Positive and negative generation effects in source monitoring

Article *in* Quarterly journal of experimental psychology (2006) · November 2007

DOI: 10.1080/17470210601025646 · Source: PubMed

David M Riefer
California State University, San Berna…

Jason F Reimer
California State University, San Berna…

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
0000, 00 (0), 1–17

Psychology Press
Taylor & Francis Group

Positive and negative generation effects in source monitoring

David M. Riefer, Yuchin Chien, and Jason F. Reimer
California State University, San Bernardino, CA, USA

Research is mixed as to whether self-generation improves memory for the source of information. We propose the hypothesis that positive generation effects (better source memory for self-generated information) occur in reality-monitoring paradigms, while negative generation effects (better source memory for externally presented information) tend to occur in external source-monitoring paradigms. This hypothesis was tested in an experiment in which participants read or generated words, followed by a memory test for the source of each word (read or generated) and the word's colour. Meiser and Bröder's (2002) multinomial model for crossed source dimensions was used to analyse the data, showing that source memory for generation (reality monitoring) was superior for the generated words, while source memory for word colour (external source monitoring) was superior for the read words. The model also revealed the influence of strong response biases in the data, demonstrating the usefulness of formal modelling when examining generation effects in source monitoring.

The *generation effect* is the observation that information that people generate themselves is remembered better than information that they receive from an external source (Slamecka & Graf, 1978). In a typical experiment, participants are shown pairs of words with a known relationship (e.g., antonyms: long–short; synonyms: rapid–fast) and are eventually tested on the second word in each pair. For some of the pairs both words are presented intact (long–short), and participants simply "read" the second word. For other pairs only the first letter of the second word is presented (long–s___), and in this case the participants must "generate" the second word themselves. This simple manipulation produces very robust generation effects, in which memory for the generated items is substantially better than memory for the read items (for a review, see Mulligan & Lozito, 2004).

As can be seen in the above example, the generation effect involves memory for information that comes from two sources—self-generated items versus externally generated items. Because of this, a few research studies have examined the generation effect from a source-monitoring perspective (Johnson, Hashtroudi, & Lindsay, 1993). In a source-monitoring task, people are presented with items from multiple sources (in the case of two sources, Source A and Source B) and are instructed to remember not only the items

Correspondence should be addressed to David M. Riefer, Department of Psychology, California State University at San Bernardino, 5500 University Parkway, San Bernardino, California 92407, USA. E-mail: driefer@csusb.edu

We thank Francisco Debhran for his assistance in running the participants in the experiment, William Batchelder for his helpful comments on an earlier version of this manuscript, Karl Christoph Klauer for showing us how interactions between model parameters can be accomplished using Hu's (1996) computer program, and Ute Bayen and two anonymous reviewers for their helpful comments.

http://www.psypress.com/qjep DOI:10.1080/17470210601025646

themselves, but also the source of each item. Source-monitoring tasks therefore tap into two different types of memory—item memory and source memory. Most studies on the generation effect have focused on item memory—that is, the typical generation effect demonstrates that memory for an item is better when that item is self-generated. But a separate question can be asked about the effects of generation on source memory. Is memory for the source of information better for self-generated than for externally generated information?

Some theorists (e.g., Marsh, Edelman, & Bower, 2001) have suggested that, like item memory, source memory should be better for self-generated information. This is referred to as a positive generation effect for source. The common explanation for the generation effect is that the cognitive operations involved in generating items enhances the encoding of semantic, item-specific, or relational aspects of the stimulus, which results in superior memory for those items over nongenerated items (Hirshman & Bjork, 1988; McDaniel, Wadill, & Einstein, 1988). It seems logical to speculate that this superior encoding enhances the source memory for an item in addition to memory for the item itself. In contrast, other researchers (e.g., Jurica & Shimamura, 1999) have proposed that source memory should be worse for self-generated than for externally generated information: in other words, a negative generation effect for source. One possible explanation is that the extra cognitive operations involved with self-generated items tend to draw mental resources away from the processing of other information, such as the context or source of the items (Jurica & Shimamura, 1999).

A number of experiments have been conducted on this issue, but with mixed results. Some studies have found a positive generation effect for source (e.g., Geghman & Multhaup, 2004; Kinjo & Snodgrass, 2000; Riefer, Hu, & Batchelder, 1994), others have found no effect (Voss, Vesonder, Post, & Ney, 1987), while others have found negative generation effects (e.g., Jurica & Shimamura, 1999; Mulligan, 2004; Mulligan, Lozito, & Rosner, 2006; Rabinowitz, 1990). An example of a positive generation effect for source can be found in a study by Geghman and Multhaup (2004). In their experiment, participants read general-information questions (e.g., *The thick layer of fat on a whale is called ...*) in which the correct answer was either scrambled (ubberbl) or missing vowels (bl*bb*r). For some questions the answer was provided by a face on a computer (external source), and for others the participant had to provide the answer (internal source). Geghman and Multhaup found that source recognition (presented by computer vs. generated by participant) was superior for questions in which people generated the answers themselves, a positive generation effect for source memory that parallels the generation effect for item memory.

In contrast, other studies have observed a negative generation effect for source. For example, Jurica and Shimamura (1999) presented people with sentences in the form of either a statement or a question. Each sentence came from one of three faces on a computer screen. On a subsequent memory test participants were asked to identify which face was the source of the sentence. Jurica and Shimamura observed that source memory for the faces was poorer when participants had to answer a question (self-generated response) than when they merely read a statement (externally generated item)—a negative generation effect for source.

In addition to the general question of how generation influences source memory, other studies have focused more specifically on the role that contextual memory plays in the generation effect. As Johnson, et al. (1993, p. 3) have pointed out, the context of a memory event is one way that the source of information can be determined, and many source-monitoring experiments have manipulated source by varying the background or spatiotemporal context of to-be-remembered information (e.g., List 1 vs. List 2, male voice vs. female voice). A recent series of studies on the effects of generation on context memory has been conducted by Mulligan (2004; see also Mulligan et al., 2006). Participants were presented with word antonyms, in which the second word in the pair was either read (hot−cold) or generated (hot−c___). Half of the word pairs were presented in red print and half in green print. When participants were tested on their recognition memory for the words, the standard generation effect was obtained, with the generated words recognized better than the words that were read. But when the context memory for the colour of the words was tested a negative generation effect was found, with participants displaying poorer colour identification for the generated words than for the read words.

Reality monitoring versus external source monitoring

Taken as a whole, it seems puzzling that research examining the role of context or source memory in the generation effect has obtained such diverse results, including both positive and negative effects. In this article, we propose and test the hypothesis that many of these divergent results can be accounted for by two factors that have often been overlooked in this research area: (a) the type of source memory being tested, and (b) response biases.

Consider first the type of source memory that is tested. Studies that find positive generation effects for source are often those in which the self-generated versus externally generated items are themselves those tested on source memory—that is, Source A = self-generated and Source B = externally generated. This type of source-monitoring task is referred to as *reality monitoring* (Johnson & Raye, 1981) and basically covers memory for internal versus external sources. Geghman and Multhaup (2004), for example, tested memory for answers that people produced themselves (internal) or those answers that were provided for them (external). In contrast, negative generation effects for source are usually observed in research studies that examine a different form of source monitoring called *external source monitoring* (Johnson et al., 1993). In those designs, different external sources are tested, which are then examined separately for the self-generated and externally generated items. For example, in the Jurica and Shimamura (1999) experiment, participants were tested on their source memory for different speakers (i.e., external sources). Studies that have examined the role of context memory in the generation effect also fit within this framework. In Mulligan (2004) and Mulligan et al. (2006), for example, participants were tested on contextual source memory for the colour of stimuli (red or green) or their location (left or right of screen). These types of stimulus attribute can contribute to the conditions under which the source of information is determined.

The contrast between the experimental designs involved with reality monitoring and external source monitoring can be illustrated by comparing the data sets created by each design. In general, a source-monitoring experiment involving two sources (A and B) results in a 3 × 3 table of data (cf. Batchelder & Riefer, 1990, p. 549):

$$\begin{array}{cc} & \text{Response} \\ & \begin{array}{ccc} \text{A} & \text{B} & \text{New} \end{array} \\ \text{Item} \begin{array}{r} \text{A} \\ \text{B} \\ \text{New} \end{array} & \begin{bmatrix} X_{AA} & X_{AB} & X_{AN} \\ X_{BA} & X_{BB} & X_{BN} \\ X_{NA} & X_{NB} & X_{NN} \end{bmatrix} \end{array}$$

In those studies that examine generation effects using the reality-monitoring paradigm, generation itself is the tested source dimension (i.e., read vs. generate). In this case a single 3 × 3 table is constructed with, say, Source A = read and Source B = generated:

$$\begin{array}{cc} & \text{Response} \\ & \begin{array}{ccc} \text{read} & \text{generate} & \text{new} \end{array} \\ \text{Item} \begin{array}{r} \text{read} \\ \text{generate} \\ \text{new} \end{array} & \begin{bmatrix} X_{rr} & X_{rg} & X_{rn} \\ X_{gr} & X_{gg} & X_{gn} \\ X_{nr} & X_{ng} & X_{nn} \end{bmatrix} \end{array}$$

In this design, source memory comparisons between read and generated items are carried out within the single 3 × 3 table.

In studies that examine the generation effect using external source monitoring, a second external source dimension besides generation is created and tested. This external source dimension could involve the origin of the stimuli (e.g., room location, with Source A = Room 1 and Source B = Room 2 as in Marsh, et al., 2001) or a contextual detail of the stimuli (e.g., stimulus colour, with Source A = red and Source B = green as in Mulligan, 2004). Regardless of whether one is testing source memory or context memory, this external source-monitoring design results in two separate 3 × 3 tables, one for the read items and one for the generated items:

$$\text{Read}\quad \begin{array}{c c} & \begin{array}{ccc} A & B & \text{New} \end{array} \\ \begin{array}{c} A \\ B \\ \text{New} \end{array} & \begin{bmatrix} X_{AA} & X_{AB} & X_{AN} \\ X_{BA} & X_{BB} & X_{BN} \\ X_{NA} & X_{NB} & X_{NN} \end{bmatrix} \end{array}$$

$$\text{Generate}\quad \begin{array}{c c} & \begin{array}{ccc} A & B & \text{New} \end{array} \\ \begin{array}{c} A \\ B \\ \text{New} \end{array} & \begin{bmatrix} X_{AA} & X_{AB} & X_{AN} \\ X_{BA} & X_{BB} & X_{BN} \\ X_{NA} & X_{NB} & X_{NN} \end{bmatrix} \end{array}$$

Source memory for the secondary, external source (A vs. B) is computed for each 3 × 3 table, but the effect of generation on source is then examined by comparing performance on the read and generated items across the two tables.

As can be seen, the separate methodologies inherent in reality monitoring and external source monitoring result in very different designs. It is already well established that experimental design can play an important role in some aspects of the generation effect. For example, between-list versus within-list designs have been shown to be factors in producing positive and negative generation effects in free-recall memory tasks (Steffens & Erdfelder, 1998). It stands to reason that differences in source-monitoring designs may similarly play a central role in positive and negative generation effects for source memory.

The influence of response biases

It has been well documented that response biases and guessing strategies can have a strong influence in source-monitoring experiments (e.g., Anderson, 1984; Johnson & Raye, 1981; Johnson, Raye, Foley, & Foley, 1981). A number of theorists (e.g., Batchelder & Riefer, 1990; Murnane & Bayen, 1996) have pointed out that traditional empirical measures used in source monitoring, such as identification-of-origin (IDO) scores, are confounded by response bias and other cognitive factors and thus are not pure measures of source memory. One solution is to employ some type of mathematical model capable of measuring memory processes separately from response biases in an experiment. In particular, a class of models that have been successfully employed in the source-monitoring field are *multinomial processing tree* (MPT) models (Riefer & Batchelder, 1988). These are relatively simple, mathematically tractable models that are capable of separately measuring different cognitive processes as parameters of the model. Since the original Batchelder and Riefer (1990) article, a number of MPT models for source monitoring have been developed and tested (see Batchelder & Riefer, 1999, for a review), and they provide separate estimates for item detection, source discrimination, and response bias.

To illustrate how response bias can influence and confound the effect of generation on source memory, consider an experiment by Voss et al. (1987). In their experiment, participants worked in pairs to learn a shared list of words and then alternated their recall of the words in the list. This produced words that each person recalled himself or herself (self items) or words that the other person recalled (other items). In a final memory test, all participants were given a recognition test for the words and were asked to indicate which words were recalled and by whom. Voss et al. (1987) observed that source memory, as measured by IDO scores, was not significantly different for the self versus the other items. However, in a follow-up experiment, Riefer et al. (1994) explored the possibility that response biases may have contributed to the original results from Voss et al. (1987). In a replication and extension of that study, Riefer et al. (1994) supplemented an empirical analysis with a modelling analysis using a version of Batchelder and Riefer's (1990) MPT model for source monitoring. Similar to Voss et al. (1987), Riefer et al. (1994) found no significant difference between self and other items as measured by IDO scores. But the MPT model, which provided parameter estimates for source memory separately from estimates for various response biases, showed that the source-discrimination parameter was in fact significantly higher for the self than for the other items. Thus, the model revealed a positive generation effect for source that was not evident when examined only by traditional empirical measures of source memory.

Another example of strong response biases in source memory for generated items can be found in an experiment by Rabinowitz (1990). This study was concerned with memory for read versus generated words when both types of item were presented at the testing phase as well as during initial encoding. Among other findings, Rabinowitz (1990) found that, given correct recognition, the probability of responding "read" to a read word was greater than the probability of responding "generate" to a generated word. This is the only example we have found of a negative generation effect for source in an experiment using the reality-monitoring paradigm, and on the surface it appears to be inconsistent with the trend of positive generation effects in such paradigms. However, Rabinowitz (1990) also reported strong response biases in his experiment as measured by the false-alarm rates for new items, which were heavily biased towards the read items (e.g., see Table 2 in Rabinowitz, 1990). Although no mathematical modelling analysis was performed on these data, it is certainly possible that these strong biases may have contributed to the negative generation effects observed in this study.

THE CURRENT EXPERIMENT

The distinction between reality monitoring and external source monitoring can account for many of the diverse results of prior research concerning the effects of generation on source memory. Reality-monitoring paradigms lead to positive generation effects for source. This is consistent with the theory that the extra cognitive operations that are associated with internally generated information not only improve item memory but improve source memory as well (Marsh et al., 2001). Thus, when people are specifically asked which items they have produced themselves and which items they have not, they are better able to remember the source of items that they have internally generated. The few occurrences of negative effects (Rabinowitz, 1990) or null effects (Voss et al., 1987) can be accounted for by the influence of response bias.

Conversely, negative generation effects for source, when they occur, tend to occur within external source-monitoring paradigms. In these designs, generated items often lead to poorer source memory when the sources consist of outside events or the contextual details of items. A number of plausible explanations have been offered for these types of negative generation effect. For example, Jurica and Shimamura (1999) have proposed a *trade-off hypothesis*, which states that the mental operations that enhance the memory for generated items also inhibit the processing of related source information. Mulligan et al. (2006) have advocated a *processing hypothesis*, which states that generation results in poorer perceptual or data-driven processing of stimuli. This results in weaker encoding of the perceptual features of generated items, leading to poorer source memory for certain contextual details.

The dual hypothesis that we are proposing here—that positive generation effects tend to occur in reality-monitoring paradigms, and negative generation effects tend to occur in external source-monitoring paradigms—is theoretically tractable and does a good job of accounting for most of the diverse results found in prior research. But to date, all previous experimental studies on this issue have used either one of the above paradigms or the other. A direct test of our hypothesis is only possible, however, by incorporating both types of source-monitoring task into a single experimental design. The current experiment accomplishes this by replicating and extending the experiment by Mulligan (2004, Exp. 1) described earlier. In that experiment, participants memorized antonym word pairs that were in red or green print and that were either read or generated. However, on the final memory test, participants were only asked if items were presented in red or in green, and thus there were only three response categories: RED, GREEN, and NEW. In the current experiment we attempted to replicate the Mulligan (2004) experiment, but we extended the memory task to test participants' memory for which items were read and which were generated. Thus, for the experiment reported here, we gave participants five response categories: read/red, read/green, generate/red, generate/green, and new.

As can be seen, this experimental design involves multidimensional source information. Participants are required to keep track of which words were read and which were generated, but they also must monitor the colour of each word. Meiser and Bröder (2002) refer to this as *crossed source information*, and it is a more complex design than that of traditional source-monitoring studies, where the information may be multidimensional but only one source is tested. Fortunately, Meiser and Bröder have also developed an MPT model for source monitoring specifically designed to analyse data in this type of paradigm. Their model provides separate measures of item detection and source discrimination for each of the two source dimensions, as well as parameter estimates for a full range of guessing and response biases.

Given the complexity of studying source monitoring across multiple source dimensions, we used Meiser and Bröder's (2002) model for the analysis of our data as a supplement to the more traditional analyses using IDO scores. Considering the role that response biases can play in source-monitoring tasks, it is crucial to use some type of modelling analysis that is capable of separately measuring and factoring out these biases. This should provide more valid estimates for the levels of item detection and source discrimination in the experiment. If our dual hypothesis is correct, then the effect of generation on the model's source memory parameters should exhibit a crossover interaction, depending on the type of source memory being examined. Specifically, we predict that memory for the colour of the stimuli (external source monitoring) should be superior for the read items over the generated items. In contrast, memory for which items were read or generated (reality monitoring) should be superior for the generated items.

Method

Participants

A total of 120 students from psychology courses at California State University, San Bernardino, volunteered for the experiment. Each received extra course credit for their participation.

Materials

The stimuli consisted of 64 antonym pairs (e.g., boy–girl, true–false). Of these pairs, 60 were adopted from Masson and MacLeod (1992, Appendix B), the same source as that used by Mulligan (2004). An additional 4 antonym pairs were developed by the experimenters. The length of all words ranged from three to eight letters. From this set of 64, 4 antonym pairs were selected to appear as practice items, and the remaining 60 constituted the main stimuli for the experiment. For each participant, 40 of these antonym pairs were presented for memorization, and 20 served as new distractors on the final memory test.

For the 40 presented items, half were in the read condition and half in the generate condition. In the read condition, both words in the pair were complete (e.g., true–false). In the generate condition, the first word in the pair was complete but the second word contained only the first letter followed by four continuous underscore spaces (e.g., true–f____). Participants therefore saw more of the word's colour for the read items. Mulligan (2004) acknowledged this as a potential confound but ruled it out in a follow-up experiment (Exp. 5; also see Mulligan et al., 2006, Exp. 3). In our replication, it may have been possible to eliminate this confound—for example, by having only the first word in each pair in colour. Our main purpose, however, was to replicate Mulligan's (2004) Experiment 1 as thoroughly as possible, which is why we decided to have both parts of each stimulus pair appear in colour.

For purposes of counterbalancing, 12 different versions of the list were created, with 10 participants randomly selected for each version. Across these lists, each antonym pair appeared as a read item, a generate item, or a new item an equal number of times. Print colour was also counterbalanced across the 12 lists, so that each of the 60 antonym pairs appeared half of the time in green print and half in red print. For any individual presentation trial, half of the read items were in red print, and half were in green print. The same was true for the generate items.

Procedure

Participants were run in small groups ranging from 1 to 3 people, with each participant sitting in front of his or her own computer monitor. Participants were informed that they would see a series of antonym pairs and that half would be in red print and half in green print. They were also told that for some of the word pairs they would see both words of the pair. For these stimuli, they were to write down the second word of each pair on a response sheet provided (read condition). For other word pairs they were told that only the first letter of the second word would appear. For these stimuli they were instructed to determine what the correct antonym should be and to write that word down on the response sheet (generate condition). Consistent with Mulligan (2004), participants were informed that their memory would eventually be tested for both the type of word presented (read or generated) and the word's colour.

A practice list of four stimuli was presented first. Once it was clear that participants understood the task and had no questions, the main list of 40 antonym pairs was presented in random order. Each pair was shown one at a time for 7 s, with a 200-ms interval between stimuli. After presentation of the main list, participants were given a 3-min distractor task that was the same as the one used by Mulligan (2004, Exp. 1). Specifically, participants completed the names of U.S. cities having been given a three-letter word stem (e.g., NAS ___ for Nashville). Although this task is similar in some respects to the experimental task, we decided to use the same distractor task as that of Mulligan (2004) in order to replicate his experiment as completely as possible. Also, Mulligan et al. (2006, Footnote 5) have indicated that this particular task does not lead to different results from those of other distractor tasks.

Immediately following the distractor task, participants were tested on their memory for the words from the main list shown previously. Only the second word from each antonym pair was tested. The final test consisted of the 40 target words from the list plus 20 new distractors. Participants were informed that some of the words on the test were new words that did not appear on the previous list. All 60 test items were presented one at a time on the computer monitor, in random order and in black print. Presentation was self-paced, with each test item remaining on the screen until the participant made his or her response. Participants were required to indicate which of five categories each test item belonged to and to make their response by pressing one of five buttons on a response box in front of the computer monitor. The first category was represented by "word" written in red print followed by "I read the complete word on the computer and it was red". The second category was represented by "word" written in green print followed by "I read the complete word on the computer and it was green". The third category was represented by "w___" written in red print followed by "I generated the word myself from the first letter and it was red". The fourth category was represented by "w___" written in green print followed by "I generated the word myself from the first letter and it was green". The fifth category was represented by "new" written in black print followed by "This is a new word". As in Mulligan (2004), it was made clear to participants that the colour response should be based on the entire stimulus display—which was always the same colour for both the first word (which was complete) and the second word (which was either complete or only contained the first letter).

Results

Model-free analysis

During the study phase, 92.5% of the incomplete words were correctly generated, and 99.9% of the complete words were correctly copied. For all analyses that we report here, we excluded those items for which participants gave no response or an incorrect response during the study phase. For comparison, we also conducted the same analyses using the full set of items, based on participants' responses to all of the words on the final memory test, and these analyses produced the same pattern of results as did the analyses that excluded items.

For the main empirical analysis we computed recognition and identification-of-origin (IDO) scores in the same manner as Mulligan (2004) to ensure that our analysis was comparable to his. Specifically, recognition memory was computed for each type of item as a corrected hit rate, which equals the proportion of correct hits minus the proportion of false alarms. The IDO scores for each item were computed as the probability of correct source identification given that the item was correctly identified as old.

Mulligan (2004, Exp. 1) observed in his study that generated items were recognized better than read items. This is the standard generation effect, and this empirical result was observed in our experiment as well. The corrected hit rates for the generated and read items in the current experiment are presented in Figure 1, and they show that recognition for the generated items was higher than that for the read items, $F(1, 119) = 390.17$, $MSE = 0.032$, partial $\eta^2 = .77$, $p < .001$. There was no significant difference in the recognition rates based on the colour of the items (red vs. green), and item colour did not significantly interact with the generation effect (both Fs < 1).

The IDO scores are also presented in Figure 1. These scores measure the accuracy of source memory for the read compared to the generated items. Two sets of IDO scores are presented in

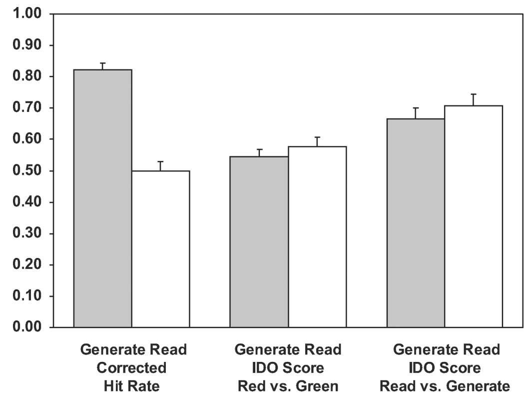

Figure 1. *Empirical results from the experiment. Corrected hit rates, IDO scores for correctly identifying red versus green items, and IDO scores for correctly identifying read versus generated items are plotted separately for the read and generated items. Vertical lines indicate 95% confidence intervals.*

Figure 1. The first set represents participants' ability to identify which colour each item appeared in. This is the external source-monitoring task, and this part of the analysis matches the analysis conducted by Mulligan (2004, Exp. 1). The second set represents the reality-monitoring task—that is, the ability of participants to identify which items were read and which were generated. An examination of Figure 1 shows that overall source identification was more accurate in the reality-monitoring task than in the external source-monitoring task, $F(1, 119) = 61.31$, $MSE = 0.031$, partial $\eta^2 = .34$, $p < .001$. But the main finding of interest is the observation of a negative generation effect for both types of task. Source memory for the read items was superior to the generated items for reality-monitoring tasks as well as the external source-monitoring tasks, and this main effect was statistically reliable, $F(1, 119) = 5.33$, $MSE = 0.030$, partial $\eta^2 = .04$, $p < .05$. Moreover, the source memory advantage of the read over the generated items was generally the same for both tasks, as indicated by a nonsignificant type of item by type of task interaction ($F < 1$).

The results of our experiment replicate those found by Mulligan (2004, Exp. 1). We observed a positive generation effect for item recognition, which matches not only Mulligan's results but also countless other studies on the generation effect. However, we observed negative effects for source memory, which occurred for both external source monitoring and reality monitoring. The negative generation effect for external source monitoring indicates that identification of an item's colour is better for the read items, a result that replicates Mulligan's. It is also consistent with the first part of our dual hypothesis, which states that any source memory advantage for read items over generated items should occur within external source-monitoring tasks. However, there was also a negative generation effect for the reality monitoring task—that is, participants' ability to identify which items were read or generated was actually superior for the read items themselves. This negative generation effect for the reality-monitoring task is in the opposite direction to that of the second part of our hypothesis, which predicts a positive generation effect for reality monitoring.

On the surface, the empirical analysis of the data from our experiment would indicate only partial support for our hypothesis. However, the above conclusions are based on IDO scores as measures of source memory. As we stated earlier, a number of theorists have criticized these types of empirical measure because they have special problems when it comes to accurately measuring source memory. For example, some theorists (e.g., Batchelder & Riefer, 1990; Murnane & Bayen, 1996) have pointed out that, because IDO scores are conditional on correct item recognition, these scores may be difficult to interpret when recognition rates differ across conditions. Indeed, in our experiment there are very strong differences in the recognition rates between read and generated items.

Other theorists (e.g., Rabinowitz, 1990; Riefer et al., 1994) have warned about interpreting IDO scores when there are strong response biases evident in the data. To explore this factor for the current experiment, we examined the false-alarm rates for the new items. This analysis revealed that when participants falsely indicated that a new item was "old", the probability that they identified it as a "read" item (.62) was significantly higher than the probability that they identified it as a "generated" item (.38), $t(119) = 3.54$, $p < .001$. Thus, participants in our experiment exhibited a strong bias in favour of the read items. This matches the response biases observed in previous studies (e.g., Rabinowitz, 1990; Riefer et al., 1994; Voss et al., 1987).

Meiser and Bröder's (2002) source-monitoring model

Because of the potential problems in interpreting the IDO scores in the current experiment, we also analysed the data using an MPT model. Specifically, we used a model developed by Meiser and Bröder (2002), which has been empirically validated as a useful model for investigating crossed dimensions of source information (e.g., Meiser, 2005; von Hecker & Meiser, 2005). There are two separate source dimensions examined in the experiment (colour and generation), which creates five types of items on the final memory test: red items that were read, red items that were generated, green items that were read, green items that were generated, and new items. This creates a 5 × 5 data table, representing the five possible responses for each of these five types of item (cf. Table 1 in Meiser & Bröder, 2002). This data table for the current experiment is presented in Table 1, which gives the total frequency for each response aggregated over the 120

Table 1. *Response frequencies for the red and green words, read and generated words, and new distractors*

		Responses				
		Red		*Green*		
Item		*Read*	*Generate*	*Read*	*Generate*	*New*
Red	Read	284	97	198	83	537
	Generate	154	330	144	318	168
Green	Read	178	83	290	104	544
	Generate	132	266	193	350	164
New		111	81	134	68	2006

participants in the study. Response frequencies across the diagonal of this table represent correct responses.

Figure 2 presents the tree structure for Meiser and Bröder's (2002) model as applied to the current experiment. In this tree, colour (red vs. green) is the first source dimension, and generation (read vs. generate) is the second source dimension, with an assumption of stochastic independence between the source retrieval processes

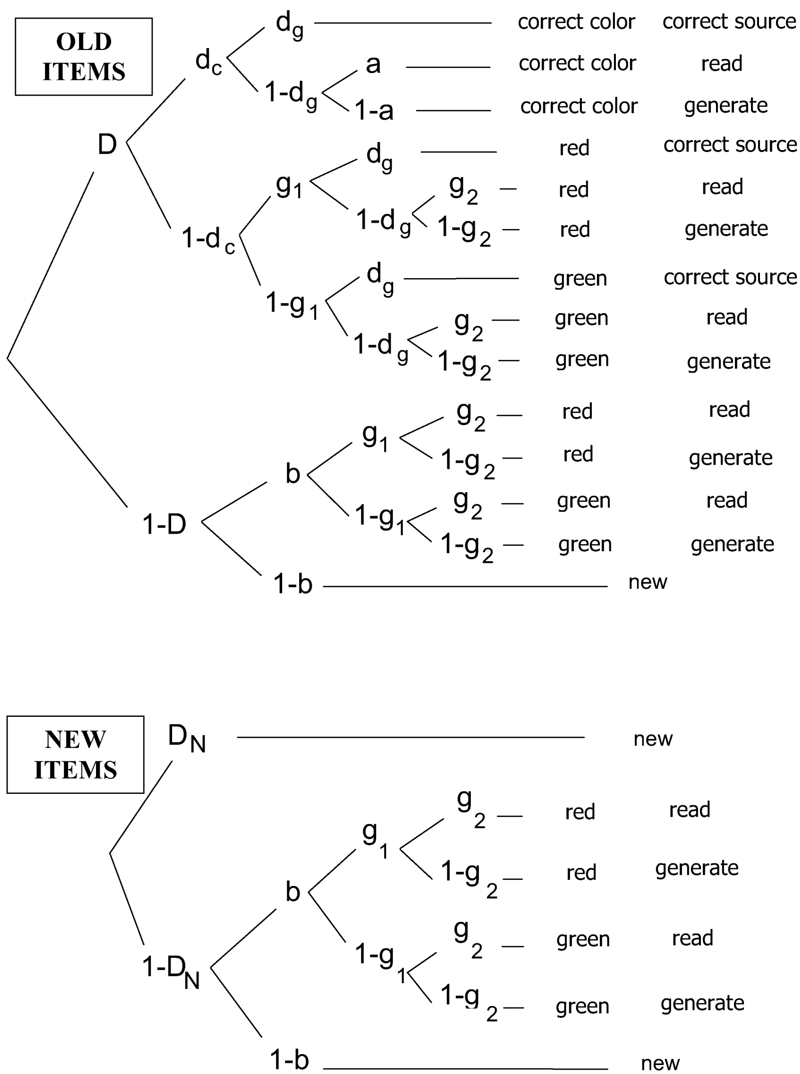

Figure 2. *Meiser and Bröder's (2002) multinomial processing tree model for crossed source information, with colour (red vs. green) crossed with generation (read vs. generated).* D = *the probability of correctly detecting a previously presented item as old;* D_n = *the probability of identifying a new item as new;* d_c = *the probability of correct discriminating red versus green items;* d_g = *the probability of correctly discriminating read versus generated items;* a = *the probability of responding "read" for detected items when colour is correctly identified but generation is not;* g_1 = *the probability of responding "red" when participants fail to remember the colour of an item;* g_2 = *the probability of responding "read" when participants guess the colour of an item; and* b = *the probability of responding "old" to nondetected items.*

for the two dimensions. We also explored a version of the model with generation as the first source dimension and colour as the second source dimension, and this version produced the same pattern of results.

As can be seen in Figure 2, the model contains a number of parameters for describing the data in Table 1. Parameter D measures item memory and represents the probability of correctly detecting a previously presented item as old. Parameter d measures source memory and is the probability of correctly discriminating the source of old items that have been correctly detected. Because there are two source dimensions, there are also two source parameters. We denote d_c as the probability of correct colour discrimination and d_g as the probability of correctly discriminating read versus generated items.

There are also a number of parameters representing various response biases when detection fails (with probability $1 - D$) or when source discrimination fails (with probability $1 - d$). Parameter b is the probability of responding that an undetected item is "old". Parameter g_1 is the probability of responding "red" when participants fail to remember the colour of an item, g_2 is the probability of responding "read" when participants fail to identify the colour or generation of an item, and a is the probability of responding "read" for detected items when colour is correctly identified (d_c) but generation is not ($1 - d_g$). Parameters a and g_2 are important parameters because they measure the extent that response bias influences the generation effect. We should also note that our assumption regarding parameter a differs from the one made by Meiser and Bröder (2002) because we are assuming that response bias for the second source dimension (read vs. generate) can differ depending on whether participants know the correct colour (a) or merely guess it (g_2). In contrast, Meiser and Bröder assumed that response biases for the second dimension are influenced by the response on the first dimension regardless of the origin of that response. Given the importance of response biases in the current study, we felt it was important to relax this assumption. However, to make sure that our conclusions do not depend on any particular set of assumptions, we analysed the data using Meiser and Bröder's original assumptions. The pattern of results was the same under either set of assumptions.

Figure 2 presents the basic tree structure for both old items and new items. However, it is important to realize that there are actually four separate trees for old items, one for each of the four types of stimulus. Thus, for the set of seven parameters described above (D, d_c, d_g, a, g_1, g_2, b), a separate set is needed for the red/generate, red/read, green/generate, and green/read items. This creates a total of 28 parameters in the model. There are 20 degrees of freedom (*df*) in the data structure in Table 1 (four *df*s in each of the five rows), and thus without further restrictions on the parameters the model is overidentified. We first assume that all guessing and response-bias parameters are equal across the four types of stimulus. Second, we assume that there are no a priori memory differences for the red versus green items. For item detection, this implies that $D_{(\text{red, generate})} = D_{(\text{green, generate})} = D_{(\text{generate})}$ and that $D_{(\text{red, read})} = D_{(\text{green, read})} = D_{(\text{read})}$. The same assumption applies to the source memory parameters d_c and d_g. With these assumptions, there are 10 total parameters for the model: $D_{(\text{generate})}$, $D_{(\text{read})}$, $d_{c(\text{generated})}$, $d_{c(\text{read})}$, $d_{g(\text{generated})}$, $d_{g(\text{read})}$, a, g_1, g_2, and b. The validity of all the assumptions for this model is supported if the model provides a satisfactory fit to the data.

The final parameter of the model is D_n, which appears in the tree for the new items and represents the high-threshold assumption that new items can be correctly identified as new. In their original source-monitoring model, Batchelder and Riefer (1990) assumed that $D_n = 0$ (referred to as the one-high threshold assumption). In contrast, Bayen, Murnane, and Erdfelder (1996) have advocated the use of two-high threshold models of source monitoring, on which D_n is nonzero and set equal to the value of D for the old items. Meiser and Bröder (2002) also used the two-high threshold assumption for their model. However, the issue for the application of their model to the current data set is whether the value of D_n

should be set equal to $D_{(generate)}$ or $D_{(read)}$. For our analysis we will explore all three assumptions ($D_n = 0$, $D_n = D_{(generate)}$, and $D_n = D_{(read)}$) in order to determine which assumption is most appropriate for the data set in Table 1.

Model-based analysis

Analysis of the data in Table 1 was conducted using the MBT computer program developed by Hu (1996). Goodness-of-fit tests for the model, as well as hypothesis tests, were conducted using the log-likelihood ratio statistic G^2, which is asymptotically distributed as a χ^2 variable (see Riefer & Batchelder, 1988). All of these tests used the .05 significance level, and for tests involving $df = 1$ the critical value is 3.84. When the model was applied to the data in Table 1, the one-high threshold assumption of $D_n = 0$ produced the best fit to the data: $G^2(10) = 6.91$, critical value = 18.31. The two-high threshold assumption of $D_n = D_{(read)}$ also produced a comparably good fit: $G^2(9) = 6.96$. In contrast, the fit under the assumption of $D_n = D_{(generate)}$ was acceptable but not as good as the other fits: $G^2(9) = 13.57$. Consistent with Meiser and Bröder (2002), we present the remainder of the model's analyses under the two-high threshold assumption that $D_n = D_{(read)}$. However, based on the comparable fit, we also conducted all analyses separately under the one-high threshold assumption that $D_n = 0$. With one exception (noted below) both assumptions yielded the same pattern of results.

Figure 3 presents the parameter estimates for D, d_c, and d_g separately for the read and generated items. Matching the empirical analysis, there is a strong and positive generation effect for item detection: parameter D, $G^2(1) = 513.88$. But in contrast to the analysis based on IDO scores, the pattern of results for the model's source memory parameters tells a different story. A negative generation effect occurs for colour discrimination (parameter d_c). However, the model reveals that there is a positive generation effect for discriminating read versus generated items (parameter d_g).

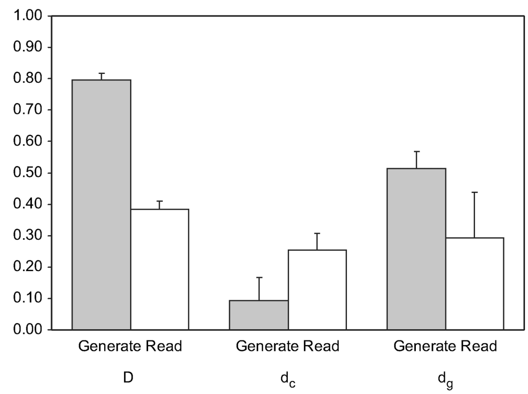

Figure 3. *Modelling results from the experiment. Item recognition (as measured by* D*), source memory for colour (as measured by* d_c*), and source memory for generation (as measured by* d_g*) are plotted separately for the read and generated items. Vertical lines indicate 95% confidence intervals.*

It is possible to test the statistical reliability of this 2 (read vs. generate) × 2 (d_c vs. d_g) crossover interaction. This is done by reparameterizing the MPT model under the assumption that the parameters $d_{c(generated)}$, $d_{c(read)}$, $d_{g(generated)}$, and $d_{g(read)}$ can be described by two stochastically independent factors. Let α_i ($i = 1, 2$) represent the factor of colour memory (α_1) versus generation memory (α_2). Also let β_i ($i = 1, 2$) represent the factor of generation (β_1) versus reading (β_2). The MPT model can then be reparameterized as $d_{c(generated)} = \alpha_1\beta_1$, $d_{c(read)} = \alpha_1\beta_2$, $d_{g(generated)} = \alpha_2\beta_1$, and $d_{g(read)} = \alpha_2\beta_2$. This is a more restrictive model that assumes stochastic independence between these two factors, and it can be compared to the full model without this assumption. The results reveal that the assumption of independence can be rejected, $G^2(1) = 15.41$, indicating that the interaction is statistically significant. (See Klauer, Ehrenberg, & Wegener, 2003; and Klauer, Wegener, & Ehrenberg, 2002, for other examples of testing interactions with MPT models.)

Exploring the details of this interaction, colour discrimination was significantly better for the read items than for the generated items, $G^2(1) = 12.20$. With regards to source discrimination based on generation, the advantage of the generated items over the read items was in the hypothesized direction but was marginally nonsignificant under the two-high threshold assumption of $D_n = D_{(\text{read})}$, $G^2(1) = 2.99, p < .10$. However, this positive generation effect for source was statistically significant under the one-high threshold assumption that $D_n = 0$, $G^2(1) = 6.52$.

The above analyses were conducted using the data in Table 1, which were pooled across participants. This is a common practice in MPT modelling; however, it assumes that responses are independent and identically distributed, an assumption that is violated if there are individual differences between the participants. Ideally, one would want to estimate parameters individually for each participant. This is not practical for the current experiment because, with only 60 observations distributed across the 5 × 5 response table, many of the category frequencies are zero. Instead, we computed parameter estimates for each of the 12 different list structures, with 10 participants contributing data for each version of the list (see Materials section). We then took the 12 sets of parameter estimates and analysed them using standard t tests with $df = 11$. Each of these t tests produced the same conclusions as the log-likelihood analyses based on the aggregated data in Table 1. In particular, parameters D and d_g were higher for the generated items ($t = 17.20$ and 3.43, respectively), and d_c was higher for the read items ($t = 2.85$).

We also estimated the response-bias parameters of the model, and these estimates are presented in Table 2, along with the values of G^2 testing if these parameters significantly differ from .5. For parameters a and g_2, values above .5 reflect a bias to respond "read". For parameter g_1, a value above .5 indicates a bias to respond "red". As can be seen in Table 2, both of the response-bias parameters based on generation (a and g_2) exhibit a strong and statistically reliable tendency to respond "read" to nondiscriminated items. As it turns out, there is also a smaller bias to respond "green" when colour discrimination fails ($g_1 < .5$), and this response bias was also statistically reliable. It is possible that this could reflect a handedness bias, given that "green" responses were always to the right of "red" responses on the response box.

Table 2. *Estimates for the response-bias parameters from Meiser and Bröder's (2002) model*

Parameter	*Estimate*	$G^2(1)$
a	.80 (.10)	7.10
g_1	.47 (.01)	7.99
g_2	.62 (.02)	23.65

Note: Parameter a is the probability of responding "read" for detected items when colour is correctly identified but generation is not; g_1 is the probability of responding "red" when participants fail to remember the colour of an item; and g_2 is the probability of responding "read" when participants guess the colour of an item. $G^2(1)$ is the log-likelihood ratio statistic based on one degree of freedom. Numbers in parentheses are the estimates of the standard deviation for each parameter.

Discussion

The purpose of this experiment was to examine a dual hypothesis for the role that source memory plays in the generation effect. We hypothesized that the type of source-monitoring task is an important factor in determining whether positive or negative generation effects will occur for source memory. The first part of our hypothesis states that negative generation effects—the memory advantage for externally generated information over self-generated information—tend to occur in external source-monitoring paradigms. The experimental results solidly supported this hypothesis. Memory for word colour was superior when the words were read than when they were generated. This finding occurred for the empirical analysis based on IDO scores, and it replicates the results found by Mulligan (2004, Exp. 1). The negative generation effect for word colour was also confirmed by the analysis of Meiser and Bröder's (2002) MPT model for crossed source dimensions. The parameter from this model measuring source memory for colour was higher for the read words than for the generated words.

The second part of our hypothesis states that positive generation effects—the memory advantage for self-generated information over externally generated information—occur in reality-monitoring paradigms in which self versus external information is the source being tested. The experimental results based on the analysis of the MPT model also support this part of the hypothesis. As it turns out, the empirical analysis based on IDO scores produced results that were in the opposite direction to that of our prediction. But the modelling analysis revealed that these IDO scores were influenced by strong response biases and by differential detection rates for the read and generated words. The model is capable of separately measuring and correcting for these factors, producing an uncontaminated measure of source memory based on a parameter of the model that directly measures source memory for the read and generated words. The model results clearly show a positive source memory advantage for the generated words, consistent with our hypothesis. This positive generation effect was statistically reliable under a one-high threshold assumption, although the effect was marginally nonsignificant under a two-high threshold assumption. However, the crossover interaction represented by the dual hypothesis was statistically significant under both threshold assumptions.

The results of the model's analysis provide a powerful illustration of the potential confounding effects that response bias can have in source monitoring. If our analysis had been restricted only to the IDO scores, we would have simply concluded that our experimental manipulation resulted in consistent negative generation effects for source memory. As such, our findings would have been just another piece in a puzzling series of research results, agreeing with prior reports of negative generation effects for source but inconsistent with those studies finding positive effects. Very little would have been added to our theoretical understanding of this issue. In contrast, Meiser and Bröder's (2002) model is capable of measuring item detection, source discrimination, and response biases as separate model parameters. This produced a clearer set of results that helps to explain the diverse pattern of findings in prior research.

In fact, with the MPT model in place, it is now easy to illustrate how response bias could have obscured the results in some of the previous studies on generation effects in source monitoring. For the purposes of this demonstration we revisit the experiment by Rabinowitz (1990). In that experiment, negative generation effects were observed within a reality-monitoring paradigm, a result that on the surface appears to contradict our hypothesis. More specifically, Rabinowitz (1990) found that the probability of responding "read" to a read item, Pr("R" | R), was higher than the probability of responding "generate" to a generated item, Pr("G" | G). Because both of these measures were conditional on correct recognition, it can be assumed that each of these measures is affected by some combination of correct source identification and guessing biases. In terms of the model's parameters, each measure equals $d + (1 - d)g$: that is, the probability of correct source identification plus the probability that source memory fails but a correct guess is made.

Rabinowitz (1990) himself noted that strong response biases in favour of the read items were operating in the experiment. For this demonstration assume that such response biases are coupled with a strong *positive* generation effect for source. Specifically, let $d_{\text{generate}} = .7$ and $d_{\text{read}} = .4$ (a positive generation effect), and let $g_{\text{generate}} = .2$ and $g_{\text{read}} = .8$ (a response bias favouring read items). Under these assumptions the empirical measures can be computed as follows:

$$\Pr(\text{“R”}|\text{R}) = .4 + (1 - .4)(.8) = .88$$
$$\Pr(\text{“G”}|\text{G}) = .7 + (1 - .7)(.2) = .76.$$

Thus a negative generation effect is observed with the empirical statistics, despite our initial assumption of a strong positive generation effect in the model's parameters. In addition, the values that we chose for this demonstration were selected because they come very close to the actual empirical numbers observed by Rabinowitz (1990, cf. his Table 2). It is therefore easy to see how an analysis of source-monitoring data based on MPT modelling can produce a very different conclusion from that of more traditional empirical analyses, a result that happened here with our experiment. A similar pattern occurred in Riefer et al. (1994)

in which empirical statistics showed no significant difference between self-recalled and other-recalled information. But a modelling analysis revealed that source memory for self-recalled items was in fact superior to other-recalled items, a positive generation effect that is consistent with our hypothesis.

The hypothesis that positive generation effects occur in reality-monitoring paradigms provides a very consistent account of the results from all generation effect studies using this paradigm. In addition to our experiment, positive generation effects have been observed in reality-monitoring tasks used by Geghman and Multhaup (2004), Kinjo and Snodgrass (2000), and Riefer et al. (1994). The only times when a null effect (Voss et al., 1987) or a negative effect (Rabinowitz, 1990) have been found in this paradigm can easily be accounted for by the confounding effects of response biases. Furthermore, the notion that source memory is superior for self-generated information is generally consistent with most theories of the generation effect as it is applied to item memory. Theories of the generation effect assume that when people generate information themselves, this results in extra mental operations that improve various aspects of the encoding of that information. It is logical to theorize that this enhanced encoding not only benefits the overall recall or recognition of this information but also memory for its source as well.

As we have noted earlier, negative generation effects for source, when they occur, tend to happen within external source-monitoring paradigms. In other words, the ability to determine the source of ancillary information (such as speaker, or word colour) is often better for externally generated than for self-generated information. At least two different theories have been offered to explain such negative generation effects. Jurica and Shimamura's (1999) trade-off hypothesis can account for our results by assuming that when participants had to devote cognitive resources to the generation of words, this took mental resources away from memorizing the word's colour. In contrast, Mulligan's (2004; Mulligan et al., 2006) processing hypothesis assumes that even though generation may have resulted in better conceptual encoding of items, it also led to poorer memory for the perceptual details of stimuli, such as their colour. This theory is consistent with studies (e.g., Jacoby, 1983; Roediger & Blaxton, 1987) showing that memory for generated information is worse when tested using perceptually based implicit memory tasks, suggesting that the perceptual or surface details of stimuli are encoded more poorly for generated items. We should point out that our experimental findings do not differentiate between the trade-off and processing hypotheses, which can equally explain the negative generation effects that we observed here for colour memory. Further experimentation will need to be conducted to determine the validity of these and other theories (cf. Mulligan et al., 2006).

We should also make it clear that there are some published experiments with results that are inconsistent with our hypothesis. In particular, some studies have observed positive generation effects for source within an external source-monitoring paradigm. For example, Koriat, Ben-Zur, and Druch (1991, Exp. 1) found positive generation effects when they examined source memory for room location, and Marsh et al. (2001) observed positive generation effects for room location, left–right location memory, and even weak but positive effects for colour memory. We cannot account for these results based on our dual hypothesis. However, as Mulligan (2004) has pointed out, they represent a challenge for the trade-off and processing hypotheses as well. (See Mulligan et al., 2006, page 845, for some possible explanations for these discrepant results.)

Based on the results from the current experiment, we would make two recommendations for researchers who wish to explore generation effects in source monitoring. First, future research and theory development on this issue should take into consideration the type of source-monitoring paradigm being examined. Second, we would recommend that any empirical analysis of the results from such research be supplemented by a model-based analysis. Numerous MPT models for source monitoring now exist (e.g., Batchelder, Hu, & Riefer, 1994; Batchelder & Riefer, 1990; Bayen et al., 1996; Meiser & Bröder, 2002) that can be applied to a wide range of experimental designs. Of course, it is quite possible that model-based conclusions may generally match those based on empirical statistics, which happened in the studies by Mulligan (2004) and Mulligan et al. (2006). But given the confounding effects that item detection or response biases can play in empirical measures of source memory, MPT modelling can be a useful and often necessary addition to these more traditional analyses.

Original manuscript received 14 June 2006
Accepted revision received 10 September 2006
First published online day month year

REFERENCES

Anderson, R. E. (1984). Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General*, *113*, 594–613.

Batchelder, W. H., Hu, X., & Riefer, D. M. (1994). Analysis of a model for source monitoring. In G. Fisher & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51–65). New York: Springer-Verlag.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548–564.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215.

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

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

Hu, X. (1996). Statistical inference program for multinomial binary tree models [Computer software]. Memphis, TN: University of Memphis, Department of Psychology.

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

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

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, N. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, *94*, 37–64.

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

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

Klauer, K. C., Ehrenberg, K., & Wegener, I. (2003). Crossed categorization and stereotyping: Structural analyses, effect patterns, and dissociative effects on context relevance. *Journal of Experimental Social Psychology*, *39*, 332–354.

Klauer, K. C., Wegener, I., & Ehrenberg, K. (2002). Perceiving minority members as individuals: The effects of relative group size in social categorization. *European Journal of Social Psychology*, *32*, 223–245.

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research*, *53*, 260–270.

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

Masson, M. E. J., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General*, *121*, 145–176.

McDaniel, M. A., Wadill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, *27*, 521–536.

Meiser, T. (2005). A hierarchy of multinomial models for multidimensional source monitoring. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 2–17.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 116–137.

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

Mulligan, N. W., & 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.

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

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

Rabinowitz, J. C. (1990). Effects of repetition of mental operations on memory for occurrence and origin. *Memory & Cognition*, *18*, 72–82.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339.

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies on source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 680–693.

Roediger, H. L., III, & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus centennial conference* (pp.). New York: Q1
Springer-Verlag.

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

Steffens, M. C., & Erdfelder, E. (1998). Determinants of positive and negative generation effects in free recall. *Quarterly Journal of Experimental Psychology*, *51A*, 705–733.

Von Hecker, U., & Meiser, T. (2005). Defocused attention in depressed mood: Evidence from source monitoring. *Emotion*, *3*, 456–463.

Voss, J. F., Vesonder, G. T., Post, T. A., & Ney, L. G. (1987). Was the item recalled and if so by whom? *Journal of Memory & Language*, *26*, 466–479.

PQJE202459

Queries

David M. Riefer, Yuchin Chien, and Jason F. Reimer

Q1 Roediger & Blaxton (1987). "pp." given but no page nos.