Silence is not golden:
A case for socially-shared retrieval-induced forgetting

Alexandru Cuc
Nova Southeastern University

Jonathan Koppel
New School for Social Research

William Hirst
New School for Social Research

Word Count: 3,996 words

Address correspondence to Alexandru Cuc, Division of Behavioral and Social Sciences, Nova Southeastern University, 3301 College Avenue, Fort Lauderdale-Davie, Fl 33314, calex@nova.edu.
Abstract

The present research extends work on social contagion and explores the effect of the selective remembering and the resulting “silences” on memory. In particular, it examines whether unmentioned information is more likely to be forgotten by a listener if related information is recollected by a speaker. In a modification of the retrieval-induced forgetting paradigm, two individuals studied material, but in the practice phase, only one person selectively recalled it, while another listened. Experiment 1 employed paired-associates, while Experiment 2 used stories. Experiment 3 involved not controlled practice, but free-flowing conversation. In each case, we observed a pattern in a final memory test that 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.
Silence is not golden:
The case for socially-shared retrieval-induced forgetting

The seemingly innocent conversations people have about their shared past can distort participants’ memories. A speaker can reinforce the retention of some listeners’ memories over others, can remind the listener of forgotten information, can reshape a 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 & Ketchum, 1994), and scholar’s 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 remaining silent about often important information. Such silences 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 (Tverksy & Marsh, 2000; Rimé, 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 audience (Echterhoff, Higgins, & Groll, 2005).

In selectively remembering, a speaker may avoid a topic altogether, failing, for instance, to talk about the build-up of the Iraqi War. Alternatively, a speaker may engage
the topic, but avoid mentioning certain specifics, for instance, the claim that there were weapons of mass destruction. We explore here 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 will promote forgetting (Pasupathi, Stallworth, & Murdoch, 1998). What about instances in which 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’s say, claims about weapons of mass destruction, or is it better to engage the topic, but leave out the embarrassing parts?

We are guided here by work on retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; for a review, see Anderson, 2006). In this work, participants study category-target pairs such as animal-cat, animal-dog, vegetable-broccoli, vegetable-pea. They then receive retrieval practice by completing, for instance, animal-d_____. Practice selectively focuses on some pairs (Rp+, e.g., animal-dog), but not on other related pairs (Rp-, e.g., animal-cat) or on sets of unrelated pairs (Nrp, e.g., all the vegetable pairs). Finally, participants recall the exemplars when provided category cues.

Numerous studies have shown that in the final cued recall, the recall of Rp+ items is better than the recall of Nrp items, but more importantly, the recall of Nrp items is better than the recall of Rp- items (for a review, see Anderson, 2006). According to Anderson, when successfully completing the stem in the practice phase, participants inhibit competing responses, in this instance, the items related to the practiced pairs (but see Butler, Williams, Zacks, & Maki, 2001). Importantly, this inhibition should
automatically take place if (1) an individual attempts to retrieve an item, (2) competing responses are elicited in the process of retrieval, and (3) 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 for 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 under her control (Anderson & Green, 2001). If a listener decides not to, is instructed not to, or for any other motive, does not concurrently retrieve, then there is no reason to expect retrieval-induced forgetting on the part of the listener.

The present paper is interested in whether there are conditions under which socially-shared retrieval-induced forgetting is possible. The first experiment contrasted two monitoring conditions, one in which listeners are induced to retrieve with the demand that they monitor the speaker’s recollections for accuracy, another that makes concurrent retrieval unnecessary by asking listeners to attend to superficial features. Both modes of listening are two of the many different types of listening that occur in everyday situations. Jury members may monitor for accuracy when collectively remembering 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 Anderson et al. (1994), employed paired associates as the stimulus material, and carefully controlled the practice
undertaken by the speaker. This procedure does not allow us to assess directly our claims about silences in free-flowing conversations. We cannot assume that listeners routinely monitor for accuracy in free-flowing conversations, and hence, concurrently retrieve along with the speaker. Indeed, inasmuch as monitoring for accuracy may be an effortful task, accuracy monitoring 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, we also examined SS-RIF in free-flowing conversations in follow-up experiments.

Experiment 1

In order to create a social dimension to Anderson et al (1994), we asked two participants to study simultaneously paired associates. 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 pair. In the accuracy condition, the listener was asked to monitor for accuracy; in the superficial listening condition, the listener monitored for the fluidity of the response.

Methods

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

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 at least 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.

The four retrieval-practice lists contained pairs from the study list, only now the exemplar was presented as a stem requiring completion, e.g., *fruit-or______*. The first and last pairs in the 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+* pairs; the “unpracticed” exemplars of a “practiced” category made up the *Rp-* pairs; the exemplars from the “unpracticed” categories designated the *Nrp* pairs. We constructed the four lists by counterbalancing the pairing of practice type and items.

*Design and Procedure.* The experiment was conducted in four phases. In the first phase, participants 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, another to the left, with their position
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 the stem pairs for one of the retrieval-practice lists appearing on the screen at a rate of 10 s per pair. The speaker completed each pair aloud, while the listener either monitored for the accuracy of what the speaker said or for the “smoothness and fluidity with which the speaker completes the stem.” In both conditions, the listener was told to record their judgment in a booklet, using a 1 to 7 scale. The experimenter sat in a corner of the experimental room and recorded the speaker’s responses.

In the third phase, as a distracter task, participants talked for five minutes about a recent movie they had both seen. Finally, participants were asked to recall the exemplars in a booklet in which a category cue appeared on the top of each page. Participants were told to write down the originally studied exemplars.

Results and Discussion

We wanted to determine whether both speaker and listener replicated Anderson et al. (1994) and under which monitoring condition. In order to streamline the discussion, we will only present the most relevant analyses. We conducted a three-way ANOVA, with one within-subject factor (retrieval type – Rp+, Rp- and Nrp) and two between-subject factors (observer function -- speaker and listener – and instructions -- accuracy and superficial listening). See Figure 1. We found a practice effect. Not only was there a main effect for retrieval type, $F(2,144) = 65.16, MSE = .02, \eta_{p}^{2} = .48, p < .001$, but the proportion of Rp+ items recalled in the final memory test [$p(Rp+)$] was greater than both $p(Nrp)$ and $p(Rp-)$, respectively, $t(75) = 8.04, d = .69, p < .001$, and $t(75) = 8.59, d = 1.03, p < .001$. Interestingly, there was a slight trend for $p(Rp+)$ 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), $t(36) = 1.80, d = .33, p < .08$. We did not find even a trend for the accuracy condition.

Critically, there was also a three-way interaction among retrieval type, observer function, and instructions, $F(2,144) = 3.55, MSE = .02, \eta_p^2 = .04, p < .05$. In order to explore this result, we undertook separate two-way ANOVAs for the two observer functions, now focusing exclusively on Nrp and Rp- trials. In the speakers’ analysis, regardless of instructions, $p(Npr) > p(Rp-)$. Thus, we found main effects for retrieval type, $F(1,36) = 21.87, MSE = .01, \eta_p^2 = .38, p < .001$, and for instructions, $F(1,36) = 14.73, MSE = .08, \eta_p^2 = .29, p < .001$, but not an interaction between retrieval type and instruction. These results suggest that speakers demonstrated WI-RIF regardless of the monitoring condition of the listener.

As for the listeners’ analysis, now instructions mattered, in that we found an interaction between retrieval type and instructions, $F(1,36) = 7.94, MSE = .01, \eta_p^2 = .19, p < .01$. For the accuracy instructions, $p(Rp-) < p(Nrp)$, $t(18) = 3.39, d = .43, p < .01$, indicating that these instructions produced SS-RIF. SS-RIF was not present in the superficial listening condition, in that we found no significant difference between $p(Rp-)$ and $p(Nrp)$.

Following Anderson et al., we calculated two measures of impairment associated with retrieval: absolute impairment $[p(Nrp) – p(Rp-)]$ and proportional impairment $[(p(Nrp) – p(Rp-))/p(Nrp)]$. Both measures did not differ between the speaker and the listener with the accuracy instructions. A significant difference was observed for the superficial listening condition [absolute impairment: $t(36) = 3.00, d = .97, p < .01$;
proportional impairment: \( t(36) = 3.51, d = 2.16, p < .01 \). 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 for accuracy, but not for superficial features.

Experiment 2

If we are going to extend these findings to free-flowing conversations, we need to examine material more suitable for a conversation than paired-associates. Experiment 2 follows the procedure of Experiment 1, but employs 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-events rather than category-exemplars pairs (Goldman, Graesser, van den Brock, 1999). In the practice phase, we employed fragment completion of a phrase instead of stem completion of a word. Inasmuch as we wanted to extend our positive finding of SS-RIF to stories, we employed only the accuracy instructions here.

Method

Participants. The 56 paid participants were evenly and randomly assigned to serve either as speakers or listeners. Participants were students at the New School or from Craig’s List. One pair’s final memory was more than two standard deviations below the mean. We eliminated them 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. In order to ensure that the story flowed naturally, the number of events varied 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 as Rp+ items, with the other half as Rp- items. The assignment of an episode set as practiced or unpracticed, and the assignment of events as Rp+ or Rp- was counterbalanced across participants. One episode was walked to school, which consisted of events such as picked up Jane. The fragment completion in the retrieval practice phase consisted of a brief description of the episode plus a fragmented description of the event, e.g., Walked to school – pi___ up J____. There were two filler items at the beginning and end constructed out of the filler 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 time for an ample 40 s. Retrieval practice immediately followed. The speaker was told to complete aloud the episode-event fragment pair presented on the computer screen with an event from the specified episode, while the listener was instructed to monitor for accuracy and record his 1-7 judgment on a scoring sheet. A 20-minute distraction phase followed in which participants 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 a page of the test booklet. Participants were told to recall the events associated with the episode.

Results and discussion
The results are similar to those of Experiment 1. Again, see Figure 1. As before, practice mattered, inasmuch as overall \( p(Rp+) \) was greater than both \( p(Nrp) \) and \( p(Rp-) \), respectively, \( t(53) < 3.34, d = .40, p < .01 \) and \( t(53) = 7.75, d = .95, p < .001 \). We also found not a trend, as we did in Experiment 1, but a significant difference between \( p(Rp+) \) of speakers and listeners, \( t(52) = 2.36, d = .65, p < .05 \). 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 MANOVA, focusing on \( Nrp \) and \( Rp- \) trials, with one within-subject factors (retrieval type – \( Rp- \) and \( Nrp \)) and one between-subject factor (observer function -- speaker versus listener). We found evidence of RIF for both speakers and listeners. There was no interaction between retrieval type and observer function, but a main effect for retrieval type, \( F(1,52) = 20.79, MSE = .35, \eta_p^2 = .29, p < .001 \). That is, for both speakers, and listeners, \( p(Rp-) < p(Nrp) \), \( t(26) = 2.17, d = .37, p < .05 \); \( t(26) = 4.37, d = .68, p < .01 \), respectively. We also failed to find a significant difference between the absolute and proportional impairments of the speakers and listeners, suggesting that WI-RIF for a speaker and SS-RIF for a listener produces equivalent levels of induced forgetting.

Experiment 3

We modified Experiment 2 so that now we asked the pair to recount jointly the story instead of completing fragments, transforming the paradigm employed in Experiments 1 and 2 into a study involving the effect of free-flowing conversations on forgetting.
Method

The 40 paid participants were either students at the New School or from Craig’s List. In order to ensure that participants would selectively recall the story in the conversation, stating some episodes and events while leaving out others, we inserted six additional episodes into Experiment 2’s stories, with each episode containing from four to seven events. In the practice phase, we asked the two participants to recount the story jointly, without any mention of how participants should monitor each other. They undertook the recounting without the experimenter present, informing the experimenter that they had completed the task when both felt they had little more to add. As in Experiment 2, a distracter task and a final cued recall task followed. We tape recorded and transcribed the group recounting.

Results

Two coders identified in each recounting the critical episodes and events from the original story and classified them as $Rp+$ (appeared in the recounting), $Nrp$ (nothing about an episode appeared in the recounting) or $Rp-$ (did not appear in the recounting, but other events from the episode appeared in the recounting). They defined an $Rp-$ event as relevant to a listener only if the listener did not recall any of the events from the episode herself and as relevant to a speaker only if the speaker remembered a related event. If both members of the pair recalled the event (A: She went to the store; B: Oh, yeah, she went to the store), we classified the event as spoken/remembered by both members. There was good reliability between the two coders, $\kappa = .82$. Disagreements were resolved. There were a few conversations in which one or both participants did not produce instances of $Nrp$ or $Rp-$s.
Once again (see Figure 1), we found a practice effect, inasmuch as $p(Rp+)$ was greater than $p(Nrp)$ and $p(Rp-)$, for speaker $[t(33) = 9.37, d = 1.83, p < .001; t(39) = 26.06, d = 4.89, p < .001]$ and listener $[t(33) = 4.05, d = .82, p < .001; t(27) = 8.03, d = 2.88, p < .001]$. Additionally, as in Experiment 2, consistent with a generation effect, $p(Rp+)$ 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, with one within-subject factor (retrieval type) and one between-subject factor (observer function). As before, we did not find an interaction between retrieval type and observer function, but discovered a main effect for retrieval, $F(1, 24) = 20.79, MSE = .11, \eta_p^2 = .37, p < .001$. That is, $p(Rp-) < p(Nrp)$ for both the speaker, $t(34) = 5.89, d = .96, p < .01$ and the listener, $t(24) = 3.62, d = 1.26, p < .01$. We failed to find a significant difference between the absolute and proportional impairments of the speaker and listener.

The more memorable the material the more likely it is to be mentioned in the 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 is no reason to expect that one type of unmentioned item, e.g., $Nrp$, should be more memorable than the other type, $Rp-$. In order to assess this claim, we obtained a measure of memorability. For each experimentally critical event, we calculated over participants the proportion of times that event was mentioned in the conversation. We then determined which events served as $Nrp$ or $Rp$- for a participant and averaged over item-type for each participant, thereby obtaining a memorability measure for each participant’s $Nrp$ and $Rp$- items. There was no significant
difference between the memorability scores for Nrp ($M = .42, SD = .12$) or Rp- ($M = .49, SD = .12$), if anything they trend in the wrong direction.

**General Discussion**

The present findings indicate that forgetting in a listener is more dramatic when a speaker is silent about some, but not all of a complex of related information than when a speaker fails to discuss the related information in its entirety. Such SS-RIF holds whether one is dealing with paired associates or stories. It may not hold for all stories, however. Anderson and McCulloch (1999) found that WI-RIF was not present when the targets of category-target pairs were integrated. An analogous situation for stories would have events within an episode closely related to each other in ways other than their relation to the episode itself. For such stories, we 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 concurrently retrieves with a speaker. As a consequence, even though SS-RIF should share many of the features of WI-RIF, they 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 occurs when listeners monitored for accuracy, but not when they monitored for fluidity.

Less predictable were our findings for conversations. Not all conversational participants will monitor for accuracy in all conversations. It 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 participants. By asking participants to recount a previously studied story, we may have indirectly
suggested that they monitor for 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 participants do not share the same conversational goal, for instance, when one participant tries to implant memories, while the other tries to resist external influences (Echterhoff, Hirst, & Hussy, 2005). SS-RIF is still possible here, with the “resisting” participant, but not the “implanting” participant 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 mentioned than when a topic is avoided altogether. If speakers want to induce listeners to forget about the claims of weapons of mass destruction, then they should discuss the build-up to the war, listing all the justifications except the presence of WMDs.

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 advice to the influence of what speakers actually say on listener’s memories, emphasizing the speaker’s ability to implant new memories or supplant old memories. The present research adds to these concerns warnings about selective remembering, the silences it produces, and the resultant induced forgetting for listeners as well as speakers.
References


Authors Note

This research was supported by National Institutes of Health grant MH 0066972.
Figure Caption

Figure 1. For Experiments 1 – 3, the proportion of items recalled in the final memory test for practiced items (Rp+), for unpracticed items related to practiced items (Rp-) and for unpracticed items unrelated to practiced item (Nrp). Memory of speakers and listeners is contrasted. Experiment 1 employed paired associates as stimulus material and contrasted two monitoring conditions: accuracy monitoring and superficial (fluidity) monitoring. Experiment 2 employed stories. Experiment 3 embedded retrieval within a free-flowing conversation.
<table>
<thead>
<tr>
<th>Speaker</th>
<th>Listener</th>
<th>Speaker</th>
<th>Listener</th>
<th>Speaker</th>
<th>Listener</th>
<th>Speaker</th>
<th>Listener</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Flidity</td>
<td>Stories</td>
<td>Conversation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Rp+ 
- Nrp 
- Rp-
Silence is not golden