

# Memory and Discredited Information: Can You Forget I Ever Said That?\*

Hollyn M. Johnson  
University of Michigan  
330 Packard Rd., Rm. 220B  
Ann Arbor, MI 48104  
hollyn.johnson@um.cc.umich.edu

Colleen M. Seifert  
University of Michigan  
330 Packard Rd., Rm. 216  
Ann Arbor, MI 48104  
seifert@um.cc.umich.edu

## Abstract

Previous research has found that when information stored in memory is discredited, it can still influence later inferences one makes. This has previously been considered as an editing problem, where one has inferences based on the information prestored in memory before the discrediting, and one cannot successfully trace out and alter those inferences. However, in the course of comprehending an account, one can potentially make inferences *after* a discrediting, which may also show influence from the discredited information. In this experiment, subjects read a series of reports about a fire investigation, and their opportunity to make inferences before a correction appeared in the series was manipulated. Subjects received a correction statement either directly following the information it was to discredit, or with several statements intervening. The results show that subjects who received the correction directly after the information it corrected made as many inferences based on the discredited information as subjects who received the correction later (and thus could presumably make many more inferences before the correction occurred). This suggests that discredited information can influence inferences made after a correction, as well as those made before. Several hypotheses accounting for this effect are proposed.

When information stored in memory is shown to be false or unfounded, ideally one would want to diminish or eliminate its effects on future reasoning and understanding processes. Some previous studies on text comprehension (Wilkes & Leatherbarrow, 1988) and jurors' use of inadmissible evidence

(Carretta & Moreland, 1983) have presented subjects with instructions to disregard previously presented information. The results show that subjects remember that instruction when queried about it directly. However, they still show influence from the discredited information when asked to make judgments or inferences (further conclusions not directly presented), relative to subjects who were never exposed to that information. Other studies providing instructions to disregard previous information have also found influence from the discredited information on judgments of personality attributes like friendliness or kindness (Wyer & Budesheim, 1987) and of success in social tasks (Ross, Lepper, & Hubbard, 1975). To understand why information that is discredited still influences inferences, one must look at how and when such inferences are generated.

Research on inferences in text comprehension has proposed that inferences can differ on two dimensions which could be helpful in understanding how discredited information influences inferences. The first, how spontaneously inferences are made, distinguishes between *on-line* inferences (made automatically in the course of comprehension) and *requested* inferences (not normally made during comprehension, but can easily be made when one is asked a question). This is similar to the distinction Hastie and Park (1986) make between on-line judgments (immediate and automatic evaluation) and memory-based judgments (unanticipated, and made only after retrieving the original information from memory). The second distinction about inferences is that they may also differ in their direction: *Forward* inferences occur as predictions or expectations about what will appear next in the text, whereas *backwards* inferences are those that link current information to preceding information, often to provide text coherence and causal connections. Studies have found that backwards inferences occur on-line for anaphoric reference (Corbett & Chang, 1983; Dell, McKoon, & Ratcliff, 1983), text coherence (Keenan, Baillet, & Brown, 1984; McKoon & Ratcliff, 1986), and for

---

\* This research was supported by the Office of Naval Research under contract N00014-91-J-1128 to the University of Michigan.

establishing causal relations (Graesser & Clark, 1985). In contrast, forward inferences tend to occur on-line only for very stereotypical or scriptlike texts (McKoon & Ratcliff, 1986; Seifert, Robertson, & Black, 1985). Kintsch (1988) has also proposed that random inferences, based on common associates of text material, are generated during comprehension. Considering the spontaneity and direction of inferences, and their timing relative to a disregard instruction, leads to different predictions about how discredited information may be used in drawing inferences.

One cause for the continued influence of discredited information may be problems in editing one's memory. That is, when one has already made inferences based on the information, and then encounters an instruction to disregard the information, one has both the original information and the inferences based on it in memory. One may successfully discredit the original information, but may not be able to trace out all the inferences that information supported and properly discount them as well. Thus, the inferences can remain in memory, and one can retrieve them when asked about them later. This could occur if, before the discrediting is introduced, one made either on-line forwards or random inferences, or on-line backwards inferences linking concepts to the information. This latter situation would require a "window" between original presentation of the information and the disregard instruction, during which intervening concepts would be linked back to the information. Finally, one may have generated and stored inferences based on the information during the course of any judgment or inference tasks before discrediting occurs.

On the other hand, if little opportunity for context-relevant from the information is provided before discrediting, then the influence the discredited information has on later inferences would be mainly due to the retrieval and use of the discredited information to generate new inferences. Such an effect may seem counter-intuitive, given findings that subjects do recall that the information is invalid when asked about it directly (Wilkes & Leatherbarrow, 1988; Carretta & Moreland, 1983). In the cases where an inference is made when questioned even after discrediting, or when on-line backwards inferences link later information to the discredited information, the comprehension process must "jump over" the correction notice, illicitly retrieving and using the discredited information.

Little previous research has looked at whether continued influence from discredited information is due to problems in editing prestored inferences, or whether it also can involve illicit retrieval and use of the discredited information itself. Some social psychology research suggests pre-discrediting inferences are difficult to edit (Anderson, Lepper, &

Ross, 1980; Anderson, New, & Speer, 1985; Hastie & Park, 1986; Wyer & Budesheim, 1987). Wilkes and Leatherbarrow (1988) demonstrated the influence effect in a text comprehension experiment, but remained neutral as to its cause. In their original experiment, subjects read a series of reports on a fire investigation, with an original statement that some volatile materials (cans of paint and pressurized gas cylinders) were stored in a closet. Several messages later, a correction occurred, stating that the closet was empty and thus did not contain volatile materials. Subjects who received this statement to disregard the information about the closet's contents still reported inferences consistent with or mentioning the volatile materials more often than did control subjects, who never received the information about possible storage of volatile materials. This occurred whether the correction directly repeated the information that one was to disregard (direct edit condition) or whether it was just indirectly referred to (indirect edit condition).

The experiment reported here uses a modified set of materials taken from Wilkes and Leatherbarrow (1988), where subjects are presented with a series of reports, one to a page, which they read through at their own pace. In the delayed correction condition, which replicates the direct edit condition in the original Wilkes and Leatherbarrow (1988) paper, subjects read about volatile materials stored in a closet early in the report series, and received a correction five messages later. In the no-mention control group, subjects hear no mention of any volatile materials stored in the closet. This replicates the control group in the original experiment, and is consistent with control conditions used in many belief persistence studies (e.g., Anderson, Lepper, & Ross, 1980; Ross, Lepper, & Hubbard, 1975). Finally, in the immediate correction group, no information intervenes between presentation of the information about the volatile materials and the correction, so subjects have little opportunity to make inferences based on the volatile materials, before hearing that those materials do not exist. In this immediate correction condition, the information about the volatile materials would still be in working memory (along with any random inferences constructed from common associates of the message propositions, and remaining after integration with surrounding context (Kintsch, 1988)) when the discrediting occurred. This should make it easier for subjects to determine that the first and the second, discrediting message about the closet share reference. Other research has proposed that proximity facilitates establishing coreference (Cirilo, 1981), and that establishing coreference is important for detecting contradictions (Epstein, Glenberg, & Bradley, 1984). Subjects could then potentially resolve the contradiction in working memory, and use that representation for further interpretation of subsequent

information. Also, for all conditions, the messages prior to the correction are written so as to limit opportunities for forwards and backwards inferences linking the information to other content. Thus, in the immediate correction condition, one might have only those random inferences that one could generate based on the volatile materials message and that happened to be context-relevant, and so survived an integration process, whereas in the delayed correction condition, one would have much more opportunity for inferencing (both random and more strategic, bridging inferences) before the discrediting occurred.

If the influence from discredited information occurs because subjects have difficulty editing inferences made before the disregard instruction, or correction, one would expect more influence from the discredited information to be evident in the delayed correction group. Subjects would have a window within which they could make additional, strategic, causal and coherence-maintaining inferences prior to the correction, which they might not be able to successfully track down and alter when the correction occurs. In the immediate correction condition, subjects will not have the opportunity to make as many context-relevant inferences before the correction occurs. If the problem lies in editing pre-stored inferences, one would not expect as much influence from the discredited information for the immediate correction condition, since fewer inferences could be formed. However, if making illicit post-correction retrievals leads to continued influence of discredited information, then one would expect influence from the discredited information in the immediate correction group, and in the delayed correction group, compared to the no-mention control. Thus, the immediate and delayed correction conditions, together with the control condition having no discrediting, will determine whether influence of discredited information occurs just due to problems in editing pre-correction inferences, or can also occur due to illicit post-correction retrieval and use of the discredited information.

Further, to test whether the effect depends on the wording of the correction message, two versions of the message are used. One version presents the correction directly but in a complex clause, similar to the version used by Wilkes and Leatherbarrow (1988). A second version presents the correction as a direct assertion that no volatile materials were stored in the closet, rather than embedding the reference to the materials in a subordinate clause.

## Method

**Subjects.** Sixty-four University of Michigan undergraduates participated in a single session lasting approximately 50 minutes. They received course

credit in an introductory psychology class for participating. Subjects were run in groups of 8 to 10.

**Materials.** The materials were modified versions of a series of reports used by Wilkes and Leatherbarrow (1988), describing the investigation of a warehouse fire. The series consisted of 13 individual messages, each 2-4 sentences long. The messages were combined into a booklet, with one message per page. The critical messages concerned the contents of a storage closet on the premises. For the no-mention control group, the fifth message in the series stated that this closet was empty, and this information was not controverted later. For the two correction groups, the fifth message stated that the closet contained cans of oil paint and pressurized gas cylinders. Then, for the immediate correction group, Message 6 stated that the previous message regarding the closet's contents was incorrect and that the closet was empty. For the delayed correction group, this statement appeared as Message 12. Half the subjects in both the delayed and the immediate correction groups received a complexly worded correction message, similar to that in Wilkes and Leatherbarrow (1988); the other half received a more direct wording (see Table 1).

Two memory tests were also prepared: a free recall summary of the reports' contents and a questionnaire adapted from Wilkes and Leatherbarrow (1988), including ten questions on facts directly presented in the messages, ten other questions requiring the subjects to make inferences about the event, and two final questions assessing whether subjects were aware of any correction or contradiction in the series (sample questions are shown in Table 2). All questions appeared in the same order for each subject, with all fact questions appearing before any inference questions to prevent the latter from introducing biases, and the two contradiction questions appearing at the end.

Direct message version:

10:40 a.m. A second message received from Police Investigator Lucas regarding the investigation into the fire. It stated that there were no cans of paint or gas cylinders in the closet that had reportedly contained them; the closet had actually been empty before the fire.

Complex message version:

10:40 a.m. A second message received from Police Investigator Lucas regarding the investigation into the fire. It stated that the closet reportedly containing cans of paint and gas cylinders had actually been empty before the fire.

Table 1: *Style of Corrections*

What was the possible cause of the toxic fumes?  
 What could have caused the explosions?  
 Why do you think the fire was particularly intense?  
 For what reason might an insurance claim be refused?

Table 2: *Sample Inference Questions from the Memory Questionnaire*

**Procedure.** Each subject received a booklet of reports and was instructed to read through it at his or her own pace, but not to go back and reread any of the messages. Subjects were also told that they would be asked to recall the information later. When individual subjects had finished reading, they were given the free recall test. Then all subjects did an unrelated distractor task for 10 minutes. After this time had elapsed, subjects received the memory questionnaire and were instructed to answer each question based on their understanding of the reports.

## Results

A complete description of the results is presented in Johnson and Seifert (1992). A coder (blind to the experimental conditions) scored the responses to the inference questions as consistent with either a "negligence" theme or a "supplies" theme. The negligence theme encompassed responses that were consistent with believing that the warehouse contained carelessly stored volatile materials, as would be reasonable if the information about the volatile materials had not been discredited. References to the presence of gas cylinders and paint, carelessness, or the closet itself without indications that it was empty were coded with this theme. The supplies theme was coded if the responses presented a reasonable inference about the fire that was not included in the previous categories, such as references to stored stationery at the warehouse or the structure of the building. One would expect control subjects to make inferences consistent with this theme, because they received no information about the volatile materials.

To assess specific influences of the information one was to disregard, all uncontroverted references to paint and gas cylinders in either memory test were counted and analyzed in a 3 x 2 analysis of variance, with Group (no-mention, delayed correction, immediate correction) and Message (complex or direct) as factors. The mean number of references to the stored volatile materials for each cell is shown in Table 3. The results showed a main effect of group,  $F(2, 59) = 10.21, p < .0001$ , with both the correction groups showing more influence than the no-mention control group. The main effect of message showed a trend towards significance,  $F(1, 59) = 9.87, p < .09$ , with

the direct version resulting in fewer references to the stored volatile materials. However, the interaction between group and message was not significant, indicating that the same overall pattern was found in both message conditions. A post-hoc comparison, collapsed over message, showed a significant difference between the control group, which never heard about the stored volatile materials, and the two groups that did,  $t(62) = 4.11, p < .0001$ .

To determine whether the groups differed in the number of inferences consistent with the different possible themes, further 3 x 2 analyses of variance were done. The inference categories were considered separately because the scores are not statistically independent of one another. The mean number of inferences per subject for the negligence and supplies themes by group are shown in Table 3. The main effects of group showed a significant difference on both the "negligence" and the "supplies" themes;  $F(2, 59) = 16.99, p < .0001$  for negligence;  $F(2, 59) = 10.31, p < .0001$  for supplies. Planned comparisons revealed significant differences between the correction groups and the control group, with the control group making significantly more inferences consistent with the supplies theme than the correction groups did. They also made significantly fewer inferences consistent with the negligence theme, relative to the correction groups;  $t(62) = 4.31, p < .0001$  for supplies;  $t(62) = 5.52, p < .0001$  for negligence. There were no significant differences in either the number of responses consistent with the arson theme, or in number of questions left blank. No other effects or interactions were significant in analyses of the inference variables.

A subject was scored as noticing the correction if it was referred to accurately in either of the memory tests. Both the correction groups showed high levels of recall of the correction, with 100% of the delayed correction group and 90.9% of the immediate correction group recalling it. Analyses of the inference categories, omitting subjects who did not recall the correction, showed the same patterns of significance as reported above.

The free recall summaries were scored for component idea units, using an adaptation of procedures described in Kintsch (1974). Only

Groups	Delayed	Immediate	Control
Supplies theme	3.3	2.9	5.4
Negligence theme	3.5	4.0	1.3
References to volatile materials	2.7	3.5	1.0

Table 3: *Number of Inferences Consistent with Story Themes, by Group*

messages common to all three conditions were scored (i.e., no correction information was included in this measure), and a unit was scored as recalled if the subject reproduced a recognizable portion of its content. Fact questions from the questionnaire were also scored for accurate content. There were no group differences in summary recall or in fact recall ( $F < 1$ ). However, there were main effects of message for both variables:  $F(1,59) = 4.87, p < .03$  for summary recall, and  $F(1,59) = 10.81, p < .002$  for fact recall, with subjects receiving the direct message showing somewhat poorer recall of the rest of the passage. Complex message subjects recalled 14.3 summary units whereas direct message subjects recalled 12.2 units; the number of facts correctly recalled was 9.2 and 8.3 for complex and direct message subjects, respectively.

## Discussion

The results show that the two correction groups made more negligence inferences based on the volatile materials, and more direct references to those materials, than did the control group. Further, there was no difference between the delayed and immediate correction groups on either of these measures. Thus, when subjects saw the incorrect information, whether corrected early or late in the sequence, they showed influence from it, relative to the control group. This replicates Wilkes and Leatherbarrow (1988), and is consistent with work on belief perseverance (Ross, Lepper, & Hubbard, 1975). Here, the effect occurred even in the immediate correction group, where subjects had little opportunity to make, and therefore little need to edit, inferences. These effects cannot be accounted for by other failures of memory: none of the groups differed in recall of the reports, and over 90% of those in the correction groups recalled the correction. Thus, subjects in both correction groups had the raw materials available to be able to make correct inferences, yet they did not use it with optimal success. Additionally, the fact that the immediate correction group used the discredited information provides some evidence that influence from discredited information can also occur due to illicit post-correction retrieval and use of the discredited information, as well as due to problems in editing prestored inferences (as in the delayed correction condition).

The results of the message manipulation generally support the interpretation advanced here, with some limitations. There were no main effects of message for the inference variables, so type of message did not lead to significant differences in the number of negligence theme inferences or direct references to the stored volatile materials, which suggests that the effect does not entirely depend on correction style.

However, as the subjects receiving the direct message also showed significantly fewer free recall units and facts recalled, further work on message effects is needed.

Because the experiment does not directly assess whether subjects in the immediate correction condition made any inferences before the correction occurred, the result could still be due to problems editing prestored inferences; however, there are several reasons why this interpretation is not plausible. First, due to the fact that the two correction groups did not differ in the number of inferences, it seems unlikely that random inferences generated before the discrediting message account for the effect. One would have to argue that all or most of the context-relevant inferences could be generated by this process in both groups. Kintsch (1988) distinguishes between this method of inference generation and says that often one must make additional, more strategic bridging inferences to augment this process and come up with a coherent text. Because the delayed correction condition allows for both processes to occur before the discrediting, one might expect it to show a higher number of prestored (and unedited) inferences, but this was not the case.

Second, the messages limited the number of backward inferences that could be made upon encountering the volatile materials information because they did not present any characteristics of the fire that could be linked with the volatile materials via backward inferences once one heard about them. The characteristics of the fire, which the stored volatile materials could potentially explain, were all mentioned *after* the correction for the immediate correction group. Also, because the earlier messages just mentioned the existence of a fire, it is unlikely that subjects would make a lot of forward inferences. Van den Broek (1990) argues that forwards inferences are more likely to be made when constrained by necessary and sufficient causal conditions; in this case, the mention of the fire gives few clues to its specific characteristics, and so one might not expect many predictions until more information comes in. Lastly, the questionnaire only asked about characteristics of the fire mentioned after the correction, and so would be more likely to catch backwards, bridging inferences subject may have made, which should favor the delayed correction group, where these could be made before the discrediting occurred.

Thus, overall, the results suggest that discredited information can influence inferences due to processes occurring *after* a disregard instruction or correction occurs. One explanation is that subjects may make illicit backward inferences following the correction. That is, as subjects read post-correction statements, they may make connecting inferences to make the text coherent. This may involve the discredited

information simply because it fills the need to find causal antecedents. A second explanation for the effect may be that simply mentioning the stored volatile materials acts to make that information more available in memory. Subjects may not make inferences involving the discredited information on-line after the correction, but may instead retrieve any available information at the time of question and use it in further inferencing.

The possibility that subjects make illicit backward inferences to maintain the account's coherence raises some interesting issues. Making backwards inferences involving the discredited information in order to establish causal connections and preserve coherence in an account presents a serious problem for comprehension accuracy. The results here suggest that asserting information results in its propagation through later inferences despite direct, immediate correction. Thus, in all understanding contexts, such as those involving reports of news events, discrediting alone appears to be an insufficient method for removing the traces and influence of incorrect information. Further studies may ascertain what factors lead to illicit post-correction retrieval, when it occurs, and whether some forms of correction might overcome the persistent influence of discredited information.

## References

- Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39(1-6), 1037-1049.
- Anderson, C. A., New, B. L., & Speer, J. R. (1985). Argument availability as a mediator of social theory perseverance. *Social Cognition*, 3(3), 235-249.
- Carretta, T. R., & Moreland, R. L. (1983). The direct and indirect effects of inadmissible evidence. *Journal of Applied Social Psychology*, 13(4), 291-309.
- Cirilo, R. K. (1981). Referential coherence and text structure in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, 20(3), 358-367.
- Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory and Cognition*, 11, 283-294.
- Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior*, 22, 121-132.
- Epstein, W., Glenberg, A. M., & Bradley, M. M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition*, 12, 355-360.
- Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93(3), 258-268.
- Johnson, H. M., & Seifert, C. M. (1992). Memory and discrediting information. Unpublished manuscript.
- Keenan, F. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 115-126.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 162-182.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 82-91.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32(5), 880-892.
- Seifert, C. M., Robertson, S. P., & Black, J. B. (1985). Types of inferences generated during reading. *Journal of Memory and Language*, 24, 405-422.
- van den Broek, P. (1990). Causal inferences and the comprehension of narrative texts. In Graesser, A. C., & Bower, G. (Eds.), *The Psychology of Learning and Motivation (Vol. 25): Inferences and Text Comprehension*. San Diego: Academic Press, Inc.
- Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology*, 40A, 361-387.
- Wyer, R. S., & Budesheim, T. L. (1987). Person memory and judgments: The impact of information one is told to disregard. *Journal of Personality and Social Psychology*, 53, 14-29.