CrossMark

A Positive Generation Effect on Memory for Auditory Context

Amy A. Overman[1] · Alison G. Richard[1] · Joseph D. W. Stephens[2]

Abstract Self-generation of information during memory encoding has large positive effects on subsequent memory for items, but mixed effects on memory for contextual information associated with items. A processing account of generation effects on context memory (Mulligan in *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 838–855, 2004; Mulligan, Lozito, & Rosner in *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 836–846, 2006) proposes that these effects depend on whether the generation task causes any shift in processing of the type of context features for which memory is being tested. Mulligan and colleagues have used this account to predict various negative effects of generation on context memory, but the account also predicts positive generation effects under certain circumstances. The present experiment provided a critical test of the processing account by examining how generation affected memory for auditory rather than visual context. Based on the processing account, we predicted that generation of rhyme words should enhance processing of auditory information associated with the words (i.e., voice gender), whereas generation of antonym words should have no effect. These predictions were confirmed, providing support to the processing account.

Keywords Memory · Generation · Context

✉ Amy A. Overman
  aoverman@elon.edu

[1] Department of Psychology, Elon University, Elon, NC 27244, USA

[2] Department of Psychology, North Carolina A&T State University, Greensboro, NC, USA

It has long been known that memory for words improves when the words are self-generated during study, compared to simply reading them (Slamecka & Graf, 1978). However, generation can have varying effects on memory for contextual or source information associated with words. Initial research on generation and context memory suggested that generation's effects were negative (e.g., Jurica & Shimamura, 1999). That is, generating information during study (compared to simply reading the information) was found to reduce memory for contextual details of the study episode. Such findings implied that there was a resource trade-off between item and context memory, such that any encoding activity that enhanced memory for items would necessarily reduce memory for contexts.

Subsequent research has cast doubt on the notion of item-context trade-offs, with some studies finding positive generation effects on context memory (e.g., Geghman & Multhaup, 2004; Marsh, Edelman, & Bower, 2001; Riefer, Chien, & Reimer, 2007), and other studies reporting that negative generation effects were observed only for some types of contexts (e.g., Mulligan, 2004; Mulligan, 2011; Mulligan, Lozito, & Rosner, 2006). As an alternative, Mulligan (2004) proposed a processing account of generation effects on context memory, based on Jacoby's (1983) hypothesis that generation during encoding shifts *what* is encoded rather than *how much* is encoded. The processing account explains generation effects in context memory as an outcome of the degree to which the item generation task also entails processing of context features for which memory is subsequently tested. Mulligan (2004; Mulligan et al., 2006) has reported a number of experimental findings consistent with this account. For example, generation of antonyms during encoding reduces memory for visual features of generated words (e.g., typeface, color) relative to reading, because antonym generation (a semantic task) is less visually oriented than the task of simply reading the words.

Rhyme generation has a similar, negative effect on memory for visual features, also interpreted as resulting from the relatively non-visual focus of processing in rhyme generation relative to reading. On the other hand, a visually oriented generation task such as letter transposition (e.g., reversing the first two letters of _ohrse_ to produce _horse_) yields no effect on context memory relative to reading. Thus, the experiments reported by Mulligan and colleagues found that the processing account accurately predicted negative generation effects on context memory in some conditions but not others.

A limitation of prior tests of the processing account is that they have focused on _negative_ generation effects on context memory, while overlooking the ways in which the processing account also predicts _positive_ generation effects on context.[1] Specifically, there should be a positive generation effect on context memory in cases where the generation task involves more extensive processing of the relevant context features than does reading. For example, generating rhyming words is likely to require greater processing of phonetic information than simply reading pairs of rhyming words. According to the processing account, we should expect this processing difference to result in _better_ memory for contextual features that are auditory and/or phonetic in nature. Thus, it should be possible to reverse the direction of rhyme generation's effect on context memory (relative to reading) by testing memory for auditory rather than visual context.

In contrast, antonym generation should not produce a positive effect on memory for auditory context, because antonym generation is not likely to emphasize auditory/phonetic processing to a greater extent than reading. To the extent that the semantic nature of antonym generation may shift processing away from perceptual features, the processing account should predict either a negative effect or no effect on memory for auditory context.

The present study provides a novel test of the processing account of generation effects on context memory by using a new type of context. The design closely parallels the methods of two prior experiments by Mulligan and colleagues, with one change: instead of testing the effects of antonym and rhyme generation on memory for visual context (e.g., the color of each word), we tested their effects on memory for auditory context (the gender of the voice that spoke each word). If the processing account is correct, then antonym and rhyme generation should have differing effects on memory for auditory context, even though they have similar effects on memory for visual context.

---

[1] Positive effects have been reported (e.g., Geghman & Multhaup, 2004; Marsh et al., 2001; Riefer et al., 2007) but only for source or visual contexts, and the effects have not been interpreted in terms of the processing account. Mulligan (2004) presented the positive generation effect of Marsh et al. (2001) as conflicting with the processing account, and conducted several experiments that failed to replicate it.

Method

Participants Thirty-one undergraduates (mean age = 18.7 years) at Elon University participated in exchange for course credit. This sample size was chosen based on sample sizes used in comparable prior experiments (Mulligan, 2004; Mulligan et al., 2006). The most similar prior experiment for which effect size data were available was Experiment 2A of Mulligan et al. (2006), which found a negative effect of rhyme generation on memory for word color with $d = .51$. Based on those data, the a priori power for the predicted effect of rhyme generation on memory auditory context with a sample size of $n = 31$ was estimated to be approximately .81. All participants provided informed consent and all procedures were approved by the Institutional Review Board of the university.

Materials Stimuli consisted of 100 antonym word pairs and 100 rhyming word pairs. Within each pair, one word was designated as the cue and other as the target. All target words were between four and seven letters long and varied in Kučera–Francis frequency from 1 to 1599 (Kučera & Francis, 1967). Female and male spoken versions of each pair were created using the AT&T Natural Voices Text-To-Speech System (www.wizzardsoftware.com) and saved digitally as .wav files.

Each set of 100 word pairs (i.e., antonym pairs and rhyme pairs) was used to create a 52-item study list and a 96-item test list. Each study list consisted of 4 initial buffer pairs and 48 regular study pairs. Each test list consisted of the target words from each of the 48 study pairs and the target words from each of the 48 remaining unused word pairs in the applicable stimulus set. At runtime for each participant, stimuli were randomly assigned to the study list, and stimuli within each study list were randomly assigned to the generate and read conditions, and to the female and male voice conditions.

Procedure The experiment was carried out as two study-test sequences. The task structure was identical across both study-test sequences, except that one used antonym pairs (and antonym generation during study) and the other used rhyming pairs (and rhyming generation during study). All participants completed both study-test sequences, and the order of the sequences was randomly determined for each participant.

At the beginning of each study task, participants were instructed that they would be studying cue–target pairs, and that they would need to remember the target words later. They were also informed that they would hear each cue–target pair spoken in either a female or male voice, and that they should try to remember which voice gender went with each target word. On each trial, participants viewed the cue–target pairs on the computer screen, and typed the target word followed by Enter. In the read condition, the cue and target words in each pair were both presented in their entirety (e.g., _before–after,_ or

hook–book), and the participant's task was to re-type the target word. In the generate condition, the cue word was presented alongside the first letter of the target, followed by a continuous underscore five spaces long (e.g., before–a_____, or hook–b_____). The participant's task was to type the target word that would belong in the provided space based on the type of cue–target association being used in the present task (i.e., antonyms or rhymes).

After the target word was typed, the entire cue–target pair was presented on the screen for 4.9 s, and the corresponding audio file was played that presented the word pair spoken by a female or male voice. The onset of the audio file was simultaneous with the visual presentation of the full cue–target pair. The audiovisual presentation of the cue–target pair occurred regardless of whether the participant successfully generated the intended target word. Presentation of the cue–target pair was followed by a blank screen for 350 ms. Study trials were randomly ordered within 12 blocks, each of which contained the four possible combinations of encoding condition (generate/read) and voice context (female/male).

Each study task was followed by a distractor task in which participants completed a set of simple arithmetic problems for 3 min before continuing to the recognition test. Within each test task, studied and new words were presented visually, one at a time, on the computer screen. Participants were instructed to judge whether each word had been studied as a target word, and if so, to identify the gender of voice that had spoken it. Responses were made using the 7, 8, and 9 keys on the keyboard to indicate "male," "female," and "new," respectively.

Results

Study phase accuracy Two participants failed to correctly follow study instructions, leading to incorrectly typed responses on all study trials in at least one of the study tasks. These participants were excluded from further analyses. The remaining 29 participants produced the intended target words on the large majority of generate trials, in both the antonym (M

$= .89$, SD $= .10$) and rhyme ($M = .95$, SD $= .04$) tasks. Subsequent analyses of recognition responses were carried out including only those trials for which the corresponding study item had a correct response (i.e. conditionalized on study responses; e.g., Mulligan, 2004). Averaged results for conditionalized analyses were found to be virtually identical to the results that included all test trials. Consequently, for brevity only the unconditionalized analyses are presented below.

Item and context memory For ease of comparison with prior studies, we generally followed the same analysis approach as Mulligan (2004) and Mulligan et al. (2006). Item memory was analyzed using hit and false alarm rates, with "female" and "male" responses scored as "old." Context memory was analyzed using identification-of-origin scores (Johnson, Hashtroudi, & Lindsay, 1993), computed as the proportion of item hits for which the correct voice gender was chosen. Table 1 displays the hit and false alarm rates, and identification-of-origin scores, for the present experiment alongside comparable tasks reported in prior studies.

For item memory, corrected hit rates (i.e., hits minus false alarms) were compared across conditions using a 2 (generation: generate vs. read) × 2 (task: antonym vs. rhyme) repeated-measures ANOVA. There was a significant effect of generation, $F(1,28) = 83.5$, $p < .001$, $MSE = 1.35$, $\eta_p^2 = .75$, indicating that generation led to improved old/new discrimination of target words, compared to reading. Task was also significant, $F(1,28) = 22.7$, $p < .001$, $MSE = .458$, $\eta_p^2 = .45$, such that memory was better for targets in antonym pairs versus rhyming pairs, consistent with the commonly-found memory advantage for semantic encoding tasks (e.g., Craik & Lockhart, 1972). No interaction was found, $F(1,28) = 2.44$, $p = .13$, $MSE = .024$.

Hit and false alarm rates for items were also used to compute sensitivity (d') and response bias (C). Separate measures of d' were computed for generate and read conditions in each task, using each study condition's hit rate along with the overall false alarm rate to new items. Figure 1a displays average d'

Table 1 Item recognition as measured by hit and false alarm rates, and context recognition as measured by identification of origin scores, for the current study in comparison to prior experiments

Experiment	Task	Context	Proportion "old" responses					ID of origin scores			
			Generate	Read	FA	*SE*	Effect	Generate	Read	*SE*	Effect
Mulligan (2004), Exp. 1	Antonym	Color	.85	.64	.18	nr	+	.58	.66	nr	-
Current study	Antonym	Voice	.85	.61	.12	.026	+	.62	.65	.031	
Mulligan et al. (2006), Exp. 2A	Rhyme	Color	.73	.57	.18	.040	+	.60	.70	.036	-
Current study	Rhyme	Voice	.77	.58	.19	.030	+	.65	.59	.034	+

Standard errors are presented in the manner of Mulligan et al. (2006), as the means of the standard errors among the respective measures of item memory and context memory

FA False alarms, *nr* not reported
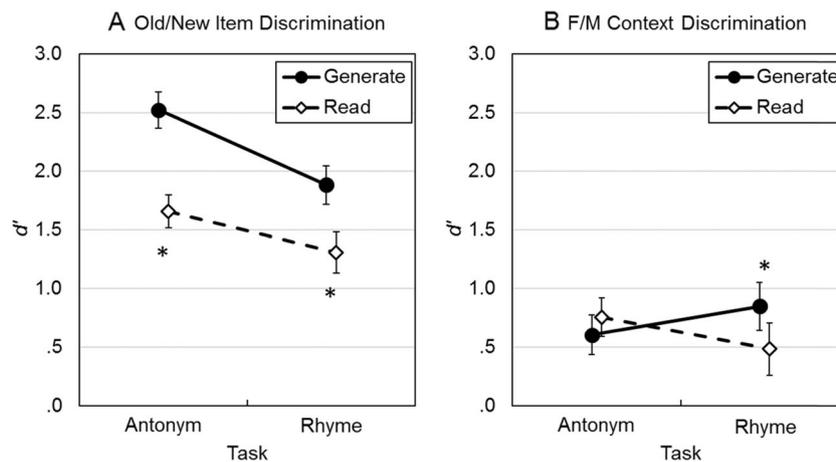
Fig. 1 Item and context memory performance across generate and read conditions, for the antonym and rhyme generation tasks. *Error bars* standard error of the mean. *Asterisks* statistically significant generation effects in pairwise comparisons within tasks

values across conditions. Consistent with the results using corrected hit rates, a 2 (generation: generate vs. read) × 2 (task: antonym vs. rhyme) repeated-measures ANOVA on d' found a significant effects of generation, $F(1,28) = 122.6$, $p < .001$, $MSE = 15.1$, $\eta_p^2 = .81$, and task, $F(1,28) = 20.7$, $p < .001$, $MSE = 7.18$, $\eta_p^2 = .425$. There was also a modest interaction, $F(1,28) = 5.06$, $p = .033$, $MSE = .608$, $\eta_p^2 = .15$, which reflected a larger effect of generation in the antonym task than in the rhyme task. This result suggests that the semantic processing involved in antonym generation provided an additional benefit to item memory (relative to rhyme generation), beyond the general advantage for antonym pairs overall. Response bias C was computed at the task level for each participant, using overall hit and false alarm rates within each task. Mean C was slightly higher (i.e. stricter response criterion) in the antonym task ($M = .33$, $SD = .35$) than in the rhyme task ($M = .25$, $SD = .38$), but this difference was not statistically significant, $t(28) = 1.14$, $p = .27$, $SE = .068$.

For context memory, identification-of-origin scores were analyzed using a 2 (generation: generate vs. read) × 2 (task: antonym vs. rhyme) repeated-measures ANOVA. There was no main effect of generation, $F(1,28) = .43$, $p = .52$, $MSE = .004$, and no main effect of task, $F(1,28) = .28$, $p = .60$, $MSE = .007$. However, the interaction was significant, $F(1,28) = 7.27$, $p = .012$, $MSE = .071$, $\eta_p^2 = .21$. Paired-samples t tests comparing generate versus read conditions within each task indicated that this interaction was driven by a positive effect of generation on context memory in the rhyming task, $t(28) = 2.85$, $p = .008$, $SE = .022$, $d = .53$, with no effect of generation on context memory in the antonym task, $t(28) = -1.28$, $p = .21$, $SE = .030$.

It was also possible to apply a signal-detection approach to participants' discrimination between the two contexts in memory, by designating as a hit any correct

"female" response, and as a false alarm any "female" response to any item originally presented in a male voice.[2] Average d' values for context discrimination are displayed in Fig. 1b. Consistent with the results for identification-of-origin scores, a 2 (generation: generate vs. read) × 2 (task: antonym vs. rhyme) repeated-measures ANOVA on d' found no main effect of generation, $F(1,28) = 1.57$, $p = .22$, $MSE = .326$, or task, $F(1,28) = .006$, $p = .94$, $MSE = .006$, but a significant interaction, $F(1,28) = 9.30$, $p = .005$, $MSE = 1.94$, $\eta_p^2 = .25$. Paired-samples t tests also indicated a positive effect of generation on context memory in the rhyming task, $t(28) = 3.59$, $p = .001$, $SE = .102$, $d = .67$, and no effect of generation in the antonym task, $t(28) = -1.12$, $p = .27$, $SE = .136$. Because this approach provided separate false alarm rates for each condition, it was also possible to analyze response bias using C in a 2 (generation: generate vs. read) × 2 (task: antonym vs. rhyme) repeated-measures ANOVA. Interestingly, there were significant main effects of generation, $F(1,28) = 28.4$, $p < .001$, $MSE = 3.41$, $\eta_p^2 = .50$, and task, $F(1,28) = 4.31$, $p = .047$, $MSE = .347$, $\eta_p^2 = 13$, such that participants had a stricter criterion to respond "female" (i.e. gave fewer "female" responses overall) in read conditions than in generate conditions, and in the rhyme task than in the antonym task. The interaction was not significant, $F(1,28) = .68$, $p = .42$, $MSE = .079$. Given that item recognition was inferior in read conditions and in the rhyme task, we interpret the overall decreased tendency to use the "female" response option in those conditions as possibly reflecting the use of "male" as the default "old" response on trials in

---

[2] Identical values of d' are obtained if correct "male" responses are counted as hits and "male" responses to items studied with female voice as false alarms. Values of C are also identical, but reversed in polarity, such that the positive values of C we obtained would be the corresponding negative values if correct "male" responses were counted as hits.

which participants were guessing about whether an item had been studied.

Discussion

The present experiment provided a novel test of the processing account of generation effects in memory for context (Mulligan, 2004; Mulligan et al., 2006), by using auditory context (voice gender) instead of visual contexts that were used in previous studies. Specifically, we tested the prediction that antonym and rhyme generation would have *differing* effects on memory for auditory context, even though they have previously been found to have *similar* effects on memory for visual context.

The results were entirely consistent with the prediction of the processing account. In particular, the interaction of generation and task in identification-of-origin scores demonstrated that the two types of generation differentially affected memory for voice gender. The fact that rhyme generation yielded a *positive* effect on auditory context memory is also highly consequential because it is exactly what the processing account predicts, given that the read condition was non-auditory in nature. Similarly, the lack of difference in context memory between antonym generation and reading can be interpreted as a consequence of both conditions being non-auditory. In this regard, it is important to note that "generation effects" are not defined relative to an entirely neutral baseline condition, but relative to a specific non-generation task (in this case, reading) that carries its own cognitive processing demands, which may differ from the generation task in more than one aspect. Thus, the pattern of findings reinforces the view that generation effects are a product of the processing differences between generate and non-generate conditions, and thus the processing characteristics of both conditions are relevant to the outcome.

It is also important to note that performance on item memory in the present experiment closely aligned with the results of prior experiments using the same generation tasks with visual context (see Table 1). Only the pattern of results for context memory differed meaningfully from the results of the comparison studies. This supports the interpretation that our current design did not fundamentally alter the underlying memory processes being studied, but rather merely created a different set of task–context relationships.

Interestingly, while rhyme generation benefited both item and context memory, the interaction observed in the d' analysis for item memory indicated that the antonym generation task provided a greater benefit in item memory than the rhyme generation task. This finding also fits well with the processing account of generation effects. In comparison to rhyme generation, antonym generation's

semantic focus should produce a greater increase in encoding of semantic features that are diagnostic in discriminating studied versus new words.

Although the purpose of the current experiment was to test a specific prediction of the processing account, the results also provide further evidence against the theory that generation induces a trade-off between item and context memory (cf. Jurica & Shimamura, 1999). Contrary to the prediction that an increase in item memory must be accompanied by a decrease in context memory, the rhyme generation task yielded positive effects on memory for both items and contexts, in line with other studies that have found positive generation effects for both items and contexts (Geghman & Multhaup, 2004; Marsh et al., 2001). The current findings also cast doubt on other alternate hypotheses proposed in the literature. For example, Riefer et al. (2007) suggested that generation specifically causes negative effects in memory for externally presented contextual information, yet we observed a positive effect with externally presented acoustic information. Similarly, Mulligan (2011) suggested that generation may specifically disrupt memory for contextual information that is intrinsic to items, such as font or color, yet voice gender in the current experiment would seem to be just as intrinsic to the items as font or color in a purely visual task.

A further point of interest in the present results is that the experimental design required generation processing to precede the presentation of the relevant context. That is, the auditory cues for voice gender were provided *after* the participant had already submitted a generated (or read) response. Thus, from the processing point of view, we must consider how rhyme generation could have enhanced the encoding of *subsequent* auditory context. We suggest two possibilities. The first is that the processing involved in rhyme generation entailed a shift in attention toward phonological features, and that this attentional shift provided some residual facilitation of acoustic/phonetic processing for long enough to affect encoding of the voice gender. A second possibility is that, after the participant submitted each generated response, the subsequent presentation of the "correct" response caused some reactivation of the cognitive processing that had occurred in generation, such that this reactivated generation processing facilitated encoding of related context features. Further research should seek more precisely to identify the mechanisms by which generation influences encoding of temporally adjacent contexts.

Finally, although the present results provide strong evidence for a processing account of generation effects in context memory, this account and others are not necessarily mutually exclusive. In particular, resource-based trade-offs may be an additional component of such effects under circumstances of heavy cognitive load (Nieznański, 2011; 2012). Future research should further examine the interplay of these factors for a more complete understanding of generation effects in memory.

References

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Geghman, K. D., & Multhaup, K. S. (2004). How generation affects source memory. *Memory & Cognition, 32*(5), 819–823.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22*(5), 485–508.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28.

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition, 27*(4), 648–656.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition, 29*(6), 798–805.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 838–855.

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *The Quarterly Journal of Experimental Psychology, 64*(8), 1543–1562.

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 836–846.

Nieznański, M. (2011). Generation difficulty and memory for source. *The Quarterly Journal of Experimental Psychology, 64*(8), 1593–1608.

Nieznański, M. (2012). Effects of generation on source memory: A test of the resource tradeoff versus processing hypothesis. *Journal of Cognitive Psychology, 24*(7), 765–780.

Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *The Quarterly Journal of Experimental Psychology, 60*(10), 1389–1405.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592–604.