INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313 761-4700  800 521-0600
Discriminating information source: Inference versus observation

Arnoult, Lynn H., Ph.D.

Rice University, 1990
RICE UNIVERSITY

DISCRIMINATING INFORMATION SOURCE:
INFERENCE VERSUS OBSERVATION

by

LYNN H. ARNOULT

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE

Michael J. Watkins, Professor of
Psychology, Director

John W. Brelsford, Jr., Professor of
Psychology

William C. Howell, Professor of
Psychology

Stephen L. Klineberg, Professor of
Sociology

Houston, Texas

May, 1990
Discriminating Information Source:
Inference Versus Observation

Lynn H. Arnoult

Abstract

People rely extensively on inference as a source of information, and sometimes they confuse inference with observation. Specifically, inferred information is sometimes mistaken for observed information. Such confusion of inference with observation can be problematic, especially if the inferred information is erroneous. One factor that might affect the probability of mistaking inference for observation is the degree of consistency between inferred information and subsequently encountered information. The present research was designed to test this possibility. In three experiments subjects made inferences on the basis of presented information, and then were given additional information that was varied in consistency with the information they had inferred. Finally, subjects were tested for accuracy in discriminating the source (inference vs. observation) of the inferred information. As expected, accuracy was lower when subsequently presented information was relatively consistent with inferred information than when it was relatively inconsistent with inferred information. This effect did not vary with delay between making an inference and attempting to discriminate information source. It is concluded that consistency of inferred information with subsequently encountered information can affect the probability of mistaking inference for observation, with the probability of error increasing as consistency increases.
Acknowledgements

I thank John W. Brelsford, Jr., William C. Howell, and Stephen L. Klineberg for serving on the committee for this thesis.

I acknowledge, with gratitude, the contributions of Michael J. Watkins, director of the thesis committee. I am especially grateful for the generosity with which he gave his time to this project and the thoughtfulness with which he reviewed each phase of the work.
# Table of Contents

Abstract ........................................................................................................ ii

Acknowledgements ....................................................................................... iii

List of Tables ................................................................................................. v

Introduction .................................................................................................... 1

Experiment 1 ................................................................................................. 20
  Method ......................................................................................................... 21
  Results and Discussion .............................................................................. 27

Experiment 2 ................................................................................................. 30
  Method ......................................................................................................... 31
  Results and Discussion .............................................................................. 33

Experiment 3 ................................................................................................. 36
  Method ......................................................................................................... 37
  Results and Discussion .............................................................................. 40

Summary and General Discussion ................................................................. 47

References .................................................................................................... 56

Appendices ................................................................................................... 62
List of Tables

Table 1. Perceived Consistency of Additional Information Statements as a Function of Consistency Condition (Experiment 1) ..................... 28

Table 2. Accuracy of Stated/Inferred Judgments for Inferences as a Function of Consistency Condition (Experiment 1) ..................... 30

Table 3. Accuracy of Stated/Inferred Judgments for Inferences as a Function of Consistency Condition (Experiment 2) ..................... 35

Table 4. Mean Ratings for Inference Items on Second True/False Test by Consistency Condition (Experiment 3) ................................. 43

Table 5. Accuracy of Stated/Not Stated Judgments for Inferences as a Function of Consistency Condition and Rating on First True/False Test (Experiment 3) ..................................................... 46
**Discriminating Information Source:**

**Inference Versus Observation**

Inference is a fundamental part of everyday life. People routinely derive more information from their experiences than that provided by sensory observation alone (Bartlett, 1932; Bruner, 1957; Clark, 1977). Bartlett (1932) demonstrated the importance of inference in behaviors ranging from performance of the most basic perceptual tasks to memory of complex materials. And Bruner (1957) pointed out that, except for the fact that people perceive, the most distinctive feature of mental life is the fact that people constantly infer information beyond that which is given.

**Types of Inference**

Several psychologists have begun their studies of inference with attempts to develop a typology. Bruner (1957) described three forms of inference: categorization, prediction of covariates, and organization of information according to certain formal coding rules (e.g., rules about transitive relations or serial completions). Other inference typologies have been proposed in more recent work. One distinction often found in the current literature is between logical and pragmatic inference (e.g., Brewer, 1977; Harris & Monaco, 1978). According to those who draw this distinction, information derived through logical inference follows necessarily from given information, whereas that derived through pragmatic inference does not.

Inferences have also been classified according to topics of particular interest to those working in certain areas of specialization within psychology. For instance, in a review of social inference literature, Hastie (1983) distinguished four types of inferences: category membership, causal, moral,
and structural balance inferences. And in reporting a study of inferences generated while reading, Seifert, Robertson, and Black (1985) mention several inference types frequently encountered in the cognitive literature: anaphoric, causal, action, state, goal, plan, instrument, and thematic inferences.

**Varied Research Perspectives**

Inference has been the subject of much research, and has been studied from a number of different perspectives. Indeed, research on inference is common to virtually every area of psychological study.

Several cognitive psychologists, for example, have investigated the role of inference in verbal communication. Particular attention has been given to demonstrating that specific inferences have in fact been made, to determining precisely at what point in the course of comprehension inferences are made, and to establishing the reliability with which inferences will be made (e.g., Harris & Monaco, 1978; McKoon & Ratcliff, 1989; Singer & Ferreira, 1983).

In social psychology the emphases have been on determining how evidence is used in making inferences and how inferences affect social perceptions and relations. Researchers interested in causal inference, for example, have conducted hundreds of studies on how causal relations are determined and how causal attributions affect social behavior. (See, for example, classic papers by Jones and Davis, 1965; Kelley, 1967; and Weiner, 1974.)

**Accurate and Erroneous Inferences**

For the most part inferences are valid, as evidenced by the generally successful behavior of people in carrying out their daily activities. Without valid inference one would be poorly equipped to understand past events, to respond appropriately to present environmental demands, and to make reasonable
predictions about the future. When inference errors do occur, a person may be less well-appointed for negotiating the world.

Inference errors have been the central concern of many studies of inference, particularly in the area of judgment. An important feature of much of this research is the comparison of human inference performance to normative models of inference. Repeatedly, human inference has been found to deviate from normative models, resulting in inference error. Although such comparisons are controversial (Funder, 1987; Hogarth, 1981), the large body of research in this area clearly shows that inferred information is not always correct. Investigators have documented errors in inference tasks such as judging contingency (Jenkins & Ward, 1965), estimating probability (Eddy, 1982), evaluating risk (Slovic, Fischhoff, & Lichtenstein, 1982), predicting category membership (Kahneman & Tversky, 1973), and judging causality (Ross & Anderson, 1982). Inference errors, it is thought, are often due to a combination of factors, including: (a) failure to utilize normative inferential strategies, such as those used in logical or statistical inference; (b) inappropriate use of preexisting beliefs or knowledge (e.g., ideas about typical relations among people, objects, and events that are either incorrect or not relevant to the judgment at hand); and (c) inappropriate application of generally helpful judgmental heuristics. (See, for example, Nisbett & Ross, 1980; Tversky & Kahneman, 1974).

Inference errors, whatever their causes, can leave people with incorrect information. This problem might be compounded if an inference is based, not on observation, but rather on another inference, which in turn carries its own risk of error (Schum, 1980; Wigmore, 1931).
After an inference error has occurred, there remains a possibility that the error will be corrected, perhaps through consideration of some additional information, or through recognition of one's own faulty logic. This possibility is diminished if the inference is mistaken for an observation: Common experience shows that people generally don't question their own observations. Mistaking an inference for an observation, then, can be a further complication of an inference error, and evidence of such confusion can be found in several studies involving inferences. To illustrate the variety of conditions under which experimental subjects have mistaken inference for observation, representative studies are reviewed in the following section.

**Confusion of Inference With Observation**

In one of the earliest studies of memory for narratives, Bartlett (1932) tested subjects for memory of an American Indian legend, "The War of the Ghosts." In free recall subjects introduced changes in story content, and some of these changes were inferences. For example, one part of the story read, "His face became contorted." A subject's recall read, "... his face became very pale. Then he writhed and shrieked...." Because much of the content of this story is culture specific, it was not consistent with the prior knowledge of Bartlett's subjects, who were not Indians. Many of the inferences that occurred in recall were culture related. That is, they resulted in changes of story content that made the story more compatible with the subjects' own culture. It appears that subjects who included inferences in their recall thought that the inferred information had been part of the presented story: They remembered a particular item of information, but were confused about its origin. This sort of confusion is also evident in several later studies.

Spiro (1980) conducted an experiment similar to Bartlett's classic study
and obtained similar results. Subjects read a story about a couple who were planning to marry, and were given a recall test at a later session. In one condition the story described a disagreement between the potential marriage partners concerning the desirability of having children. Before being dismissed from the first session, subjects in this condition were told that the couple had carried out their plan to marry and were happy. Spiro predicted that these subjects would find that outcome inconsistent with the disagreement about having children and, in recall, would change the story to reduce this inconsistency. As predicted, subjects made accommodative errors in recall, often involving inferences. For example, one subject recalled that, "They underwent counseling to correct the major discrepancy," and another recalled that, "They discussed it and decided they could agree on a compromise: adoption." Neither counseling nor adoption had been mentioned in the presented story.

An experiment by Bransford, J. R. Barclay, and Franks (1972) utilized pairs of sentences like the following:

(1a) Three turtles rested on a floating log and a fish swam beneath them.

(1b) Three turtles rested on a floating log and a fish swam beneath it.

(2a) Three turtles rested beside a floating log and a fish swam beneath them.

(2b) Three turtles rested beside a floating log and a fish swam beneath it.

Subjects were presented with either sentence 1a or sentence 2a and were tested for recognition of implication sentence 1b or 2b, respectively. False recognition of the implication sentence was much greater when the presented sentence contained the word on than when it contained the word beside. This
indicates that subjects confused presented sentences with similarly worded implications when the implications were logically consistent with the presented sentences.

A series of experiments by J. R. Barclay (1973) showed that subjects presented with sets of related sentences would infer information not derivable from any one sentence alone. For example, in one experiment sentences like 3a and 3b, presented together, would lead to production in free recall of an unpresented sentence like 3c.

(3a) The artist is taller than the salesman.
(3b) The salesman is taller than the welder.
(3c) The artist is taller than the welder.

Inference is the source of sentence 3c, but subjects thought it had been included in the study list.

Johnson, Bransford, and Solomon (1973) presented subjects with sentences that implied the use of certain objects, like sentence 4a, which implies the use of a car.

(4a) ...the man drove to work.

They found that subjects hearing this sentence would later falsely recognize sentence 4b.

(4b) ...the man drove his car to work.

Likewise, when a presented sentence contained an action that implied a consequence, subjects would falsely recognize test sentences explicitly stating the consequence. For example, sentence 5a led to false recognition of sentence 5b.

(5a) The spy threw the secret document into the fireplace.
(5b) The spy burned the secret document.

These false recognition responses show, again, a confusion of implication with observation.

This experiment differs from the two experiments immediately above in an important way: The implications falsely recognized in this experiment were not necessarily true. The man going to work could have driven some vehicle other than a car, and the secret document was not necessarily burned. These implications are of the type often called pragmatic (not necessarily correct) as opposed to logical or necessary implications, which cannot be incorrect.

Singer (1980) found false recognition of implications in his study of the role of inferences in reading comprehension. He focused on the effects of three kinds of implications: agent, patient, and instrument implications. An agent is an actor in a situation. A patient is a recipient of an action. An instrument is an object used to perform an action. In sentence 6, for example, boy is the agent, ball is the patient, and bat is the instrument.

(6) The boy hit the ball with the bat.

Singer presented subjects with passages that implied an agent, a patient, or an instrument. Subjects later falsely recognized sentences that included the implied elements. For instance, subjects hearing passage 7a falsely recognized sentence 7b.

(7a) The worker drove the nail. The tool was too small for the task.

(7b) The worker drove the nail with the hammer.

The hammer had only been implied in passage 7a, but subjects thought that sentence 7b had been presented verbatim in the passage.

Dichotomous and continuous antonyms were used by Brewer and Lichtenstein (1975) in an experiment on recall of implied information. Whereas
the negation of a dichotomous antonym necessarily implies its opposite (e.g., not true necessarily implies false), the negation of a continuous antonym does not necessarily imply its opposite (e.g., not hot does not necessarily imply cold). In recalling presented sentences, subjects made errors involving both kinds of antonyms (e.g., not true recalled as false, not hot recalled as cold). The authors interpreted these results as evidence that subjects hearing sentences containing implications, whether or not the implications are necessarily true, tend to make the suggested inferences. In recall, they do not remember that they have made the inferences and erroneously recall the implications in place of the presented sentences.

Schweller, Brewer, and Dahl (1976) found that subjects hearing a sentence containing a reported utterance would confuse this sentence with another sentence containing either a potential effect of the utterance on hearers or the speaker's possible intent. For example, subjects presented with sentence 8a falsely recognized sentence 8b, which assumingly contains the effect of 8a on the hearers.

(8a) The English professor told his students a dull story about Jane Austin.
(8b) The English professor bored his students with a story about Jane Austin.

Similarly, presentation sentence 9a was recalled as sentence 9b, which assumingly contains the speaker's intent.

(9a) The housewife spoke to the manager about the increased meat prices.
(9b) The housewife complained to the manager about the increased meat prices.
(For related work on inferring communication intent and utilizing intended meaning see Clark and Lucy, 1975; Haviland and Clark, 1974.)

In an experiment by Harris (1974), subjects heard a list of complex sentences like sentence 10a, each of which contained a complement sentence as the object of the main verb.

(10a) Miss America said that she played the tuba.

Subjects were later asked to judge the truth value (true, false, or indeterminate) of each complement sentence, for example, sentence 10b.

(10b) Miss America played the tuba.

"Comprehension subjects" made their judgments immediately after each sentence. "Memory subjects" made their judgments after the entire list. Results showed that accuracy in discriminating between implied and presented information decreased with time. Complement (implication) sentences that were not necessarily true were usually judged as having indeterminate truth value by comprehension subjects, whereas they were usually judged as true by memory subjects. Harris concluded that distinctions between different types of remembered information, such as assertions and inferences, are sometimes obscured in memory.

Harris and his colleagues extended his work on inference to simulations of real-life settings. In one experiment (Harris, Teske, & Ginns, 1975), subjects heard mock courtroom testimony and then judged the truth value of test sentences, some of which contained information that had been implied rather than asserted in the testimony. Sentences 11a and 11b were presentation and test sentences, respectively.

(11a) That absent-minded Herman didn't have his walkie-talkie either.
(11b) Herman Lyons, the night watchman, had lost his walkie-talkie. Although the implied information was not necessarily true, subjects generally judged it to be true, even if they had been specifically instructed to resist being affected by implications. Similarly, Harris (1977) found that subjects judged as true information implied rather than asserted by advertisements. In this case, warnings about misleading implications helped subjects avoid errors, but only when testing occurred immediately after presentation. When testing was delayed for five to ten minutes, instructions were not helpful.

Brewer (1977) tested subjects' recall of a list of sentences like sentence 12a.

(12a) The clumsy chemist had acid on his coat. Subjects frequently failed to recall the sentences as presented, producing instead sentences that included implications, such as sentence 12b.

(12b) The clumsy chemist spilled acid on his coat. The percentage of sentences in which implications were included in recall was greater than the percentage of sentences correctly recalled, indicating considerable confusion of implied with presented information.

Brewer noted that the probability of including an implication in recall was greater for some of the sentences than for others. He hypothesized that the surface structure of the sentences might affect the probability that subjects would make inferences and produce implications in recall. Specifically, he thought that the amount of surface information (i.e., number of words) that must be deleted or added to change the sentence might be important. Analyses revealed that the amount of surface information that must be deleted was not related to the inclusion of implications in recall. The amount of surface information that must be added, however, was a reliable predictor of the
inclusion of implications: The more new information required, the lower the probability of including an implication.

Brewer also found that when the nature of the implication contained in the presented sentence was ambiguous, different subjects made different inferences. For example, in sentence 13 "didn't have" can lead to a number of different inferences.

(13) The absent-minded professor didn't have his car keys.

In recalling this sentence Brewer's subjects variously wrote that the professor, "forgot his car keys," "lost his car keys," and "left his keys in the car."

Brewer's (1977) recall paradigm and some of the same sentences, along with some new sentences, were used by C. R. Barclay, Toglia, and Chevalier (1984) in their study of inferences. In their experiment, two factors were varied, the first of these being the experimental task. All subjects were instructed to remember the sentences and told that their memory of the sentences would be tested. One group of subjects received no further instructions. A second ("semantic task") group was instructed to answer a question after each sentence relating to sentence meaning. A third ("nonsemantic task") group was instructed to answer a question after each sentence relating to sentence length.

The second factor varied was the consistency between character descriptors in the presented sentences and outcomes implied by those sentences. Sentence 14 illustrates three versions of the same sentence, with character descriptor varied.

(14) The (chemist/clumsy chemist/capable chemist) had acid on his coat. The outcome that might be inferred in this case is that the chemist had spilled acid on his coat. It was hypothesized that subjects presented with "clumsy
chemist" would be more likely than would other subjects to infer that the chemist had spilled the acid on his coat.

After all the sentences were presented, subjects were given the recall test. Results showed that the semantic task led to the greatest number of correctly recalled sentences as well as the greatest number of inferences. Character descriptors consistent with implied outcomes led to more inferences than did character descriptors inconsistent with outcomes. For instance, "clumsy chemist" led to more "spilled" responses than did "capable chemist." These results, like the results of experiments discussed above, indicate a confusion of inference with observation. It seems unlikely that subjects would include inferences in recall, unless they thought that the inferred information had been presented.

Seifert et al. (1985) presented subjects with stories designed to lead to four kinds of inferences: goal, plan, action, and state inferences. For example, a story about two women going out to lunch contained sentence 15a, which in the context of the entire passage implies the plan in sentence 15b.

(15a) Mary's pocketbook was empty.

(15b) She decided to ask Kate for a loan.

Subjects were tested for recognition of implication sentences like sentence 15b. False recognition rates were high for goal, plan, and action implications, but not for state implications. Seifert and colleagues concluded that goal, plan, and action inferences made while reading a narrative become an indistinguishable part of one's long-term memory of the narrative.

Similar results were obtained by Bower, Black, and Turner (1979). In one experiment they presented subjects with stories about routine activities, such as attending a movie, or visiting the dentist. A person's general
knowledge about the sequence of events that comprise one of these routine activities has been called a "script." (See, for example, Abelson, 1981; Schank and Abelson, 1977.) After reading several stories, subjects were given a recall test, with the story titles as cues. Recall protocols often included script-consistent actions that had not been stated in the stories. These responses showed that subjects mistook actions they had inferred for actions stated in the stories. The recall results were replicated using a recognition test. Subjects falsely recognized script-related actions that had not been stated.

Another study involving scripts (Abbot, Black, & Smith, 1985) yielded similar results. Each script used in this study contained several scenes. Each scene included a scene header, like sentence 16a, and a scene action, like sentence 16b.

(16a) They ordered their meal.

(16b) They discussed what they wanted to eat.

For each subject some scenes were presented without the scene header. Only the scene action was presented for these scenes. Recognition test results showed that when only the scene action was presented, subjects often falsely recognized the corresponding scene header. These results indicate that subjects often thought that the scene header had been presented rather than only implied.

Scripts were used by Owens, Bower, and Black (1979) to study the effect of prior knowledge about a script's main character on story memory. They hypothesized that knowledge about a character's motives would affect understanding of the meaning of story actions, of their importance, and of their interconnections. To test this idea they presented subjects with a dull story about the activities of a character (e.g., making coffee, going to the doctor,
attending a lecture). Some subjects were given additional information just before reading the story. This information concerned the character's motives. In one experiment the character was a woman, and the subjects who were given additional information were told that she was worried about a possible pregnancy, about how she would tell the professor she had been seeing, and about money. All subjects were later asked to recall the story as closely to verbatim as possible. In recalling the dull events of the story, subjects who knew about the character's motives often produced motive-relevant inferences. For example, the presented story stated that the doctor's nurse "...went through the usual procedures." These "usual procedures" were sometimes recalled as "pregnancy tests." Following the recall test, subjects were given a recognition test in which those subjects who knew about the character's motives falsely recognized motive-relevant implications. In both recall and recognition tasks, these subjects' responses revealed that they thought information had been presented, when actually, it had only been implied.

To summarize, a number of researchers have provided evidence that people sometimes confuse inference with observation. This phenomenon has been found using different kinds of study materials (e.g., differing in length and in content), different study instructions, different modes of presentation, different delays between study and test, and different kinds of tests. This variety is convincing: People clearly do have difficulty discriminating the source of information that has, in fact, been inferred.

The confusion of inference with observation is but one instance of a larger phenomenon: confusion about the origins of remembered information in general. Several researchers have documented confusion about information
origin in cases not involving inference. Two examples from this research are described in the following section.

**Confusion About the Source of Information**

**Other Than Inferred Information**

Sullin and Dooling (1974) found that people sometimes confuse previously acquired information with newly presented information. In their experiment control subjects read a passage describing the behavior problems of a child called Carol Harris. Experimental subjects read the same passage, but with the child's name changed to Helen Keller. Both groups were later given a recognition test on old and new sentences. One of the new sentences read, "She was deaf, dumb, and blind." Subjects given the name Helen Keller tended to report that this was an old sentence more than did subjects given the name Carol Harris, indicating that experimental subjects had difficulty discriminating between their preexisting knowledge about Helen Keller and the information presented in the passage.

Another example of confusion about the origin of remembered information was provided by Loftus (1975). In one experiment subjects confused information acquired through their own observations as they viewed an event and information supplied later by the experimenter. Subjects viewed a videotape of an automobile accident and then were asked a series of questions, varied between subjects, about the event. In the experimental condition one of the questions was misleading, incorporating some inaccurate information: "How fast was the white sports car going when it passed the barn while traveling along the country road?" (There had been no barn on the videotape.) In the control condition a similar question without mention of a barn was substituted: "How fast was the white sports car going while traveling along the
country road?" One week later subjects' memory of the videotape was tested with a list of questions including, "Did you see a barn?" Compared to control subjects, experimental subjects more often reported seeing a barn. They apparently remembered something about a barn, but did not distinguish the source of that memory.

Such errors have prompted some researchers to formulate theories concerning the cause(s) of confusion about the source of information. Two of these theories are discussed in the following section.

Theories Concerning Confusion About the Source of Information

One explanation of confusion about the source of information involves the integration of information from various sources. The experiments described above clearly show that memory is not always a record of sensory observation alone. People think about information gained through observation, relating it to their preexisting knowledge, combining it with new information from later experiences, extending it with inferences, and so on. Some theorists have proposed that these activities result in a high degree of integration of information from various sources. It is this integrated information that is remembered. The sources of particular items of information that have been integrated are forgotten (Bartlett, 1932; Cofer, 1973).

A more extreme version of the integration explanation has been proposed by Loftus and her associates (e.g., Loftus, 1979, 1982). According to this version, once integration has occurred people not only fail to discriminate sources of information, but they are not capable of doing so. This is because integrated, or altered, information has replaced original information in memory, so that the original information no longer exists (Loftus & Palmer, 1974).
Recently, the adequacy of the evidence on which this position is based has been debated (Loftus, Schooler, & Wagenaar, 1985; McCloskey & Zaragoza, 1985a, 1985b). As Loftus (1979, p. 117) has noted, "It is possible to design experiments that might demonstrably show that one's original memory for an event has remained intact. It is not possible to design an experiment that can conclusively prove that an alteration in memory has occurred." Whenever subjects fail to remember original information, there remains the possibility that, under some conditions, they would be able to remember it. Inability to prove the loss of original information through integration does not, however, indicate that integration does not affect memory.

A different approach to explaining confusion about the source of remembered information has been suggested by Johnson and her colleagues (see Johnson & Raye, 1981). They assume that information from different sources can coexist in memory, and they propose that confusion about information origins may reflect this coexistence as well as information integration. Johnson and colleagues are particularly interested in the confusion of externally generated memories (those derived through perception) with internally generated memories (those derived through activities such as thought, imagination, and reasoning). Consequently, they have focused research on ways in which people might distinguish between these two classes of memories, and they have reported that the two classes have different characteristics. For instance, externally generated memories are accompanied by more contextual information than are internally generated memories (Johnson, Raye, Foley, & Kim, 1982). Internally generated memories are accompanied by more information about the cognitive operations that occurred
when the memories were formed (Johnson, Raye, Foley, & Foley, 1981). According to the reality monitoring model proposed by Johnson and Raye (1981), these distinguishing characteristics can serve as cues in deciding memory origins. Confusion about the source of a memory is likely to occur when a memory is not typical of its class, when a memory of one class is similar (e.g., semantically similar) to one or more memories of the other class, or when there are failures in reasoning about memory origin (e.g., failure to consider relevant information).

The theories concerning confusion about information origins might be applied to the specific case of mistaking inference for observation. The degree of integration of inferred information with observed information as well as the presence/absence of distinguishing memory characteristics might influence the probability of mistaking inference for observation. Accordingly, a variable that influences the probability of accurately discriminating the source of inferred information might do so through its effects on information integration or through its contributions to distinguishing memory characteristics.

**Consistency of Inferred Information**

**With Subsequently Observed Information**

One variable that might affect the probability of correctly discriminating the source of inferred information is the degree of consistency between the inferred information and subsequently observed information. Findings from a variety of studies suggest this possibility.

Several experiments have shown that memory of events can be influenced by subsequently encountered information. The additional information sometimes improves memory. For instance, when subjects were given a new perspective on a previously recalled event, they recalled
previously unrecalled event information (Anderson & Pichert, 1978). On the other hand, subsequently encountered information has led subjects to confuse presented information with prior knowledge (Dooling & Christiaansen, 1977), with an experimenter's question (Loftus & Palmer, 1974), and with their own inferences (Spiro, 1980).

Regarding information consistency, a number of experiments have demonstrated consistency effects on memory. For example, subjects who had a prior impression of a person falsely recognized personality characteristics (Cantor & Mischel, 1977) and behaviors (Snyder & Uranowitz, 1978) that were consistent with the impression. Also, subjects having a prior impression recalled impression-consistent behaviors better than they recalled impression-inconsistent behaviors (Rothbart, Evans, & Fulero, 1979). Sometimes, however, behaviors were more likely to be recalled if they were inconsistent, rather than consistent, with a prior impression (Hastie & Kumar, 1979; Srull, Lichtenstein, & Rothbart, 1985). It has been proposed that these conflicting findings can be explained by taking into account the completeness of impression formation at the time memory is tested: When an impression is well developed, impression-consistent information is more likely to be remembered; when an impression is not yet well formed, inconsistent information is more likely to be remembered (Belmore, 1987; Higgins, & Bargh, 1987; Srull, & Wyer, 1989).

Concerning consistency effects on confusion of inference with observation, several of the experiments cited above show that the degree of consistency between implied information and previously observed information affects the probability that an implication will be mistaken for an observation. For example, in the experiment by Bransford et al. (1972), subjects falsely
recognized implied information that was consistent with previously presented information, whereas they tended to reject implied information that was inconsistent with previously presented information. A parallel effect was found by C. R. Barclay et al. (1984) using a recall paradigm. In their experiment subjects were more likely to include inferred information in recall when it was consistent, rather than inconsistent, with previously presented information.

Given that consistency between inferred information and previously observed information can affect confusion of inference with observation, it seems reasonable to think that consistency between inferred information and subsequently observed information might have a comparable effect. Specifically, the probability of mistaking an inference for an observation might be expected to increase with exposure to new information that is consistent with the inferred information. In contrast, the probability of mistaking an inference for an observation might be expected to decrease with exposure to new information that is inconsistent with the inferred information. As an example, suppose that I have inferred that Jane is a competent worker. I might be more likely to mistake my inference for an observation if Jane is promoted than if her salary is decreased. The present experiments were designed to explore this possibility.

Experiment 1

The purpose of Experiment 1 was to learn whether consistency between inferred information and subsequently observed information would affect judgments of the source (observed/inferred) of the inferred information. It was predicted that when subsequent information was inconsistent with inferred information, the subjects would be more accurate in recognizing that the inferred information had, in fact, been inferred than they would be, when
subsequent information was consistent with the inferred information.

Method

Although there were some variations among the experiments reported here, each had four basic phases:

Phase 1. Study of presented information.

Phase 2. Opportunity to make inferences based on the presented information.

Phase 3. Exposure to information consistent/inconsistent with the inferences that might have been made.

Phase 4. Judgment of the sources of information, including both information that had been observed and information that had been inferred.

In Phase 1 of this experiment the subjects read a text describing a criminal case. In Phase 2 they judged statements about the case as either true or false. This true/false test included, in addition to true items and false items, a number of inference items (i.e., items containing information that had been implied rather than explicitly presented in the text). A "true" response to an inference item was regarded as evidence that an inference had been made. In Phase 3 the subjects read a list of statements giving additional information about the criminal case. Each of these statements was related to one of the inference items from the true/false test. Half of the statements were consistent with their related inference items, and half were inconsistent with their related inference items. After reading each of these additional information statements, the subjects rated it for consistency with their previous ideas about the case. In Phase 4 the subjects looked over their true/false test responses and judged whether those items to which they had responded "true" had been stated in the
case description or inferred by them from the case description.

Materials

Materials consisted of a description of a criminal case, a true/false test, and a list of statements giving additional information about the case.

Case description. A description of a fictitious criminal case was written, based on a case devised by Schum and Martin (1981). The description included information about the principals in the case, the scene of the crime, eyewitness reports, and so on. The following is an excerpt from the case description, which is given in Appendix A.

Officer Farmer arrested defendant Driver in a departure area at Houston Intercontinental Airport at 11:45 p.m. on the night of the assault on Hogan. Defendant Driver had a one way ticket to Mexico City on a flight due to depart Houston at 11:55 p.m. on the night in question. Driver was alone at the time of the arrest.

True/false test. There were 20 true/false items: 5 true items, 3 false items, and 12 inference items. True items consisted of information that had been stated in the case description. None of these true items, however, were direct quotations. False items consisted of information that conflicted with the case description. Inference items consisted of information that had not been stated in the case description, but had been implied. The truth of inference items could not be conclusively determined from the case description. The order of items followed the order of presentation of item-relevant information in the case description, with true, false, and inference items mixed throughout the
list. The true/false test items are included in Appendix A. The following are test items related to the above case description excerpt.

True item: Defendant Driver was arrested at Houston Intercontinental Airport.

False item: Defendant Driver had a ticket to Caracas, Venezuela.

Inference item: Defendant Driver had gone to the airport alone.

Additional information statements. There were 24 additional information statements, two statements relating to each of the 12 inference items from the true/false test. Of the two statements relating to each inference item, one was consistent with the inference item, and one was inconsistent with the inference item. These 24 statements were divided into 2 lists of 12 statements each. Within each list half of the statements were consistent with their respective inference items, and half were inconsistent. The order of the additional information statements within each list followed the order of presentation of statement-relevant information in the case description, with consistent and inconsistent statements mixed throughout the list. The additional information statements are included in Appendix A. The following are additional information statements relating to the above inference item.

Consistent statement: Defendant Driver said that it was his habit to go to the airport alone, as he did not like to ask anyone to drive him. His wife, Mary, said that she had occasionally offered to drive him to the airport, but that he had always refused her offer.
Inconsistent statement: Defendant Driver was seen by an airport
employee talking earnestly with another man as he waited in line to buy
a ticket. The other man stood in the line talking with defendant Driver for
a while and then left without buying a ticket for himself.

In these examples the defendant’s habit of going to the airport alone is
consistent with the possibility that he had gone to the airport alone on the night
of his arrest, but does not prove that he did: Past behavior is not, after all, a
perfect predictor of future behavior. Likewise, the defendant’s earnest
conversation while waiting in the ticket line with a man who did not buy a ticket
for himself seems inconsistent with the possibility that the defendant had gone
to the airport alone, but does not conclusively disconfirm it: Perhaps the
defendant is a gregarious sort, who engages passers-by in conversation.
Throughout this report, then, “consistency” and “inconsistency” are used in a
relative rather than an absolout sense.

Design

Consistency between inference items on the true/false test and additional
information statements was varied within subjects. Counterbalancing measures
ensured that: (a) For a given subject, six of the additional information
statements were consistent with their related inference items and six were
inconsistent with their related inference items, and (b) for a given inference item,
the additional information was consistent for half of the subjects and
inconsistent for the other half.

Subjects

The subjects were 28 Rice University students, who participated in partial
fulfillment of course requirements. They were tested in two groups of 14
subjects each.

Procedure

All experimental tasks, with the exception of intervening tasks, were self-paced, and each new task was begun when all subjects had completed the previous one. In Phase 1 of the experiment the subjects read the criminal case description. Their instructions were as follows:

The following pages contain information about a criminal case. Your task is to read all of the information carefully and to answer questions about it. You can assume that all of the information given is correct. For example, all evidence obtained from witnesses can be considered to be literally true. You do not have to be concerned about witnesses lying or being mistaken, as you might in real-life criminal cases.

In Phase 2 the subjects completed the true/false test. Instructions for the true/false test were as follows:

Please answer the following questions by circling either True or False. If you don't know the answer, please give your best guess.

A "true" response to an inference item was regarded as evidence that an inference had, in fact, been made. The true/false test was followed by a 5 min intervening task, namely, writing descriptions of personality types to be used in another experiment.

In phase three of the experiment the subjects read the additional
information statements. After reading each statement, they rated its consistency with their previous ideas about the case on a 5-point scale (1 = low consistency; 5 = high consistency). The subjects were given the following instructions for this task:

The following pages contain additional information about the case. Again, you can assume that all of the information given is correct. After reading each item, think carefully about how it relates to what you already know about the case. In particular, think about whether or not this new information is compatible with your previous ideas about the case. Mark the scale beneath each item by circling the appropriate number.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>not at all compatible with previous ideas</td>
<td>moderately compatible with previous ideas</td>
<td>highly compatible with previous ideas</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The rating task was designed to focus the subjects' attention on the consistency/inconsistency between each additional information statement and the related inference that they might have made. In addition, the rating task provided a measure of perceived consistency, as a check on the effectiveness of the consistent and inconsistent additional information statements. The rating task was followed by a 5 min intervening task, namely, writing additional personality descriptions.

In Phase 4 of the experiment the subjects looked over their true/false test responses. Focusing on items for which they had selected the "true" alternative,
they now judged whether each of those items had been (a) stated (not necessarily verbatim) in the case description they had read before the true/false test, or (b) inferred by them from the case description. They rated their confidence in each judgment on a 3-point scale (1 = low confidence; 3 = high confidence). The subjects were given detailed instructions for this task, supplemented by examples from a new fictitious criminal case. The instructions and examples are given in Appendix B. To prevent the subjects from changing any of their true/false test responses while making their stated/inferred judgments, a different colored pencil was used for each of these tasks, and the pencils were collected after each task.

Finally, the subjects were given a written debriefing.

Results and Discussion

Due to a technical error, one inference item had to be deleted from the analyses. Because the deleted data were equally divided between the two consistency conditions, there is no reason to suppose that the deletion would affect the pattern of results.

Perceived Consistency of Additional Information Statements

Based on the subjects' consistency ratings of the additional information statements, two perceived consistency scores (means of consistency ratings) were computed for each subject, one for consistent and one for inconsistent additional information statements. Mean perceived consistency scores are shown in Table 1. These scores show that the subjects' perceptions of the additional information statements were as expected.
Table 1

Perceived Consistency of Additional Information Statements as a Function of Consistency Condition

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>4.15</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>2.95</td>
</tr>
</tbody>
</table>

Note. 1 = low perceived consistency; 5 = high perceived consistency.

Memory for the Criminal Case Description

The subjects' true/false test responses for true items and false items served as a measure of memory for the case description. The percentage of correct responses was computed for each subject. There was a mean of 97% correct, indicating good overall memory of the case description.

Inferences

Judgments for inference items on the true/false test served as a measure of the inferences that had been made. Of the 11 inference items on the true/false test, the mean number that were judged as true was 8.79, with a mean of 4.43 in the consistent condition (i.e., followed by consistent additional information) and a mean of 4.36 in the inconsistent condition (i.e., followed by inconsistent additional information).

Stated/Inferred Judgments

The two confidence scales that the subjects had used in making their stated/inferred judgments (one for items believed to have been stated, and one
for items believed to have been inferred) were combined into one 6-point scale, with the end points "certain this was stated" and "certain this was inferred." This 6-point scale was used in converting the subjects' stated/inferred judgments to "accuracy scores." For true items a rating of "certain this was stated" was scored 6, and a rating of "certain this was inferred" was scored 1. Scoring was reversed for inference items, so that a rating of "certain this was inferred" was scored 6, and a rating of "certain this was stated" was scored 1.

Two mean accuracy scores were computed for each subject, one for true items and one for inference items. For true items, overall mean accuracy was 5.40, indicating good discrimination of the source of information that had, in fact, been stated.

Of more relevance to the present concern was the effect of the consistency between inferred information and subsequent information on the accuracy scores for the inference items. The accuracy scores were significantly different for the consistent and inconsistent conditions, \( t(27) = 3.54, p < .002 \). As might be expected, given that there was an effect of consistency, the mean accuracy score was higher for the inconsistent condition than for the consistent condition. The accuracy results for inference items are summarized in Table 2.
Table 2

Accuracy of Stated/Inferred Judgments for Inferences as a Function of Consistency Condition

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>3.30</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>4.13</td>
</tr>
</tbody>
</table>

*Note.* 1 = low accuracy; 6 = high accuracy.

These results show that the consistency between inferred information and subsequently observed information affected the subjects' judgments of the source (stated/inferred) of information that had, in fact, been inferred. The subjects were less likely to mistake inferences for observations when subsequently observed information was inconsistent, rather than consistent, with inferred information.

**Experiment 2**

Experiment 2 was a replication of Experiment 1, with two variations designed to explore the generalizability of the consistency effect found in that experiment: (a) A second set of materials was introduced, and (b) all materials, except the test items, were presented orally. In addition, delay between making an inference and judging the source (stated/inferred) of the inferred information was varied to test for an effect of delay on source judgments. Given that there is a well-documented decline in memory accuracy with delay, it might be expected that the accuracy of source judgments would also decline with delay.
Method

The method of Experiment 2 was the same as that of Experiment 1, with the exceptions described in the following sections.

Materials

In addition to the materials used in Experiment 1, there was a second set of materials, based on a second fictitious criminal case also drawn from Schum & Martin (1981). The new materials are included in Appendix A. This addition doubled the number of test items.

Design

Two factors were varied within subjects: (a) consistency between inferred information and subsequently observed information (consistent/inconsistent), and (b) delay between making an inference and judging the source of the inferred information (15 min/30 min).

Subjects

The subjects were 24 Rice University students, who participated in partial fulfillment of course requirements. They were tested in four groups of six subjects each.

Procedure

All materials, except the test items, were presented orally. The subjects were divided into two equal groups. For Group 1, the order of tasks was as follows:

1. Listen to Case Description A.
2. Complete true/false test on Case A.
3. Listen to Case Description B.
4. Complete true/false test on Case B.
5. Listen to additional information statements about Case A.
6. Listen to additional information statements about Case B.
7. Complete stated/inferred judgments for Case B.
8. Complete stated/inferred judgments for Case A.

The subjects in Group 1, then, performed each task first with Case A, and then with Case B, except for the final stated/inferred judgment task, which was performed first with Case B, and then with Case A. This order of tasks resulted in a longer delay between the true/false test and the stated/inferred judgment task for Case A (30 min) than for Case B (15 min). For Group 2 the order of cases was reversed, so that there was a longer delay for Case B than for Case A. The true/false tests and the stated/inferred judgment tasks were self-paced.

The subjects were not asked, as those in Experiment 1 had been, to rate the consistency between additional information statements and their previous ideas about a criminal case. They were not, therefore, forced to attend to the consistency/inconsistency between each additional information statement and the related inference that they might have made. They were, however, asked to think about the consistency, and they were given 2 s after each statement to do so. Their instructions were:

After I read each item of information, I will pause for a few seconds to give you a chance to think about how this new information relates to what you already know about the case. In particular, think about whether or not this new information is compatible with your previous ideas about the case.

The subjects were told that in making their stated/inferred judgments they
should rate an item "stated" if they thought it had been stated either in the case
description they had read before the true/false test or in the list of additional
information statements. (In Experiment 1 the subjects had been told to rate an
item "stated" only if they thought it had been stated in the case description.)
This change of instructions was made in response to participants in Experiment
1 who commented that they thought some inference items had been stated in
the additional information. Because inference items were stated neither in the
case description nor in the additional information, those subjects who thought
an inference item had been stated anywhere were mistaken. Although those
subjects clearly thought that an inference item had been stated, they had been
restricted in Experiment 1 to rating the item as inferred. Some instances of
inferences being mistaken for observations, therefore, might have gone
undocumented.

Results and Discussion
The main effect of delay between making an inference and judging the
source (stated/inferred) of the inferred information was not significant, \( p > .10 \).
Nor was the interaction of delay with consistency significant, \( p > .10 \). Therefore,
the results reported here are based on data collapsed across delay conditions.

Memory for the Criminal Case Description
The subjects' true/false test responses for true items and false items
served as a measure of memory for the case description. The percentage of
correct responses was computed for each subject. There was a mean of 94%
correct, indicating good overall memory of the case description.

Inferences
Judgments for inference items on the true/false tests served as a
measure of the inferences that had been made. Of the 24 inference items on
the two true/false tests (12 on each test), the mean number judged as true was
19.88, with a mean of 10.29 in the consistent condition (i.e., followed by
consistent additional information) and a mean of 9.58 in the inconsistent
condition (i.e., followed by inconsistent additional information).

**Stated/Inferred Judgments**

The subjects' stated/inferred responses for true items and for inference
items were converted to accuracy scores, as in Experiment 1. The possible
range of accuracy scores was 1-6 (1 = low accuracy, 6 = high accuracy). Two
mean accuracy scores were computed for each subject, one for true items and
one for inference items. For true items, overall mean accuracy was 5.18,
indicating good discrimination of the source of information that had, in fact, been
stated.

As in Experiment 1, the central concern was the effect of the consistency
between inferred information and subsequent information on the accuracy
scores for the inference items. Again, the accuracy scores were significantly
different for the consistent and inconsistent conditions, $F(1,23) = 41.63, p < .0001$, with the mean accuracy score being higher for the inconsistent condition
than for the consistent condition. The accuracy results for inference items are
summarized in Table 3.
Table 3

**Accuracy of Stated/Inferred Judgments for Inferences as a Function of Consistency Condition**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>2.74</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>4.06</td>
</tr>
</tbody>
</table>

**Note.** 1 = low accuracy; 6 = high accuracy.

These results, like the results of Experiment 1, show that the consistency between inferred information and subsequently observed information affected the subjects' judgments of the source (stated/inferred) of the inferred information. The subjects were less likely to mistake inferences for observations when subsequently observed information was inconsistent, rather than consistent, with the inferred information.

The instructions given for the stated/inferred judgment task in this experiment (i.e., to rate an item "stated" if it had been stated either in the case description or in the additional information) might have biased the results in favor of the predicted consistency effect: There were more opportunities in the consistent condition for the subjects to be confused by implications than there were in the inconsistent condition, due to the implications of the additional information statements in the consistent condition. To avoid the possibility of bias in favor of the expected consistency effect, instructions in the next experiment will correspond to those used in Experiment 1.
Experiment 3

The consistency effect found in Experiments 1 and 2 was further investigated in Experiment 3. One of the functions of the true/false tests in the previous experiments was to induce the subjects to make inferences. For the inference items included in the tests, the subjects had not been given enough information to determine truth, and a "true" response to one of these items was regarded as evidence that an inference had been made. It is possible, however, that with the forced-choice testing some "true" responses were made reluctantly and, therefore, did not represent inferences. To eliminate this possibility, responses in this experiment were made on continuous scales, giving the subjects the option of indicating uncertainty, rather than forcing them to choose between "true" and "false" alternatives.

Another issue concerning true/false test responses is the meaning of "true" responses that are made without being forced: Given that such a response represents an inference, what exactly has been inferred? In the previous experiments the test instructions stated, "If you don't know the answer, please give your best guess." The instructions did not explain that for some items the subjects had not been given sufficient information to determine truth. Nor did the instructions specify that, when the subjects were in doubt, they should base their truth judgments on the information presented in combination with their general knowledge, rather than guessing on some other basis. Under these conditions a "true" response to an inference item might represent an inference that the item had been presented (an inference based on factors such as item component familiarity), rather than an inference that the item was true (an inference based on the implications of the presented information). Because
it is the latter type of inference that is of interest here, test materials and instructions were modified in this experiment to ensure that any inferences made were of this type.

**Method**

The method of this experiment was basically the same as that of Experiment 2, with some modifications of materials and procedure, which are described in the following sections.

**Materials**

For each of the criminal cases, two inference items were deleted from the true/false test (see Appendix C). In the previous experiments these items had not been effective, rarely inducing inferences.

Two extraneous items were added to each true/false test, one of which was placed before any of the inference items on the test. These items were included to demonstrate the fact that the subjects had not been given enough information to judge the truth of some test items. Responses to these items could only be based on general knowledge, specifically, frequency knowledge. For each criminal case, there was one high-frequency extraneous item and one low-frequency extraneous item (see Appendix C). For example, the extraneous items for the Driver assault case were:

- **High frequency item**: The victim (Hogan) drinks coffee in the morning.
- **Low frequency item**: Defendant Driver has an inherited blood disorder.

**Subjects**

The subjects were 28 Rice University students, who participated in partial fulfillment of course requirements. They were tested in four groups of seven
subjects each.

Design and Procedure

The design of this experiment was identical to that of Experiment 2. The procedure was the same as that of Experiment 2, except for the following changes.

There were two identical true/false tests for each of the criminal cases. The first test was given after the presentation of the case description. As in the previous experiments, this test measured both overall memory and the inferences that had been made. The second test was given after the consistency manipulation (the additional information statements about the case). This second test was included to determine the effect of consistency on truth judgments.

True/false test responses were made on unmarked 18 cm lines, with end points designated "True" and "False." Instructions for the first true/false test were:

Please give your opinion about the truth of each of the items on this page by placing a vertical mark anywhere along the line for that item. A mark at the extreme "False" end of the line means that you are certain the item is false. A mark at the extreme "True" end of the line means that you are certain the item is true. For some of the items you have not been given enough information to know whether the item is true or false. We're asking only for your opinion. In forming your opinion you may use your general knowledge about the world as well as the information given in the Driver (Walsh) case description.
Instructions for the second true/false test were:

Now that you have more information about this case, please give your opinion once again about the truth of each item by placing a vertical mark on the scale for that item. In forming your opinion you may use your general knowledge about the world, the information given in the Driver (Walsh) case description that you heard earlier, and the additional information about the Driver (Walsh) case that you just heard. Remember, a mark at the extreme end of the line means that you are certain.

To emphasize the difference between the truth judgment task (i.e., the true/false test), in which judgments were to be based on general knowledge as well as on the information presented in the case description, and the source judgment task, in which presentation in the case description was to be the only criterion for judgments, the response choices for the latter task were "stated" and "not stated" rather than the "stated" and "inferred" options given in the previous experiments. Stated/not stated test responses were made on unmarked 18 cm lines, with end points designated "Stated" and "Not Stated." The subjects were given detailed instructions for this test, supplemented by examples from a new fictitious criminal case. The instructions and examples are given in Appendix D.

The between-conditions difference in delay between making an inference and judging the source of the inferred information was increased to 25 min (as compared to 15 min in Experiment 2). This greater between-
conditions difference increased the probability of finding a possible effect of delay on source judgements.

The order of the subjects’ tasks was changed to allow the addition of the second true/false test on each criminal case and to increase the between-conditions difference in delay. As in Experiment 2, the subjects were divided into two equal groups. For Group 1, the order of tasks was as follows:

1. Listen to Case Description A.
2. Complete first true/false test on Case A.
3. Listen to Case Description B.
4. Complete first true/false test on Case B.
5. Listen to additional information statements about Case B.
6. Complete second true/false test on Case B.
7. Complete stated/not stated judgments for Case B.
8. Listen to additional information statements about Case A.
9. Complete second true/false test on Case A.
10. Complete stated/not stated judgments for Case A.

For the subjects in Group 1, then, there was a longer delay between the first true/false test and the stated/inferred judgment task for Case A (35 min) than for Case B (10 min). For Group 2 the order of cases was reversed, so that there was a longer delay for Case B than for Case A.

Results and Discussion

As in the previous experiment, neither the main effect of delay between making an inference and judging the source (stated/inferred) of the inferred information, nor the interaction of delay with consistency, was significant, ps > .10. Therefore, the results reported here are based on data collapsed across delay conditions.
The unmarked 18 cm lines on which true/false test responses had been made were divided into 18 intervals of 1 cm each. Responses falling within the 1st interval at the "false" end of the line were scored 1, those in the 2nd interval were scored 2, and so on to the 18th interval at the "true" end of the line. Possible true/false ratings, then, ranged from 1 to 18 (1 = false; 18 = true).

**Memory for the Criminal Case Description**

The subjects' ratings of true items and false items on the first true/false test (1 = false; 18 = true) served as a measure of memory for the case description. The mean ratings were 16.34 for true items and 3.39 for false items, indicating good overall memory of the case description. Mean ratings of true items and false items on the second true/false test were similar: 16.49 for true items and 4.02 for false items. On both the first and second true/false tests each of the 28 subjects gave a higher mean rating for true items than for false items.

**True/False Judgments for Extraneous Items**

The mean rating for high-frequency extraneous items on the first true/false test (1 = false; 18 = true) was 9.63, which was significantly higher than the mean rating of 5.59 for low-frequency items, \( t(27) = 4.72, p < .0001 \). These results show, as expected, that the subjects used frequency knowledge in making their judgments of extraneous items. Mean ratings for extraneous items on the second true/false test were virtually the same: 9.73 for high-frequency items and 6.20 for low-frequency items. Again, the difference was significant, \( t(27) = 4.12, p < .001 \).
Inferences

Ratings of inference items on the first true/false test (1 = false; 18 = true) served as a measure of the inferences that had been made. The mean rating was 15.47, with a mean of 15.35 in the consistent condition (i.e., followed by consistent additional information) and a mean of 15.58 in the inconsistent condition (i.e., followed by inconsistent additional information). The 18-interval rating scale was divided into thirds, and a rating in the highest six intervals was regarded as evidence that an inference had been made. Of the 20 inference items on the two true/false tests (10 on each test) the mean number that were rated in the highest six intervals was 16.18, with a mean of 8.11 in the consistent condition and a mean of 8.07 in the inconsistent condition. Because the choice of a rating in the highest six intervals as the criterion for inferences is arbitrary, the frequency of inference item ratings at each level of the 18-interval scale will be given in a later section.

Effect of Consistency on Truth Judgments for Inference Items

Mean ratings for inference items on the second true/false test (after presentation of the consistent/inconsistent additional information statements) are given in Table 4. As expected, there was a significant difference between consistency conditions, $F(1,27) = 142.96, p < .0001.$
Table 4

Mean Ratings for Inference Items on Second True/False Test by Consistency Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>16.19</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>9.97</td>
</tr>
</tbody>
</table>

Note. 1 = false; 18 = true.

These results show that the consistency between inference items and their related additional information statements affected judgments on the second true/false test, with consistent additional information leading to higher truth ratings and inconsistent additional information leading to lower truth ratings.

**Stated/Not Stated Judgments**

The unmarked 18 cm lines on which stated/not stated test responses had been made were divided into 18 intervals of 1 cm each. This 18-interval scale was used in converting the subjects' stated/not stated judgments to accuracy scores ranging from 1 to 18 (1 = low accuracy; 18 = high accuracy). For true items, responses falling within the 1st interval at the "not stated" end of the line were scored 1, those in the 2nd interval were scored 2, and so on to the 18th interval at the "stated" end of the line. Scoring was reversed for false items, extraneous items, and inference items, so that responses falling within the 1st interval at the "not stated" end of the line were scored 18, those in the 2nd interval were scored 17, and so on to responses in the 18th interval at the "stated" end of the line, which were scored 1.
Mean accuracy scores were computed for each subject, for each type of test item. Mean accuracy for true items was 15.63, indicating good discrimination of the source of information that had, in fact, been stated.

Accuracy was also high for false items and extraneous items. Mean accuracy for false items was 15.75. For extraneous items, mean accuracy was 17.22, with a mean of 17.30 for low-frequency items and a mean of 17.13 for high-frequency items.

For inference items, accuracy scores were based only on items that had been judged "true" on the first true/false test (i.e., items rated in the upper third of the true/false rating scale). The "true" responses to these items were regarded as evidence that inferences had been made. Overall, the subjects were not very accurate in judging the source of these items. The mean accuracy score was 5.70.

As in the previous experiments, the central concern was the effect of the consistency between inferred information and subsequently observed information on the accuracy scores for the inference items. Again, the accuracy scores were significantly different for the consistent and inconsistent conditions, $F(1,27) = 47.37, p < .0001$. As might be expected, accuracy scores were higher for the inconsistent condition ($M = 8.13$) than for the consistent condition ($M = 3.58$).

Because the choice of a rating in the upper third of the true/false scale as the criterion for inferences is arbitrary, mean accuracy scores for both consistent and inconsistent conditions were computed using every possible cutoff point on the 18-interval rating scale (i.e., $18, \geq 17, \geq 16$, and so on). These means along with significance tests of the difference between each consistent/inconsistent pair of means show that the same pattern of results was produced regardless of
the cutoff point chosen. The source judgment accuracy results for inferences are summarized in Table 5. For further understanding of the results, it is useful to consider the distribution of observations along the 18-interval true/false rating scale as well as the mean accuracy of stated/not stated ratings at each level of the scale, both of which are given in Appendix E.
Table 5

Accuracy of Stated/Not Stated Judgments for Inferences as a Function of Consistency Condition and Rating on First True/False Test

<table>
<thead>
<tr>
<th>Rating on first true/false test&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Consistent</th>
<th>Inconsistent</th>
<th>df&lt;sup&gt;c&lt;/sup&gt;</th>
<th>F&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>2.73</td>
<td>7.41</td>
<td>20</td>
<td>19.19</td>
</tr>
<tr>
<td>≥ 17</td>
<td>2.80</td>
<td>7.50</td>
<td>24</td>
<td>35.89</td>
</tr>
<tr>
<td>≥ 16</td>
<td>3.01</td>
<td>7.69</td>
<td>26</td>
<td>34.33</td>
</tr>
<tr>
<td>≥ 15</td>
<td>3.28</td>
<td>7.95</td>
<td>27</td>
<td>36.30</td>
</tr>
<tr>
<td>≥ 14</td>
<td>3.46</td>
<td>8.07</td>
<td>27</td>
<td>39.25</td>
</tr>
<tr>
<td>≥ 13</td>
<td>3.58</td>
<td>8.13</td>
<td>27</td>
<td>47.37</td>
</tr>
<tr>
<td>≥ 12</td>
<td>3.72</td>
<td>8.13</td>
<td>27</td>
<td>45.13</td>
</tr>
<tr>
<td>≥ 11</td>
<td>3.72</td>
<td>8.05</td>
<td>27</td>
<td>45.12</td>
</tr>
<tr>
<td>≥ 10</td>
<td>4.16</td>
<td>8.29</td>
<td>27</td>
<td>38.43</td>
</tr>
<tr>
<td>≥ 9</td>
<td>4.27</td>
<td>8.18</td>
<td>27</td>
<td>40.94</td>
</tr>
<tr>
<td>≥ 8</td>
<td>4.33</td>
<td>8.16</td>
<td>27</td>
<td>39.15</td>
</tr>
<tr>
<td>≥ 7</td>
<td>4.32</td>
<td>8.17</td>
<td>27</td>
<td>39.30</td>
</tr>
<tr>
<td>≥ 6</td>
<td>4.35</td>
<td>8.17</td>
<td>27</td>
<td>38.52</td>
</tr>
<tr>
<td>≥ 5</td>
<td>4.35</td>
<td>8.17</td>
<td>27</td>
<td>38.52</td>
</tr>
<tr>
<td>≥ 4</td>
<td>4.33</td>
<td>8.18</td>
<td>27</td>
<td>38.91</td>
</tr>
<tr>
<td>≥ 3</td>
<td>4.33</td>
<td>8.18</td>
<td>27</td>
<td>38.94</td>
</tr>
<tr>
<td>≥ 2</td>
<td>4.32</td>
<td>8.18</td>
<td>27</td>
<td>39.36</td>
</tr>
<tr>
<td>≥ 1</td>
<td>4.67</td>
<td>8.13</td>
<td>27</td>
<td>27.76</td>
</tr>
</tbody>
</table>

<sup>a</sup><sup>1</sup> = false; 18 = true. 1<sup>b</sup> = low accuracy; 18 = high accuracy. <sup>c</sup><sup>df</sup> differ because not all of the subjects used ratings > 15. <sup>d</sup><sup>F</sup>All <sup>p</sup><sub>s</sub> < .001.
Like the results of the previous experiments, the present results show that the consistency between inferred information and subsequently observed information affected the subjects' source judgments for the inferred information. The subjects were less likely to mistake inferences for observations when subsequently observed information was inconsistent, rather than consistent, with the inferred information.

**Summary and General Discussion**

The experiments reported here were concerned with confusion between inference and observation, specifically, with the problem of inferences being mistaken for observations. The main objective of these experiments was to investigate the effect of consistency between inferred information and subsequently observed information on the probability that an inference will be mistaken for an observation.

Although the experiments were not identical, each experiment had four essential phases. In Phase 1 the subjects studied a descriptive text. In Phase 2 they were given an opportunity to make inferences based on the text. In Phase 3 the subjects were given additional information that was either consistent or inconsistent with the inferences that might have been made. In Phase 4 they judged the sources of information, including both observed and inferred information. The consistency between inferred information and subsequently observed information was varied within subjects.

In Phase 1 of Experiment 1 the subjects read a description of a criminal case. In Phase 2 they were given a true/false test on the criminal case. In addition to true items and false items the test included a number of inference items, which consisted of information that had been implied rather than explicitly
presented in the case description. The truth of these items could not be conclusively determined from the information given in the case description, and a "true" response to an inference item was regarded as evidence that an inference had been made. In Phase 3 the subjects read additional information about the criminal case. The additional information consisted of a list of statements, each of which was related to one of the inference items from the true/false test. Half of the statements were consistent with their related inference items, and half were inconsistent with their related inference items. In Phase 4 the subjects reviewed their true/false test responses. For each item to which they had responded "true" they now judged whether the item had been stated in the case description or inferred by them from the case description.

The results of Experiment 1 showed that the consistency between inferred information and subsequently observed information affected the subjects' judgments of the source (stated/inferred) of information that had, in fact, been inferred. The subjects were less likely to mistake inferences for observations when subsequently observed information was inconsistent, rather than consistent, with inferred information.

Experiment 2 was essentially a replication of Experiment 1, with two changes in method designed to explore the generalizability of the consistency effect found in that experiment: (a) There were two sets of materials, those used in Experiment 1 and a new set of materials based on a second criminal case; and (b) all materials, except the test items, were presented orally. In addition, the amount of delay between making an inference and judging the source (stated/inferred) of the inferred information was varied within subjects. This change in method was introduced to test for an effect of delay on source judgements.
The results of Experiment 2, like the results of the previous experiment, showed that the subjects were more likely to recognize their inferences as inferences when subsequently observed information was inconsistent, rather than consistent, with the inferred information. There were no effects of delay.

The design of Experiment 3 was the same as that of Experiment 2. There were, however, some changes in procedure:

1. All responses were made on continuous scales, giving the subjects the option of indicating uncertainty, rather than forcing them to choose between two alternatives.

2. To emphasize the difference between the truth judgments (i.e., true/false test responses), which were to be based on general knowledge as well as on memory of the case description, and the source judgments, which were to be based solely on memory of the case description, the response choices for the latter task (i.e., the end points of the continuous scales) were "stated" and "not stated" rather than the "stated" and "inferred" choices that subjects had been given in the previous experiments.

3. To increase the probability of finding a possible effect of delay on source judgments, the between-conditions difference in delay between making an inference and judging the source of the inferred information was increased.

The results of Experiment 3 showed, once more, that the subjects were less likely to mistake inferences for observations when subsequently observed information was inconsistent, rather than consistent, with the inferred information. Again, there were no effects of delay.

The results of the experiments reported here are in agreement with those of previous studies on inference. Like earlier studies these experiments
showed that: (a) The subjects frequently made inferences, (b) they made inferences that were potentially incorrect, and (c) they sometimes mistook inferences for observations. In the present experiments a large percentage of the inference items, each of which contained information that might have been either correct or incorrect, was judged to be true. In addition, the subjects sometimes judged that these items had been presented when they had not.

The present experiments extend previous work on inference by introducing a new procedure for studying factors that might affect the probability of mistaking inference for observation and by identifying one factor that does so, namely, the consistency between inferred information and subsequently observed information.

**Methodological Issues**

In many previous studies of inference, including a number of those cited earlier (e.g., Brewer, 1977; Johnson, et al., 1973), the primary objective was to demonstrate that people do, in fact, make inferences. In a typical experiment the subjects study a text containing both explicit and implicit information and then are tested for recognition of both kinds of information. False recognition of implied information is interpreted as evidence that an inference has been made, and the experimenters often note that the false recognition indicates a confusion of inference with observation.

In this type of experiment a correct rejection of implied information is interpreted as evidence that the inference has not been made. Actually though, clear interpretation of correct rejection responses in such experiments is not possible. Specifically, it cannot be determined whether a correct rejection means that (a) the subject did not make the inference, or (b) the subject made the inference (e.g., during study) but recognized that it was an inference and,
therefore, rejected the item on the recognition test. Because the researchers who have used this paradigm were primarily concerned with showing that inferences had been made, the distinction between (a) and (b) above was not a crucial one for them.

If, as in the present experiments, the objective is to identify a factor that will affect the probability of mistaking inference for observation, the distinction between inferences mistaken for observations and inferences recognized as inferences is crucial, and previous methods, therefore, are not adequate. A method is needed that will allow discrimination between these two classes of inferences.

This problem was solved in the present experiments by having the subjects make two kinds of judgments: truth judgments (indicating whether or not inferences had been made), and source judgments (indicating whether or not the inferences made had been mistaken for observations). This procedure allows discrimination of two distinct classes of inferences within the set of inference items judged "true" in the truth judgment task: those mistaken for observations and those recognized as inferences. And because the two classes of inferences can be discriminated, the effect of an independent variable on confusion of inference with observation can be determined.

Not all of the previous work on inference has been primarily concerned with demonstrating that inferences have been made. Some researchers have attempted to identify factors affecting the probability of inference being mistaken for observation. Generally, they have done so by using the established method described above and varying some feature(s) of the study materials (e.g., Barclay, et al., 1984; Owens et al, 1979). The question addressed in such experiments might be stated as: "What conditions lead to inferences that are
mistaken for observations?" or, to put it another way, "What kinds of inferences are likely to be mistaken for observations?" The method of varying study materials has been successful in answering this question. However, the fact that some kinds of inferences are more likely to be mistaken for observations than others is only one aspect of the problem of confusion between inference and observation.

An equally important aspect of the problem is the possibility that a given inference, with a given probability of being mistaken for an observation, sometimes will be mistaken for an observation and sometimes will not. This was the concern of the present experiments, which addressed the question: "Given that an inference has been made, under what circumstances is the inference likely to be mistaken for an observation?"

The procedure of varying study materials is not suitable for answering this question. Any variation that occurs before inferences are made, including the variation of study materials, admits the possibility of confounding between the independent variable and some feature of the inferences made. Suppose, for example, that the inferences made in one experimental condition are by chance more distinguishable from observations than the inferences made in another condition. Any difference in results between conditions, then, might be due to the difference between specific inferences rather than to the difference between experimental conditions.

In the present experiments the possibility of such confounding was precluded by placing the manipulation of the independent variables after the inferences had been made. With this procedure any difference between conditions can be attributed to the experimental manipulation. The placement
of the experimental variation after inferences are made, as well as the use of both truth judgments and source judgments, should prove useful in other experiments concerning the confusion of inference with observation.

**Effect of Information Consistency**

Several previous studies have shown that the degree of consistency between implied information and previously observed information affects the probability that an implication will be mistaken for an observation (C. R. Barclay et al., 1984; Bransford et al., 1972). These findings are paralleled by the present finding that inferences were more likely to be mistaken for observations when inferred information was consistent, rather than inconsistent, with subsequently observed information. One possible explanation for these consistency effects concerns the amount of attention given to making inferences as well as the amount of attention given to consistent versus inconsistent information.

The large percentage of inference items judged "true" in the present experiments (even though all inference items were potentially false) suggests that people do not give much attention to making inferences or to evaluating inferences once they have been made. Indeed, given that inference is an essential survival skill, people would not survive for long if they took time for careful consideration of every inference. If little attention is given to inference, then it is not surprising that people don't always remember the source of inferred information, sometimes mistaking inference for observation. Further, it seems reasonable to expect that if more attention were given to an inference, source identification would be more accurate.

Prior research suggests that more attention might be given to an inference when subsequently observed information is inconsistent, rather than
consistent, with the inference. In a study of impression formation (Belmore, 1987) subjects having a newly formed impression of a target person read a list of behaviors attributed to the target. Reading times were longer for behaviors that were inconsistent, rather than consistent, with the target's personality. Some people have suggested that behaviors inconsistent with a prior impression receive more attention than behaviors consistent with it, because it is more difficult to integrate them into the established impression (Hastie, R., 1980; Srull, et al., 1985). Similarly, information inconsistent with a prior inference may be difficult to integrate with the inferred information and, therefore, may be allotted more attention, which in turn could lead to reevaluation of the inference and to recognition of the inference as an inference. In contrast, information consistent with the inferred information would present no difficulty, and so no reevaluation would be expected.

Regardless of how the influence of information consistency is explained, the results of the present experiments suggest that inconsistent information might be applied to prevent confusion between inference and observation in areas as diverse as medical diagnosis, political prediction, and personnel evaluation. In all of these areas, as well as in everyday life, mistaking inference for observation can have serious consequences, especially when inferences are erroneous. Perhaps people could be trained to think more carefully about the source of inferred information through exposure to information that is inconsistent with their inferences (through use of devil's advocates, for example).

Future Research

In the present experiments the delay between making an inference and judging the source of the inferred information had no effect on the confusion of
inference with observation. This may have been due to the relatively small difference in delay between conditions (15 min in Experiment 1; 25 min in Experiment 2). One possibility for future research, therefore, is experimenting with longer periods of delay, perhaps a day or a week. Another possibility is investigating the effects of other variables that might influence the probability of confusion between inference and observation. Possibilities include a person's objectives at the time an inference is made, distraction while an inference is being made, and subject variables such as age or expertise. Also of interest are the possible effects of training people, as mentioned above, to resist mistaking inference for observation by exposing them to information that is inconsistent with their inferences.
References


Samuels (Eds.), *Perception and comprehension* (pp. 243-263). Hillsdale, NJ: Erlbaum.


Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as


Wigmore, J. H. (1931). The principles of judicial proof or the process of proof as given by logic, psychology, and general experience and illustrated in
judicial trials (2nd ed.). Boston: Little, Brown.
Appendix A

Materials

Driver Assault Case: Fictitious Criminal Case

Used in Experiments 1, 2, and 3

Case Description

The defendant, Michael G. Driver, is 35 years old. He is a marketing executive for a major tire company. Defendant Driver is married, and he is the father of two children. He owns a home in the Memorial area of Houston.

Defendant Driver is charged with aggravated assault. At or about 10:00 p.m. on Sunday, 4 March, 1979, Driver is alleged to have beaten Walter A. Hogan with the butt of a pistol or some other hard, probably metal, object. The beating took place in victim Hogan's garage. Hogan lives in a townhouse complex in northwest Houston. Hogan suffered a fractured skull and jaw.

Hogan is 26 years old and single. He is a tennis professional at a country club in north Houston. Defendant Driver is a member of this same country club. The victim (Hogan) said that his garage was dark when he drove in. He hadn't noticed anyone in the garage. When Hogan stepped out of his car, Driver allegedly struck him several times. At no time did Hogan see his assailant, but he remembered hearing, "You've had this coming for a long time," shouted at him.

Hogan had been giving tennis lessons to Driver's wife for three months prior to the night of the assault. Hogan had been a resident of Houston for only six months prior to the assault. He had moved to Houston from Los Angeles. He had wanted to be an actor, and he had had several small parts in made-for-television movies. He had never worked as a tennis instructor before his arrival in Houston.

Peter H. Lapin is 40 years old. He is a maintenance worker at the victim's townhouse complex. Lapin saw a 1979 Cadillac Coupe de Ville parked in front of the victim's townhouse at 10:00 p.m. on the night of the assault. Lapin said that the color of the car seemed to be a combination of blue and green. He could not decide whether it had looked more blue than green, or more green than blue. Lapin was standing on the porch of a townhouse located across the street from the victim's townhouse when he saw the car. Defendant Driver owns a blue-green 1979 Cadillac Coupe de Ville.

David L. Howard is Chief of the Criminal Investigations Division of the Houston Police Department. Howard examined a tire mark found in the mud in
front of the victim's townhouse and made a plaster cast of the mud sample directly after the incident. Howard compared marks on the plaster cast of the mud sample with marks on a plaster cast of the left front tire of defendant Driver's car (the 1979 blue-green Cadillac Coupe de Ville). Howard found that the marks on the cast of the mud sample were identical to the marks on the cast of defendant Driver's tire.

Frank P. Farmer is a detective assigned to the plain-clothes division of the Houston Police Department. Officer Farmer has 20 years experience in this division. Officer Farmer arrested defendant Driver in a departure area at Houston Intercontinental Airport at 11:45 p.m. on the night of the assault on Hogan. Defendant Driver had a one way ticket to Mexico City on a flight due to depart Houston at 11:55 p.m. on the night in question. Driver was alone at the time of the arrest.

Vincent X. Kearney is 35 years old and single. Kearney has been a bartender at the Hide-A-Way club in southwest Houston. Kearney observed defendant Driver's wife, Mary, kissing Hogan (the victim) in full view of defendant Driver as the three were seated at the Hide-A-Way bar on the night before the assault on Hogan.

True/False Test Items and Additional Information Statements

True Items

1. Defendant Driver is an executive.

2. Defendant Driver has two children.

3. The victim's skull was injured.

4. Defendant Driver's wife had taken tennis lessons from the victim (Hogan).

5. Defendant Driver was arrested at Houston Intercontinental Airport.

False Items

1. Defendant Driver works for a computer company.

2. The maintenance worker saw the car through a kitchen window.

3. Defendant Driver had a ticket to Caracas, Venezuela.
Inference Items With Consistent and Inconsistent Additional Information Statements

1. Defendant Driver lives in Houston's Memorial area.

**Consistent:** Defendant Driver has owned the home in the Memorial area of Houston for 11 years. He purchased it just before he and his wife were married.

**Inconsistent:** In addition to the house in the Memorial area of Houston, defendant Driver owns a condominium in the Galleria area of Houston. He finds that there are more interesting things to do in the Galleria area than there are in the quiet Memorial area.

2. The victim was beaten with a hard metal object.

**Consistent:** The nature of the victim's injuries clearly indicated that he was struck with a very hard object such as the butt of a pistol.

**Inconsistent:** The nature of the victim's injuries clearly indicated that he was struck with a very hard object, but it was not clear whether the object was metal or some other hard material, such as wood or stone.

3. The light in the victim's garage was turned off.

**Consistent:** The victim (Hogan) said that he never minded driving into a dark garage. With the garage dark, he would notice his headlights shining on the back wall of the garage, and that would remind him to turn them off.

**Inconsistent:** The victim (Hogan) said that his garage light had not been operating properly for some time. He was not sure whether the light had been turned off on the night of the assault, or whether it had been malfunctioning.

4. The assailant was waiting in the victim's garage when the victim drove in.

**Consistent:** The victim got out of his car immediately after driving into the garage. He didn't see or hear anything before he was attacked.

**Inconsistent:** The victim did not get out of his car immediately after driving into the garage. He took a few seconds to pick up
some fast-food bags and cups from the front seat and floor. He said that this took long enough so that his assailant might have come quietly into the garage while he was cleaning up the car.

5. The assailant shouted at the victim.

**Consistent:** The victim said that during the assault he also remembered hearing, “You deserve to die for what you’ve done,” shouted at him.

**Inconsistent:** The victim said that he was not sure that the person shouting at him and the person attacking him were the same person.

6. The victim (Hogan) had worked as an actor in Los Angeles.

**Consistent:** The victim (Hogan) had been involved in acting off and on since high school. He thought that Los Angeles would be a good place to develop his acting career.

**Inconsistent:** The victim (Hogan) had been trying to develop an acting career for several years. He had been somewhat successful in New York; in Los Angeles he had not been well received.

7. The victim had been working as a tennis instructor since he moved to Houston.

**Consistent:** The victim was offered the tennis instructor’s job on his first day in Houston. He had never thought of tennis as a means of supporting himself, but the idea was appealing to him.

**Inconsistent:** The victim was offered the tennis instructor’s job on his first day in Houston. He had never thought of tennis as a means of supporting himself, and he thought that he should try some better paying job before settling for the tennis instructor’s job.

8. The color of the car that the maintenance worker saw was some combination of blue and green.

**Consistent:** The maintenance worker said that he saw the Cadillac Coupe de Ville quite clearly, as it was parked under a street light.

**Inconsistent:** A police investigator noted that the street lights
on the victim's street were mercury-vapor lights, and that a grey or white car would look blue-green under such a light.

9. The tire marks in the mud in front of the victim's townhouse were made by one of defendant Driver's tires.

**Consistent:** The man who made the plaster casts of the mud sample and of defendant Driver's tire had extensive experience with plaster casts and was known for the consistently high quality of his work.

**Inconsistent:** The tire on defendant Driver's car was almost new and had no unique marks that would positively identify the tire. Other new tires of the same sort would leave the same kind of traces in mud.

10. The detective went to the airport departure area to arrest defendant Driver.

**Consistent:** On the night of the assault, the police learned that defendant Driver was on his way to the airport. They moved quickly to arrest him at the airport before he could leave Houston.

**Inconsistent:** When the police learned that defendant Driver was on his way to the airport, they immediately notified detective Farmer, who was already in the airport departure area. Detective Farmer was often stationed at the airport in the event such a case should arise.

11. Defendant Driver had gone to the airport alone.

**Consistent:** Defendant Driver said that it was his habit to go to the airport alone, as he did not like to ask anyone to drive him. His wife, Mary, said that she had occasionally offered to drive him to the airport, but that he had always refused her offer.

**Inconsistent:** Defendant Driver was seen by an airport employee talking earnestly with another man as he waited in line to buy a ticket. The other man stood in the line talking with defendant Driver for a while and then left without buying a ticket for himself.

12. The man who reported seeing defendant Driver's wife kissing the victim (Hogan) was the bartender at the Hide-A-Way club.

**Consistent:** The man who saw defendant Driver's wife kissing
the victim (Hogan) had worked for many years as a bartender. He had learned to observe everything that goes on in a bar, and not much escaped his notice.

**Inconsistent:** The man who saw defendant Driver's wife kissing the victim (Hogan) frequently changed bartending jobs, moving from one bar to another and then back again. The Hide-A-Way club was one of his favorite places, and when he was not working there, he often spent his nights off socializing at the Hide-A-Way.

**Walsh Murder Case: Fictitious Criminal Case**

**Case Description**

The defendant, Norman W. Walsh, is a 51-year-old bachelor. He resides alone in an apartment on the second floor of a five-story apartment building near downtown Houston. Walsh is regional sales manager for a large insurance company.

Defendant Walsh is charged with first-degree murder in the death of Thomas R. Lew, a 50-year-old bachelor. Death was caused by a single bullet through the victim's heart fired at close range from a .357 Magnum revolver. Time of the shooting/death was placed within one hour of 11:30 p.m., Saturday, 13 January, 1979. The victim, Lew, was part-owner of Logtech, a flourishing oil-well service company.

Ralph N. Moon, is a 53-year-old bachelor and life-long resident of Houston, who lives alone in a townhouse in west Houston. At about 11:50 p.m. on Saturday, 13 January, 1979, Moon had knocked at the door of defendant Walsh's apartment. There was no answer, and since the door was unlocked he entered, finding only victim Lew's body. The body was on the bedroom floor near defendant Walsh's bed. Moon immediately called the police to the scene. Moon found a .357 Magnum revolver beside the body of the victim (Lew). He recognized the revolver as belonging to defendant Walsh.

Grace W. Potts, is a 36-year-old nurse, who is no relation or acquaintance of any other person involved in this case. On the night in question she was on her way home from her work at Methodist Hospital. She lives in an apartment building across the street from the one in which defendant Walsh lives. Potts heard a "sharp crack" that sounded like it came from the window of the defendant's apartment at about 11:28 p.m. on the night in question. She had boarded the bus at 11:10 p.m. on Fannin Street in front of Methodist Hospital. Her apartment is about 18 minutes by bus from Methodist
Hospital. Defendant Walsh's bedroom is the only room in his apartment with a window facing the street.

George T. Farber is the night guard in defendant Walsh's apartment building. Farber is a 33-year-old ex-marine sargeant, who has known defendant Walsh for two years. Farber saw defendant Walsh's car leave the apartment garage at about 11:00 p.m. on the night in question.

Mary S. Parker, is a 25-year-old clerk at a convenience store located in the 10,000 block west on Westheimer Road in Houston. Defendant Walsh had purchased cigarettes from her in the store at about 11:30 p.m. every Saturday night that she worked at the store since she began her job there about four months before the murder. She works at the store on Tuesdays, Thursdays, Fridays, and Saturdays. The day of the shooting, 13 January, 1979, was a Saturday.

George W. Hampton, is a 65-year-old retired bank vice-president. Hampton occupies an apartment in the same building as the one occupied by defendant Walsh. Hampton saw a man who looked like the victim (Lew) strike defendant Walsh during an argument in the hallway outside the defendant's apartment at 10:30 p.m. on the night in question.

True/False Test Items and Additional Information Statements

True Items

1. The murder took place in the month of January.
2. The victim was a bachelor.
3. A revolver was found next to the victim's body.
4. The police were called to the scene after the body was found.
5. The body was found on the floor near defendant Walsh's bed.

False Items

1. The cause of death was several shots to the victim's head.
2. The nurse works in a nursing home.
3. The convenience store clerk works only one day a week.
Inference Items With Consistent and Inconsistent Additional Information Statements

1. The victim was successful in the oil-well service business.
   
   **Consistent:** At the time of his death the victim (Lew) had several large bank accounts and a well-managed investment portfolio.
   
   **Inconsistent:** The victim (Lew) owned only a small share of the oil-well service company, Logtech. He had been trying for some time to obtain a larger share of the company, but he had not succeeded.

2. The victim was shot in defendant Walsh's bedroom.
   
   **Consistent:** When the victim's body was found, it was noticed that a small bedroom table next to the body was out of place. There was a bruise on the victim's head made by a blunt object. The medical examiner said that the victim's head could have struck the table as he fell to the floor.
   
   **Inconsistent:** Blood stains were found on the carpet near the bar in defendant Walsh's dining area. Testing revealed that the blood was of the same type as that of the victim.

3. The murderer was not in the apartment when the victim's body was found.
   
   **Consistent:** The police arrived almost immediately after Moon found the body. They searched the premises to see if the murderer was hiding in the apartment or elsewhere in the building. They found no one suspicious.
   
   **Inconsistent:** Just after Moon found the victim's body in the bedroom he heard a noise in the apartment living room. He did not investigate, because he thought that it might be the murderer, and he was afraid of being attacked.

4. The murder weapon was defendant Walsh's gun.
   
   **Consistent:** Police records verified that the gun found next to the victim's body belonged to defendant Walsh.
   
   **Inconsistent:** Because of damage to the bullet that caused Lew's death, ballistics examination to determine whether or not the bullet had been fired from defendant Walsh's gun was
inconclusive.

5. The nurse's bus ride home took 18 minutes.

**Consistent:** The nurse was certain that her apartment was 18 minutes by bus from the Methodist Hospital, because she had been riding that bus for several years.

**Inconsistent:** Although the nurse's bus ride home from the Methodist Hospital usually took 18 minutes, the bus driver reported that on the night of the murder there had been a delay due to a stalled bus ahead, so that the bus might have been several minutes late.

6. The nurse heard a shot.

**Consistent:** John A. Kasper is a 37-year-old electrician, who occupies the apartment just below defendant Walsh's apartment. Claire G. Kasper is a 33-year-old secretary and the wife of John Kasper. The Kaspers heard something that sounded like a shot at about 11:30 on the night of the murder.

**Inconsistent:** Professor Walter N. Mudge, an acoustics expert, testified that the sound of a .357 Magnum revolver is similar to the sound of a bursting tire. He pointed out that the sound nurse Potts heard could have been something other than a shot.

7. The "sharp crack" came from defendant Walsh's apartment.

**Consistent:** The noise that the Kaspers heard seemed to come from the apartment, above (i.e., defendant Walsh's apartment).

**Inconsistent:** Professor Mudge also testified that sound sources are hard to localize in places where there are many large buildings. He said that the sound nurse Potts heard may have come from some source other than the defendant's window.

8. The night guard saw defendant Walsh leaving the apartment garage.

**Consistent:** The night guard who saw defendant Walsh's car leave the apartment garage at 11:00 was only about 15 feet away when the car passed him. He could see quite clearly, as the area was well lighted.
**Inconsistent:** The night guard who saw defendant Walsh's car leave the apartment garage at 11:00 saw the driver of the car only from the back. Therefore, he could not be certain that the driver was Walsh.

9. On the night of the shooting (Saturday) defendant Walsh bought cigarettes at a convenience store.

**Consistent:** The convenience store clerk was certain that defendant Walsh routinely bought his cigarettes at about 11:30 p.m.

**Inconsistent:** The convenience store clerk had occasionally called in sick on Saturday nights.

10. On the night of the murder (Saturday) the convenience store clerk who regularly served defendant Walsh was on duty at the store.

**Consistent:** The convenience store clerk always looked forward to the end of her workshift, especially on Saturday nights.

**Inconsistent:** The convenience store clerk sometimes switched work days with another clerk.

11. Defendant Walsh had an argument with the victim (Lew) in the hallway outside Walsh's apartment.

**Consistent:** The neighbor who saw the argument in the hallway had seen the victim (Lew) visiting defendant Walsh's apartment several times during the months before he murder. He was quite familiar with the victim's appearance.

**Inconsistent:** The neighbor who saw the argument in the hallway had never seen the victim (Lew) in person, although he had seen Lew's picture in the newspaper after the murder.

12. The victim struck defendant Walsh during the argument in the hallway.

**Consistent:** This same neighbor had seen defendant Walsh and the victim (Lew) arguing on two or three occasions prior to the night of the murder.

**Inconsistent:** This same neighbor could not be certain that Lew was the man who argued with defendant Walsh in the hallway,
because he had not gotten a good look at the man's face.
Appendix B

Instructions for Stated/Inferred Judgments

Experiment 1

In responding to the true/false items in this study there are two ways that you might have judged an item to be true:

1. The item was stated in the information you were given before the true/false test.
2. The item was not stated, but could reasonably be guessed or inferred from something that was stated in the information you were given before the true/false test.

By "stated" we mean that the idea was expressed, although the wording may have been different.

By "inferred" we mean that the idea was not expressed, although it might have been implied.

Example: The defendant was arrested by police for kidnapping, just as he was taking the child from her home.

This example sentence expresses the ideas that:
The police arrested the defendant.
The defendant was with the child when he was arrested.
The police stopped the defendant from taking the child.
These ideas would be classified as Stated.

From the example sentence one might infer that:
The defendant went into the child's home.
The defendant forced the child to go with him.
The defendant was kidnapping the child.
These ideas would be classified as Inferred.

Your next task in this study is to look over your answers to the items on the true/false test. For each item that you marked "True" decide whether the item was Stated in the information you were given before the true/false test, or whether you Inferred it from the information you were given before the True/False test.
Use this scale to indicate whether items marked "True" were Stated or Inferred.

**Stated**

- **S3**: I am certain this information was stated.
- **S2**: I am fairly confident this information was stated.
- **S1**: I am not sure, but I tend to think this information was stated.

**Inferred**

- **I3**: I am certain this information was inferred.
- **I2**: I am fairly confident this information was inferred.
- **I1**: I am not sure, but I tend to think this information was inferred.

For each item that you previously marked "True" please write the appropriate letter and number (e.g., S1, I2, S3) on the line next to the word "True."
Appendix C

Test Item Additions and Deletions

Experiment 3

Extraneous Items Added

**Driver assault case**

Low frequency: Defendant Driver has an inherited blood disorder.

High frequency: The victim (Hogan) drinks coffee in the morning.

**Walsh murder case**

Low frequency: The victim had once traveled to the south pole.

High frequency: Defendant Walsh likes to watch football games on television.

Inference Items Deleted

**Driver assault case**

The assailant was waiting in the victim's garage when the victim drove in.

The man who reported seeing defendant Driver's wife kissing the victim (Hogan) was the bartender at the Hide-A-Way club.

**Walsh murder case**

The victim was shot in defendant Walsh's bedroom.

On the night of the shooting (Saturday) defendant Walsh bought cigarettes at a convenience store.
Appendix D

Instructions for Stated/Not Stated Judgments

Experiment 3

In the preceding tests on this case you have been asked to give your opinion about the truth of the items. In this test we're interested in your memory for the information stated in the Driver (Walsh) case description that you heard before you took either of the true/false tests. An item could be true, even if it wasn't stated. What we're interested in now is your memory of what was stated.

By "stated" we mean that the idea was expressed, although the wording may have been different.

Example: The defendant was arrested by police for kidnapping, just as he was taking the child from her home.

This example sentence expresses the ideas that:
The police arrested the defendant.
The defendant was with the child when he was arrested.
The police stopped the defendant from taking the child.
These ideas would be classified as Stated.

Please indicate whether or not each of these items was stated in the Driver (Walsh) case description you heard before you took either of the true/false tests. For each item place a vertical mark anywhere along the line for that item. A mark at the extreme "Not Stated" end of the line means that you are certain the item was not stated. A mark at the extreme "Stated" end of the line means that you are certain the item was stated. We're interested in your memory for ideas, not exact words.
### Appendix E

**Accuracy of Stated/Not Stated Judgments for Inferences**

as a Function of Consistency Condition and Rating on First True/False Test

<table>
<thead>
<tr>
<th>Rating on first true/false test&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Consistent</th>
<th></th>
<th>Inconsistent</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of observations</td>
<td>Mean accuracy&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Number of observations</td>
<td>Mean accuracy&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>18</td>
<td>148</td>
<td>2.99</td>
<td>152</td>
<td>7.00</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
<td>3.20</td>
<td>23</td>
<td>7.17</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>3.35</td>
<td>21</td>
<td>10.95</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>6.65</td>
<td>12</td>
<td>10.25</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>8.70</td>
<td>11</td>
<td>8.09</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>7.10</td>
<td>7</td>
<td>8.86</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>7.40</td>
<td>6</td>
<td>3.83</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>---</td>
<td>5</td>
<td>6.60</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>8.18</td>
<td>11</td>
<td>13.82</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>5.00</td>
<td>18</td>
<td>11.89</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>13.00</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>7.50</td>
<td>5</td>
<td>10.80</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>18.00</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>---</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3.00</td>
<td>1</td>
<td>11.00</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.00</td>
<td>2</td>
<td>5.50</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.00</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>11.27</td>
<td>4</td>
<td>9.50</td>
</tr>
</tbody>
</table>

**Note.** Data are pooled over subjects.  
<sup>a</sup>1 = false; 18 = true.  
<sup>b</sup>1 = low accuracy; 18 = high accuracy.