Silence Is Not Golden
A Case for Socially Shared Retrieval-Induced Forgetting
Alexandru Cuc,1 Jonathan Koppel,2 and William Hirst2

1Nova Southeastern University and 2New School for Social Research

ABSTRACT—The present research explored the effect of selective remembering and the resulting “silences” on memory. In particular, we examined whether unmentioned information is more likely to be forgotten by a listener if related information is recollected by the speaker than if related information is not recollected by the speaker. In a modification of the retrieval-induced forgetting paradigm, pairs of individuals studied material, but in the practice phase, only one member of each pair selectively recalled it, while the other listened. Experiment 1 employed paired associates, and Experiment 2 used stories. Experiment 3 involved not controlled practice, but free-flowing conversation. In each case, results from a final memory test established not only within-individual retrieval-induced forgetting, but also socially shared retrieval-induced forgetting. The results demonstrate that listening to a speaker remember selectively can induce forgetting of related information in the listener.

The seemingly innocent conversations people have about their shared past can distort their memories. A speaker can reinforce the retention of some of a listener’s memories over others, can remind the listener of forgotten information, can reshape the listener’s existing memories, and can implant new memories (Basden, Basden, & Henry, 2000; Loftus, 2005; Meade & Roediger, 2002). An appreciation of these influences has altered the legal system’s treatment of eyewitness testimony (Wells & Olsen, 2003), the debate over recovered memories (Loftus & Ketcham, 1994), and scholars’ understanding of the formation of collective memory and, in turn, collective identity (Cuc, Ozuru, Manier, & Hirst, 2006; Weldon, 2001).

To a large extent, this work has focused on what the speaker says. But remembering is a selective process, with a speaker intentionally or unintentionally often remaining silent about important information. Such silences may arise because a speaker wishes to deceive or wants to avoid something stressful or socially taboo (Zerubavel, 2006). Silences may also appear because what is left unsaid is inconsistent with conversational goals (Tversky & Marsh, 2000; Rime, 1995), is unlikely to be readily accessible given the conversational direction (Basden, Basden, Bryner, & Thomas, 1997; Weldon & Bellinger, 1997), or is a poor fit with what the speaker may view as appropriate for his or her audience (Echterhoff, Higgins, & Groll, 2005).

In selectively remembering, a speaker may avoid a topic altogether, failing, for instance, to talk about the buildup of the war in Iraq. Alternatively, a speaker may engage the topic, but avoid mentioning certain specifics, for instance, the claim that Iraq had weapons of mass destruction. In the study reported in this article, we explored whether these two routes to silence have different mnemonic consequences. If two people avoid talking about something completely, then it is not surprising that their silence on that topic will promote forgetting of related information (Pasupathi, Stallworth, & Murdoch, 1998). But what happens when they recollect some information, while leaving related information unmentioned? Do such silences within the context of conversational remembering lead a listener to forget more than one might expect if the conversation had never taken place at all? Is it better to avoid a topic altogether if you want someone to forget, let us say, claims about Iraq having weapons of mass destruction, or is it better to engage the topic, but leave out the embarrassing parts?

Our study was guided by work on retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; for a review, see Anderson, 2005). In this work, subjects study category-target pairs such as animal-cat, animal-dog, vegetable-broccoli, and vegetable-pea. They then receive retrieval practice by completing cued words (e.g., animal-d__). Practice selectively focuses on some pairs (e.g., animal-dog), but not on other related pairs (e.g., animal-cat) or on sets of unrelated pairs (e.g., all the vegetable pairs). Finally, subjects recall the exemplars when provided category cues.

Numerous studies have shown that on the final cued-recall test, the recall of retrieval-practiced items (Rp + items) is better
than the recall of unpracticed unrelated items (Nrp items), but, more important, the recall of Nrp items is better than the recall of unpracticed items related to the RP+ items (Rp− items; for a review, see Anderson, 2005). According to Anderson, when successfully completing a stem in the practice phase, subjects inhibit competing responses, that is, the items related to the practiced pairs (but see Butler, Williams, Zacks, & Maki, 2001). This inhibition should take place automatically if (a) a subject attempts to retrieve an item, (b) competing responses are elicited in the process of retrieval, and (c) in the end, retrieval is successful.

This previous work focused on the effect of selective remembering on a speaker, that is, on within-individual retrieval-induced forgetting (WI-RIF). Do silences in the recollections of a speaker also induce forgetting of related material in the listener? That is, is socially shared retrieval-induced forgetting (SS-RIF) possible? In the WI-RIF paradigm, the conditions needed for automatic inhibition, and hence, retrieval-induced forgetting, are built into the experimental design. The instructions demand retrieval, and the stimulus material is designed to elicit response competition. Task demands imposed on the speaker, however, do not guarantee retrieval on the part of the listener. Retrieval is not automatic for the listener, but is under his or her control (Anderson & Green, 2001). If a listener decides not to, is instructed not to, or for any other reason does not concurrently retrieve the information, then there is no reason to expect retrieval-induced forgetting on the part of the listener.

In the present study, we were interested in whether there are conditions under which SS-RIF is possible. The first experiment contrasted two monitoring conditions, one in which listeners were induced to retrieve information with the demand that they monitor the speaker’s recollections for accuracy, and another that made concurrent retrieval unnecessary by asking listeners to attend to superficial features. These modes of listening are two of the many different types of listening that occur in everyday situations. For example, jury members may monitor for accuracy when listening to eyewitness testimony. Dinner guests may monitor for superficial features when judging the entertainment value of a story recounted by a fellow guest.

In the first experiment, we closely followed the methods of Anderson et al. (1994), employed paired associates as the stimulus material, and carefully controlled the practice undertaken by the speaker. This procedure did not allow us to answer directly our questions about silences in free-flowing conversations. We cannot assume that listeners routinely monitor for accuracy in free-flowing conversations and, hence, concurrently retrieve information along with the speaker. Indeed, inasmuch as monitoring for accuracy may be an effortful task, it may not be the default mode of listening, especially if listeners routinely assume that what is told to them is correct (Grice, 1975; Marsh & Tversky, 2004). Consequently, in a follow-up experiment, we examined whether free-flowing conversations also produce SS-RIF.

**EXPERIMENT 1**

To introduce a social dimension into the methodology of Anderson et al. (1994), we paired subjects and asked members of each pair to study paired associates simultaneously. During the retrieval-practice phase, one member of the pair, the speaker, received additional practice, while the other member, the listener, listened to the speaker complete the word stems. In the accuracy condition, the listener was asked to monitor the speaker’s accuracy; in the superficial-listening condition, the listener monitored the speaker’s fluidity of response.

**Method**

**Subjects**

The 76 paid subjects were evenly and randomly assigned to the accuracy and superficial-listening conditions. One half of the subjects in each group were randomly selected as speakers, and the other half served as listeners. Subjects were students at the New School or Nova Southeastern University, or were recruited from Craig’s List, a classified-ad Web site.

**Stimulus Material**

The material consisted of the 60 category-exemplar paired associates used in Experiment 1 of Anderson et al. (1994, Appendix B; e.g., fruit-orange). There were eight experimental categories and two filler categories, with six exemplars associated with each category. Within a category, none of the exemplars began with the same first two letters.

The learning list consisted of all 60 pairs. The list was divided into six blocks of eight experimental and two filler pairs. Each block contained one exemplar from each category. In the first block, the filler item appeared in the beginning, whereas in the last block, the filler item appeared at the end. No two categories appeared in sequence more than once.

We created four retrieval-practice lists containing pairs from the study list. In the retrieval-practice lists, the exemplars were presented as stems requiring completion (e.g., fruit-or ____). The first and last pairs in each list involved filler items. Each incomplete pair was presented three times, with at least three items between repetitions. In order to construct the practice lists, we randomly divided the eight experimental categories into two sets of four. We also divided the six exemplars associated with each category into two sets of three. Each practice list contained one set of the categories, as well as one set of exemplars of each category in the set. The practiced exemplars constituted the Rp+ items; the unpracticed exemplars of a practiced category made up the Rp− items; the exemplars from the unpracticed categories were the Nrp items. Across the four lists, the assignment of items to practice types was counterbalanced.

---

728
**Design and Procedure**

The experiment was conducted in four phases. In the first phase, subjects studied the learning list as it was presented, one pair at a time, 5 s per pair, in the center of a computer screen. Both the speaker and the listener sat in front of the screen, one to the right of center and the other to the left, with their positions counterbalanced. Both were asked to study the pairs so that they could subsequently remember them. After a 60-s pause, the retrieval-practice phase of the study began, with a retrieval-practice list appearing on the screen at a rate of 10 s per pair. The speaker completed each pair aloud, while the listener monitored either the accuracy of what the speaker said or the “smoothness and fluidity” with which the speaker completed the stem. In both conditions, the listener was told to record his or her judgments in a booklet, using a scale from 1 to 7. The experimenter sat in a corner of the room and recorded the speaker’s responses.

In the third phase, subjects performed a distractor task in which they talked for 5 min about a recent movie they had both seen. Finally, subjects were asked to recall the exemplars. They were told to write down the originally studied exemplars in a booklet in which each category cue appeared at the top of a separate page.

**Results and Discussion**

We wanted to determine whether both speakers and listeners demonstrated the patterns of forgetting obtained by Anderson et al. (1994), as well as whether any effects depended on the monitoring condition. To streamline the discussion, we present only the most relevant analyses. We conducted a three-way analysis of variance (ANOVA), with one within-subjects factor (retrieval type: Rp+, Rp−, or Nrp) and two between-subjects factors (function: speaker or listener; instructions: accuracy or superficial listening). We found a practice effect (see Fig. 1). Not only was there a main effect for retrieval type, \( F(2, 144) = 65.16, MSE = 0.02, \eta^2_p = .48, p < .001 \), but the proportion of Rp+ items recalled in the final memory test, \( p(Rp+) \), was greater than both the proportion of Nrp items recalled, \( p(Nrp) \),
and the proportion of $R_p$–items recalled, $p(R_p–), t(75) = 8.04, d = 0.69, p < .001, and t(75) = 8.59, d = 1.03, p < .001, respectively. There was a slight trend for $p(R_p+) to be greater for speakers than for listeners in the fluidity condition, a pattern consistent with the presence of a generation effect (Slamecka & Graf, 1978), $t(36) = 1.80, d = 0.33, p < .08$. We did not find such a trend in the accuracy condition.

Critically, there was also a three-way interaction among retrieval type, function, and instructions, $F(2, 144) = 3.55, MSE = 0.02, \eta^2_p = .04, p < .05$. To explore this result, we conducted a separate two-way ANOVA for each function, focusing exclusively on $N_p$ and $R_p$–items. In the speakers’ analysis, $p(N_p)$ was greater than $p(R_p–)$ regardless of instructions. Thus, we found main effects for retrieval type, $F(1, 36) = 21.37, MSE = 0.01, \eta^2_p = .38, p < .001, and for instructions, $F(1, 36) = 14.73, MSE = 0.08, \eta^2_p = .29, p < .001,$ but not an interaction between retrieval type and instruction. Thus, speakers demonstrated WI-RIF regardless of the monitoring condition of the listener.

For the listeners, instructions mattered, in that we found an interaction between retrieval type and instructions, $F(1, 36) = 7.94, MSE = 0.01, \eta^2_p = .19, p < .01.$ In the accuracy condition, $p(R_p–)$ was less than $p(N_p)$, $t(18) = 3.39, d = 0.43, p < .01$, indicating that the accuracy instructions produced SS-RIF. SS-RIF was not present in the superficial-listening condition, in that we found no significant difference between $p(R_p–)$ and $p(N_p)$ in this condition.

Following Anderson et al. (1994), we calculated two measures of impairment associated with retrieval: absolute impairment, $\frac{p(N_p) - p(R_p–)}{p(N_p)}$, and proportional impairment, $\frac{|p(N_p) - p(R_p–)|}{p(N_p)}.$ Neither measure differed between speakers and listeners in the accuracy condition. Both measures differed significantly between speakers and listeners in the superficial-listening condition, $t(36) = 3.00, d = 0.97, p < .01,$ for absolute impairment and $t(36) = 3.51, d = 2.16, p < .01,$ for proportional impairment. This result is consistent with our claim that superficial listening produces WI-RIF for a speaker, but not SS-RIF for a listener.

The results demonstrate that listening to someone remember can induce forgetting in the listener. Such SS-RIF occurred when listeners monitored accuracy, but not when they monitored superficial features.

**EXPERIMENT 2**

To extend these findings to free-flowing conversations, we needed to use material more suitable for conversation than paired associates are. Experiment 2 followed the procedure of Experiment 1, but employed a story as stimulus material. Following standard story grammars, we constructed the stories so that they were easily divided into a sequence of episodes, which, in turn, could be divided into a sequence of events, thereby creating episode-event rather than category-exemplar pairs (Goldman, Graesser, & van den Broek, 1999). In the practice phase, subjects completed fragments of phrases, rather than word stems. Inasmuch as we wanted to extend our positive finding of SS-RIF to stories, we employed only the accuracy instructions in Experiment 2.

**Method**

**Subjects**
The 56 paid subjects were evenly and randomly assigned to serve as speakers or as listeners. Subjects were students at the New School or recruited from Craig’s List. One pair’s final recall was more than 2 standard deviations below the mean. We eliminated the data from this pair from the analyses.

**Stimulus Material**
The 503-word story about a boy’s day contained eight experimental episodes, divided into subsets of four episodes each. To ensure that the story flowed naturally, we varied the number of events per episode, but each subset of four episodes contained 18 events. When the events in an episode set were practiced, half of the 18 events were assigned to be $R_p+$ items, and the other half were $R_p–$ items. The assignment of an episode set to be practiced or unpracticed and the assignment of events to be $R_p+$ or $R_p–$ items was counterbalanced across subjects. Each stimulus in the retrieval-practice phase consisted of a brief description of an episode plus a fragmented description of an event in that episode (e.g., for the event *picked up Jane, Walked to school– pi ____ up J*). Each retrieval-practice list began with a filler item and ended with a filler item; these filler items were constructed out of the remaining items in the story.

**Design and Procedure**
The design and procedure were similar to those of Experiment 1 in most respects. In the study phase, the story was presented on a computer screen, one episode at a time; each episode was displayed for an ample 40 s. Retrieval practice followed immediately afterward. Speakers were told to complete each event fragment presented on the computer screen with an event from the specified episode, and to say their answers aloud; listeners were instructed to monitor speakers’ accuracy and to record their judgments on a scoring sheet, using a scale from 1 to 7. In the following 20-min distraction phase, subjects wrote about their best friend or worst enemy. In the final cued-recall test, a brief description of an episode appeared at the top of each page of the test booklet. Subjects were told to recall the events associated with each episode.

**Results and Discussion**
The overall results were similar to those of Experiment 1 (see Fig. 1). As before, practice mattered, inasmuch as overall, $p(R_p+) was greater than both $p(N_p)$ and $p(R_p–)$, $t(53) = 3.34,$
d = 0.40, p < .01, and t(53) = 7.75, d = 0.95, p < .001, respectively. We also found a significant difference between p(Rp+) of speakers and of listeners, t(52) = 2.36, d = 0.65, p < .05. This difference was consistent with the presence of a generation effect: Speakers remembered more of the practiced items than did listeners, who did not have the advantage of generating the items themselves.

As in Experiment 1, we conducted a two-way ANOVA focusing on Nrp and Rp− items; this analysis had one within-subjects factor (retrieval type: Rp− or Nrp) and one between-subjects factor (function: speaker or listener). We found evidence of retrieval-induced forgetting for both speakers and listeners. There was no interaction between retrieval type and function, but there was a main effect for retrieval type, F(1, 52) = 20.79, MSE = 0.35, ℓp^2 = .29, p < .001. That is, for both speakers and listeners, p(Rp−) was less than p(Nrp), t(26) = 2.17, d = 0.37, p < .05, and t(26) = 4.37, d = 0.68, p < .01, respectively. We also failed to find a significant difference between the absolute and proportional impairments of the speakers and listeners. This result suggests that the levels of WI-RIF for a speaker and SS-RIF for a listener are equivalent.

EXPERIMENT 3

In Experiment 3, we modified the procedure of Experiment 2 so that during the retrieval-practice phase, members of each pair recounted the story jointly instead of completing fragments. Thus, the paradigm employed in Experiments 1 and 2 was transformed to study the effect of omissions in free-flowing conversations on forgetting.

Method

The 40 paid subjects were either students at the New School or recruited from Craig’s List. To ensure that subjects would selectively recall the story in their conversation, stating some episodes and events while leaving out others, we inserted six additional episodes into the story used in Experiment 2; each episode contained four to seven events. In the practice phase, we asked the members of each pair to recount the story jointly, without mentioning how they should monitor each other. They completed their recounting of the story without the experimenter present, informing the experimenter that they were done when both felt they had little more to add. As in Experiment 2, a distractor task and a final cued-recall task followed. We tape-recorded and transcribed each pair’s recounting of the story.

Results

Two coders identified in each recounting the critical episodes and events from the original story and used this information to classify all items as Rp+ (the event appeared in the recounting), Nrp (nothing about the episode appeared in the recounting), or Rp− (the event did not appear in the recounting, but other events from the same episode did) for each subject. For a listener, an item was coded as an Rp− event only if the listener did not recall any of the events from the episode him- or herself, but the speaker did. For a speaker, an item was coded as an Rp− event only if the speaker remembered a related event. For instance, if a speaker recalled event A but not event B from an episode, and the listener did not recall either A or B, we classified event B as an Rp− event for the listener (and also for the speaker) because the related event A had been recalled only by the speaker. There was good reliability between the two coders, κ = .82. Disagreements were resolved. There were a few conversations in which one or both subjects did not produce any Nrp or Rp− items.

Once again (see Fig. 1), we found a practice effect, insasmuch as p(Rp+) was greater than p(Nrp) and p(Rp−), both for speakers, t(33) = 9.37, d = 1.33, p < .001, and t(39) = 26.06, d = 4.89, p < .001, and for listeners, t(33) = 4.05, d = 0.82, p < .001, and t(27) = 8.03, d = 2.88, p < .001. Additionally, as in Experiment 2, results were consistent with a generation effect, in that p(Rp+) was greater for the speakers than for the listeners, t(39) = 8.30, p < .001. Finally, we observed both WI-RIF and SS-RIF. We conducted a two-way ANOVA, focusing exclusively on Npr and Rp− trials. This analysis had one within-subjects factor (retrieval type) and one between-subjects factor (function). As before, we did not find an interaction between retrieval type and function, but discovered a main effect for retrieval type, F(1, 24) = 20.79, MSE = 0.11, ℓp^2 = .37, p < .001. That is, p(Rp−) was less than p(Nrp) for both speakers, t(34) = 5.39, d = 0.96, p < .01, and listeners, t(24) = 3.62, d = 1.26, p < .01. We failed to find a significant difference between the absolute and proportional impairments of speakers and listeners.

The more memorable the material, the more likely it is to be mentioned in a conversation and to be subsequently remembered. Consequently, practice effects might reflect the memorability of the material, rather than practice alone. Our interest centered on the unmentioned items. There was no reason to expect that one type of unmentioned item (e.g., Nrp) would be more memorable than the other type (Rp−). To assess this prediction, we obtained a measure of memorability. For each experimentally critical event, we calculated over subjects the proportion of times the event was mentioned in the conversation. We then determined which events served as Nrp or Rp− items for a subject and averaged the proportions over item type for each subject, thereby obtaining a memorability measure for each subject’s Nrp and Rp− items. There was no significant difference between the memorability scores for Nrp (M = .42, SD = .12) or Rp− (M = .49, SD = .12) items; if anything, the trend was in the wrong direction.

GENERAL DISCUSSION

The present findings indicate that forgetting in a listener is more dramatic when the speaker is silent about some, but not all, of a
complex of related information than when the speaker fails to discuss the related information in its entirety. Such SS-RIF holds whether the information is presented as paired associates or stories. It may not hold for all stories, however. Anderson and McCulloch (1999) did not find WI-RIF when the targets of category-target pairs were integrated (i.e., participants interrelated category exemplars during study). An analogous situation would obtain if the events within an episode of a story were closely related to each other by more than the fact that they were in the same episode. For such stories, one might fail to find either WI-RIF or SS-RIF.

Our research was guided by the claim that SS-RIF depends on whether a listener retrieves information concurrently with a speaker. As a consequence, even though SS-RIF should share many of the features of WI-RIF, SS-RIF and WI-RIF need not be present simultaneously. SS-RIF is under the (intentional or unintentional) control of the listener and does not depend simply on the recollection of the speaker. From this perspective, it is not surprising that SS-RIF occurred when listeners monitored speakers’ accuracy, but not when they monitored speakers’ fluidity.

Less predictable were our findings for conversations. Not all conversational partners will monitor accuracy in all conversations. Whether or not they do probably depends in part on the goals guiding a conversation. We probably found SS-RIF in the conversations collected in this study because of the conversational goal we provided our subjects. By asking them to recount a previously studied story, we may have indirectly suggested that they monitor accuracy. Other conversational goals may not make the same demands. For instance, accuracy and concurrent retrieval may not be important in a conversation in which entertainment is highly valued, making such conversations unlikely candidates for SS-RIF (Marsh & Tversky, 2004). There may even be instances in which partners do not share the same conversational goal, for instance, when one tries to implant memories, while the other tries to resist external influences (Echterhoff, Hirst, & Hussy, 2005). SS-RIF is still possible in such cases, with the “resisting” subject, but not the “implanting” subject, potentially susceptible to SS-RIF.

Although much still needs to be understood about SS-RIF, it is clear that forgetting can be more pronounced when a topic is engaged without all the specifics being mentioned than when a topic is avoided altogether. Thus, if speakers want to induce listeners to forget about the claims that Iraq had weapons of mass destruction, then they should discuss all the justifications for the war except the presence of such weapons.

Thus, the present work uncovers an unrecognized influence of conversations on subsequent remembering, one that needs to be added to the caveats about memory routinely offered to lawyers, clinicians, and others. Until now, psychologists have limited their concern to the influence of what speakers actually say on listeners’ memories, emphasizing speakers’ ability to implant new memories or supplant old memories. The present research adds to these concerns by demonstrating that selective remembering is also a potential problem, inducing forgetting in listeners, as well as speakers.

Acknowledgments—This research was supported by National Institutes of Health Grant MH 0066972.

REFERENCES


(RECEIVED 7/12/06; REVISION ACCEPTED 9/25/06; FINAL MATERIALS RECEIVED 10/14/06)