In: Applied Memory
Editor: Matthew R. Kelley
ISBN 978-1-60692-143-2

Chapter 1

IRONIC EFFECTS OF CENSORSHIP: GENERATING CENSORED LYRICS ENHANCES MEMORY

Matthew R. Kelley[1,1], Brittany A. Goldman[1], James E. Briggs[2], and James W. Chambers[3]

[1] Lake Forest College
[2] University of Wisconsin at Milwaukee
[3] Wheaton College, IL, USA

ABSTRACT

Three experiments explored the generation effect—mnemonic advantage for self-generated information—in the applied setting of lyrical censorship. Participants listened to an original song in which a subset of nouns were either partially or completely censored and then completed a recognition memory test consisting of heard, censored, and distractor items. Overall recognition accuracy did not differ for censored and heard items, despite the fact that the censored items were never presented. More importantly, when the data were made conditional upon successful generation of the censored item during encoding, the standard generation effect was observed—ironically, recognition accuracy was significantly higher for the generated censored items compared to the heard items. Source memory did not differ from chance and participants showed a strong bias towards indicating than any 'recognized' word was 'heard.' Overall, these results suggest that by omitting certain words from songs, censors might actually make those words more memorable.

INTRODUCTION

In the human memory literature, the *generation effect* refers to the mnemonic advantage for to-be-remembered (TBR) information that is self-generated (e.g., Mulligan and Lozito, 2004; Slamecka and Graf, 1978). Typically, generation is manipulated by presenting words in

1 Corresponding author.

an incomplete or deficient format (e.g., word stems, word fragments, letter transposition) so that participants are required to actively and overtly generate the TBR material; retention of these generated words is then compared to retention of words that were originally presented intact and required only that participants read the material (e.g., Gardiner, 1988; Nairne and Widner, 1987). The mnemonic benefits of self-generation have been well documented in the laboratory using a wide variety of materials, designs, and procedures, including using pictures (e.g., Kinjo and Snodgrass, 2000; Lohaus and Lachnit, 2001), explicit recall and recognition tests (e.g., Burns, 1990; Hirshman and Bjork, 1988), implicit memory tests (e.g., Mulligan, 2002; Nicolas, Ehrlich, and Facci, 1996), within-list and between-list designs (e.g., Serra and Nairne, 1993; Slamecka and Katsaiti, 1987).

Although the vast majority of generation effect research has employed traditional laboratory materials and procedures, a growing number of researchers have begun to examine the generation effect in applied settings, such as in the classroom and the marketplace (e.g., de Winstanley and Bjork, 2004; Lutz, Briggs, and Cain, 2003; Reardon and Moore, 1996; Thompson and Barnett, 1981). For instance, Foos, Mora, and Tkacz (1994) reported a reliable generation effect for students who generated their own outlines after reading a 2,300 word text, as compared to students who were given an experimenter-generated outline. In a study examining memory for advertisements, Sengupta and Gorn (2002) omitted key elements from certain advertisements and found elevated recall for the ads that required generation. Given the robust nature and broad influence of the generation phenomenon, one may wonder whether generation also is relevant within other contexts, such as with lyrical censorship.

Musicians and the music industry commonly practice self-censorship in which they proactively censor words or lyrics that may be deemed profane, obscene, or offensive (Corbett, 1994). Typically, lyrics are censored by completely or partially omitting the profane material; however, other techniques such as a backward masking (presenting the word in reverse) are increasingly common (Corbett, 1994). Regardless of the specific censorship technique employed, one could imagine that listeners would be able to generate some of the censored words on their own using the context of the song as a guide. If indeed this is the case, then one might expect a generation effect for the censored lyrics in which the words that were self-generated (censored words) would be remembered better than the other non-generated (heard) words from the song. That is, in contrast to its original intent, censorship might actually encourage substantial cognitive processing and promote retention of censored material.

To date, no existing research has attempted to examine the generation phenomenon within the context of lyrical or musical censorship. However, evidence from two disparate lines of research indicates that memory enhancement following lyrical censorship is plausible. First, McDaniel (1984) explored the generation effect using written story stimuli. Participants read stories in which some idea units were degraded because certain letters from important words were deleted, whereas other idea units remained intact. Across three experiments, he showed that participants recalled significantly more idea units when they required generation. To the extent that song lyrics are similar to stories, people should be able to use contextual information to generate missing lyrics and their retention of this information may be enhanced.

Moreover, extant evidence in the form of the phonemic restoration effect suggests that humans are able to naturally generate missing auditory information (Samuel, 1981; 1996). For instance, in a study by Warren and Warren (1980), participants heard several sentences in which a critical phoneme within a word was replaced by the sound of a cough. Despite the missing phoneme, participants' comprehension of the sentences was unaffected. Furthermore, participants rarely realized that a sound was missing; they reported having heard the entire word. Similarly, Cooper et al. (1987) examined whether comprehension of a text was affected by missing phonemic information. In their study, participants listened to several paragraphs in which certain key content words were omitted and replaced by silence. Although these key words were critical to the accurate comprehension of the text, Cooper et al. (1987) found that comprehension of the paragraphs was not impaired by the missing words. Based on the convergence of McDaniel's (1984) demonstration of a generation effect in story recall and the phonemic restoration effect (e.g., Cooper et al., 1987), a positive generation effect for censored lyrics seems reasonable, if not likely.

The current experiments were designed to determine whether people are able to generate and subsequently remember censored lyrics within the context of an aurally presented song. In each experiment, participants listened to an original song that contained a variety of censored nouns (either partially censored or completely censored). Whereas most censored lyrics represent a small pool of emotional or controversial words (e.g., profanities, drug- or violence-related terms), the current study used a variety of common everyday nouns as the target stimuli. We reasoned that the use of common nouns would provide a stronger test of the generation effect (i.e., set of censored items is not known a priori; larger pool of potential targets) and would allow easier comparisons between the current study and traditional investigations of the generation effect using similar stimuli. In addition, we were able to minimize any potential ethical conflicts associated with the presentation of profane material.

Following the song, participants completed a three-alternative, forced-choice recognition test in which participants were asked to identify whether (a) they remembered "hearing" the word in the song, (b) they remembered that the word was "censored" from the song, or (c) the word did not appear in the song. In other words, participants made a source judgment along with their recognition judgment. According to the leading explanations of the generation effect—the multifactor account (e.g., Hirshman and Bjork, 1988; Hunt and McDaniel, 1993; Mulligan, 2001; Nairne, Riegler, and Serra, 1991) and the processing account (e.g., Jacoby, 1983; Mulligan, 2004)—source memory should be impaired for censored items relative to not censored items. The multifactor account suggests that generation encourages extra item-specific processing, which enhances item retention, while disrupting the processing and subsequent retention of other types of information (e.g., order, intertarget relational, context, source). Although the processing account offers the same prediction, the explanation is markedly different as it does not appeal to a processing trade-off. Instead, generated and non-generated information are assumed to encourage different types of processing during encoding; the former promotes greater item processing while the latter promotes greater perceptual processing. This enhanced perceptual processing should benefit source recollection solely for non-generated (heard) items.

EXPERIMENT 1

Experiment 1 explored the influence of lyrical censorship on memory using intentional (expected) and incidental (unexpected) memory tests. In the incidental condition, participants were told that they would hear a song and would be asked to rate the song on a variety of characteristics, such as pleasantness of lyrics, vocals, and instrumentals. We reasoned that the incidental condition would be roughly analogous to how people normally listen to a song outside the laboratory (i.e., determining whether they like a song's music and lyrics), and hence, would have high ecological validity and realism. In the intentional condition, on the other hand, participants were informed of both the pleasantness ratings and the subsequent recognition test—a manipulation more congruous with traditional explorations of the generation effect (e.g., Slamecka and Graf, 1978). Of central interest were the recognition and source judgments for the censored and heard words, as well as the influence of test expectation.

Method

Participants. Eighty students from an introductory psychology course at Lake Forest College earned extra-credit for their participation in this task. Groups of 4 or fewer participants were tested in sessions lasting approximately 15 minutes. Stimuli were presented and controlled with an IBM-compatible computer.

Materials and Design. The authors created an original song about a budding relationship between a man and a woman. The song consisted of 306 words, of which 30 non-offensive, high-frequency nouns were drawn from the Kucera and Francis (1967) norms (average frequency 241.72; range 107-787; average word length 5.38; range 4-7). In a given version of the song, 15 of these nouns were censored (e.g., market, voice, family) and 15 were presented intact (e.g., island, hand, friend); the words were counterbalanced such that each word served as a censored and non-censored words equally often across participants. Along with the 30 nouns used in the song, 30 other matched, high frequency nouns were selected to serve as distractors in the recognition memory test (average frequency 178.72; range 101-470; average word length 5.81; range 4-7). The song was recorded using a male vocalist who was accompanied by an acoustic guitar and a keyboard.

We factorially manipulated the type of censorship (partial vs. complete) and whether the participants expected a memory test or not (intentional vs. incidental) using a between-subjects design. In the completely censored conditions, the entire word was eliminated from the song (silence), whereas in the partially censored conditions, the word's initial phoneme remained audible (e.g., the 't' sound in 'tree'). In the intentional conditions, participants were instructed that, once the song was complete, they would 1) complete a short questionnaire about how much they enjoyed the song, and 2) complete a recognition test for the words in the song. In the incidental condition, participants were instructed only about the short questionnaire, but did not expect a memory test.

Procedure. Participants began the experiment by reading a brief set of instructions (intentional or incidental). Following the instructions, participants listened to the song (approximately 6 minutes) and then recorded their opinions about certain aspects of the song. They were asked to rate, on a 5-point Likert scale, how interesting, pleasant, and realistic the song was, as well as whether the presence of censored items affected their enjoyment of the song. Next, participants completed the three-alternative, forced-choice recognition test, which consisted of viewing 60 words, one at a time, and deciding whether 1) they heard the word in the song, 2) they knew that the word was censored from the song, or 3) the word was neither heard nor censored. Of these 60 words, 30 were new distracting items, 15 were censored during the original song presentation, and 15 were heard normally. Following the test, participants were debriefed as to the exact nature of the experiment.

Results and Discussion

Overall Recognition and Discrimination. Overall recognition performance was analyzed with a 2 (intent: intentional vs. incidental) x 2 (generation type: partial vs. complete) x 3 (stimulus type: not censored, censored, neither) mixed-factor analysis of variance (ANOVA). Figure 1 displays the average performance on the recognition test in each of these experimental conditions. The ANOVA revealed a significant main effect of stimulus type, $F(2, 152) = 222.97, p < .01$. Consistent with past research showing equivalent performance for intentional and incidental learning tasks (e.g., Hyde and Jenkins, 1973; Postman, 1964), there was no significant main effect of intent. The main effect of generation type also failed to reach significance, as did all of the interactions (all $p > .05$).

A Newman-Keuls post-hoc analysis revealed a common pattern of results across every intent and generation condition—recognition accuracy was statistically equivalent for censored and non-censored words and performance in both conditions far exceeded the false alarm rate for distractor items. This pattern held true for the intentional-partial condition [(.75=.72)>.38], the intentional-complete condition, [(.73=.72)>.43], the incidental-partial condition [(.69=.69)>.40], and the incidental-complete condition [(.68=.66)>.40]. Thus, it appears that the experiment did not produce a traditional generation effect, yet participants were able to accurately recognize censored words at the same rate as words they actually heard. Importantly, overall recognition accuracy for these censored items (.70) was significantly higher than the false alarm rate for distractors (.40).

In addition to examining recognition accuracy, participants' discriminability (d') was measured using the signal detection method. The d' measure considers the relationship between recognition accuracy (hit rate) and false alarm rate, such that higher d' values reflect better discrimination performance. Figure 2 displays the d' statistics as a function of intent, generation type, and stimulus type (not censored vs. censored). An ANOVA revealed no significant main effects or interactions, which suggests that participants were able to discriminate distractor items from censored and not censored items equally well in each condition. This finding is not surprising given the nearly identical recognition accuracy for censored and not censored words.

Source Discrimination. A separate 2 (intent) x 2 (generation type) x 3 (stimulus type) x 2 (source judgment: censored vs. not censored) ANOVA examined the pattern of source judgments across the various conditions (see Figure 1). The ANOVA revealed a reliable main effect for stimulus type, once again showing that performance was higher for not censored and censored words than for false alarms, $F(2, 152) = 222.971, p < .01$. The main effect of source judgment was also statistically significant, $F(1, 76) = 360.432, p < .01$, which shows that participants were much more likely to report that the words they 'recognized' were words that were 'not censored' in the song. The analysis also yielded a reliable interaction between stimulus type and source judgment, $F(2, 152) = 37.393, p < .01$. A Newman-Keuls analysis revealed that participants made a greater proportion of 'not censored' judgments to not censored stimuli (.58) than to censored stimuli (.52) and distractor stimuli (.29).

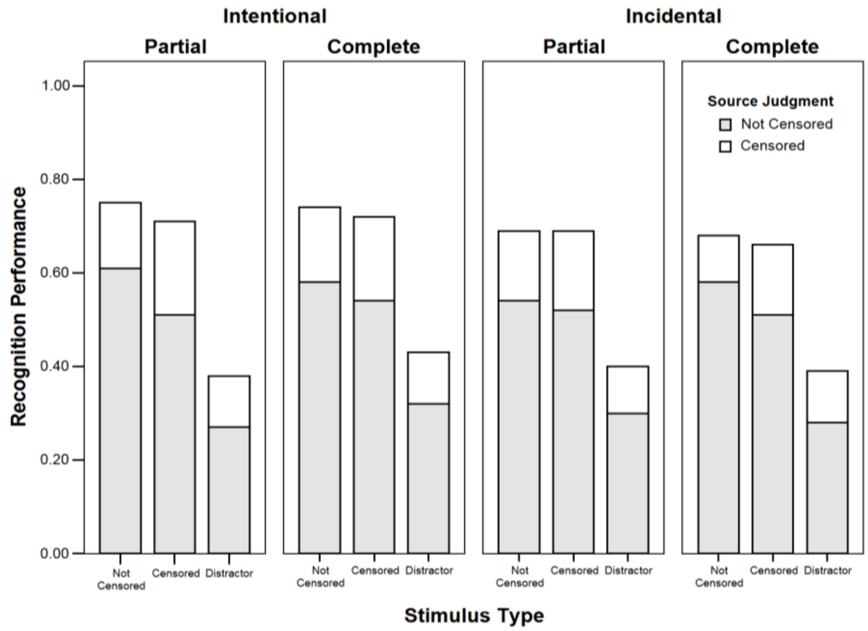

Figure 1. Recognition performance and source judgments as a function of intent condition, censorship type, and stimulus type in Experiment 1.

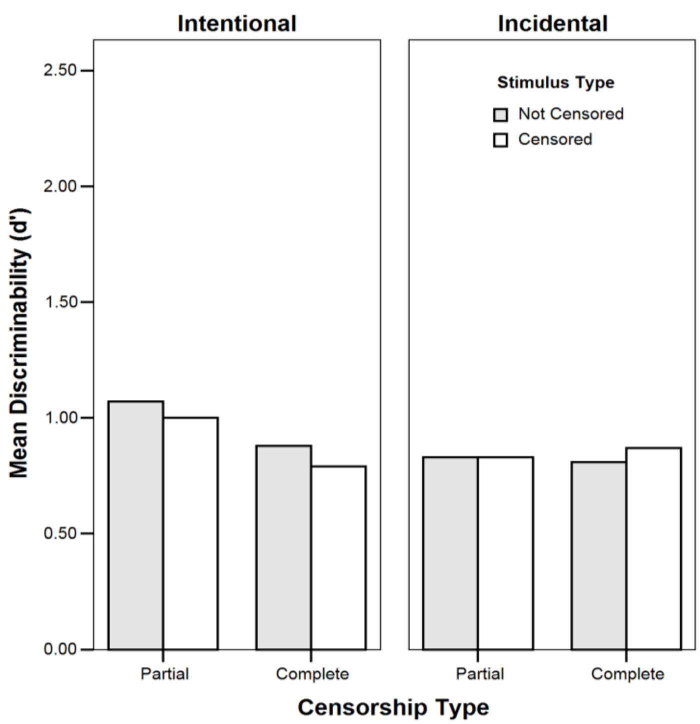

Figure 2. Discriminability (d') as a function of intent condition, censorship type, and stimulus type in Experiment 1.

Alternatively, the proportion of ‘censored’ judgments did not differ for censored and not censored stimuli (.17=.14), but both were significantly higher than the proportion of ‘censored’ judgments made for distractor stimuli (.11). All remaining main effects and interactions failed to reach significance (all $p>.05$).

Consistent with the predictions of the multifactor and processing accounts, participants appeared to ascribe source information more accurately for the not censored (heard) stimuli. However, with participants exhibiting such a strong source bias toward ‘not censored’ responses, the issue is not so clear cut. Indeed, in cases with such robust bias, one encounters the problems of determining whether overall source memory exceeded chance levels and then disentangling source memory from source bias. The former issue can be addressed by computing the ACSIM measure (Murnane and Bayen, 1996) which is determined by calculating the sum of the censored items called ‘censored’ and the not censored items called ‘not censored’ and dividing this score by all censored and not censored items that were ‘recgonized.’ The ACSIM measures were .51, .54, .56, and .52 for the intentional-partial, intentional-complete, incidental-partial, and incidental-complete conditions, respectively. Given that a series of one-sample t-tests, using Bonferroni correction for a family of tests, revealed that none of these measures was differed significantly from chance, it appears that participants were unable to effectively use source information in this task. Moreover, the aforementioned differences in the proportions of source responses are easily explained by assuming that participants (a) were unable to distinguish among the two sources, (b) assumed all recognized items were ‘not censored,’ and (c) occasionally made ‘censored’ source judgments because it was an available option.

Limitations. In traditional laboratory studies of the generation effect, experimenters ensure that each to-be-generated item, in fact, is generated by the participant (e.g., by using talk aloud protocols or providing the participants with a generation rule). In an attempt to maintain high ecological validity in the present study, participants were *not* asked to explicitly generate the missing information during the encoding/listening phase, leaving them free to spontaneously and privately self-generate the censored words. An unfortunate consequence of this design, however, is that we cannot know (a) which words, if any, were generated during encoding; (b) whether these generated words were remembered or forgotten at test; and (c) what percentage of the recognized words were simply guessed at test.

If participants successfully generated a portion of the censored words during encoding, then the mnemonic benefit of generation effect may have been underestimated in Experiment 1. For instance, a participant who remembered 11 of the 13 words that she correctly generated during encoding would enjoy a conditional accuracy of 85%, as compared to an absolute accuracy of 73% (11 out of 15 total stimuli). Thus, by making the data conditional upon generation during encoding, one might find a generation advantage.

EXPERIMENT 2

Experiment 2 was designed to address the major limitation of Experiment 1 by having participants indicate when they generated a censored word during the encoding phase. Participants began the task by listening to the same original song that contained a selection of censored nouns (partial or complete omission) used in first experiment. Then, they listened to each line of the song individually and were asked to indicate with a yes/no button press (a) whether the line contained a censored item, and (b) whether they were able to determine the censored item's identity. Thus, participants did not actually report the identity of the censored item, but instead stated whether they believed that they successfully generated the missing word. Although we cannot be certain that a given word was generated correctly, we reasoned that this processes maintained a better sense of realism in the generation process (e.g., spontaneously and privately generating censored lyrics while listening to music) than other potential methods (e.g., having participants shadow each non-censored word of the song and generate each censored word). All participants were informed of the ensuing forced-choice recognition test prior to the song's presentation. The conditionalized data were of primary importance in this study, as they allowed estimates of how often words were generated during encoding, how often generated words were remembered at test, and the extent to which guessing contributed to overall performance.

Method

Participants. Sixty-eight students from an introductory psychology course at Lake Forest College earned extra-credit for their participation in this task. Groups of 4 or fewer participants were tested in sessions lasting approximately 15 minutes. Stimuli were presented and controlled with an IBM-compatible computer.

Materials and Design. The materials and design were identical to those employed in Experiment 1, with the one exception—the incidental condition was eliminated.

Procedure. The procedure was identical to that used in Experiment 1, with the following additions. Upon hearing the entire song once, participants listened to the song again, one line at a time. After each line, participants were asked, "Did the phrase contain a censored (or omitted) word?" and pressed either a yes or no button. Following a yes response, they were asked, "Were you able to determine the identity of the censored item? That is, could you still tell what the word was supposed to be, even though it was censored?" At the conclusion of the line-by-line presentation, participants completed the opinion survey and the forced-choice recognition test, as detailed in Experiment 1.

Results and Discussion

Overall Recognition and Discrimination. Overall recognition performance was analyzed with a 2 (generation type: partial vs. complete) x 3 (stimulus type: not censored, censored, neither) mixed-factor ANOVA (see left half of Figure 3). As in the first experiment, only the main effect of stimulus type reached statistical significance, $F(2, 132) = 351.581, p < .01$. A Newman-Keuls post-hoc analysis revealed that recognition accuracy was statistically equivalent for censored and non-censored words and performance in both conditions far exceeded the false alarm rate for distractor items, which held true for the partial condition [(.65=.67)>.23] and the complete condition [(.61=.58)>.23].

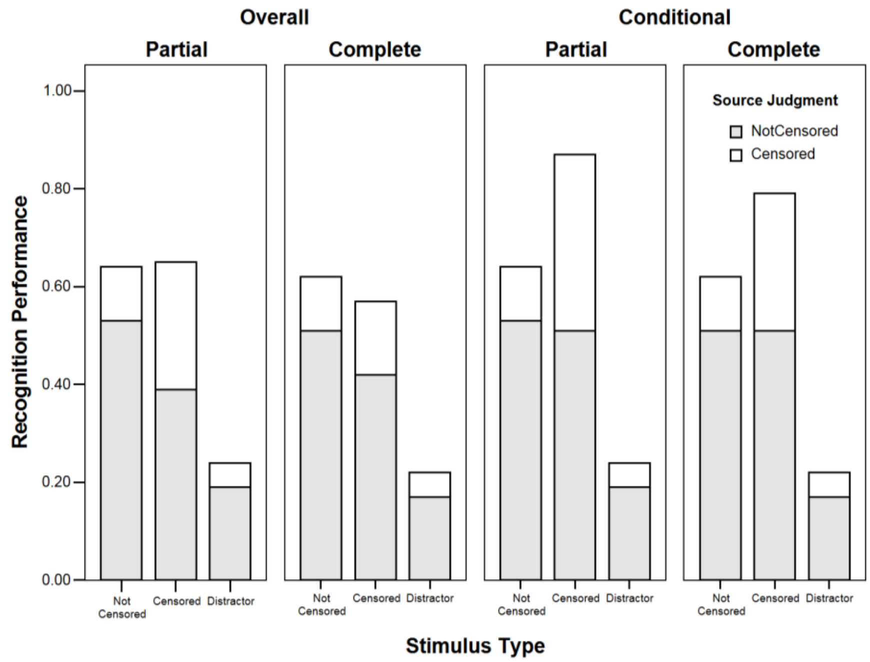

Figure 3. Overall and conditional recognition performance and source judgments as a function of censorship type and stimulus type in Experiment 2.

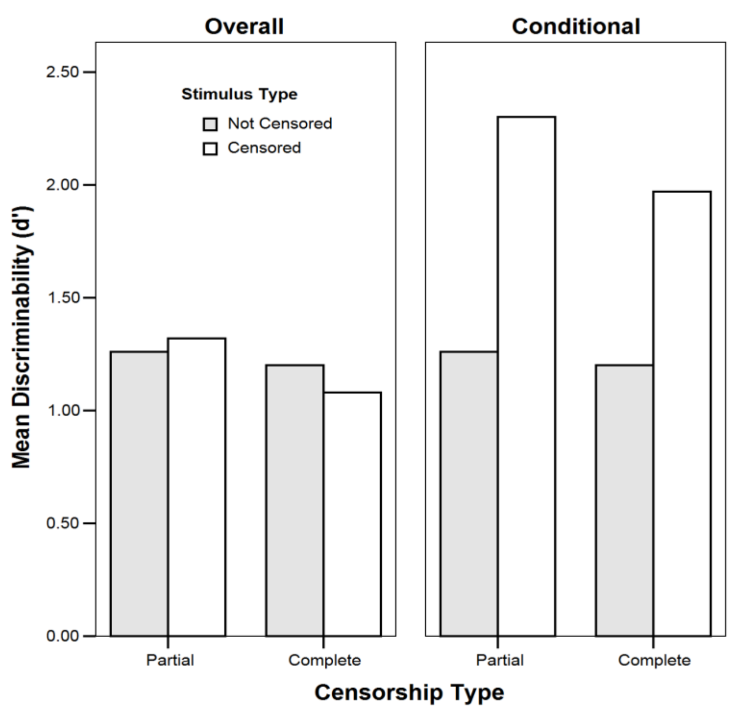

Figure 4. Overall and conditional discriminability (d') as a function of censorship type and stimulus type in Experiment 2.

Figure 4 (left half) shows the overall d' statistics as a function of generation type and stimulus type (not censored vs. censored). As in Experiment 1, the present ANOVA revealed no significant main effects or interactions, which suggests that participants were able to discriminate distractor items from censored and not censored items equally well in each condition. The current patterns of results are consistent with those reported in Experiment 1, albeit with reduced hit and false alarm rates and enhanced discrimination ability in Experiment 2. Overall, participants were able to accurately recognize censored words at the same rate as words they actually heard, but did not produce a traditional generation effect.

Overall Source Discrimination. The overall pattern of source judgments was examined using a 2 (generation type) x 3 (stimulus type) x 2 (source judgment: censored vs. not censored) ANOVA (see left half of Figure 3). The ANOVA revealed reliable main effects of stimulus type, $F(2, 132) = 351.581, p < .01$, and source judgment, $F(1, 66) = 197.178, p < .01$. As in Experiment 1, participants were much more likely to designate a 'recognized' word as having a 'not censored' source. The ANOVA also indicated a reliable interaction between stimulus type and source judgment, $F(2, 132) = 48.536, p < .01$. A Newman-Keuls analysis revealed that participants made significantly more 'not censored' judgments to not censored stimuli (.52) than to censored stimuli (.41) and distractor stimuli (.18). Alternately, participants made significantly more 'censored' judgments to censored stimuli (.21) than to not censored stimuli (.11) and distractor stimuli (.05). This latter pattern differed from Experiment 1, which showed similar proportions of 'censored' judgments for censored and not censored stimuli. All remaining main effects and interactions failed to reach significance (all $p>.05$). Consistent with Experiment 1, the ACSIM measures were not statistically different than chance (.61 and .55 for partial and complete, respectively), which suggests that participants were unable to accurately discriminate among the two sources.

Conditional Recognition and Discrimination. In an attempt to more closely examine the direct influence generation on retention, data inclusion was made conditional upon participant reports of successful item generation during encoding. During encoding, participants in the partial condition indicated that they were able to determine the identities of 60% of the censored items, as compared to 37% in the complete condition.

To calculate the conditional recognition performance for censored words, we determined the proportion of censored words 'generated' during encoding that was subsequently remembered on the recognition test. Conditional data were not necessary for the 'not censored' and 'neither' stimulus types. Recognition performance was analyzed with a 2 (generation type: partial vs. complete) x 3 (stimulus type: not censored, censored, neither) mixed-factor ANOVA. Figure 4 (right half) displays the average performance on the recognition test in each of these experimental conditions.

The ANOVA revealed that only the main effect of stimulus type was significant, $F(2, 132) = 461.744, p < .01$. A Newman-Keuls post hoc analysis revealed a traditional generation effect—recognition accuracy for generated censored items *exceeded* that of not censored items by .21, and both exceeded the false alarm rate. Similar patterns of performance emerged for both the partial [.88>.65>.23] and complete [.80>.61>.23] censorship conditions. Moreover, a 2 (generation type: partial vs. complete) x 2 (stimulus type: not censored vs. censored) ANOVA comparing d' performance indicated a significant main effect of stimulus type, $F(1, 66) = 81.798, p < .01$, which reflects superior discrimination of censored items as compared to not censored items (see right half of Figure 4). These results demonstrate an ironic effect of censorship, namely, that omitting words from a song can make those words more memorable and more discriminable.

Conditional Source Discrimination. The pattern of source judgments for the conditional data was examined using a 2 (generation type) x 3 (stimulus type) x 2 (source judgment: censored vs. not censored) ANOVA (see right half of Figure 4). The ANOVA showed reliable main effects of stimulus type, $F(2, 132) = 461.744$, $p < .01$, and source judgment, $F(1, 66) = 172.992$, $p < .01$. As in the overall analysis, participants were much more likely to report that the words they 'recognized' were words that were 'not censored' in the song. The ANOVA also revealed a reliable interaction between stimulus type and source judgment, $F(2, 132) = 21.516$, $p < .01$. A Newman-Keuls post-hoc analysis revealed no difference between the proportions of 'not censored' judgments made to not censored stimuli (.52) and to censored stimuli (.51), while both exceeded the proportion made to distractor stimuli (.18). Participants made significantly more 'censored' judgments to censored stimuli (.32) than to not censored stimuli (.11) and distractor stimuli (.05). All remaining main effects and interactions failed to reach significance (all $p>.05$). The ACSIM measures of .54 and .47 for the partial and complete conditions did not differ significantly from chance performance, which again reflects a lack of source discrimination.

Limitations. As mentioned previously, in an attempt to maintain a sense of realism in the generation process, participants in Experiment 2 were not asked to report the identities of the censored items. Instead, they simply stated whether they believed that they successfully generated the censored item. By not identifying the words, we cannot be sure that participants generated the appropriate items during song presentation, which may influence estimates of both the proportion of items generated during encoding and subsequent recognition performance. While generation rate may have been overestimated in Experiment 2, conditional recognition performance for censored items may have been underestimated. That is, if a participant erroneously generated 'food' instead of 'friend' during encoding, she would indicate successful generation but would fail to recognize 'friend' on the test and conditional performance would wrongly decline. Of course, two wrongs could also make a right (i.e., erroneous generation of 'food' during encoding, false alarm to 'friend' at test; appears to be successful generation and recognition). Clearly, relying solely on participant self reports of internal generation is problematic.

EXPERIMENT 3

Experiment 3 was designed to address the major limitation of Experiment 2 by having participants shadow the song's lyrics during the presentation phase. That is, participants were asked to repeat every 'heard' word and generate every 'censored' word that they encountered. Accordingly, conditionalized data were computed for both censored and not censored items based on shadowing/generating accuracy. Although shadowing solves the problem of relying on participant self reports of successful generation, the process is more intrusive, less ecologically valid, demands more resources (which may negatively impact generation rate), and the generation is less realistic (e.g., less spontaneous, not private). Still, the shadowing manipulation should provide an unambiguous answer to the question of central interest—namely, are generated censored words remembered better than not censored words?

Method

Participants. Thirty-two students from an introductory psychology course at Lake Forest College earned extra-credit for their participation in this task. Participants were tested individually in sessions lasting approximately 15 minutes. Stimuli were presented and controlled with an IBM-compatible computer.

Materials and Design. The materials and design were identical to those employed in Experiment 2, with the two additions. First, a 30 second clip of the theme song from the television show *Cheers* was used in a shadowing practice session. Also, an experimenter was in the room with the participant and recorded shadowing/generating accuracy on a sheet of paper that displayed the song lyrics.

Procedure. The procedure was similar to that used in Experiment 1, with the following additions. Prior to hearing the experimental song, participants were given shadowing practice. Then, participants were asked to shadow the experimental song while an experimenter recorded shadowing/generating accuracy. At the conclusion of the shadowing task, participants completed the opinion survey and the forced-choice recognition test, as detailed in Experiment 1.

Results and Discussion

Overall Recognition and Discrimination. Overall recognition performance was analyzed with a 2 (generation type: partial vs. complete) x 3 (stimulus type: not censored, censored, neither) mixed-factor ANOVA (see left half of Figure 5). Consistent with first two experiments, only the main effect of stimulus type reached statistical significance, $F(2, 60) = 139.112, p < .01$. A Newman-Keuls post-hoc analysis revealed that recognition accuracy was statistically equivalent for censored and non-censored words and performance in both conditions far exceeded the false alarm rate for distractor items, which held true for the partial condition [(.66=.63)>.29] and the complete condition [(.69=.67)>.33]. Figure 6 (left half) shows the overall d' statistics as a function of generation type and stimulus type (not censored vs. censored). Similar to Experiment 1 and 2, this ANOVA revealed no significant main effects or interactions. Participants were able to discriminate distractor items from censored and not censored items equally well in each condition. Overall, participants were able to accurately recognize censored words at the same rate as words they actually heard, but did not produce a traditional generation effect.

Overall Source Discrimination. The overall pattern of source judgments was examined using a 2 (generation type) x 3 (stimulus type) x 2 (source judgment: censored vs. not censored) ANOVA (see left half of Figure 5). The ANOVA revealed reliable main effects of stimulus type, $F(2, 60) = 139.112, p < .01$, and source judgment, $F(1, 30) = 133.820, p < .01$. As in the first two experiments, participants were much more likely to designate a 'recognized' word as having a 'not censored' source. The ANOVA also indicated a reliable interaction between stimulus type and source judgment, $F(2, 60) = 36.241, p < .01$. A Newman-Keuls analysis revealed that participants made significantly more 'not censored' judgments to not censored stimuli (.58) than to censored stimuli (.49) and distractor stimuli

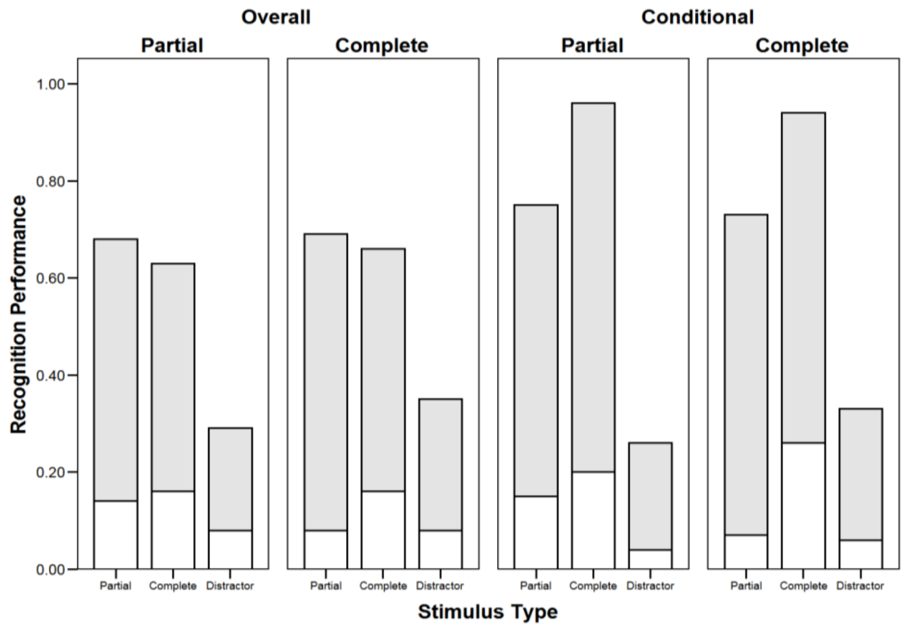

Figure 5. Overall and conditional recognition performance and source judgments as a function of censorship type and stimulus type in Experiment 3.

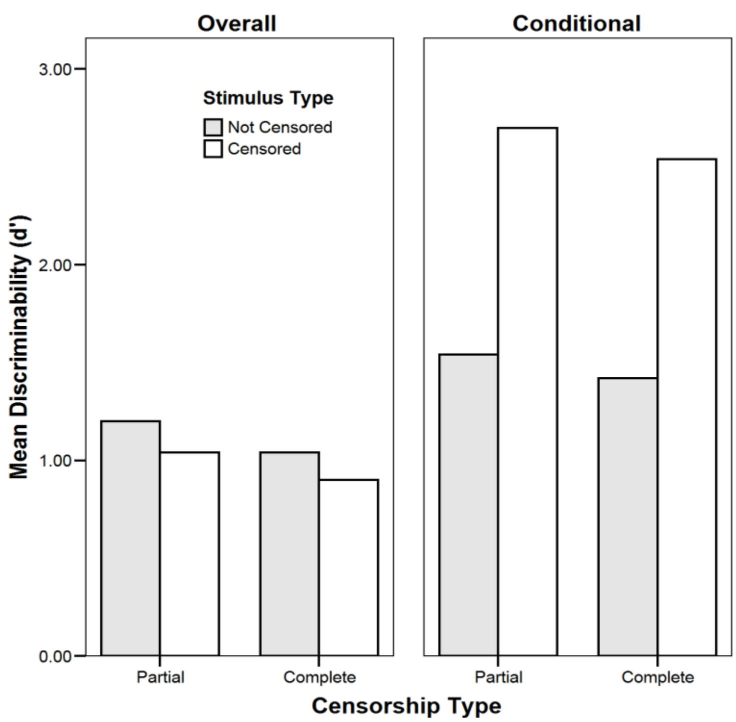

Figure 6. Overall and conditional discriminability (d') as a function of censorship type and stimulus type in Experiment 3

(.24). In contrast, participants made significantly more 'censored' judgments to censored stimuli (.16) than to not censored stimuli (.11) and distractor stimuli (.08). All remaining main effects and interactions failed to reach significance (all $p>.05$). Consistent with Experiments 1 and 2, the ACSIM measures of .53 and .56 for the partial and complete conditions did not differ significantly from chance, which suggests that participants failed to effectively utilize any available source information in this task.

Conditional Recognition and Discrimination. In the present analyses, data inclusion was made conditional upon successful shadowing of non-censored items and successful generation and shadowing of censored items. The shadowing task proved challenging for participants, as evidenced by not censored shadowing rates of 73% and 70% and generation rates of 46% and 31% for the partial and complete conditions, respectively.

To calculate the conditional recognition performance for not censored and censored words, we determined the proportion of not censored and censored words shadowed/generated during encoding that were subsequently remembered on the recognition test. Conditional data were not necessary for the 'neither' stimulus type. Recognition performance was analyzed with a 2 (generation type: partial vs. complete) x 3 (stimulus type: not censored, censored, neither) mixed-factor ANOVA. Figure 6 (right half) displays the average performance on the recognition test in each of these experimental conditions.

The ANOVA revealed that only the main effect of stimulus type was significant, $F(2, 60) = 195.770, p < .01$. A Newman-Keuls post hoc analysis revealed a traditional generation effect—recognition accuracy for generated censored items *exceeded* that of not censored items by .21, and both exceeded the false alarm rate. Similar patterns of performance were evident for both the partial [.96>.75>.29] and complete [.94>.73>.34] censorship conditions. A 2 (generation type: partial vs. complete) x 2 (stimulus type: not censored vs. censored) ANOVA comparing d' performance indicated a significant main effect of stimulus type, $F(1, 30) = 46.651, p < .01$, which reflects superior discrimination of censored items as compared to not censored items (see right half of Figure 6). These results confirm the ironic effect of censorship noted in Experiment 2—that is, censored words that were generated during encoding were both more memorable and more discriminable than comparable not censored (heard) words.

Conditional Source Discrimination. The pattern of source judgments for the conditional data was examined using a 2 (generation type) x 3 (stimulus type) x 2 (source judgment: censored vs. not censored) ANOVA (see right half of Figure 5). The ANOVA showed reliable main effects of stimulus type, $F(2, 60) = 50.817, p < .01$, and source judgment, $F(1, 30) = 166.890, p < .01$. As in the overall analysis and previous experiments, participants were much more likely to report that the words they 'recognized' were words that were 'not censored' in the song. The ANOVA also revealed a reliable interaction between stimulus type and source judgment, $F(2, 60) = 12.907, p < .01$. A Newman-Keuls post-hoc analysis revealed no difference between the proportions of 'not censored' judgments made to not censored stimuli (.72) and to censored stimuli (.64), while both exceeded the proportion made to distractor stimuli (.25). Participants made significantly more 'censored' judgments to censored stimuli (.23) than to not censored stimuli (.10) and distractor stimuli (.05). All remaining main effects and interactions failed to reach significance (all $p>.05$). As with all the previous studies, the ACSIM measures of .47 and .55 for the partial and complete conditions were not statistically different from chance.

CONCLUSION

College-aged students were able to successfully generate censored words from an aurally presented song, with a higher rate of generation for partially censored as compared to completely censored items. Although overall recognition accuracy did not differ for censored and heard items, when the data were made conditional upon successful generation of the censored item during encoding, an ironic effect of censorship was revealed—recognition accuracy was significantly higher for the generated censored items compared to the heard items. In other words, the present study demonstrated a standard generation effect within the applied context of lyrical censorship. These results are consistent with past research which showed that, despite deleting and obscuring single phonemes (e.g., Warren et al., 1980) or entire words (e.g., Cooper et al, 1987), people were able to effectively generate the missing information as evidenced by no reduction in performance on subsequent comprehension tests.

Although generated items enjoyed a mnemonic benefit in this study, the act of generating the censored lyrics proved to be a challenging task for participants, who produced generation rates of 60%, 46%, 37%, and 31% for the partial and complete conditions in Experiments 2 and 3, respectively. With such low rates of generation, the conditional analyses were necessary to unmask the direct influence of generation on memory for censored lyrics. Of course, this procedure opens the door to item selection artifacts—that is, one could argue that successfully generated items were intrinsically easier than unsuccessfully generated items and, therefore, the mnemonic benefit was a function of ease and not due to the generation process. To address this possibility, a series of separate item analyses were performed on the conditional data from the partial and complete conditions of Experiment 2 and 3. Three values were calculated for each of the 30 nouns: (a) the item's generation rate, (b) the proportion correct when the item was not censored, and (c) the proportion correct when the item was censored.

Intrinsic ease was operationalized by performing a quartile split on the generation rate data, where items in highest and lowest quartiles represented the intrinsically easy and difficult ends of the continuum, respectively. Sign tests were used to compare not censored and censored recognition performance across the items. Quartile membership and sign test results were remarkably similar across the partial and complete conditions of each experiment. For the sake of simplicity and to minimize repetition, we combined all of the data into one grand item analysis; the results are displayed Table 1. Of the 30 words, 24 showed a standard generation effect (censored > not censored), four showed the opposite, and there were two ties ($p<.01$). Positive generation effects were clearly the modal response in each quartile, with the four opposite performance patterns spread equally across the upper-middle and lower quartiles, and one tie each in the highest and upper-middle quartiles. Hence, there is no evidence that the present data represent item selection artifacts.

With regard to source memory, when participants recognized a word in the present experiments, they were more likely to report that they "heard" the word, regardless of whether the word was actually heard or censored. Analyses of the raw proportions of correct source judgments offered evidence that the sources of non-censored items were remembered better than censored items. However, ACSIM measures suggested that overall source memory did not differ from chance and that the non-censored advantage may have been driven by the 'heard' bias.

Table 1. Item analysis for conditional data, collapsed across partial and complete conditions from Experiments 2 and 3

		Proportion Correct	
	Item Generation Rate	Not Censored	Censored
Face	.88	.81	.94
Heart	.79	.86	.94
Couple	.73	.93	.93
Town	.73	.64	.73
Voice	.69	.52	.91
Feeling	.65	.83	.95
Hand	.63	.55	.84
Memory	.58	.55	.67
Friend	.54	.78	.82
Woman	.52	.92	1.00
Story	.50	.84	.84
Family	.48	.90	.85
Wife	.44	.85	.91
Island	.42	.60	.46
Company.	.42	.37	.74
Father	.38	.46	1.00
Mother	.38	.81	.85
Years	.37	.80	1.00
Need	.33	.53	.72
Hour	.31	.71	.75
Market	.31	.58	.63
Music	.31	.35	.79
Person	.29	.64	.79
World	.29	.47	.64
Faith	.27	.53	.33
Country	.25	.66	.71
Volume	.25	.24	.47
Week	.23	.45	.49
History	.19	.57	.90
Union	.15	.50	.42

FUTURE DIRECTIONS

If the primary goal of censorship is to discourage, or prohibit, cognitive awareness or processing of profane or objectionable material, then these current findings certainly suggest that lyrical censorship does not achieve always its desired results. Of course, lyrical censorship is not designed with college-aged students in mind and it is not surprising that college students are able to generate omitted information from context alone (e.g., Cooper et al, 1987; McDaniel, 1984). Of greater interest is whether younger children are able to generate the censored information. Indeed, previous research with children has shown that generation ability (for word fragments) increases substantially from age 7 to 11 (McFarland, Duncan, and Bruno, 1983). Given that the standard generation effect generalizes to younger populations, one might expect a similar mnemonic benefit for generated censored items, although with a reduced generation rate. Other variables also might influence the generation and retention of censored items, including: type of censored material (e.g., common words, emotional words, traditionally censored words), type of memory test (e.g., recognition, recall, relearning); type of censorship (e.g., omission, over-dub, backward); level of familiarity with the song; and the availability of context information. Clearly, further research with different populations and procedures is required before we can assess the generality of these ironic effects of censorship.

AUTHOR NOTE

This research was sponsored by a Psi Chi Faculty Advisor Research Grant (Psi Chi National Office; Chattanooga, TN) and the Hotchkiss Fellows Program for Junior Faculty at Lake Forest College. Special thanks to Emily Blegen, Joanna Bovee, Jerrica Cerda, Dustin Koch, Stacey Parrot, & Gretchen Yehl for their assistance with data collection.

REFERENCES

Burns, D. J. (1990). The generation effect: A test between single- and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1060-1067.

Corbett, J. (1994). *Extended play: Sounding off from John Cage to Dr. Funkenstein*. Durham, NC: Duke University Press.

Cooper, W. E., Tye-Murray, N., and Nelson, L. J. (1987). Missing words and the comprehension of spoken text. *Acta Psychologica, 65*, 13-24.

de Winstanley, P. A., and Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory and Cognition, 32*, 945-955.

Foos, P. W., Mora, J. J., and Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology, 86*, 567-576.

Gardiner, J. M. (1988). Generation and priming effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 495-501.

Hirshman, E., and Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484-494.

Hunt, R. R., and McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421-445.

Hyde, T. S., and Jenkins, J. J. (1973). Recall of words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior, 12,* 471-480.

Kinjo, H., and Snodgrass, J. G. (2000). Does the generation effect occur for pictures? *American Journal of Psychology, 113,* 95-121.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 495-508.

Lohaus, D., and Lachnit, H. (2001). Generation effects with pictorial material. *Psychologische Beitrage*, *43,* 787-802.

Lutz, J., Briggs, A., and Cain, K. (2003). An examination of the value of the generation effect for learning new material. *Journal of General Psychology, 130,* 171-188.

McFarland, C. E., Duncan, E., and Bruno, J. M. (1983). Developmental aspects of the generation effect. *Journal of Experimental Child Psychology, 36,* 413-428.

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory and Cognition*, *12*, 46-51.

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 436-450.

Mulligan, N. W. (2002). The effects of generation on conceptual implicit memory. *Journal of Memory and Language, 47,* 327-342.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 838-855.

Mulligan, N. W., and Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.) *The psychology of learning and motivation: Advances in research and theory, Vol 45* (pp. 175-214). San Diego, CA: Elsevier Academic Press

Murnane, K., and Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory and Cognition, 24,* 417-428.

Nairne, J. S., and Widner, R. L. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 164-171.

Nairne, J. S., Riegler, G. L., and Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17,* 702-709.

Nicolas, S., Ehrlich, M., and Facci, G. (1996). Implicit memory and aging: Generation effect in a word-stem completion test. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *15,* 513-533

Postman, L. (1964). Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press.

Reardon, R., and Moore, D. J. (1996). The greater memorability of self-generated versus externally presented product information. *Psychology and Marketing, 13,* 305-320.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474-494.

Samuel, A. G. (1996). Phoneme restoration. *Language and Cognitive Processes, 11*, 647-653.

Sengupta, J., and Gorn, G. J. (2002). Absence makes the mind grow sharper: Effects of element omission on subsequent recall. *Journal of Marketing Research, 39*, 186-201.

Serra, M., and Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory and Cognition, 21,* 34-40.

Slamecka, N. J., and Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592-604.

Slamecka, N.J., and Katsaiti, L.T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26,* 589-607.

Thompson, C. P., and Barnett, C. (1981). Memory for product names: The generation effect. *Bulletin of the Psychonomic Society, 18(5),* 241-243.

Warren, R. M., and Warren, R. P. (1980). Auditory illusions and confusions. *Scientific American*, *223,* 30-36.