

19

TEXTUAL INFERENCES: MODELS AND MEASURES

Gail McKoon and Roger Ratcliff

Northwestern University

INTRODUCTION

The inferences that are made during reading stimulate intriguing questions for research. Some inference processes seem to be automatic and effortless yet they yield quite complex kinds of information. Other inference processes seem to be dependent on the goals, strategies, and contextual situations of the readers. Some inferences are concerned with the relatively small units of reading represented by words; others are concerned with much larger units like event structures or story outlines. Since about 1970, all of these kinds of inferences have been the subject of investigation, and all of these investigations have shared a common problem: finding empirical measures that can be used to investigate the processes and the products of inference. Different investigators have used different measures, but all the measures have eventually come under criticism, and as a result, progress in understanding inference has been less than impressive.

The reason all the different measures have come under criticism is that they are often viewed as just that: empirical measures. Until recently, there has been little effort to determine a theoretical basis from which to relate them to the processes and structures they are intended to measure. This situation is beginning to change. As models in the several areas that impinge on text processing become more sophisticated, accounting for wider and more complex ranges of data, we can begin to look to these models for an understanding of how specific tasks are performed. For each model in any particular area, the implications

that performance on some task has for conclusions about inference processes can be evaluated. Sometimes, conclusions will be the same across all the models in an area, and sometimes they will be different and model specific. Most beneficially, sometimes the models will give new interpretations or hypotheses about inference processes that would not otherwise have been considered. The important point is that, because of the possibility of different conclusions from different classes of models, it is necessary to evaluate inference processes through all available models.

The issue of relating tasks and data to models arises because the kinds of information that are involved in reading cannot be measured directly; instead, they are mapped onto tasks that require the identification of words, the comprehension of sentences, or the recognition that some piece of information was previously presented. Obviously, this mapping is not one-to-one; a word cannot be identified in isolation from its context, a sentence cannot be comprehended without the involvement of meta-level decisions, and memory for a single word from a text will be embedded in the mental representation of the text as a whole. Thus, models of word identification, sentence comprehension (including syntax and semantics), and memory retrieval become essential.

The models to be considered in this chapter are models of word identification and memory retrieval, because these models have been developed in such a way that they can be useful in evaluating tasks that have been designed to measure inference processes. Currently, there are few models of semantic processes, syntactic processes, or meta-level processes that can be used to directly understand such tasks. For example, in some models of word identification, it is clear how inference processes are supposed to impact word identification processes and, in turn, how word identification processes impact specific tasks (e.g., lexical decision). There are no equivalently specific models to show how inference processes interact with syntactic processes to affect some task that is used to measure syntactic structures. However, even without a complete set of models, progress can be made, and this chapter outlines what can and cannot be done with current models.

WORD IDENTIFICATION

Many of the measures that have been used to investigate inference processes were originally developed to investigate the identification of single words. These include lexical decision latency, naming latency, and gaze duration. The idea behind these tasks is to catch word identification processes as they occur in real time, perhaps as they are affected by perceptual variables (e.g., stimulus degradation) but more often as they are affected by contextual variables such as preceding words or sentences. Typically, a word is presented immediately after a context (i.e., on-line), and the time spent looking at the word (gaze duration

in fairly normal reading), the time to decide that the string of letters is a word (lexical decision), or the time to name the word is measured. Word identification processes are also involved when a word is tested on-line for a recognition decision about whether or not it has appeared in the immediately preceding context. All of these tasks have been used to investigate inference processes under the assumption that inferences will affect word identification and therefore performance on these tasks (see McKoon, 1988, for discussion).

Models that account for how it is that inference processes affect word identification fall into two general classes. The first, older, class views word identification as a series of component processes; this class includes the models of Becker (1979), Forster (1981), and Morton, (1969) and, following Balota (this volume), can be labeled magic moment models. The second class presents a more interactive view of word identification, and includes models by Norris (1986; see also Kintsch, 1988) and Cottrell and Small (1983; see also Kawamoto, 1988). The two classes of models share some, but not all, implications for the use of word identification tasks in measurements of inference.

In the first set of models (e.g., Forster, 1981), word identification proceeds by a series of subprocesses, which can be divided into those that affect lexical access and those that occur postlexically. The subprocesses that affect lexical access occur when the context is presented, before presentation of the target word for which processing time is to be recorded. These processes can be speeded by the prior presentation in the context of the target word itself or a high associate of the target word. Access is speeded because the criterion for recognition of the target word is lowered (Morton, 1969) or because the target word is moved forward in a search list (Becker, 1979; Forster, 1981).

The important point to stress for the purposes of this chapter is that processes that affect lexical access need have nothing to do with inference processes. From the point of view of the models, lexical access effects occur in the lexicon and can occur independently of whatever might be going on in the processes of constructing a representation of the meaning of the context. For example, a context sentence about a child's birthday party might facilitate lexical access for the target word *candles*, but this would not indicate anything about the presence of the concept *candles* as an inference in the mental representation of the context sentence. Alternatively, words in the birthday party sentence might not be highly enough associated to *candles* to facilitate lexical access, but the concept *candles* might later be inferred.

In contrast, postlexical processes are processes of word identification that do not occur until after the target word has been presented (Forster, 1981; see also Balota & Lorch, 1986; Lorch, Balota, & Stamm, 1986). In Forster's model, a word to be identified (either in normal reading or in a test situation) is checked against the preceding context for its compatibility or coherence with the preceding text. Context checking is meant to be an inescapable part of word identification and so must be taken into account in any on-line test. In models proposed by Balota

and Lorch (1986) and Lorch et al. (1986), postlexical processes are not mandatory but are a function of decision biases set by the conditions in an experiment.

Because postlexical processes are triggered by the presentation of the target word, they obviously cannot directly reflect inference processes that occur during reading of the context sentence (before the target is presented). For example, identification of *candles* might be facilitated by postlexical processes that found it compatible with the birthday party context, but this could happen whether or not *candles* had been inferred during reading of the sentence. However, if identification of *candles* was not facilitated by postlexical processes, then we might want to say that *candles* was not inferred during reading. The reasoning rests on the assumption that whatever relations between the target and context underlie postlexical processes, they are at least as strong as those that underlie inference processes. If the relations are not compelling enough to affect compatibility checking, then they probably are not compelling enough to generate an inference. In the example, if the relation between *candles* and the birthday party sentence does not affect word identification, then it seems unlikely that it can generate an inference.

This last point is an important difference between the class of models just reviewed and the second class (Cottrell & Small, 1983; Norris, 1986). In the second class, the relation between a target word and its context sentence can take so much time to compute that it is not available in time to affect word identification. So the absence of an effect of context on identification of a target does not necessarily indicate that the relation is not involved in inference processing.

In this second class of models, context never affects lexical access; perceptual processes produce the same candidates for identification in every context. But context does affect identification, via the decision process. In Norris's model, candidate words are checked against context and the criterion for identification of compatible words is lowered. This criterion change allows faster identification. In Cottrell and Small's model, a word is identified when the amount of evidence in favor of that word reaches a threshold. Faster identification is produced when context adds evidence towards the threshold.

The processes proposed by these models could occur either when a context sentence was presented or when the target was presented, and in fact proceed continually from presentation of context through a decision on the target. In Norris's model, the criterion for identification could be lowered either by an inference generated during reading of the context or it could be lowered only after the target word was available to relate back to the context. Similarly, in Cottrell and Small's model, context could add evidence toward a threshold as the result of an inference generated from the context or only as the result of working backward from the target. So if the relation between the sentence about the birthday party and the target word *candles* affected word identification, it might or might not reflect an inference generated by the context alone.

Both of these models contain an explicit assumption that the computations that produce context effects vary in the amount of time they require. Some, like the connections between strong associates, can be computed very quickly. Others may take more time. In fact, they may take so much time that identification of a target may be accomplished before they finish. Thus, as mentioned earlier, the absence of an effect of context on word identification is not necessarily an indication of the absence of an inferred relation.

The two classes of models taken together provide a guide to interpretation of results of experiments that use word identification tasks to investigate inference processes. Any result must be checked against all the possible explanatory mechanisms of the models. To continue the birthday party example, facilitation of the target *candles* might be due to facilitation of lexical access (Becker, 1979; Forster, 1981; Morton, 1969), postlexical context checking (Balota & Lorch, 1986; Lorch et al., 1986), or information computed when the target was presented (Cottrell & Small, 1983; Norris, 1986). In all three cases, inference processes would not be implicated. If the birthday party context did not facilitate identification of *candles*, it might be because the computation of the necessary relation was too slow (Cottrell & Small, 1983; Norris, 1986). So again, a conclusion about whether an inference was constructed from the context would not be warranted.

All of the mechanisms of the models appear to apply to all of the tasks mentioned. The pre- and postlexical processes proposed in models like Forster's (1981) were constructed for reading words in context, and apply to lexical decision, reading of single words, and recognition of single words. It has been argued that postlexical processes are less likely to be involved in naming, but more recently it appears that they can be (Balota, Boland, & Shields, 1989). The more interactive processing systems of Norris (1986) and Cottrell and Small (1983) are also intended to apply to all of the tasks.

Finally, it should be mentioned that these models are not complete. None of them explicitly allows for other kinds of information to affect tasks that measure word identification. For example, it has been shown that syntactic processes can affect the speed of word identification (cf. West & Stanovich, 1986). Target words presented in different syntactic contexts could have different response times due to differences in syntactic processing; a target presented at the end of a sentence might have faster or slower response times than a target presented in the middle of a phrase. Also, the models do not deal with such discourse effects as anaphora. To use the phrase *the clothes* to refer to some previously mentioned *clothes* might require processing to establish the joint reference even though exactly the same word, *clothes*, was repeated; there is nothing in the models to compare this kind of processing to inference processing (e.g., it might take more or less time). Neither do the models include meta-level processes. Subjects can translate their (lack of) motivation, surprise, or bewilderment into response time differences, but these variables are outside the scope of the current models.

MEMORY RETRIEVAL

In the many experiments in which inference processes have been investigated by presenting a test item immediately after a context, the relevant information is assumed to be available in a short-term or working memory. But in other experiments, the goal is to find some indication of inferred information in the memory representation of a text, and test items are presented so that the text is no longer immediately available. To interpret these latter experiments, models for retrieval of information from memory are required.

Currently, there are several such models, all impressive in the range and detail of experimental results that they can explain. Global memory theories (Gillund & Shiffrin, 1984; Hintzman, 1986; Murdock, 1982; Ratcliff, 1978; Ratcliff & McKoon, 1988a) have explained data from recognition, recall, frequency judgments, categorization, and various reaction-time paradigms. The models differ in such respects as whether they assume that there are different nodes for each concept (Gillund & Shiffrin, 1984) or assume that concepts are distributed across a number of nodes (Hintzman, 1986; Murdock, 1982), and whether they assume that information is kept separate for each item (Gillund & Shiffrin, 1984; Hintzman, 1986) or assume that all information is collapsed into a single memory trace (Murdock, 1982). But the models converge in several respects relevant to the use of memory tasks to investigate inference.

First, the models all postulate one underlying mental representation for an encoded event, the same representation to be used for all memory retrieval tasks. However, different tasks require different processes to operate on this representation. The tasks most often used in inference research are recognition and recall. For recognition, the models all assume that a decision about a test item is based on parallel (global) access to all the items stored in memory. The recognition decision reflects the overall familiarity or match of the test item to all the items in memory (Ratcliff, 1978). Recall and cued recall are assumed to be based on an iterative search that is slower than the parallel comparison used for recognition, and involves some degree of serial processing (Metcalf & Murdock, 1981; Raaijmakers & Shiffrin, 1981). The second point of convergence of the models is that retrieval is assumed to be cue dependent (Tulving, 1974, 1983). A test item is not matched against memory in isolation but instead is matched as part of a compound of information made up from the item plus its retrieval context. Thus, a test item may match information in memory quite well in one context but poorly in another context (see Ratcliff & McKoon, 1988a, 1988b, for discussion). Finally, the models all define the familiarity or strength of retrieval of an item as a matter of degree. A test item presented for recognition or a cue presented for recall will not match other information in memory in an all-or-none fashion but instead will match that information to some degree.

Unlike the models for word identification, the memory retrieval models are in remarkable accord with respect to the general aspects of the

retrieval processes that have implications for the use of memory tasks to investigate inference. Even ACT* (Anderson, 1983), which might at first glance seem very different, embodies cue dependent retrieval, global memory matching, and a distinction between recognition and recall processes. Thus, it is a relatively straightforward task to apply the models to retrieval of possibly inferred information. However, the implications of the memory models, as well as the word identification models, for research on inference processing are not necessarily obvious. The best way to make them clear is to give specific examples. The final two sections of this chapter provide case studies of application of both kinds of models to the investigation of particular kinds of inferences.

CASE STUDY I: ELABORATIVE INFERENCES

Elaborative inferences have been studied extensively since the beginning of current interest in inference processes. These inferences have been of special interest because they go beyond what is actually required to connect the explicitly stated ideas in a text, and so could begin to test the limits of inference processes. Bransford, Barclay, and Franks (1972) and Bransford and Franks (1971) argued that a reader constructs a mental model of the situation described in a text, adding information to complete the model and combining the elements of a text into an integrated whole.

For this chapter, one kind of inference from our own research has been chosen as a case study with which to examine the implications of various experimental techniques for addressing questions about the construction during reading of elaborative inferences. The kind of elaborative inference is one that we have investigated in a series of experiments, and concerns predictable events; an example is shown in Table 19.1. The predicting sentence was written to predict that the actress would die, and this inference was expressed by the test word *dead*.

TABLE 19.1

Examples of Stimuli Used in Elaborative Inference Study.

Predicting:

The director and the cameraman were ready to shoot closeups when suddenly the actress fell from the 14th story.

Test word: *dead*

Control:

Suddenly the director fell upon the cameraman, demanding closeups of the actress on the 14th story.

Test word: *dead*

The research question that has been addressed in most studies of elaborative inferences is whether or not the predictable event is inferred during reading of the predicting sentence. This question has been addressed using most of the tasks already discussed: the on-line tasks of recognition, lexical decision, and naming, and the delayed memory tasks of cued recall and recognition. Interpretations of results from these tasks can be discussed in light of the models of word identification and memory retrieval reviewed previously.

On-line Tasks

One possible experiment would be to present a predicting sentence of the kind shown in Table 19.1, and then follow it immediately with a word representing the predicted event (*dead*) for lexical decision. There would be two possible results; either the sentence could facilitate the lexical decision or the sentence could have no effect on the lexical decision (the third possibility, inhibition, seems unlikely).

According to the models of word identification, facilitation could arise for several reasons. First, there might be individual words in the predicting sentence that were highly semantically associated to the target word. These words could speed lexical access for the target (Becker, 1979; Forster, 1981; Morton, 1969), lower the threshold for identification (Norris, 1986), or increase the amount of information leading to identification of the target (Cottrell & Small, 1983). None of these mechanisms would require that an inference about the predicted event was constructed. However, these interpretations of facilitation can be ruled out by using a control sentence like that shown in Table 19.1. The sentence includes all the words from the predicting sentence that might possibly be associated to the target word, yet the sentence does not predict the target. So if facilitation in lexical decision is obtained for the predicting sentence relative to the control sentence, then it cannot be due to word-to-word associations.

A second reason that facilitation might arise according to the models would be that a relation between the target word and the predicting sentence was constructed at the time the target word was presented. This could come about through postlexical processes (e.g., Forster's model, 1981) or through criterion or threshold changes (Norris's model or Cottrell and Small's model). In neither case would facilitation indicate that information about the inference was constructed during reading of the predicting sentence.

The alternative possible result is that the predicting sentence would *not* facilitate a lexical decision on the target word. In the interactive models, this could occur if the inference was not constructed at all or if it was constructed but too slowly to affect the lexical decision. From the point of view of models of the pre- and postlexical type, lack of facilitation would indicate that the inference was not constructed, because if it were constructed it should affect postlexical processes.

The question that arises from all of these possible interpretations of different possible results is whether there is some consistent subset of them that can fit data. The answer appears to be that there are in fact two such subsets. The results of several experiments are shown in Table 19.2. When presentation of the predicting sentence is slow (or a number of words intervene between the target word and the point in the predicting sentence where the inference could be generated), then facilitation is observed. When presentation is faster and no extra words intervene, then there is no facilitation. This pattern of data is observed both for lexical decision and for recognition.

TABLE 19.2

Results from On-line Experiments with Elaborative Inference.

Lexical Decision (Slow); (Potts, Keenan, & Golding, 1988).

Predicting	980 ms (3%)	
Control	1034 ms (11%)	sig.

Lexical Decision (Fast); (McKoon, 1988).

Predicting	651 ms (4%)	
Control	645 ms (6%)	n.s.

Recognition (Slow); (McKoon & Ratcliff, 1986).

Predicting	883 ms (7%)	
Control	853 ms (6%)	sig.

Recognition (Fast); (McKoon & Ratcliff, 1989c)¹

Predicting	768 ms (18%)	
Control	748 ms (16%)	n.s.

Naming latency (Slow); (Potts, Keenan, & Golding, 1988).

Predicting	423 ms	
Control	422 ms	n.s.

Note: Slow and Fast refer to the amount of time available for generating the inference between presentation of context material and test. Response times and error rates are both shown; *sig.* indicates that results of analyses of variance were significant for at least one of the measures.

¹ The same result was obtained by Till, Mross, & Kintsch, 1988.

One way to look at this whole pattern of data and the models is to argue that the pattern is consistent with the interactive models and the assumption that the inferences are relatively slow to generate. The slow inferences affect decisions when there is enough time before the target word, and they do not affect decisions when the target is presented too quickly. This way of looking at the data would not be consistent with the pre- and postlexical models; if postlexical processes were operating with slow presentation, then they should also be operating with fast presentation. These models could be made consistent with the data by adding an assumption that inferences are slow to generate, but unlike the interactive models, they currently have no mechanism by which to implement such an assumption.

The data from naming latency could be fit into this pattern if it were assumed that lexical information enters a naming production system faster than it is available for the other tasks, and enters production without the benefit of slower, constructed, semantic kinds of information (Seidenberg, this volume). In this way, naming latencies could be free from any effects of inferred information. As an aside, it should be noted that there is one case where facilitation of naming was obtained for a target word that represented an inference (Potts, Keenan, & Golding, 1988). In this case, the target word was preceded by two sentences, a predicting sentence and a second sentence that required the target inference. With the second sentence, very specific information about the inference could have been constructed, and so an effect on naming latency might be expected. The information might have been articulatory (the subject "saying" the target word), or it might have been only conceptual (in which case it might not affect performance on a Stroop task; see Keenan, Potts, Golding, & Jennings, this volume).

So far, one way of looking at the pattern of data illustrated in Table 19.2 has been discussed; this way involved assuming that the target inferences are generated in response to the predicting sentences. A second way to look at the pattern of data is to suppose that the elaborative inferences are not generated during reading of the predicting sentences at all. The positive results in lexical decision and recognition would be due to postlexical processes (e.g., Forster, 1981) or to relations between the target word and predicting sentence computed when the target was presented (Cottrell & Small, 1983; Norris, 1986). The absence of facilitation with fast presentation would be a problem for the postlexical processes models, but they could add an assumption that these processes require more time than available with fast presentation.

Obviously, these two ways of viewing the data are contradictory; one assumes that elaborative inferences are generated during reading (albeit slowly) and the other assumes that they are not. The conclusion therefore is that some other method of investigating these kinds of inferences is needed. Later, it is argued that a memory task, speeded recognition, can provide such a method.

Memory Tasks

To investigate whether the inference represented by *dead* is included in the mental representation of the predicting sentence, subjects could be asked (among other tasks) to recall the sentence without a cue (free recall), recall the sentence given the cue *dead*, or decide whether the word *dead* had appeared in the sentence (recognition). Global memory models suggest interpretations for possible results in each case.

In cued recall, a list of sentences, some predicting and some control, is presented. Then after the end of the list, single word cues are given and the subject's task is to produce a studied sentence for each cue. According to the models, the cue is used to generate an iterative search of memory, a process that may take some time. If the cue word, *dead*, was encoded with the predicting text, then at some point in the search process, the text should be retrieved. However, even if the cue word was not encoded with the text, it could still be the case that self-generated cues from the meaning of the word *dead* would be used to probe memory and eventually produce the text (cf. Corbett & Doshier, 1978; McKoon & Ratcliff, 1986). Thus, according to the models, if the target word was a better cue for the predicting than the control sentence (as shown in Table 19.3), it could be either because the inference was generated and encoded during reading of the predicting sentence or because it was generated later as part of the memory retrieval process.

TABLE 19.3

Results from Delayed Memory Experiments with Elaborative Inference.

Cued Recall; (McKoon & Ratcliff, 1986).			
Correct recall rate:	Predicting	23%	sig.
	Control	4%	
Speeded Recognition; (McKoon & Ratcliff, 1986). ¹			
Error rate:			
Neutral Prime	Predicting	36%	n.s.
	Control	27%	
Prime from Text	Predicting	56%	sig.
	Control	24%	

¹ This pattern of data has been replicated in McKoon, 1988, and McKoon and Ratcliff, 1989a, 1989b, 1989c.

In free recall, the same list of sentences could be presented and then subjects could be asked to recall as many as they remembered, without experimenter-provided cues. An experiment like this with materials of the type shown in Table 19.1 has not been done. Interpretation of results would not be straightforward because of the possibility of all kinds of inferential processes that might take place during recall, as retrieval of some information led to the retrieval and generation of more information. Whether information inferred during reading would be produced at retrieval is not clearly predictable from the models; they have been formulated only to deal with explicitly presented information.

Recognition can be tested either by allowing subjects to respond in their own time or by giving them a deadline to require them to respond quickly. In the first case, processing may become slow enough that the mechanisms are the same as in cued recall, and any evidence of inferences could be due to generation of the inference information at the time of the test.

If recognition is speeded, with response times around 600 ms or less, then according to the models, the recognition decision is based on a global parallel match between the target word in its context and all information in memory. The models assume a fast parallel matching process because of direct empirical evidence about the time course of retrieval (Doshier, 1982, 1984; Gronlund & Ratcliff, 1989; Ratcliff & McKoon, 1982, 1989; Reed, 1973; Wickelgren, Corbett, & Doshier, 1980) and because of indirect evidence that such a process is required to account for large ranges of data. The parallel matching process produces a goodness of match value that determines the speed and accuracy of the decision. For a target word that represents an inference, a positive match must be based on a relation between the target word and information encoded when the predicting text was read. The global parallel match process would not allow for the generation of new information in the 600 ms before a response (neither would the spreading activation/production system of ACT*). If there is a positive match between the target and memory for the predicting text, a correct (negative) decision will tend to be more difficult.

The global memory models stress the cue dependent nature of recognition, and the recognition of targets representing elaborative inferences exhibits this cue dependence. In Table 19.3, when the target is primed by a word from the studied sentence, the match between the target and the predicting sentence is large enough to inhibit the decision process, relative to the control condition. But when the target is presented by itself, with only a neutral word for a prime, the decision process is not inhibited. This pattern of results can be interpreted as showing some degree of partial match or partial encoding of the inference (see McKoon & Ratcliff, 1986, 1988, 1989b, 1989c, in press, for discussion).

An alternative interpretation of this pattern of results was proposed recently by Potts et al. (1988). They suggested that a prime from a predicting sentence activates the predicting sentence, and then compatibility is computed between the target word and the activated sentence.

If the target word is compatible, then the response to it will be inhibited, even though no inference about it was made during reading. This proposal can be rejected on several grounds. First, the computation of compatibility within 600 ms is not consistent with a large body of data on the time course of information processing in recognition (cf. Doshier, 1984; Gronlund & Ratcliff, 1989; Ratcliff & McKoon, 1982, 1989); all of these data show that information about relations among items, information similar to what would be needed for the computation of inferences, is not available early enough in processing to affect speeded recognition. Second, McKoon and Ratcliff (1989a) have shown that recognition performance is not predicted by compatibility ratings in the way the compatibility hypothesis would require. According to the compatibility hypothesis, ratings should be more highly correlated with recognition performance in the condition with a prime from the text than in the neutral condition, because the prime from the text makes it more likely that compatibility will be calculated against the correct one of the several studied texts. In fact, ratings did not correlate more highly in the condition with the prime from the text. Finally, McKoon and Ratcliff (1989a) showed that there are inference target words that are compatible with their predicting texts but that do not show inhibition in speeded recognition. For all these reasons, the compatibility hypothesis cannot account for recognition data of the kind shown in Table 19.3.

Summary

The pattern of data in Table 19.3 can be made consistent with models of word identification and models of memory retrieval if it is assumed that elaborative inferences are partially encoded during reading and that the encoding is a relatively slow process. The inferences were not encoded to such a high degree that they were equally available under all retrieval conditions (see the neutral vs. prime from text conditions in Table 19.3), but they must have been encoded to some degree because they did inhibit recognition when combined with a cue from their text. This evidence from the memory tasks rules out one interpretation of the on-line data in Table 19.2, the interpretation that the inferences were not encoded at all. As a consequence, the remaining interpretation of the on-line data, slow processing, is supported (see McKoon & Ratcliff, 1986, 1988, 1989b, 1989c, in press, for further discussion).

CASE STUDY II: ANAPHORIC REFERENCES

The conclusion from the case study of elaborative inference is that progress is made by the method of combining converging empirical results with careful consideration of all possible theoretical accounts of the results. This is the same conclusion that can be drawn from recent work on anaphoric reference. Examples in this section are limited to work from our laboratory on category names used as anaphors, but the

empirical and theoretical points are based also on work by Corbett and Chang (1983).

The questions that have been raised in this work concern, first, the process of connecting an anaphor to its antecedent, and second, the time course of this inference process. These questions can be illustrated using the example text shown in Table 19.4. In the first version of the text, the final sentence contains the anaphor *the criminal*, which should be connected by inference to its referent *burglar*. The second version of the text ends with a control sentence that mentions some word that is not supposed to refer to the burglar mentioned in the first sentence. To investigate whether *criminal* is understood to refer to *burglar*, both on-line and memory tasks can be used.

TABLE 19.4

Examples of Texts Used in Anaphoric Reference.

Version 1

A burglar surveyed the garage set back from the street.
 Several milk bottles were piled at the curb.
 The banker and her husband were on vacation.
 The criminal slipped away from the streetlamp.

Version 2

A burglar surveyed the garage set back from the street.
 Several milk bottles were piled at the curb.
 The banker and her husband were on vacation.
 A cat slipped away from the streetlamp.

Test words:

Referent - *bottles*

From the same proposition as burglar - *garage*

Control word - *bottles*

One way to test whether *the criminal* is understood to refer to *the burglar* might be to present the target word *burglar* both before and after the referent, and compare response times (this has been suggested by O'Brien, Duffy, & Myers, 1986). The idea would be that response times to the target would be speeded for the after test relative to the before test, because the target concept would be processed as part of the inference connecting the anaphor to the referent. However, this possible test has problems. If the target concept were in short-term memory, then response times might be at floor even in the before test. And if response times were reduced in the after condition, it could be because of syntactic effects on decision times (as previously mentioned,

even though word identification models do not address the interaction of syntactic and lexical processes directly, they would still allow them to affect response times). For example, the before position is in the middle of a noun phrase and might have slow response times relative to the after position, which is at the end of a noun phrase.

Another way to test whether *the criminal* is understood to be the same person as *the burglar* might be to present the anaphoric and control texts for reading, and then immediately after the final sentences, present the word *burglar* for test. If the inference connecting *criminal* to *burglar* was constructed during reading of the final sentence of the anaphoric text, then a decision about the test word *burglar* should be facilitated. However, this facilitation could have several alternative interpretations other than that the inference was constructed during reading of the anaphoric sentence. The facilitation could have come about because the preexisting semantic relation between *burglar* and *criminal* speeded lexical access for *burglar*. Or, it could be that the facilitation was the result of a more compatible relation (constructed when the target was presented) between the target word and the anaphoric text than between the target word and the control text. In neither of these cases would inference processes during reading of the anaphoric sentence be implicated.

These alternative interpretations are all dependent on the preexisting relation between *criminal* and *burglar*. So to rule out these interpretations, a test word without such a relation is required. One such target is *garage*. If *the criminal* is connected to *burglar* by some inference process, then this process might also connect propositions about *criminal* and *burglar* and so involve other concepts in the propositions, such as *garage*. Then response times for *garage* would be facilitated after the anaphoric sentence relative to the control sentence. Because there are no preexisting semantic relations between *criminal* and *garage*, the facilitation could not be due to a speed up of lexical access for *garage*. Furthermore, the relations that could be computed between the target *garage* and the anaphoric sentence do not seem more compatible than those that could be computed between the target and the control sentence.

Using *garage* as a target word rules out the alternative interpretations most obviously suggested by word identification models. However, there might be other kinds of processes, at the syntactic or discourse levels, that differed between the anaphoric and control texts. Instead of speculating about what these possible differences might be, it is easier to measure them by the use of a control target word. In the example in Table 19.4, the word *bottles* has no particular relation to either *the criminal* or *a cat*, so response times to *bottles* used as a target should be the same whether it is tested with the anaphoric sentence or the control sentence. If response times for *bottles* were not the same in these two conditions, it would be an indication that the two sentences were different in some way that had nothing to do with the anaphor. Then response times for the target of interest, *garage*, would have to be evaluated against the response times for *bottles*.

The logic outlined here was followed by McKoon and Ratcliff (1980) and Dell, McKoon, and Ratcliff (1983). Using on-line recognition, they showed that both the target words *burglar* and *garage* were facilitated by the anaphoric sentence relative to the control and that response times for the control word, *bottles*, did not differ for the two sentences. Thus, it seems reasonable to suppose that an inference about the relation between *the criminal* and *burglar* was constructed during reading of the anaphoric sentence. Additional evidence that the inference was constructed during reading was provided by a recognition memory experiment that showed that responses to a word from the same proposition as *the criminal*, *streetlamp*, were facilitated by the prime *burglar* when the anaphoric sentence had been read relative to when the control sentence had been read.

This case study shows the power of using models as guides to experimental design. With anaphors of the type examined here, all of the mechanisms proposed by the models to be involved in word identification can be taken into consideration, and sources of facilitation for a target word that are not due to inference processes can be eliminated. Additionally, the memory models allow recognition data to confirm the results of the on-line experiments.

CONCLUSIONS

The aim of this chapter has been to illustrate how essential theoretical models are in the interpretation of data from empirical tasks. None of the tasks that has been used to investigate inference can be taken as an unequivocal measure of inference processes. Only when both theory and experimental design converge can conclusions be ventured and progress toward an understanding of inference be claimed.

ACKNOWLEDGMENT

This research was supported by NSF grant 85-16350 to Gail McKoon and NSF grant 85-10361 to Roger Ratcliff.

REFERENCES

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Balota, D., Boland, J., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, 28, 14 - 36.

- Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 336 - 345.
- Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 252 - 259.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3, 193 - 209.
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331 - 350.
- Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition*, 11, 283 - 294.
- Corbett, A. T., & Doshier, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 17, 479 - 491.
- Cottrell, G. W., & Small, S. (1983). A connectionist scheme for modeling word sense disambiguation. *Cognition and Brain Theory*, 6, 89 - 120.
- Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior*, 22, 121 - 132.
- Doshier, B. A. (1982). Effect of sentence size and network distance on retrieval speed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 173 - 207.
- Doshier, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, 16, 519 - 555.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, 33A, 465 - 495.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1 - 67.
- Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 846 - 858.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411 - 428.
- Kawamoto, A. H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolution in the comprehension of human language* (pp. 195 - 228). San Mateo, CA: Morgan Kaufman.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163 - 182.
- Lorch, R. F., Balota, D. A., & Stamm, E. G. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre- or postlexical access? *Memory & Cognition*, 14, 95 - 103.
- McKoon, G. (1988). *Word identification and elaborative inference: The relation between theory and empirical measurement*. Unpublished manuscript.
- McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, 19, 668 - 682.

- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 82 - 91.
- McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 331 - 343.
- McKoon, G., & Ratcliff, R. (1989a). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language*, 28, 547 - 563.
- McKoon, G., & Ratcliff, R. (1989b). Inferences about contextually-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1134 - 1146.
- McKoon, G., & Ratcliff, R. (1989c). Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 326 - 338.
- McKoon, G., & Ratcliff, R. (in press). Dimensions of inference. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25). Orlando, FL: Academic Press.
- Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 20, 161 - 189.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165 - 178.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609 - 626.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, 22, 93 - 136.
- O'Brien, E. J., Duffy, S. A., & Myers, J. L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 346 - 352.
- Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, 27, 399 - 415.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93 - 134.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59 - 108.
- Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Human Learning and Memory*, 8, 16 - 36.
- Ratcliff, R., & McKoon, G. (1988a). Memory models, text processing, and cue-dependent retrieval. In F. I. M. Craik & H. L. Roediger (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 73 - 92). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ratcliff, R., & McKoon, G. (1988b). A retrieval theory of priming in memory. *Psychological Review*, 95, 385 - 408.
- Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, 21, 139 - 155.
- Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, 181, 574 - 576.

- Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, *16*, 283 - 298.
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist*, *62*, 74 - 82.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- West, R. F., & Stanovich, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory & Cognition*, *14*, 104 - 112.
- Wickelgren, W. A., Corbett, A. T., & Doshier, B. A. (1980). Priming and retrieval from short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, *19*, 387 - 404.