Suppression of Misinformation in Memory

Hollyn M. Johnson
University of Michigan
330 Packard Road
Ann Arbor, MI 48104-2994
hollyn_johnson@um.cc.umich.edu

Colleen M. Seifert
University of Michigan
330 Packard Road
Ann Arbor, MI 48104-2994
seifert@umich.edu

Abstract

Agents in a dynamic world must continue to comprehend and reason about events, even after they learn that previously encoded information about an event is incorrect. As a result, some mechanism is needed to modify incorrect information in memory, and allow one to use new, superceding knowledge instead. How is misinformation suppressed in human memory? A study using a text understanding paradigm and a standard anaphoric inference task investigates this problem of updating memory. Subjects read a set of stories, half of which contained a correction, and were asked to make a speeded word-recognition judgment for a probe word appearing after an anaphor sentence. Subjects in a short delay condition showed *slower* reaction times to correct referents in correction stories than in control stories that did not contain misinformation. Those in the longer delay condition showed no difference in reaction times to correct referents, but more priming for *invalidated* items in correction stories. These results suggest that misinformation can interfere with accessing correct information, but that an additional comprehension process, possibly suppression-like, may facilitate access to correct information after delay.

Processing Corrections during Comprehension

In learning about events in the world, one may encounter misinformation that is corrected later. In many cases, people may have to do inferential reasoning after the correction occurs. This presents a challenge, because one must integrate the new, post-correction information only with *valid* pre-correction information. If the new information was linked to the misinformation instead, an inaccurate understanding of the event would result. The present research examines whether people involve misinformation in their post-correction inferences, and what memory processes might decrease or prevent its use.

A number of studies, in various domains, have found that misinformation can still influence one's inferences or judgments, even though it is no longer relevant for them. Belief perseverance research (Ross, Lepper, & Hubbard, 1975) has found that subjects rate their true ability at an experimental task consistent with feedback they'd received, even after a debriefing indicates that the feedback was false. Other studies (Wyer & Budesheim, 1987; Wyer & Unverzagt, 1985) have had subjects form person impressions, based on a list of behaviors, and then instructed subjects to disregard some of them. Under some conditions, forget-cued subjects still make judgments similar to those of subjects not instructed to forget. Finally, several text comprehension studies (Johnson & Seifert, in press; Wilkes & Leatherbarrow, 1988) have had subjects read a series of reports that contained a correction, and found that misinformation influenced subjects' answers to questions that required inferring beyond the facts presented.

Johnson and Seifert (in press) also found evidence suggesting that misinformation influences post-correction , as well as pre-correction, inferences. They manipulated whether misinformation was immediately corrected, or was corrected only after subjects had learned several more facts about an event. An immediate correction would limit the chances to involve misinformation in pre-correction inferences, because subjects would have learned fewer facts the misinformation could help explain. The results showed that subjects in both the immediate and delayed correction groups made significantly more influenced inferences than did subjects in a control group not exposed to the misinformation. More importantly, the two correction groups did not differ, even though the immediate correction group allowed subjects to make fewer pre-correction inferences based on the misinformation. This suggests that subjects were making a number of post-correction inferences. Subjects could be involving misinformation in post-correction inferences at two points: either on-line, as they continue to process the account, or on demand, as they answer questions posed later.

The current research will test whether misinformation influences anaphoric inferences made after a correction has occurred. Subjects make anaphoric inferences to establish coreference and coherence in a text; for example, if one reads about a "burglar" and then that "the criminal was apprehended," one will infer that "burglar" (the referent) and "criminal" designate the same entity. Much evidence suggests that such inferences are typically made on-line (Dell, McKoon, & Ratcliff, 1983; O'Brien, Duffy, & Myers, 1986). In texts that do not contain misinformation ("normal" texts), subjects show more priming for proper referents (e.g., "burglar") after they read an anaphor (e.g., "criminal") than they do for equated control terms that are inappropriate referents (e.g., "cat") (Dell, McKoon & Ratcliff, 1983; O'Brien, Duffy, & Myers, 1986). McKoon and Ratcliff (1980) hypothesized that one makes anaphoric inferences by searching memory for possible referents for the anaphor, and then reactivating the proper one.

One might expect this pattern of "normal" inferencing if misinformation suffers inhibition when people read a statement that presents correct information. Other previous research has found evidence that irrelevant information becomes inhibited. MacDonald and Just (1989) had subjects read disjunctions (e.g., bread but not butter) and found decreased priming for the negated member (e.g., butter). Intentional forgetting research has found that subjects recall fewer items from the first of two word lists if they are explicitly told to forget it, even when the items are incidentally learned (Geiselman, Bjork, & Fishman, 1983). This suggests that retrieval inhibition accounts for the effect, rather than differential encoding. If misinformation also becomes inhibited after a correction, it would be less active in memory and harder to reactivate in later inferencing. In this case, misinformation would be unlikely to influence post-correction inferencing.

On the other hand, corrected misinformation within an account may still influence later inferencing in a number of ways First, it may impede people's ability to make inferences involving the correct information, by inhibiting its reactivation. When processing the correction, one might establish an inhibitory relation between the invalid and correct concepts. This could inhibit the invalid information, so one becomes less likely to produce it, but could also potentially lead to interference with correct information, relative to situations that do not include invalidated information.

Second, the misinformation itself may become reactivated during post-correction inferencing. According to a minimal inference theory (McKoon & Ratcliff, 1992), subjects make relatively few inferences in constructing a representation, and do not necessarily make all the inferences needed to maintain coherence in an account. Instead, new information may be appended to the current representation, with little elaboration or interconnection with other elements. In such a case, a correction might be incorporated into a representation without being linked to the original, now invalid information, so the two would be independent. Then, when an anaphoric reference is made later, both the incorrect and correct information may be activated.

Even if misinformation becomes reactivated, however, post-retrieval comprehension processes may prevent it from permanently influencing one's understanding of post-correction information. Previous research on comprehension of normal texts suggests that people may immediately activate inappropriate lexical meanings (Swinney, 1979) or associations (Gernsbacher, 1990) as they read; however, with an increased delay, this activation drops for inappropriate, but not appropriate, information. Gernsbacher (1990) has proposed that suppression and facilitation processes operate to inhibit inappropriate information and enhance appropriate information. Such processes might also operate in correction situations, to decrease influence of invalid information and enhance that of correct information.

The purpose of the experiment is to determine the extent to which subjects reactivate invalid and correct information when they read stories containing corrections. Subjects will read correction stories and control stories, which contain no misinformation. They will then do speeded recognition (McKoon & Ratcliff, 1990) of either appropriate referents (the correct information in both types of stories) or inappropriate ones. Probe onset will occur either 300 ms or 850 ms after anaphor sentence offset. If subjects simply inhibit the misinformation, they should respond faster to appropriate than inappropriate probes in both story conditions, and at both delay periods. On the other hand, if misinformation interferes with the inferential process, one might expect faster responses to inappropriate probes, and slower responses to appropriate ones, when subjects read correction stories than when they read control stories. However, this may only occur at the short delay. Subjects may still show more activation for correct than invalid information at a longer delay, if they can use post-retrieval comprehension processes such as suppression and facilitation.

Method

Design

The design was a 2 x 2 x 2 factorial, with Delay (short, long), Probe (appropriate, inappropriate), and Correction (yes, no) as factors. Probe and Correction were manipulated within subjects, and Delay was a between subjects factor. In the long delay condition, the interval between the anaphor sentence offset and the probe onset was 850 ms. In the short delay condition, this interval was 300 ms.

Subjects

One hundred and thirty-three University of Michigan students participated in the experiment, as part of an introductory psychology course requirement.

Materials

Subjects read 35 target stories, styled after news events, each 15 sentences long. Half of the stories contained corrections; the remaining, control stories were normal texts containing no misinformation. Each story contained 3 critical sentences: one sentence providing either misinformation or a control statement (line 4), one sentence giving the correct fact (line 8), and one assessment sentence making anaphoric reference to the correct fact (line 12) (see Table 1). Line 4 in the correction versions presented an alternative in a causally important context, whereas in the control versions, this sentence also mentioned the alternative, but as an incidental fact. For example, a story about an athlete's new contract had alternatives of Boston and Dallas, which are both teams for which the athlete could now play. Thus, subjects in both conditions saw the same words, which they had to identify later. The alternative mentioned in line 4 will be referred to as the *inappropriate referent*, and the one mentioned in line 8 will be referred to as the *appropriate referent*, because it is the one that subjects in both groups should end up with as the "correct" fact. Lines 8 and 12 were the same in both the correction and control version of each story. The stories appeared in a random order, with the constraint that no more than three correction or control stories appeared in a row.

Table 1: Sample story materials.

Athlete's new contract

(1) A famous pro athlete has been offered a lucrative new contract.
(2) It will make him one of the highest paid sports figures.
(3) As a free agent, he has negotiated with others for more money.

(4) *The athlete will accept a generous offer to play for Boston.* (misinformation)
or
(4) *He wants a contract like that just given to a player for Boston.* (control)

(5) The management he currently plays for won't pay what he is asking.
(6) They argue paying him that much will alienate the other players.
(7) They also say the athlete is too old to receive a 7-year contract.

(8) *His agent says Dallas just signed the pro to play there next year.* (correct info)

(9) He will be able to play with an exciting rookie prospect there.
(10) The athlete will hold a press conference to discuss the deal.
(11) He will also meet with the new staff he will play for.
(12) *The athlete says he is excited about playing for his new team.* (anaphor)
(13) He adds he immensely respects the management he will now play for.
(14) Fan reaction to the deal has mainly been favorable.
(15) They feel it will make their championship hopes more realistic.

For each story, subjects saw either an inappropriate or an appropriate referent (or distractor items) presented as the first probe presented immediately after the anaphor sentence. Both correction and probe conditions were counterbalanced across subjects, so both versions of each story were tested with both probes. Three other probes followed the critical items during each test sequence, to serve as distractors. These items were balanced so that for each ordinal position two through four, true and false answers occur equally often. Also, for each story, two true/false items were constructed, to test overall story comprehension. Lines 4 and 8, which introduced the probe alternatives, were never used as comprehension items. The presentation order of the two comprehension items for each story, as well as whether they were true or false, was chosen randomly, with the constraint that each possible TF response pattern occurred equally often across the story set.

Procedure

All materials were presented using the MEL program (Schneider, 1988), on a Zenith computer, and all responses were made via keyboard. Each subject completed two training sets, designed to familiarize the subject with reaction time tasks. After training, subjects were told they would have to do two tasks. One was to read carefully, and try to comprehend, a series of stories based on live-on-the-scene news reports and to answer test questions after each story. The second task was to verify whether probe words presented during the story had appeared earlier in the story. After presentation of the anaphor sentence (line 12) in each story, the subject saw a "???????" cue, lasting for 300 ms in the short delay condition, or 850 ms in the long delay condition. The first word in each probe series was either the inappropriate referent probe (presented in line 4), the appropriate referent probe (presented in line 8) for the story currently being read, or else a distractor item not appearing in the story. A "**" symbol appeared under the probe after 650 milliseconds, and the word disappeared after 2 seconds if no response was made. Subjects were told to respond as quickly and as accurately as they could, and to try to beat the "**" deadline. Accuracy feedback appeared after each trial. After each story, subjects' comprehension of that story was tested with two true/false statements.

Results

For each story for each subject, reaction times were included in analyses if they were for correct identifications and if the reading time for the immediately preceding anaphor sentence was greater than 900 milliseconds. Reaction times were then trimmed to be within 2.5 SDs of the mean; thus, times shorter than 300 milliseconds or longer than 1261 milliseconds were eliminated. The data were analyzed in 2 x 2 x 2 analyses of variance, with Change (yes, no), Probe (appropriate, inappropriate), and Delay (long, short) as factors, using both subjects and materials as random factors. The pattern of results can be seen in Figure 1. Reaction times showed a significant interaction of Change and Probe under a subjects analysis, $F(1, 131) = 6.75$, $p = .01$, and with materials as a random factor, $F(1, 27) = 6.4$, $p = .018$. Planned comparisons showed a significant difference between appropriate probes in the control and correction conditions in the short delay condition, $F(1, 67) = 6.93$, $p = .011$; $F(1, 27) = 2.56$, $p = .12$ under subjects and materials analyses, respectively. This was not significant in the long delay condition. However, planned comparisons showed a significant difference between inappropriate probes in the control and correction conditions in the long delay condition, $F(1, 64) = 5.13$, $p = .03$; $F(1, 27) = 6.53$, $p = .017$. This was not significant in the short delay.

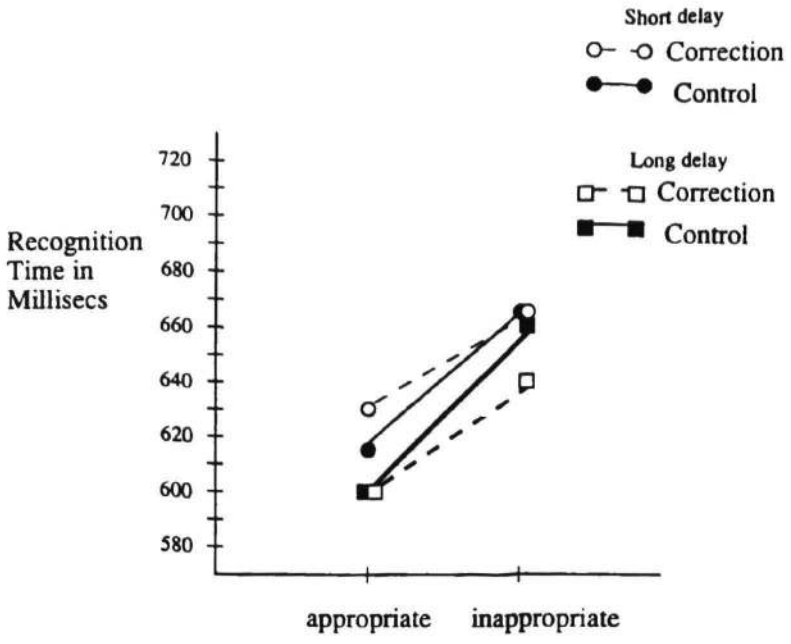

Figure 1: Mean reaction times, by story type, probe type, and delay.

The results of subjects and materials analyses also showed a significant interaction of Change and Delay, $F(1, 131) = 5.38$, $p = .023$ and $F(1, 27) = 5.5$, $p = .026$, respectively. Planned comparisons of reactions times to appropriate correction story probes by delay showed a significant difference under both subjects and materials analyses, $t(1, 131) = 2.41$, $p = .017$; $t(1, 27) = 4.45$, $p = .0001$. The difference in reaction times to inappropriate correction story probes by delay was significant in a materials analysis, $t(1, 27) = 4.15$, $p = .0001$, and showed a trend toward significance in a subjects analysis, $t(1, 131) = 1.54$, $p = .13$.

Analysis of the number of inaccurate responses showed a significant interaction of Change and Delay under subjects analysis, and a trend under materials analyses, $F(1, 131) = 5.38$, $p = .02$; $F(1, 27) = 4.0$, $p = .054$, respectively, with more errors in the control conditions with the short delay. Post-hoc comparisons showed a significant difference in reaction times to appropriate probe control stories by delay under both subjects and materials analyses, $t(131) = 2.67$, $p = .009$; $t(27) = 2.88$, $p = .008$. There were no significant differences in reaction times to appropriate or inappropriate correction story probes by delay. Subjects got a mean of 91% of the true/false items correct. There was a significant main effect of delay, $F(1, 131) = 6.18$, $p = .015$, with subjects in the long delay condition showing 92% correct, versus 90% in the short delay condition. Subjects correctly identified 82.8% of the total recognition items; they were correct 89.5% of the time on just the target probes.

Discussion

The results suggest that both inhibition and post-comprehension facilitation contribute to making appropriate post-correction inferences. The fact that appropriate information is reactivated more slowly in the correction than the control condition, at a short delay, suggests that one has established an inhibitory link between the correct and invalid information. One may have encoded the eventually-invalidated information in a causal context, but not done so for the same information when mentioned only incidentally. Then, one inhibits the invalid information when one processes a correction, resulting in no apparent difference in activation for the two inappropriate probes. This also interferes with one's ability to activate the proper referent. Other studies (Basden, Basden, & Gargano, 1993; MacDonald & Just, 1989) have also found evidence for inhibition based on instructions to ignore particular information.

However, overcoming the initial difficulty in activating an appropriate referent would require a second process that facilitates its activation later. Over delay, activation for appropriate information from correction stories increased, but the activation of appropriate information from control stories did not change. This second process also contributes to subjects' ability to make appropriate post-correction inferences. It also has the apparent side effect of releasing inhibition on invalid information, making it more likely to continue to be involved in one's text representation than incidental information is. Thus, misinformation indirectly influences subsequent inferencing by inhibiting access to the correct information initially, but facilitation processes soon overcome this effect.

The results found here contrast with a number of studies that have found that misinformation influences one's judgments (Ross, Lepper, & Hubbard, 1975; Wyer & Unverzagt, 1985) and inferences (Wilkes & Leatherbarrow, 1988). One factor that could account for this is a discrepancy in the type of

correction used. In the present research, the correction provided an alternative to the misinformation, rather than merely negating it, as in most of the previous work. This also suggests that creating a link between correct and invalid information plays a role in being able to make inferences that do not involve invalid information. Several previous studies have found that providing an alternative to discredited information decreases its influence (Anderson, 1982; Johnson & Seifert, in press).

In summary, the findings here imply that subjects are competent to meet the challenge of post-correction comprehension. They make inferences that involve appropriate information, and do not involve invalid information directly. The presence of invalid information initially affects how easily one can activate appropriate information, but post-correction processes can compensate. This eventually results in a level of activation similar to that of appropriate information in stories that do not contain corrections. The findings also imply that this competence occurs in the context of trying to understand the story, and may be relatively automatic. The fact that subjects can respond so quickly to appropriate information, and that activation levels change for appropriate information in correction stories, suggests that subjects are not using conscious, strategic processes to determine correct referents.

These results also impose constraints on models of language processing regarding their integration with semantic memory. Increasingly, AI models that comprehend language incorporate conceptual structures already in memory into the understanding process. As a consequence, information in memory can be directly accessed, and potentially changed, as a consequence of comprehension processes (Martin, 1990; Martin & Riesbeck, 1986). While these models have focused on the role of activation from related concepts in resolving meaning (Charniak, 1983), the same models must be expanded to account for the effects of correction. In particular, the present results suggest that the presence of a prior correction changes the processing of target information. From these results, additional inhibitory processes may be needed to account for the successful resolution of meaning.

Related research on resolving conflicting information within belief systems within the theory revision framework (e.g., Gick & Matwin, 1991) and on the effects of conflicting information on one's beliefs (e.g., Chinn & Brewer, 1993) suggest that the correction processes evident in the current studies may have a broad range of applicability outside of comprehension tasks. While detecting and resolving contradictory beliefs is typically addressed through deliberative, conscious reasoning in these models, there may be some role for low-level, automatic suppression and facilitation processes in ensuring the use of preferred information in memory.

Psychological models of text comprehension are beginning to provide results about such suppression mechanisms (see Gernsbacher, 1990), and further work will provide a detailed depiction of the role of these mechanisms in comprehension. Because information in realistic settings is both dynamic and unstable, comprehension processes must allow for the correction and updating of concepts already in memory. Only through systematic studies incorporating misinformation can the mechanisms that maintain veracity and allow change be unveiled.

Acknowledgments

This work was supported by the Office of Naval Research under contract N00014-91-J-1128 to the University of Michigan. Thanks to Seema Shastri for assistance with materials, and to an anonymous reviewer for comments on an earlier draft.

References

Anderson, C. A. (1982). Inoculation and counterexplanation: Debiasing techniques in the perseverance of social theories. *Social Cognition, 1*(2), 126-139.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 603-616.

Charniak, E. (1983). Passing markers: A theory of contextual influence in language comprehension. *Cognitive Science, 7*, 171-190.

Chinn, C. A., & Brewer, W. F. (1993). Factors that influence how people respond to anomalous data. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 318-323). Hillsdale, NJ: Lawrence Erlbaum.

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.

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112*(1), 58-72.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum.

Gick, M., & Matwin, S. (1991). The importance of causal structure and facts in evaluating explanations. In L. A. Birnbaum and G. C. Collins (Eds.), *Machine Learning: Proceedings of the Eighth International Workshop* (pp. 51-54). San Mateo, CA: Morgan Kaufman.

Johnson, H. M., & Seifert, C. M. (in press). Sources of the continued influence effect: When misinformation in memory affects later inferences. To appear in *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 633-42.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures in anaphoric reference. *Journal of Verbal Learning & Verbal Behavior, 19*, 668-82.

McKoon, G., & Ratcliff, R. (1990). Dimensions of inference. In Graesser, A. C.,& Bower, G. (Eds.), *The Psychology of Learning and Motivation (Vol. 25): Inferences and text comprehension*. San Diego: Academic Press, Inc.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440-466.

Martin, C. E.. (1990). Direct memory access parsing (Tech. Rep. 93-07). Doctoral dissertation, Chicago: University of Chicago, Dept., of Computer Science.

Martin, C. E. & Riesbeck, C. (1986). Uniform parsing and inferencing for learning. In *Proceedings of the Fifth Annual Conference on Artificial Intelligence*, Philadelphia, PA.

O'Brien, E. J., Duffy, S. A., & Myers, J. L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 346-52.

Ross, L., Lepper, M. A., & 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.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments and Computers, 20*, 206-17.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior, 18*, 645-59.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40A*(2), 361-387.

Wyer, R. S., & Budesheim, T. L. (1987). Person memory and judgments: The impact of information that one is told to disregard. *Journal of Personality & Social Psychology, 53*(1), 14-29.

Wyer, R. S., & Unverzagt, W. H. (1985). Effects of instructions to disregard information on its subsequent recall and use in making judgments. *Journal of Personality & Social Psychology, 48*(3), 533-549.