such as a color TV set or ten-speed bicycles might be of more interest to a burglar than would physical attributes of the house such as a leaky roof, attractive grounds, and a damp and musty basement, which would be noticed by a real-estate agent.

Informational elements relevant to the adopted perspectives were recalled better than irrelevant elements. Furthermore, a perspective shift yielded information previously unrecallable in the initial adopted perspective. Informational elements not recalled under one adopted perspective were recalled under the complementary adopted perspective. Pichert et al. (1977) argued for an inference-at-retrieval interpretation of their findings, since the perspective shift occurred after initial encoding.

In a related experiment where personal knowledge served as contextual information, Owens, Bower, & Black (1979) presented subjects with simple narratives about a main character which outlined such mundane activities as making a cup of coffee, visiting a doctor, attending a lecture, shopping in a grocery store, and attending a cocktail party. Four groups of subjects received differential prior information about this main character's motives, plans, problems, and goals. However, all subjects read the same skeletal narratives with only name changes for the main characters.

Specifically, before reading the passage, one group of "problem" subjects was told about Nancy, the main character who was worried about pregnancy after an illicit affair with her professor-lover. Subjects in the "Nancy control" group read the same narrative about Nancy with no prior information. Another group of "problem" subjects received prior information about Jack, who was trying to pass a chemistry course and gain enough weight in order to play on the football team. "Jack control" subjects read the narrative without any prior information.

Delayed recall tests of verbatim memory showed that "problem" subjects produced more motive-related intrusions than "control" subjects. A recognition memory test was also administered following the recall test and contained "neutral" trues and falses based on the actions of the main character, and character-appropriate and character-inappropriate inferences, based on the prior

contextual information. Test statements were rated on a scale ranging from 1 (old--absolutely certain that the statement was in the text) to 7 (new-- absolutely certain that the statement was in the text). Although problem and control subjects' ratings of neutral items were not significantly different, problem subjects rated character-appropriate inferences as significantly more likely to have appeared in the text than did control subjects. These results suggested that memory for a character's motives was highly influenced by contextual knowledge.

Research on Conversations

<u>Psycholinguistic Studies.</u>

Of the few studies using conversational materials, all have examined issues previously investigated using sentence stimuli. The three studies to be reviewed here focused on memory for certain aspects of conversation: memory for reference (Bates, Kintsch, & Masling, 1978), memory for pragmatic information (Keenan, MacWhinney, & Mayhew, 1977), and effects of contextual knowledge on memory (Chmielewski, 1978).

<u>Memory for reference in conversations</u>. Bates, Kintsch, & Masling (1978) assessed incidental memory for types of reference in dialogues videotaped from a daytime television soap opera episode. Following the soap opera episode, subjects were instructed to circle the items that they had actually heard in the program. The more explicit and informative items (i.e., definite descriptions) were retained better than the less explicit and informative ones (i.e., elliptical clauses) for verbatim memory measure but not for the gist memory measure.

<u>Pragmatic factors in memory for conversation</u>. In a closely related study, Keenan, MacWhinney, & Mayhew (1977) assessed subjects' incidental recognition memory for sentence-utterances transcribed from a spontaneous faculty research discussion. These sentence utterances varied in "interactional content", a dimension ostensibly reflecting the speaker's beliefs, intentions, and attitudes toward the listener, including the amount of wit, affect, sarcasm, and irony.

An incidental recognition memory test, administered about 30 hours after the discussion, consisted of low and high interactional content statements (sentences (16) and (17) respectively, as shown in Table 1), and distractors preserving the illocutionary force of the interactional content statements. For example, the distractor for an interactional content statement which was

a request for information was also a request for information. Subjects rated all items as either "old" or "new".

Meanings and surface forms of high interactional-content statements were recognized significantly better than those of low interactional-content statements. Additional control experiments ruled out alternative explanations of these memorability differences.

Interpretation of Bates et al. and Keenan et al. In contrast to studies employing lists of unrelated sentences which have demonstrated the forgetting of surface structure information and the memorability of semantic information (e.g., Sachs, 1967; Brewer, 1975), the central finding of both conversation studies described above was the memorability of surface structure information over time under certain conditions. In Bates et al. (1978), subtle stylistic changes in surface form were associated with varying levels of retention. Keenan et al. (1977) found significant levels of discrimination between high interactional-content target utterances and true paraphrases even 30 hours after the luncheon discussion. This paradox may be resolved by noting that surface form merely conveys semantic structure in a list of unrelated sentences presented in isolation. However, in the studies described here, where target utterances are extracted from an extended discourse in a natural context, surface forms may have functioned in Bates et al. (1978) to highlight information, and in Keenan et al. (1977) to communicate the speaker's relations with the hearer independent of the semantic content of the utterance. In summary, these two studies have underlined the psychological reality of pragmatic information, demonstrating that it is often manifested in surface forms, and is encoded, stored, and retrieved along with syntactic and semantic information.

Contextual knowledge and inferences from conversations. Chmielewski

(1978) examined recognition memory for statements from a simulated conversation

which was ambiguous with respect to topic; the dialogue was either about raising

children or cultivating plants.

Contextual cues were manipulated by asking videotaped actresses to enact

the same conversational script twice; the first time they would keep in mind

that the discourse topic was plants, the second time children, Presumably, the

actresses' nonverbal and intonational cues would vary across "children" and

"plant" conditions, but this was never empirically determined.

After viewing one of the two videotaped sequences, subjects completed a

recognition-memory test consisting of old test items lifted verbatim from the

script, and new test items which consisted of context-relevant, context-

irrelevant, and completely unrelated items. New test items were constructed

from old test items by substituting a specific noun phrase related to either

the plants topic or the children topic for the original demonstrative pronoun.

For example, as shown in Table 1, (18) is an old item; (19) is a "child"

context item; (20) is a "plant" context item; and (21) is an unrelated item.

Consistent with the results previously reviewed in the Prose Research section

on context effects, Chmielewski demonstrated that contextual knowledge affects

memory for conversational materials. Subjects falsely recognized statements

which included topic-appropriate references. For example, more "plant" items

were recognized than "children" items when subjects viewed the "plant" tape

and vice-versa for the "children" tape, i.e., subjects hearing (18) responded

to it as (20) more often than not, when viewing the "plant" tape.

Alternative Approaches to Conversations

Diverse basic research has come from such areas as ethnomethodology, artificial intelligence (AI), and sociolinguistics. Approaches in ethnomethodology and AI have been largely theoretical, rather than experimental. In contrast, approaches from sociolinguistics have been empirical in nature, lacking a theoretical framework. What follows is a brief description of some representative research within each area.

Ethnomethodological theory. Ethnomethodologists are interested in the organization of social interaction. The behaviors they study are everyday practical activities that are often "taken for granted", such as shaking hands or navigating a crowded street. They are interested in specifying the tacit knowledge required for successful performance of these behaviors. One implicit assumption is that conversational interaction is organized and structured. Accordingly, theorists have developed implicit rules and conditions on rules or have examined functions of various devices in maintaining conversational structure. The implicit rules formalized by theorists which govern such banal actions may often be as complex as transformational rules in generative grammar. Rules have been specified to account for the opening and closing of conversations (Schegloff & Sacks, 1973); sequences of turntaking in dialogues (Mohan, 1974); and the location of interruptions (Lycan, 1977).

Artificial intelligence (AI) research. AI researchers have applied conversational analysis to the computer simulation of natural language comprehension. Their major goal is to write programs which model the cognitive structures necessary for interactive communication.

In developing a model for rules and topics in man-machine conversational interaction, Schank (1977) used the notion of "script" to generate classes of probable machine responses to human input utterances.

To have a conversation about, for example, dining in a restaurant, the dialogue system would contain such "scriptal" information as types of "props" (i.e., tables, food); "roles" (i.e., customer, waiter); "entry conditions" (customer hungry, customer has money) and "results" (customer has less money and is not hungry, owner has more money), along with schematic outlines of prototypical events in the restaurant situation (i.e., how one goes about entering, ordering, eating, and exiting). These outlines may be instantiated to comprehend the large variety of individual restaurant experiences.

Schank's theorizing has had a great impact on psychological approaches to learning, memory, and comprehension. Recently, Bower, Black, & Turner (1979) reported a series of experiments testing the psychological reality of scripts in comprehending simple texts about stereotyped events. Not only was there consensus about what actions would be in the script, but also about the hierarchical structure of these actions. Subjects also used their script knowledge to infer unstated yet highly probable script actions, organize memory for scrambled instances of a script-based text, and to generate expectations about likely script events.

Sociolinguistic studies. Although there are few, if any, theories of conversational interaction in sociolinguistics, models have been put forth to be used as empirical tools in decomposing conversation into functional utterance classes and relating these to the speakers, discourse topics, content, or purpose of the communication. For example, Weiner & Goodenough (1977) developed a "games" model of conversational interaction where the unit of analysis is the "conversational move", a "speech act" type of unit applicable to both comprehension and production. Using a model similar to Weiner et al.'s, Mishler (1975, 1978) analyzed the distribution of different types of "moves" in a series of teacher-student classroom interactions. He found that these

distributions of utterance types were highly correlated with status relationships between students and teachers.

Overview of the various approaches. Sociolinguists and ethnomethodologists approach conversation as an organized social behavior, isolating rules and functional devices which contribute to its systematic structure. AI scientists and psycholinguists approach conversation as an organized linguistic behavior, often incorporating the notion of "scripts". Furthermore, where AI scientists are equally concerned with production and comprehension and developing cognitive structures and programs prerequisite to higher-order processing, psycholinguists emphasize comprehension and mental processes involved in going beyond the information given.

Present Research

The present research used a novel combination of a recognition-of-information task with conversational materials, filling a gap left by extant psycholinguistic research. In addition, the contextual manipulation to be employed here is a great improvement over the one of Chmielewski (1978). The major weakness of that study is that its internal validity is open to question; the contextual manipulation was never precisely operationally defined or even empirically verified. In the proposed research, subjects will summarize each dialogue under each one of the various context conditions in efforts to verify that the a priori interpretation had actually been computed.

This study may also be viewed as an extension of Owens' at al. (1979) work on the effects of knowledge of characters' motivations on memory. Information about characters and the situations they find themselves in is established through dialogues between two persons in the present study, rather than through short texts. In addition to recognition memory responses to neutral, inappropriate, and appropriate inferences, latency to respond will also be examined.

Finally, recognition memory will be assessed for propositional as well as pragmatic inferences from these dialogues. Previous studies have used well-formed "sentence-utterances" lifted verbatim from the discourse as target stimuli. The use of a five-point truth response scale with the "don't know" response option should be a considerable advance over the binary response scales previously used.

EXPERIMENT 1: NORMATIVE SCALING

Experiments 1 and 2 collected normative scaling data used to establish ratings of topical ambiguity for eight ostensibly ambiguous dialogues. Eight conversational dialogues were written, designed to be ambiguous with respect to topic. Experiment 1 was undertaken to determine the relative salience of each alternative interpretation of the topic for each dialogue. Two methods used by MacKay (1973) to rate lexically ambiguous sentences were employed here; one involved the use of dominant meaning ratings and the other of subjective likelihood ratings.

Method

Subjects. Forty-six undergraduate psychology students at Kansas State University served as subjects. All were native English speakers and received course credit for participation. Subjects were run in groups of 15 to 25. The experiment lasted about 20 to 30 minutes.

Materials. Each dialogue set consisted of two dialogue contexts and an ambiguous "base" dialogue, as shown in Appendix 1. Each dialogue context preceding the base resulted in alternative interpretations of the dialogue topic. The following themes comprise the alternative interpretations induced by each context preceding the ambiguous base: astronaut's first trip to the moon or helicopter policeman's experience at a political demonstration; buying a waterbed or a car; first day on the job for a beautician or a stagehand; cultivating plants or raising children; gossip about a teacher or a student; a doctor diagnosing a patient or a doctor and an intern conferring; buying a dress or observing a new art exhibition; hunting with a bow and arrow or playing the violin. For most of the dialogues, this topical ambiguity was due to ambiguity and vagueness in word meaning.

For a given ambiguous base, its two dialogue contexts were roughly matched on number of words, but precisely matched on number of turns (the point at which a given speaker has the floor during a conversation). The difference in words between contexts for a given base ranged from 0 to 5 (M = 2.875), and the number of turns was identical, ranging from 1 to 5 (M = 1.87). For all ambiguous bases, the number of turns ranged from 8 to 14 (M = 11.12) and the number of words ranged from 110 to 203 (M = 132.38).

All ambiguous bases and dialogue contexts were tape recorded by a male and female speaker, who played the roles of the two speakers in each dialogue. Both actors read the dialogues in their natural midwestern U.S. accents as if they were actually carrying on a conversation. However, in Experiment 1, the dialogue contexts were not presented to subjects. They heard only the base dialogues.

Procedure. After hearing the instructions shown below, subjects supplied successive dominant meaning and subjective likelihood ratings following each ambiguous dialogue base.

You will be hearing a series of dialogues which are ambiguous; what the two people are talking about can be interpreted in more than one way. After each dialogue has been played, the experimenter will signal you to turn the page. You will find two alternative interpretations of the dialogue topic on the page. First, you will circle which one of the two alternative interpretations you understood first while listening to the dialogue. Next, you will indicate the likelihood of each alternative interpretation independent of which interpretation you understood first. You will use percentages ranging from 0% to 100% to indicate likelihood. If you think two interpretations are equally likely, you would give each a rating of 50%. If you think one interpretation is much more likely than the other, you might rate one 90% and the other 10%, and so forth.

For example, after hearing the first dialogue, subjects circled either "astronaut talking about blast-off to the moon" or "policeman talking about political protest he saw from helicopter", the two alternative interpretations of the dialogue topic. Immediately after making this rating, subjects generated subjective likelihood ratings.

Results and Discussion

In principle, the alternative interpretations would have received equally likely subjective likelihood ratings (50%) if the dialogues had been perfectly ambiguous. However, as shown in Table 2, the most "unbiased" ambiguity ratings (those ratings of the two interpretations closest to 50%) were in the 60-40 range.

Not only did ratings differ according to dialogue, but also according to the method used, with ratings of the two alternative interpretations being more "biased" using the dominant meaning method than the subjective likelihood method. These findings are consistent with those of MacKay (1973), who attributed this discrepancy to differences in reliability of the two methods. Insofar as each subject providing a subjective likelihood rating contributed more information, this is inherently more reliable than the dominant meaning method. In the former method, the subject not only indicated a dominant meaning by assigning a greater percent to one interpretation than the other, but also indicated the magnitude of this dominance by assigning percents on a continuous scale. In the dominant meaning method, the subject merely provided a dichotomous preference judgment.

EXPERIMENT 2: REPLICATION WITH DIFFERENT RECORDINGS

Following unexpected damage to the stimuli rated in the previous experiment, all materials were re-recorded using a different male and a different female speaker. Since these two different speakers read the materials with their own idiosyncratic intonation patterns, and it has been shown, at least using advertising materials, that intonation patterns contribute to the implications of an utterance (Coleman, in press), ambiguity ratings were obtained for these new recordings.

Method

Subjects. Thirty-two undergraduate psychology students at Kansas State University served as subjects. All were native English speakers and received course credit for participation. Subjects were run in two large groups in sessions ranging from 30 to 45 minutes.

Materials. Aside from a change in speakers, these were identical to the materials used in Experiment 1.

Procedure. With the exception of using only the subjective likelihood rating method, the procedure was identical to that of Experiment 1.

Results and Discussion

With the exception of the "theatre-beauty parlor" dialogue, results were very similar across the two sets of recordings, as shown in the third column of Table 2. That the results were very similar even with two independent groups of subjects suggests that the two sets of ratings reliably reflect properties of the materials and not those of the speakers or the raters.

The new set of ratings, shown in the third column of Table 2, was used as the basis for classifying dialogues for later analyses in which effects of ambiguity upon drawing certain kinds of inferences was examined.

EXPERIMENT 3: DRAWING INFERENCES FROM CONVERSATIONS

The research that follows examined the processing of two classes of
inferences: "Neutral" and "Context" inferences, from conversational materials.
This two-fold classification is related to whether or not comprehension of the
inference is dependent upon preceding contextual information. In the ambiguous
dialogues described above that have been constructed for this research, Neutral
inferences are comprehensible without reference to a preceding context, whereas
Context inferences are based on information cued by the preceding dialogue context.

Subjects listened to each of the eight dialogues as in Experiments 1 and 2,
but these were presented either with or without one of the two dialogue contexts.
After each dialogue, subjects judged the truth values of a series of inference
statements. Truth response rating as well as speed of response was assessed.

Method

Subjects. Sixty-eight undergraduate psychology students at Kansas State
University who had not participated in either of the previous experiments were
run individually in sessions lasting from 45 to 60 minutes. All subjects were
native English speakers with no hearing defects; they received class credit for
participation.

Materials. As mentioned earlier, two classes of inferences were examined.
The various types of Neutral inferences were true propositional, false proposi-
tional, true pragmatic, and false pragmatic inferences. The two types of Context
inferences were "Inappropriate" and "Appropriate" inferences, to be explained
below.

What follows is a discussion of the various inference statements with
illustrative examples from test statements from the "astronaut-policeman"
dialogue (in Appendix One) to show how they were constructed. Neutral inferences,

whose descriptions follow, were based solely on information in the ambiguous sequences and not the preceding contexts. For each dialogue, four test statements were written for each of the inference types mentioned below. Examples are shown at the bottom of Table 1, and also in Appendix 2.

a) True and false propositional inferences: A true propositional inference (22) is essentially a paraphrase of an utterance in the dialogue while a false propositional inference (23) is a contradictory paraphrase.

b) True and false pragmatic inferences: A true pragmatic inference (24) doesn't necessarily or logically follow from an utterance in the dialogue, but is nevertheless highly probable. Conversely, a false pragmatic inference (25) is not a logical contradiction of a dialogue utterance, but is nevertheless highly improbable.

c) Inappropriate and appropriate context inferences: Since there were two dialogue contexts for each dialogue, a given context-inference pair consisted of an item tapping information following from one alternative interpretation, and an item tapping information from the other interpretation. Four pairs of context inferences were written for each dialogue. Most pairs were balanced such that for each member of the pair, the only difference was in the wording of the critical terminal phrase appropriate to one interpretation and inappropriate to the other.

In dialogue one, for example, the alternative interpretations were an astronaut relating the gala events leading to his blast-off to the moon, and a helicopter policeman discussing the riot he observed from his helicopter. The context inference pair comprised by (26) and (27) corresponds to these respective former and latter interpretations. For subjects hearing the "astronaut" context with dialogue one, (26) would be an appropriate inference and (27) would be inappropriate. Conversely, (26) would be inappropriate and (27) would

be appropriate after having heard the "policeman" context. Thus every context inference was appropriate for half the subjects and inappropriate for the other half.

In constructing all of the above types of test statements, special care was taken to avoid the use of negatives, which would have presumably complicated processing (cf. Gough, 1965; Meyer, 1973; Wason & Jones, 1963), i.e., false statements were contradictions of information in the dialogues, not true statements to which negatives had been added.

Test statements following each dialogue were randomly ordered with the constraints that no members of the same context inference pair were immediately adjacent and that no inappropriate inference occurred as the first test statement. Initial pilot work showed that subjects were confused after first hearing an inappropriate inference subsequent to a dialogue. All test statements for the dialogues are shown in Appendix 2.

Three taped lists of the eight ambiguous dialogues combined with the dialogue contexts were dubbed from master recordings. Each taped list contained the eight dialogues in the same order. However, across the three lists, the ambiguous sequence of each dialogue occurred only once under each one of the three context conditions, i.e., each was preceded by either one of the two alternative contexts, or by no context. Dialogue lines from the contexts were spliced precisely to the ambiguous dialogues so that the listener heard one continuous dialogue, not two successive sequences.

Apparatus. All test statements and subjects' responses to them were recorded using an Ampex Model 120A stereo tape recorder. The resulting tape was later played during manual transcription of subjects' latencies. Test statements, which were the same for each dialogue regardless of its context condition, were presented using a second tape recorder. In addition, taped lists of dialogues were presented using a cassette recorder.

All latencies to respond to test statements were manually transcribed using a Lafayette Model 6602A Voice Reaction Time Control and a Series D Hunter KlockKounter. A recording of each subject's vocal responses to test statements along with the test statements themselves was played at 3 3/4 i. p. s. through an Ampex Model 414 speaker to a microphone connected to a voice key input. To transcribe latencies, a push-button switch connected to the other voice key input was manually triggered by the experimenter immediately following the offset of the test statement, thus activating the timer. The subject's subsequent vocal response stopped the clock. Comprehension time was defined as the interval between the offset of the test statement and the onset of the vocal response.

Procedure. As explained to subjects, the purpose of the experiment was to understand how people process information in conversations. Subjects were told that after hearing each one of eight dialogues, they would give either a summary or a title for it, and then respond to test statements about information in it. Requiring subjects to give a summary or title presumably encouraged deep semantic processing of the dialogues and enabled subjects' interpretations of dialogue topics to be checked against a priori intended interpretations. Subjects supplied truth responses to 24 test statements (16 Neutral inferences and 4 Context inference pairs). Context conditions were counterbalanced across dialogues. Instructions were as follows:

I'd like to present to you a series of eight dialogues or conversations, each of which takes place between a male speaker and a female speaker. After you hear each dialogue, I'd like you to summarize it or give me a title for it. After that, I'll play a series of sentences about information in the dialogue you will have just summarized or given a title for.

Using the scale in front of you, please respond as quickly and accurately as you can concerning the truth of each sentence. Your truth responses will be timed, starting at the end of the sentence. Respond with 1 if the statement is definitely false based on the conversation, 2 if it is probably false, 3 if the statement is indeterminate in truth value (i.e., the conversation did not make clear whether this would be true or false and it could be either), 4 if the statement is probably true, and 5 if it is definitely true based on the conversation.

If you have absolutely no idea as to what number to respond with, say "don't know" into the microphone. If you do make a mistake in responding, feel free to correct yourself, and I'll make a note of your correction. If you speak too softly during the experiment, I'll hold up a small sign that says, "Please talk louder," since a certain voice volume is necessary to activate the timer.

In the test statements that follow each dialogue, any mention of "the man" is a reference to the male speaker and any reference to "the woman" is a reference to the female speaker. You will have ample time to respond to each sentence after you hear it. Are there any questions?

Results

Selection of subjects. Data from eight of the sixty-eight subjects were not used in any of the analyses. These eight subjects had made greater than eight "don't know" responses to the entire set of test statements for all eight dialogues. Specifically, these subjects provided many more "don't know"

responses (M = 25,88), than the other sixty subjects (M = 0.93). Including these eight subjects would have resulted in mean scores based on very few data points, since neither truth responses nor latencies could be computed from "don't know" responses.

Transcription of latencies. As mentioned earlier, each subject's entire set of responses to test statements, along with the test statements themselves from all eight dialogues, was tape-recorded. Latencies were manually transcribed from this tape after all the data had been collected.

To achieve greater reliability, latencies to each test statement for each subject were transcribed three times. The mean of these three latencies was the "raw" score which itself was averaged across all test statements of the same item type across all eight dialogues to produce final means upon which analyses of variance were performed.

Three types of latency responses were not transcribed: "extraneous noise" responses (occasionally subjects touched the microphone, prematurely stopping the clock), "self-correction" responses (e.g., a subject responded "4" to a test statement, but quickly changed it to "3"), and "filled pause" responses (the time interval was taken up by a speech sound like "ah", "er", or "uh" which prematurely stopped the clock). Calculation of truth responses was not affected by these types of responses. Raw truth response scores were obtained by averaging across all test statements of the same item type across all eight dialogues.

Accuracy of manual transcription. The experimenter manually triggered the timer as soon as possible immediately subsequent to offset of each test statement. In this procedure, the experimenter's simple reaction time (triggering the clock after hearing the end of the test statement) was included in the subject's reaction time to respond to the test statement. To attenuate this problem,

the experimenter himself became highly practiced at the transcription task until he was performing at an asymptotic level.

To determine the magnitude of the experimenter's simple reaction time, latencies were manually transcribed from a test segment of known duration. This test segment was constructed by splicing in 1000 msec. of blank "silent" tape within a pre-recorded segment of conversation. Thus, there was an interval of conversation, 1000 msec. of silence, and then another interval of conversation.

The experimenter triggered the clock immediately after the offset of the trial conversation. This was analogous to triggering the clock after the offset of the test statement. The timer stopped after onset of the subsequent interval of the trial conversation, just as it had stopped during the subject's verbal response to the test statement. Duration of the interval of silence was transcribed ten times in succession with results for each transcription as follows: 1022, 1061, 1114, 1061, 1026, 1077, 1062, 1032, 1047, and 1065 msec., with M = 1056.7 msec., and SD = 27.24. Thus, an estimate of the experimenter's mean simple reaction time in transcribing 1000 msec. latencies was approximately 56.7 msec., (i.e., 1056.7-1000 msec.).

Throughout the transcription procedure, the experimenter was blind to whatever context condition a given dialogue occurred under. Latencies were transcribed from tapes containing only test statements and subjects' responses to them.

Correlational data. To determine whether latencies were solely a function of the subjectively rated truth responses to the test statements, correlations were computed between latency and truth responses for each one of the Neutral and Context inference types, as shown in Table 3. The only statistically significant correlations were for true pragmatic, true propositional, and context-

appropriate inferences. Nevertheless, considering the large sample size and small proportion of variance accounted for, these significant correlations are still very modest.

Overview of analyses of variance.

Since there were two conceptually distinct classes of inference types, i.e., Neutral and Context inferences, analyses on each inference type will be reported separately. Within each class of inferences, parallel analyses were performed on the two dependent measures of truth response and latency. Only the analyses of variance for the Neutral inferences included Truth value as a factor. Context inferences were appropriate or inappropriate, and not true or false.

Analyses of Neutral inferences: Context effects. Neutral pragmatic and propositional inferences were constructed to be understood without reference to either preceding context. To determine if the effects of preceding context on inference type were indeed negligible, as was intended, 2 x 2 x 2 completely within-subjects analyses of variance were performed with Context (preceding context vs. no preceding context), Inference type (pragmatic vs. propositional), and Truth value (true vs. false) serving as factors. Subjects' mean responses under a given context condition served as input to the analyses. Specifically, responses were separately averaged over all conditions containing a preceding context and over all "no context" conditions.

As hypothesized, there were no significant main effects of or interactions with the Context factor for either dependent measure. Cell means are shown in Table 4.

Differences among Neutral inferences. For both dependent measures, each subject contributed four scores, each of which was the mean of 32 responses (i.e., one response for four Neutral inference item types for eight dialogues). A completely within-subjects 2 x 2 analysis of variance performed on mean

latencies to pragmatic and propositional inference test statements of true and false truth values collapsed across all context conditions indicated significant main effects of Truth value, $\underline{F}$ (1, 59) = 58.90, $\underline{MSe}$ = .03, $\underline{p}$ < .001, Inference type, $\underline{F}$ (1, 59) = 21.90, $\underline{MSe}$ = .024, $\underline{p}$ < .001, and also a significant Inference type x Truth value interaction, $\underline{F}$ (1, 59) = 4.40, $\underline{MSe}$ = .02, $\underline{p}$ < .05.

As expected, mean latencies were longer to false than true inferences, and to pragmatic than propositional inferences, as shown in Table 5. This true-false difference was greater for pragmatic than propositional inferences.

A parallel analysis on truth responses revealed a significant main effect of Truth value, $\underline{F}$ (1, 59) > 100, $\underline{MSe}$ = .113, $\underline{p}$ < .001, but no significant main effect of Inference type, although the Inference type x Truth value interaction was highly significant, $\underline{F}$ (1, 59) = 103.09, $\underline{MSe}$ = 4.63, $\underline{p}$ < .0001.

In general, true inferences were rated more true than false inferences, but this true-false difference was greater for propositional than pragmatic inferences, as shown in Table 5. There were also no significant differences between pragmatic and propositional inferences in terms of overall truth ratings.

Analyses of Context inferences: Deletion of dialogues. Recall that subjects were required to give a summary or title after hearing each dialogue. This was done to verify the effectiveness of the contextual manipulation as well as induce deep semantic processing of the materials. In all analyses of Context inferences in which Context condition was a factor, items in dialogues that subjects misinterpreted were deleted. In addition, all dialogue items for which there was no record of an interpretation (due to mechanical error, experimenter error, or if the subject just didn't know what to say) were also deleted. In total, 16% of subjects' responses to dialogues were deleted for any and all of the above reasons. This procedure closely resembles the customary deletion of incorrect responses in traditional reaction time tasks. Misinterpretation was defined here as giving an interpretation of the dialogue not

consistent with the preceding context. Insofar as test statement verification required understanding of the dialogue, truth and latency responses to test statements from a misinterpreted dialogue had little meaning. All data concerning subjects' interpretations of the dialogues are shown in the first two lines of each row in Table 6.

Validity of the "no context" condition as a control condition. Several prose processing studies using a contextual manipulation to influence comprehension have been criticized on the basis that the stimuli were incomprehensible without the context. In this research, the "no context" condition served as a baseline against which the preceding context conditions were compared in terms of truth and latency responses. The suitability of this condition as a baseline would be open to question if responses reflect non-comprehensibility. However, as shown by subjects' interpretations of "no context" dialogues in the third line of each row in Table 6, in only five instances out of 160, subjects could not generate an interpretation for a dialogue.

Effects of preceding context. To assess whether the preceding context influenced the latency with which appropriate and inappropriate inference test statements were verified, a one-way analysis of variance was performed on subjects' mean latencies to items preceded by an appropriate context, an inappropriate context, and no context.

Due to incomplete counterbalancing, the number of items contributing to each subject's three scores varied according to whatever list each subject had been assigned to. Specifically, for respective appropriate, inappropriate, and control conditions, the number of items per score was 24, 24, and 16 for list one; 20, 20, and 24 for list two; and 20, 20, and 24 for list three.

As hypothesized, there was a significant main effect of Context on both latency, $F (2, 118) = 4.63$, $\underline{MSe} = .117$, $\underline{p} < .025$, and truth responses, $F (2, 118) = 83.45$, $\underline{MSe} = .30$, $\underline{p} < .0001$. Latency responses increased ordinally across appropriate, control, and inappropriate levels of the Context factor. In addition, while appropriate inferences were rated truer than inappropriate inferences, the mean truth response to control inferences was 3.00, the center of the 5-point scale, suggesting that subjects were using the scale appropriately. All cell means for the two dependent measures are shown in Table 7.

Scheffé tests performed on these data showed that, for truth responses, appropriate inferences were rated as truer than inappropriate inferences, $F (2, 177) = 176.10$, $\underline{p} < .001$, which were, in turn, rated as less true than control inferences, $F (2, 177) = 39.06$, $\underline{p} < .001$. Latencies to appropriate inferences were significantly shorter than those to inappropriate inferences, $F (2, 177) = 7.482$, $\underline{p} < .05$, but these, in turn, were not significantly longer than control inferences, $F (2, 177) < 1$, $\underline{p} > .05$.

Effects of dialogue ambiguity on context inferences. Since the ambiguity ratings of the dialogues obtained in Experiment 2 varied considerably, the effects of rated ambiguity on the speed and accuracy of processing context inferences was examined. Particular dialogues were divided into "balanced" and "skewed" levels of a "Bias" factor. Classification as balanced or skewed was based upon deviation of the dialogue ratings from 50-50 subjective likelihood method values.

Dialogue five (Student-Teacher) was not included in any subsequent analyses since there were so many misinterpretations of the "José as teacher" version (see Table 6), resulting in few usable data points. To equate the number of dialogues within levels of the bias factor, dialogue four (Plants-Children), the least balanced of the dialogues within the balanced level, was also deleted, resulting in three dialogues per bias level. The mean likelihood for alternative

interpretations was 60.09-39.91 for balanced dialogues (i.e., dialogues 1, 2, and 8), and 68.80-31.20 for the skewed dialogues (i.e., dialogues 3, 6, and 7).

For a given dialogue, one topic interpretation was always assigned a higher subjective likelihood than the other. In other words, one alternative interpretation was always preferred over the other to some degree, with this difference in preference more acute for skewed than balanced dialogues. Thus, for each dialogue, responses to context inferences were examined for "preferred" vs. "non-preferred" interpretations induced by the preceding contexts. These were the two levels of the "Preference" factor. Since responses to Context inferences were of interest, a third factor in the analysis was Inference type, with levels of appropriate and inappropriate inferences.

It was expected that Appropriate inferences would be rated truer than Inappropriate inferences, consistent with results from the single-factor analysis using the entire set of dialogues. Furthermore, it seemed plausible that this difference in truth response ratings would be greater for preferred than non-preferred interpretations. It should be easier for subjects to respond to topic-appropriate and topic-inappropriate items from dialogues with plausible than non-plausible topics. This effect may be more substantial for skewed than balanced dialogues, since there might be a greater difference in plausibility between preferred and non-preferred interpretations for the skewed dialogues.

At best, one might expect the pattern of results for the latency data to mirror the truth response results. However, this seemed doubtful in light of the modest correlations between truth response and latency for the Context inferences, as shown in Table 3. Since there were no well-motivated predictions for the latency data, only the results from analyses of truth responses will be reported here.

In a 2 x 2 x 2 analysis of variance with Inference type as the only within-subjects factor, the truth response data showed significant main effects of Inference type, F (1, 203) = 276.64, MSe = .71, p < .0001; Bias, F (1, 203) = 46.28, MSe = .55, p < .001, and a significant Bias x Preference x Inference type interaction, F (1, 203) = 6.02, MSe = .71, p < .015. Skewed dialogues received higher truth ratings than balanced dialogues. Consistent with previous results, context-appropriate inferences received higher truth ratings than context-inappropriate inferences.

To determine if the significant triple interaction may have been due to differences across levels of Bias of the theoretically interpretable Inference type x Preference interaction, simple interaction tests were performed. Results showed that this two-way interaction was nonsignificant for balanced and skewed dialogues. All cell means are shown in Table 8.

Discussion

The major findings will be discussed separately for Neutral and Context inferences, since these item types are conceptually distinct. The major results discussed below include latency differences in verifying pragmatic and propositional Neutral inferences, truth-response differences for the same item types, and the facilitating and inhibiting effects of context on appropriate and inappropriate inferences.

Neutral Inferences

Latency differences. The shorter latencies to verify propositional than pragmatic inferences suggest that comprehension of pragmatic inferences requires reconstructive processes beyond those needed for propositional inferences. While this is consistent with findings from studies using sentences as stimuli (e.g., Kintsch, 1974; Singer, 1974, 1977, 1979), the present results were obtained using conversational materials. Some understanding of the dialogue as a whole (i.e., not mere memorization of speakers' individual statements) was required to verify each one of the inference statements. These results extend the pragmatic-propositional distinction to conversational materials and, through the use of a latency measure, complement the findings obtained with memory measures alone (cf. Harris & Monaco, 1978).

Supplementary support for the propositional-pragmatic distinction is provided by the correlational data. To eliminate the alternative that latency differences were totally dependent upon rated truth or falsity of the particular inference types being compared, correlations between truth and latency responses were performed for each one of the inference types. If the separate correlations for each one of the inference types were not highly signi-ficant, then observed significant mean differences in latencies could not be attributed to differences in rated truth responses. As shown earlier, the latency-truth response correlations were modest, at best, for two of the four Neutral inferences.

An alternative interpretation of the latency difference between propositional and pragmatic inferences is that these data support an inference-at-retrieval view. In Singer's (1979) analysis, if inferential processing occurs during retrieval, then verification of true and false pragmatic inferences should take longer than true and false propositional inferences, due to additional inferential activities

required at the time of response. Furthermore, false pragmatic inferences should take the longest to verify, since false responses generally take longer than trues. On the other hand, if inferences are drawn during encoding, both propositional and pragmatic false inferences will take equally longer than their true counterparts, because the memory representations are equal in both cases. In summary, a retrieval model predicts significant main effects of Inference type and Truth value, while an encoding model predicts a significant main effect of Truth value alone.

Singer's (1979) results and those obtained here are consistent with an inference-at-retrieval interpretation. However, the inference-at-retrieval versus inference-at-encoding dichotomy may not be very useful theoretically. It is another one of those research questions cast in the form of a binary opposition which, according to Newell (1973), may inhibit scientific progress. Insofar as research is designed to support or refute one or the other of these opposing concepts, the result is oversimplification, leading to the neglect of such phenomena as details of processing (cf. Lachman, Lachman & Butterfield, 1979).

Resolving the issue of inference-at-encoding versus inference-at-retrieval has been difficult to achieve experimentally. In Singer's previous (1974, 1977) work, for example, failing to reject the null hypothesis was used to support the encoding view. Similarly, Loftus and her colleagues (Loftus & Palmer, 1974; Loftus, 1975) have shown how difficult it may be to find a methodologically "pure" test of the retrieval view, since the nature of the retrieval cue itself (e.g., a question with false presuppositions such as "Did you see the Stop sign?" when there was a Yield sign), may alter the pre-existing memory structure through integration of this false new information with the old information. In fact, one prevalent view is that memory is both constructive and reconstructive; the stored representation is continually being revised based on information processed before and after a target stimulus.

Truth response differences. In previous research (e.g., Harris, 1974, 1977), true pragmatic inferences were rated less true than true propositional inferences during an immediate test. These differences were not found during a delayed test, presumably because the stored representations in long-term memory for the inference types were equivalent, which led subjects to respond to pragmatic and propositional inferences in the same way.

In the present research, truth responses to true and false pragmatic and true and false propositional inferences were examined separately. True pragmatic inferences were rated as less true than true propositional inferences; false pragmatic inferences were rated as less false than false propositional inferences. In other words, both true and false pragmatic inferences were rated closer to "indeterminate", (the theoretical logically "correct" value of a pragmatic inference), than their propositional inference counterparts.

Although the present research was a memory study, the time interval between the end of a given dialogue and the test statements may have been too short to effectively alter subjects' discrimination of these two inference types, leading subjects to rate pragmatic inferences closer to indeterminate than propositional ones.

Context Inferences

Effects of dialogue topic ambiguity. One consistent finding replicated by analyses of the full and partial dialogue sets was the significant main effect of Inference type. In general, appropriate inferences were rated as truer and responded to more quickly than inappropriate inferences, consistent with results of Owens et al. (1979).

"Script" theory and context effects. Latencies were longer to inappropriate than appropriate inferences, with intermediate latencies to control inferences.

Similarly, truth responses were highest for appropriate inferences, intermediate for control, and lowest for inappropriate inferences.

These results are consistent with a view of inference verification which derives from Bower et al.'s (1979) "spread of activation" explanation for script comprehension. A script is defined by Bower et al. (1979) as "a generic knowledge representation of everyday stereotyped events or situations." This knowledge structure sets up expectations to guide comprehension by specifying default values, necessary components, and slots to be filled.

Each dialogue in the present research detailed a commonplace activity (e.g., going to an art show, visiting a doctor) which subjects are well acquainted with and have readily accessible knowledge about. The script conceptualization is relevant to the present research insofar as subjects accessed the necessary scripts in understanding each dialogue and accessed a different one for each version of a given dialogue. In the research of Owens et al. (1979), subjects who received prior information about the character's motives did not access a script different from those who received no such information.

According to Bower et al. (1979), script comprehension begins when memory nodes representing major components are activated by hearing, in the context of this experiment, the preceding dialogue lines. This activation spreads to nearby related nodes, activating them in turn. Thus, "expected stimulus patterns" are identified more rapidly due to a priming effect of the accumulating activation, resulting in shorter latencies to appropriate than inappropriate inferences with intermediate latencies to control inferences. Since the appropriate inferences are consistent with the expectations set up by the accessed script, they should benefit more from the previously accumulating

activation than the inappropriate inferences which are counter to expectations. Control inferences which are neither wholly inconsistent or wholly consistent with expectations (they are theoretically indeterminate), should be verified more quickly than inappropriate inferences but less quickly than appropriate inferences. Although, in the present research, inappropriate inferences were verified no more quickly than control inferences, they were verified significantly more slowly than appropriate inferences.

In summary, the present research has extended findings on memory and comprehension of inferences to conversational materials, used two dependent measures and a five-point scale to avoid some of the methodological pitfalls of earlier work, and has supported an explanation of the Context inference results based on Script theory.

REFERENCES

Anderson, R.C., & Pichert, J.W. Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 1-12.

Bates, E., Kintsch, W., & Masling, M. Recognition memory for aspects of dialogue. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 187-197.

Bower, G.H., Black, J.B., & Turner, T.M. Scripts in memory for text. Cognitive Psychology, 1979, 11, 177-220.

Bransford, J.D., & Franks, J.J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Bransford, J.D., & Johnson, M.K. Consideration of some problems of comprehension. In W.G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973, pp. 383-438.

Brewer, W.F. Memory for ideas: Synonym substitution. Memory & Cognition, 1975, 3, 458-464.

Brewer, W.F. Memory for the pragmatic implications of sentences. Memory & Cognition, 1977, 5, 673-678.

Brewer, W.F., & Lichtenstein, E.H. Recall of logical and pragmatic implications in sentences with dichotomous and continuous antonyms. Memory & Cognition, 1975, 3, 315-318.

Bruno, K.J., & Harris, R.J. The effect of repetition on the discrimination of asserted and implied claims in advertising. Applied Psycholinguistics, 1980, in press.

Chmielewski, D. Memory for information inferred from conversations. Paper presented at the American Psychological Association Convention, Toronto, Canada, August, 1978.

Clark, H.H. The language-as-fixed-effect fallacy: A critique of language
    statistics in psychological research. Journal of Verbal Learning and
    Verbal Behavior, 1973, 12, 335-359.

Coleman, L. Semantic and prosodic manipulation in advertising. In R.J. Harris
    (Ed.), Information-processing approaches to advertising. Hillsdale, NJ:
    Lawrence Erlbaum Associates, Inc., in press.

Dooling, D.J., & Christiaansen, R.E. Levels of encoding and retention of prose.
    In G.H. Bower (Ed.), The psychology of learning and motivation. Vol. 11.
    New York: Academic Press, 1977, pp. 1-39.

Dooling, D.J., & Mullet, R.L. Locus of thematic effects in retention of prose.
    Journal of Experimental Psychology, 1973, 97, 404-406.

Glanzman, D.1., 7 Pisoni, D.B. On verifying semantic inferences. Indiana
    Mathematical Psychology Program, Report #73-5, 1973.

Gough, P.B. Grammatical transformations and the speed of understanding.
    Journal of Verbal Learning and Verbal Behavior, 1965, 5, 107-111.

Harris, R.J. Memory and comprehension of implications and inferences of complex
    sentences. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 626-637.

Harris, R.J. Comprehension of pragmatic implications in advertising. Journal of
    Applied Psychology, 1977, 62, 603-608.

Harris, R.J., Dubitsky, T.M., & Thompson, S. Learning to identify deceptive
    truth in advertising. In J.H. Leigh and C.R. Martin, Jr. (Eds.), Current
    Issues and Research in Advertising, Ann Arbor, Michigan: University of
    Michigan Graduate School of Business Administration, Division of Research,
    1979.

Harris, R.J., & Monaco, G.E. The psychology of pragmatic implications: Infor-
    mation processing between the lines. Journal of Experimental Psychology:
    General, 1978, 107, 1-22.

Hildyard, A., & Olson, D.R. Memory and inference in the comprehension of oral and written discourse. Discourse Processes, 1978, 1, 91-117..

Keenan, J.M. Inferring causal connections in prose comprehension. Paper presented at the Annual Meeting of the American Psychological Association, Toronto, August 28, 1978.

Keenan, J.M., MacWhinney, B., & Mayhew, D. Pragmatics in memory: A study of natural conversation. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 549-560.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc., 1974.

Kozminsky, E. Altering comprehension: The effect of biasing titles on text comprehension. Memory & Cognition, 1977, 5, 482-490.

Lachman, R., Lachman, J.L., & Butterfield, E.C. Cognitive Psychology and Information Processing: An Introduction. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc., 1979.

Loftus, E.F. Leading questions and the eyewitness report. Cognitive Psychology, 1975, 7, 560-572.

Loftus, E.F., & Palmer, J.C. Reconstruction of automobile destruction: An example of the interaction between language and memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 585-589.

Lycan, W.G. Conversation, politeness, and interruption. Papers in Linguistics, 1977, 10, 23-53.

MacKay, D.G. Aspects of the theory of comprehension, memory, and attention. Quarterly Journal of Experimental Psychology, 1973, 25, 22-40.

Meyer, D.E. Verifying affirmative and negative propositions: Effects of negation on memory retrieval. In S. Kornblum (Ed.), Attention and Performance, Vol. 4. New York: Academic Press, 1973, pp. 379-394.

Mishler, E.G. Studies in dialogue and discourse: II. Types of discourse initiated by and sustained through questioning. Journal of Psycholinguistic Research, 1975, 4, 99-121.

Mishler, E.G. Studies in dialogue and discourse: III. Utterance structure and utterance function in interrogative sequences. Journal of Psycholinguistic Research, 1978, 7, 279-305.

Mohan, B. Do sequencing roles exist? Semiotica, 1974, 12, 75-96.

Monaco, G.E. The locus of inference construction. Kansas State University Human Information Processing Institute Report #76-16A. April, 1977.

Newell, A. You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W.G. Chase (Ed.), Visual Information Processing. New York: Academic Press, 1973.

Owens, J., Bower, G.H., & Black, J.B. The "soap opera" effect in story recall. Memory & Cognition, 1979, 7, 185-191.

Paris, S.G. The development of inference and transformation as memory operations. In P. Ornstein (Ed.), Memory development in children. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Pichert, J.W., & Anderson, R.C. Taking different perspectives on a story. Journal of Educational Psychology, 1977, 69, 309-315.

Posner, M.I. Abstraction and the process of recognition. In J.T. Spence and G.H. Bower (Eds.), Advances in Learning and Motivation, Vol. 3. New York: Academic Press, 1969.

Rubinstein, H.R., Garfield, L., & Millikan, J.A. Homographic entries in the interval lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Sachs, J.S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Schallert, D.L. Improving memory for prose: The relationship between depth of processing and context. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 621-632.

Schank, R. Rules and topics in conversation. Cognitive Science, 1977, 1, 421-441.

Schegloff, E.A., & Sacks, H. Opening up closings. Semiotica, 1973, 8, 289-327.

Singer, M. Context inferences in the comprehension of sentences. Canadian Journal of Psychology, 1976, 30, 39-46.

Singer, M. The temporal locus of inference in language comprehension. Unpublished manuscript, 1977.

Singer, M. Processes of inference during sentence encoding. Memory & Cognition, 1979, 7, 192-200.

Sperling, G.A. The information available in brief visual presentations. Psychological Monographs, 1960, 74, No. 498.

Springston, F.J., & Clark, H.H. And and or, or the comprehension of pseudo-imperatives. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 258-272.

Sternberg, S. Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 1975, 27, 1-32.

Wason, P.C., & Jones, S. Negatives: Denotation and connotation. British Journal of Psychology, 1963, 54, 299-307.

Weiner, S.L., & Goodenough, D.R. A move towards a psychology of conversation. In R.O. Freedle (Ed.), Discourse production and comprehension. N.J.: Ablex Publishing Corporation, 1977.

Table 1

Sentence Examples

(1a) All collies are dogs.
(1b) Lassie, a collie, is a dog.
(2a) John forgot to shut the door.
(2b) John didn't shut the door.
(3a) Kathy is taller than Mary.
(3b) Mary is shorter than Kathy.
(4a) John was pounding the nail to fix the birdhouse.
(4b) John was using a hammer to fix the birdhouse.
(5a) The dock workers talked about their high taxes.
(5b) The dock workers complained about their high taxes.
(6a) The barnacles clung to the sides.
(6b) The barnacles clung to the ship.
(7a) Joe was not tall.
(7b) Joe was short.
(8a) The safecracker put the match to the fuse.
(8b) The safecracker lit the fuse.
(10a) The clumsy chemist had acid on his coat.
(10b) The clumsy chemist spilled acid on his coat.
(11a) The small girl spent the gleaming penny.
(11b) The penny was new.
(12a) The hungry python caught the mouse.
(12b) The hungry python ate the mouse.
(12c) The hungry python caught the mouse, but the hungry python did not eat
      the mouse.
(12d) The hungry python did not eat the mouse.
(13) The discarded cigarette caused the fire.
(14) A burning cigarette was carelessly discarded. The fire destroyed
      many acres of virgin forest.
(15) A carelessly discarded burning cigarette started a fire. The fire
      destroyed many acres of virgin forest.
(16) Do you always use CRT displays?
(17) Do you always put your foot in your mouth?
(18) They don't really know what it's like to raise them themselves.
(19) They don't really know what it's like to raise kids themselves.
(20) They don't really know what it's like to raise plants themselves.
(21) They don't really know what it's like to raise fish themselves.
(22) Plainclothesmen attempted to restore order.
(23) The man's outfit was old.
(24) The crowd was bored by the speeches.
(25) The photographer was shooting portraits.
(26) Some musicians played military fanfares.
(27) Some musicians played folk songs of protest.

Table 2

Ambiguity Ratings: Preferences for Alternative Interpretations*

Alternative Interpretations	Experiment One		Experiment Two
	Dominant Meaning[1] $\frac{N}{}$	Subjective Likelihood[2]	Subjective Likelihood[3]
1. Astronaut – Policeman	51.72-48.28 (30)	57.14-42.86	58.75-41.25
2. Waterbed – Car	32.25-67.75 (32)	59.65-40.35	59.18-40.82
3. Beauty Parlor – Theatre	37.50-62.50 (31)	34.86-65.14	18.78-81.22
4. Plants – Children	66.67-33.33 (30)	61.86-38.14	66.81-33.19
5. Student – Teacher	90.32- 9.68 (31)	61.91-38.09	67.31-32.69
6. Intern – Patient	25.81-74.19 (31)	34.62-65.38	32.97-67.03
7. Fashion Boutique – Art Gallery	81.25-18.75 (32)	75.89-24.11	73.12-26.88
8. Bow and Arrow – Violin	70.00-30.00 (30)	58.65-41.35	62.34-37.66

[1]Number of subjects contributing data varied by dialogue.

[2]N = 46

[3]N = 32

* % choosing preferred interpretation - 1st . . . 2nd

Table 3

Truth Response by Latency Correlations (N = 60)

INFERENCE TYPE	Pearson $\underline{r}$
Neutral Inferences	
True propositional	$-.421$***
False propositional	$-.094$
True pragmatic	$-.355$**
False pragmatic	$-.082$
Context Inferences	
Appropriate	$-.265$*
Inappropriate	$.094$
Control	$-.185$

* $p < .05$

** $p < .01$

*** $p < .001$

Table 4

Mean Responses to Neutral Inferences Under Different Contexts

Inference Type	Truth Responses (1 = false, 5 = true)		Latency (in seconds)	
	Preceding Context	No Context	Preceding Context	No Context
True Pragmatic	3.781	3.864	1.728	1.736
False Pragmatic	2.530	2.531	1.910	1.917
True Propositional	4.136	4.110	1.661	1.650
False Propositional	2.335	2.360	1.807	1.801

Table 5

Mean Responses to Neutral Inferences

	True Pragmatic	False Pragmatic	True Propositional	False Propositional
Truth Response (1=false, 5=true)	3.82	2.56	4.13	2.32
Latency (in seconds)	1.714	1.925	1.659	1.792

Table 6

Dialogue Interpretations

Context Heard	Interpretation A	Interpretation B	A and B	Neither	Miscellaneous*
Astronaut (A)	20	0		0	0
Policeman (B)	4	12		4	0
No Context	11	2		6	1
Waterbed (A)	20	0		0	0
Car (B)	0	17		2	1
No Context	1	3		14	2
Theatre (A)	20	0	0	0	0
Beauty Parlor (B)	0	16	0	2	2
No Context	8	4	1	3	4
Plants (A)	13	1	1	4	1
Children (B)	1	11	6	2	0
No Context	6	3	4	6	1
Student (A)	15	1		4	0
Teacher (B)	13	3		3	1
No Context	9	1		9	1
Doctor-Patient (A)	19	0	0	0	1
Doctor-Intern (B)	0	20	0	0	0
No Context	13	3	1	2	1
Clothing (A)	16	1	0		3
Art work (B)	8	9	3		0
No Context	13	4	2		1
Bow and Arrow (A)	19	0	0		1
Violin (B)	0	19	1		0
No Context	18	1	1		0

*Includes subject's non-response, mechanical error, experimenter error.

Table 7

Mean Responses to Context Inferences

Dependent Measure	Appropriate	Control	Inappropriate
Truth Response (1=false, 5=true)	3.664	3.002	2.377
Latency (in seconds)	1.673	1.828	1.846

Table 8

Mean Truth Responses (1=false, 5=true) to Context Inferences
for Balanced and Skewed Dialogues

Inference Type	Balanced		Skewed	
	Preferred	Nonpreferred	Preferred	Nonpreferred
Appropriate	3.201	3.615	4.177	3.950
Inappropriate	2.159	2.201	2.412	2.629

APPENDIX 1:

CONTEXTS AND AMBIGUOUS BASES FOR EACH DIALOGUE

Dialogue One Contexts

Astronaut Context

A: What sticks out in your mind concerning the early days of space exploration?

B: Let me tell you about my first lift-off.

Police Context

A: What was it like to have been a helicopter patrolman?

B: Let me tell you about one incident . . .

Dialogue One Ambiguous Base

B: I was watching the whole thing from about 3000 feet up. Do you smoke?

A: All I have are Menthols . . .

B: That's fine . . . The view was breathtaking--from the window everything looked extremely small from a distance.

A: What was the crowd like?

B: At first there was a great deal of activity. Everybody was jumping up and down! One guy with a portable TV camera was getting too close. You'll never believe how those streets were littered . . .

A: When did things quiet down?

B: After plainclothesmen had to hold the crowd back, some of whom must have had tears in their eyes. Soon afterwards, the speeches started.

A: Did the speeches stir up the crowd again?

B: No, I think they were boring; everyone seemed pleased when the music started .

A: Your outfit--were you wearing the newest version designed by the space people complete with thermal padding, glare-proof visor, and Velcro snaps?

B: Sure, do I look like the type of guy who likes to take chances?

A: When you got down to the ground--what was the atmosphere like?

B: Excluding the fact that the entire force was there, at first it was better than we had expected . . . No particular precautions had to be taken until later, when we contacted the station . . .

Dialogue Two Contexts

Auto Context

A: So you finally decided on an economy compact instead of a station wagon?

Waterbed Context

A: So you finally decided on a waterbed instead of a regular box spring mattress?

Dialogue Two Ambiguous Base

B: Yes, this is the kind I wanted.

A: Were you interested in any particular style or color?

B: Yes--within a restricted price range.

A: What's that?

B: The cheapest . . .   But my real concern is with motion sickness. My wife is highly susceptible.

A: I'm sure you'll find it smooth, comfortable, and with very little rocking. Perhaps you'd prefer a week-long trial period before payment?

B: Fine, but I'm also worried about--well, how rugged is this anyhow? What if a slight accident should occur? How about the dangers of a puncture?

A: This stuff is triply reinforced! Under this outer layer are two coatings of fiberglass!

B: O.K., well, what was the cheapest color?

A: Neon pink . . .

B: I'll take it! And for a week-long trial period as well!

Dialogue Three Contexts

Beauty Parlor Context

A: Hi, I'm Bill Bartlett, the manager. We're primarily a hair styling salon, but now have expanded facilities to include manicures, pedicures, facial electrolysis, and tatoo removals.

Stagehand Context

A: Welcome to the Stage Crew! We're mostly a group of aspiring actors who will do anything that needs to be done--as long as it's connected with theatre and drama!

Dialogue Three Ambiguous Base

B: I hope you realize that I've never worked in this kind of place before . . .

A: Don't worry about that. Just make yourself at home here--fiddle around with the equipment, ask questions, explore dark corners . . . Find out how things work.

B: What are those things over there?

A: Just different colored lights. You'll have to get used to the idea of switching off hot lights when they're not being used. Just don't stare straight into them.

B: Where are make-up supplies stored?

A: On the bottom shelf alongside the curtain . . .

B: Over here?

A: No--don't open that. If you're going to peek behind the curtains, do it discreetly . . . I meant that black rear curtain . . . There are some patrons waiting--I'd better go . . . Please be careful!

Dialogue Four Contexts

Children Context

A: I know George and Joan Miller go camping every summer and have a basketball hoop over the garage door . . . Don't they have four kids?

Plants Context

A: The Millers go camping every summer, enjoy bird-watching, eat organic foods . . . I'll bet they have their picture-window filled with plants.

Dialogue Four Ambiguous Base

B: No, the Millers don't have any.

A: They don't really know what it's like to raise them themselves.

B: Yes, it certainly is fulfilling to have a house full of them.

A: But they must get adequate outdoor experience--even at young ages . . . And plenty of sunlight. Sunlight is very healthy for all living things.

B: And good nutrition . . .

A: Of course it's easier if they're fed at the same time each day, especially if you have more than one or two . . .

B: Yes, and they should have some water daily . . . After all, by the time they're full-grown adults they drink at least a glassful of water each day.

A: I give mine special vitamins too!

B: Hopefully, they'll grow up tall and strong.

Dialogue Five Contexts

Teacher Context

A: Sure, I've been on the faculty of San Palmetto Community College for twelve years now.

B: Which department?

A: English.

B: Is it very large?

A: There are just four of us.

Student Context

A: Sure, I teach at San Palmetto Community College.

B: Which department?

A: English.

B: Do you teach many classes?

A: Just two in Freshman English Composition.

Dialogue Five Ambiguous Base

B: You must know José--José of the Mendez family?

A: Well, I know him from the College.

B: He lives down the street from here. By the way, is he feeling any better?

A: Today was the first time he had been back in classes in over a week . . .

B: So I guess he is feeling better.

A: He looked kind of anxious though . . .

B: Wouldn't you be if you were out for a week?

A: But I think merely being back made him feel better . . .

B: He certainly likes it . . .   Do you know his mother?

A: Yes, but not very well . . .

B: Did you know she was a teacher?

A: No . . .   Anyhow, seeing friends and familiar faces after an absence is uplifting. Even Dr. Smithers, the Dean, came over to say hello to José.

B: That bald-headed, pigeon-toed old coot?

A: You can't deny that all of the students and faculty really like him . . . José is no exception.

Dialogue Six Contexts

Patient Context

A: Good morning, I'm the resident cardiologist, Dr. Fenster. I hope you had a restful sleep.

B: Yes, thanks. I dozed off right in the middle of the news.

A: I'm still concerned about the trouble you had falling asleep two nights ago.

Intern Context

A: Having an internship in a city hospital is nothing like going to medical school full-time . . .

B: You can say that again.

A: Are you still having trouble with that fellow in the corner, name of Bransford? He was having trouble falling asleep two nights ago.

Dialogue Six Ambiguous Base

A: Was there any pain last night?

B: In the upper area of the arms. I took your advice--there's been no change . .

A: Same symptoms?

B: Yes--and choking spells . . .

A: You should get a blood test done and have a urine sample taken.

B: What's your diagnosis?

A: Probably only anemia accompanied by an infection. If so, it will show up on the blood test . . .

B: What if it doesn't show up?

A: Then you'll get some other tests done . . .

B: X-rays?

A: Perhaps--I wouldn't rule out the possibility of a tumor, but get the results of the blood test and the urinalysis first. What about fever?

B: The nurse told me that the temperature went down about a half-hour ago.

A: At best, I'd recommend some additional days off--to get more rest.

Dialogue Seven Contexts

Artwork Context

B: This must be the exhibit on 20th century European art . . .

Clothing Context

B: What about your line of imported European blouses?

Dialogue Seven Ambiguous Base

A: Here are some new prints by that man everyone in France is talking about!

B: Who's that?

A: Jean-Paul Deschamps!

B: Hmmm . . .   I don't like the way the shades match on this one . . .

A: What about this? It's an earlier style . . .

B: Hmmm . . .   Nice--although it does clash with my blouse . . .

A: His chromatic synthesis is far from conventional. Here's a most recent creation!

B: Very elegant! What's your estimate?

A: I'm not able to quote the exact figures, but you must consider that it is a Paris original!

B: The texture is somewhat coarse . . .

A: Unfortunately, it hasn't been hung very carefully. There are some wrinkles and creases on the bottom . . .

Dialogue Eight Contexts

<u>Violin Context</u>

A: The violin is one of the most difficult instruments to play in tune . . .

<u>Bow and Arrow Context</u>

A: Hunting with a bow and arrow is not an easy sport . . .

Dialogue Eight Ambiguous Base

B: I'd agree with that. But experts sometimes forget.

A: I knew two beginners once. They were in a very large place. Open space all around. Their leader signaled them to raise their bows. They were ready to begin . . .

B: Did they follow through with proper timing?

A: Yes, but they kept missing the mark. I'm glad you didn't hear the terrible noises they were making!

B: Holding the bow the correct way may take years to learn . . .

A: An additional factor is tension . . .

B: A loose bow is a sure guarantee for failure . . .

A: Except when it's being stored--I almost ruined one by forgetting to loosen it up after I was done using it . . .

B: And it should be drawn back evenly, with uniform motion . . .

A: That's not too hard if your bow is well-balanced and not warped. Otherwise there's not much that can be done . . .

APPENDIX 2

NEUTRAL AND CONTEXT INFERENCE TEST STATEMENTS FOR EACH DIALOGUE

Dialogue 1

Neutral Inferences

True Propositional

Plainclothesmen attempted to restore order.
The woman had menthol cigarettes.
The man was 3000 feet up in the air.
The crowd was pleased when the music started.

True Pragmatic

The man smoked a menthol cigarette.
The crowd was bored by the speeches.
The man didn't like to take chances.
The streets were congested with litter.

False Propositional

The man thought the speeches were exciting.
The National Guard had to hold the crowd back.
The man's outfit was old.
Only some of the force was present when the man returned.

False Pragmatic

The photographer was shooting portraits.
Neither man nor woman could light their cigarettes.
The crowd came only to hear the music.
The plainclothesmen were not expecting trouble.

Context Inferences

Some of the persons in the crowd were crying because they were emotionally moved.
teargassed.

The crowd was joyously excited at first.
angry and hostile

Some musicians played military fanfares.
folk songs of protest.

The man wore a space-travel suit.
riot-protection

Dialogue 2

Neutral Inferences

True Propositional

The customer did not want to spend very much money.
The customer's wife often felt sick if she was moving too rapidly.
The customer found the kind he wanted.
The customer asked about the possibility of a puncture.

True Pragmatic

The customer was worried about the product's ruggedness.
The saleswoman said that the product was smooth, comfortable, with little rocking.
The saleswoman said that the probability of an accident occurring was very low.
The customer bought the product on impulse.

False Propositional

The most inexpensive color was green.
The customer decided to try the product out for a month.
The outer surface was two times stronger.
He was interested in the darkest color.

False Pragmatic

The customer's wife enjoys roller coasters, rocking chairs, and merry-go-rounds.
The customer was financially well-off.
The customer was completely confident about his new purchase.
The customer valued the looks of the product more highly than its performance.

Context Inferences

The customer decided to buy a compact rather than a   station wagon.
                                                        conventional bed.

The customer said that his wife often got   carsick.
                                             nighttime indigestion.

The customer asked about the durability of the outer   tire layer.
                                                         mattress cover.

He was also worried about the danger of a   collision.
                                             flood.

Dialogue 3

Neutral Inferences

True Propositional

The man had to attend to some patrons.
The woman was encouraged to satisfy her curiosity.
The man told the woman to be careful when peeking behind curtains.
The man had to leave in a hurry.

True Pragmatic

The hot lights used much electricity and overheated easily.
The woman looked behind and not alongside the curtain for the makeup.
The hot lights were very bright.
The man knew that the woman was unfamiliar with the equipment.

False Propositional

The women asked about the length of the curtain.
The woman looked alongside the curtain for the makeup.
The man said that makeup was stored on the top shelf alongside the curtain.
The man made the woman worry about her minimal previous experience.

False Pragmatic

The woman had held similar jobs before.
The woman was very confident of her vocational abilities.
Tools and supplies were easy to find.
There was nothing behind the curtain that the woman opened.

Context Inferences

The woman had never worked in a theatre / beauty parlor before.

The man advised the woman to fool around with the theatrical / hair-styling equipment.

The woman asked about the spotlights. / sunlamps.

Patrons were waiting to see the play. / get their hair done.

Dialogue 4

Neutral Inferences

True Propositional

Those that are full-grown drink at least a glassful daily.
The man said that it was nice to have a house full of them.
According to the woman, sunlight and outdoor experience are essential for the young.
The Millers never raised any themselves.

True Pragmatic

Water will help them grow.
The man said that they need more water as they grow older.
Vitamins will help them grow.
The woman asked the man a question about the Miller family.

False Propositional

The man said that sunlight is healthy.
The woman commented that it was rewarding to have many of them.
The woman gives hers liquid nutrients.
The man said that it's good to feed them at different times every other day.

False Pragmatic

It is rare to have more than one of them.
The man prefers to keep them outside as much as possible.
The woman spends lots of time feeding one or two.
The Millers used to have some.

Context Inferences

If you have more than one, it's easy to water / make meals for them.

The woman commented that sunlight is good for all living things,

especially seedlings. / infants.

The man hopes that they will be nice and green. / healthy men and women.

The woman said they need to be planted outdoors. / play outside.

Dialogue 5

Neutral Inferences

True Propositional

The professor knew José from College.
José liked Dr. Smithers.
José Mendez is the woman's neighbor.
The woman commented that Dr. Smithers was an old coot.

True Pragmatic

José found it uplifting to see familiar faces once again.
José's friends came over to say hello.
José felt overwhelmed while thinking about the work he had to do.
The professor wasn't interested in talking about José's mother.

False Propositional

Dr. Smithers, the Dean, had long dark hair.
José had been out of school for two weeks.
José's mother was a secretary.
The woman asked if José had gotten married.

False Pragmatic

José had been on vacation.
Dr. Smithers usually greeted everyone hello.
The woman liked Dr. Smithers.
José's mother taught English.

Context Inferences

The professor said that he knew José from the English composition class.
Department.

José had not attended / taught class in over a week.

All of the students / faculty including José liked Dr. Smithers.

José was anxious at school because he had a lot of papers to write. / grade.

Dialogue 6

Neutral Inferences

True Propositional

The diagnosis was anemia accompanied by an infection.
It was also suggested that X-rays be taken.
It was possible that the cause of illness was a tumor.
The doctor recommended recuperation to the woman.

True Pragmatic

The doctor's advice had not helped.
The new symptoms were choking spells.
The woman said that the fever had subsided.
The X-rays were suggested to test for a tumor.

False Propositional

The blood test and urinalysis were advised to test for diabetes.
The doctor asked whether there had been pain during the preceding week.
The doctor gave the reason for the urinalysis.
The woman said that the nurse took the temperature an hour ago.

False Pragmatic

The blood test is a conclusive test.
The doctor said that the illness would persist even with more rest.
The doctor makes hospital rounds twice a day.
The doctor was more confident in his diagnosis of tumor rather than anemia.

Context Inferences

The doctor recommended to the woman that a blood test should be performed

on $\begin{cases} \text{her} \\ \text{the patient in the corner.} \end{cases}$

The doctor asked whether $\begin{cases} \text{the woman} \\ \text{Bransford} \end{cases}$ had trouble falling asleep.

The woman told the doctor that $\begin{cases} \text{she} \\ \text{the patient} \end{cases}$ felt pain in the upper area of the arm

If the blood test was inconclusive, $\begin{cases} \text{the woman} \\ \text{Bransford} \end{cases}$ would have to undergo further tests

Dialogue 7

Neutral Inferences

True Propositional

The man was knowledgeable about Deschamps' work.
The man said that Deschamps' color mixtures were unusual.
The texture of the most recent of the items was rough.
The most recent item had been poorly cared for.

True Pragmatic

The woman was very wealthy.
The man forgot the exact price.
The woman preferred colors that matched.
Deschamps' creations were very expensive.

False Propositional

Deschamps' creations were London originals.
The woman liked the first item she viewed.
The Frenchman's name was Jean-Claude Deschamps.
The woman asked the man how old Deschamps was.

False Pragmatic

The woman didn't mind having the colors clash with her blouse.
The woman had previously heard of Deschamps.
Deschamps was very well-known in America.
The items were poorly displayed.

Context Inferences

The Frenchman, Deschamps, is a contemporary painter. / fashion designer.

Wrinkles and creases were caused by uneven hanging of the garments. / stretching of the canvas.

The woman preferred blouses / paintings of an earlier style.

The man served as a gallery tour guide. / fashion consultant.

Dialogue 8

Neutral Inferences

True Propositional

Bows should be loosened only when being stored.
A warped bow cannot be drawn back evenly and uniformly.
One time the woman forgot to loosen her bow.
The woman didn't like the noises the beginners were making.

True Pragmatic

A bow should be loosened before being stored.
A well-balanced bow is easy to draw back evenly.
One's bow must be raised in order to begin.
The man and woman are experts.

False Propositional

The beginners were successful.
The woman spoke about her observations of three beginners.
It is easy to learn how to hold the bow correctly.
The beginners and their leader were in a densely populated area.

False Pragmatic

Proper timing in itself is sufficient for success.
It's necessary to have a well-balanced but not a straight bow.
The woman still associates with beginners.
The woman is unsympathetic concerning beginners.

Context Inferences

The woman observed beginners in a large concert hall.
open field.

Their leader signalled them to start by raising his baton.
hunting cap.

The beginners never hit the right notes.
wild animals.

With a loose bow, you won't be able to play music well.
shoot game

THE EFFECTS OF CONTEXTUAL KNOWLEDGE ON DRAWING
INFERENCES FROM CONVERSATIONS

by

TONY MATTHEW DUBITSKY

B.S., State University College of New York at Oswego, 1977

———————————————

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Psychology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1980

ABSTRACT

The present research investigated the effects of contextual knowledge on response latencies and the remembered truth values of inferences drawn from ambiguous and non-ambiguous conversational dialogues.

The effects of contextual knowledge on comprehension previously reported in the prose-processing literature were replicated here for dialogue materials, thus extending the generalizability and ecological validity of these findings. Faster comprehension was found for inferences consistent with the context than for context-inconsistent inferences, and an explanation based on "script" theory was put forth to account for the results.

In addition, reliable processing differences were found among various types of Neutral inferences which were independent of whatever contextual knowledge subjects had. With respect to Neutral inferences, both true and false types based on implicit information in the conversations were rated closer to "indeterminate" in truth value and took longer to verify than their counterparts based on explicit dialogue information.