

Source-monitoring judgments about anagrams and their solutions: Evidence for the role of cognitive operations information in memory

MARY ANN FOLEY AND HUGH J. FOLEY
Skidmore College, Saratoga Springs, New York

Generating solutions to anagrams leads to a memory advantage for those solutions, with generated words remembered better than words simply read. However, an additional advantage is not typically found for solutions to difficult anagrams relative to solutions to easy ones, presenting a challenge for the cognitive effort explanation of the generation effect. In the present series of experiments, the effect of manipulating anagram difficulty is explored further by introducing two new source-monitoring judgments. These studies demonstrate that when attention is directed at test to the operations activated during encoding (by way of source-monitoring judgments focused on solving vs. constructing anagrams), a source advantage is observed for difficult anagrams. However, when attention is directed to the anagrams themselves, asking participants to remember the kinds of anagrams generated or solved (based on kind of rule rather than subjective impressions of difficulty), a similar source advantage is not observed. The present studies bring a new perspective to the investigation of difficulty manipulations on memory for problem solving by illustrating the impact of a shift in focus from the effort mediating cognitive operations to specifics about the cognitive operations themselves.

The beneficial effect of involving individuals (whether participants in experiments or students in classes) in the creation of materials to be remembered has long been recognized in both theoretical (Johnson, Raye, Foley, & Foley, 1981; Mulligan, 2001; Slamecka & Graf, 1978; Tacconat & Isingrini, 2004) and educational contexts (deWinstanley, 1995; Ross & Balzer, 1975; Ross & Killey, 1977). In a prototypical study investigating the enhancing effects of involvement, participants generate words in response to prompts presented by another person (e.g., an experimenter or partner). On subsequent memory tests, the materials one has generated are often better recognized or recalled than are nongenerated control materials (Johnson et al., 1981; Slamecka & Graf, 1978), and these self-generated materials are also identified as such in source-monitoring judgments (e.g., Johnson et al., 1981, Experiment 1). Thus, this memory advantage for self-generated information, referred to by Slamecka and Graf (1978) as the *generation effect*, is observed for both item and source memory. Although it is intuitively reasonable to expect additional memory advantages for more difficult generative acts, this expectation is often not confirmed (Foley, Foley, Wilder, & Rusch, 1989; McNamara & Healy, 2000; Zacks, Hasher, Sanft, & Rose, 1983). The purpose of the present series of studies is to explore the conditions under which additional memory advantages might be observed for difficult versions of problems.

Generation effects are observed across a variety of materials, test conditions, and retention intervals (e.g., Foley, Foley, Durley, & Maitner, 2006; Foley & Ratner, 1998; Greene, 1992; Johnson et al., 1981; Mulligan, 2001, 2002, 2004). Advantages in item memory follow the use of several kinds of encoding rules defining the basis for generating items, including requests to produce antonyms, category instances, or unrelated words (Johnson et al., 1981; Mulligan, 2001; Rabinowitz, 1989; Slamecka & Graf, 1978), requests to solve anagrams (Foley et al., 1989), requests to solve numerical problems (McNamara & Healy, 1995), and requests to identify incomplete pictures (Kinjo & Snodgrass, 2000). Frequently reported for explicit memory tasks (Glisky & Rabinowitz, 1985; Hirshman & Bjork, 1988; McElroy & Slamecka, 1982), the advantage is evident for some implicit tasks as well (e.g., Blaxton, 1989; Srinivas & Roediger, 1990, Experiment 1) and is maintained over relatively long retention intervals for both item memory (7 days; Gardiner, Ramponi, & Richardson-Klavehn, 1999, Experiment 2) and source memory (10 days; Johnson et al., 1981). Similar facilitative effects are also reported for order (Kelley & Nairne, 2001) and contextual information (E. J. Marsh, Edelman, & Bower, 2001).

One explanation for these beneficial effects of generating materials emphasizes the role of the cognitive operations that lead to the formulation of questions about materials (e.g., Ross & Balzer, 1975) or to the generation of the

M. A. Foley, mfoley@skidmore.edu

materials themselves (e.g., Johnson et al., 1981). Broadly defined within the source-monitoring framework guiding our work, *cognitive operations* refer to processes activated during the encoding of information, such as those guiding the reading of material presented by an experimenter or those guiding the search and decision strategies leading to the generated materials (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 2000; Johnson et al., 1981). Cognitive operations could contribute to generation effects in any one of a number of ways. The effort associated with generating materials could lead to stronger (or more distinctive) traces for generated than for nongenerated materials, making generated materials more readily available at test. Furthermore, if participants remember the operations themselves, these operations could later serve as retrieval strategies for reactivating the generated information.

Early tests of the role of cognitive operations information on both item and source memory focused on effort manipulations, and, thus, not surprisingly, were referred to by many researchers as tests of the *cognitive effort* explanation (e.g., McDaniel, Einstein, & Lollis, 1988; Slamecka & Graf, 1978; Tyler, Hertel, McCallum, & Ellis, 1979; Zacks et al., 1983). According to this view, the relative magnitude of the generation effect (as measured by the difference in memory for generated and nongenerated materials) should be in proportion to the effort required to produce the generated responses (Johnson et al., 1981; Slamecka & Graf, 1978). Consistent with this expectation, the manipulation of cognitive effort by way of different kinds of encoding rules led to predicted variations in the strength of the generation effect when comparing generated materials with their respective control materials (e.g., Johnson et al., 1981). Despite their robustness, the facilitative effects of effort manipulations are far from ubiquitous (for reviews, see Greene, 1992; Mulligan, 2001; Schmidt & Cherry, 1989; Steffens & Erdfelder, 1998). Two lines of research on generation effect failures are particularly problematic for the cognitive effort explanation—research focusing on the effects of generating nonwords and the effects of difficulty manipulations on problem sets.

Evidence from letter transformation tasks plays a prominent role in the evaluation of the cognitive effort explanation (e.g., Johns & Swanson, 1988). If cognitive effort (or the act of generation per se) is responsible for generation effects, then memory advantages should be observed for nonwords as well. However, reports of failures to find an advantage in item memory for generating nonwords cast doubt on the cognitive effort explanation (e.g., McElroy & Slamecka, 1982; Nairne, Puse, & Widner, 1985; for an exception, see Johns & Swanson, 1988).

These findings on memory for nonwords are relevant for the present series of studies because, along with other investigators (Kinoshita, 1989; Mulligan, 2002), we conceptualize anagram solving as one of several types of letter transformation tasks. When asked to follow simple transposition rules to create nonwords, participants rearrange a string of letters to generate nonwords by switching two letters in a string that are denoted by underlining (e.g., changing ZERPIK to PERZIK; McElroy & Slamecka, 1982; Nairne et al., 1985). Letter-switching

rules sometimes result in the creation of words (e.g., Kinoshita, 1989; Nairne & Widner, 1987). For example, when asked to “correct spelling errors,” by switching two letters denoted by underlining (e.g., JDAE), participants produce words (e.g., JADE; Kinoshita, 1989; Mulligan, 2002). Item memory generation effects are observed in letter transformation tasks only when the rearranging of letter strings results in word productions (e.g., Kinoshita, 1989; Nairne et al., 1985).

Also damaging for the cognitive effort explanation were the demonstrations of failures to find additional memory advantages for solutions to difficult versions of problems in comparison with easy versions. Although a traditional generation effect is observed when comparing memory for words generated as anagram solutions to memory for nongenerated words (Foley et al., 1989; Jacoby, 1991; Taconnat & Isingrini, 2004, Experiment 2), an additional advantage is not found for words generated as solutions to difficult anagrams in comparison with words generated as solutions to easy ones (Foley et al., 1989; R. L. Marsh & Bower, 1993; McNamara & Healy, 2000; Zacks et al., 1983). (Based on norms for anagram construction, a difficult anagram is a letter string produced by rearranging as much as possible the ordering of letters in a word, whereas an easy anagram is a letter string produced by transposing an adjacent pair of letters in a word.) Neither solving difficult anagrams (e.g., Foley et al., 1989) nor discovering words embedded in more difficult letter arrays patterned on the game Boggle (e.g., R. L. Marsh & Bower, 1993) lead to memory advantages for those solutions relative to control versions of the tasks purposefully made to be easy to solve.

Further weakening the persuasiveness of the cognitive effort explanation, McNamara and Healy (2000) predicted and confirmed an advantage for solutions to simpler versions of multiplication problems in comparison with more complex versions. This advantage for simpler versions led McNamara and Healy (2000) to conclude that the operations involved in solving the simpler versions were easier to remember (or at least more likely to be activated at test), contributing to the memory advantage. Referring to the basis of their predictions as the *procedural reinstatement view*, McNamara and Healy (2000) offered this view as an alternative to the cognitive operations explanation, presumably because their finding of an advantage in memory for easy problems over difficult ones is not predicted by an explanation emphasizing the role of cognitive effort.

However, from our perspective, rather than replacing explanations based on cognitive operations information, the procedural reinstatement view shifts attention appropriately from effort (or any general feature of cognitive operations) to the operations themselves (e.g., those driving different kinds of problem-solving and, subsequently, retrieval strategies). As such, both the procedural reinstatement view and the source-monitoring framework led us to expect that source-monitoring judgments that directed attention to the cognitive operations involved in anagram solving might be more likely to reveal advantages for difficult in comparison with easy anagram solving.

In part, this expectation arose from demonstrations in other contexts of the sensitivity of source judgments to cognitive operations information (Finke, Johnson, & Shyi, 1988; Foley, Durso, Wilder, & Friedman, 1991; Foley, Foley, & Korenman, 2002; Foley & Ratner, 1998, Experiment 3; Johnson et al., 1981). After generating materials in response to experimenters' prompts, source judgments about the origin of the items (e.g., self or experimenter) are better when the generations are more difficult to complete (e.g., Johnson et al., 1981). Indeed, source judgments about materials that one has generated are particularly sensitive to cognitive operations information, leading us to predict that an anagram difficulty manipulation on decisions about who generated anagram solutions might lead to more accurate performance on words associated with difficult anagrams, and this expectation was confirmed (Foley et al., 2006). Participants alternated turns with a partner to generate solutions to anagrams. When they were asked to remember the person who generated solutions (self or partner), participants' source accuracy was better for solutions to difficult anagrams than for solutions to easy ones. Presumably, when participants are asked to remember who generated particular solutions, if problem-solving strategies come to mind at test, these strategies would serve as cues about who solved the problems (self or partner). Thus, participants seem to use cognitive operations information to guide the retrieval of problem solutions (McNamara & Healy, 2000) and to render source-monitoring judgments about who generated problem solutions (i.e., self or partner, Foley et al., 2006).

The results from a letter transformation experiment in which attention was directed to the transformation rules also lends support to our perspective (Nairne & Widner, 1987, Experiment 1). When participants were asked to identify the letters that they had transposed, a memory advantage was observed for participant-generated nonwords in comparison with control conditions (Nairne & Widner, 1987, Experiment 1). Most other studies using letter transformation tasks, including anagram solving, have focused exclusively on item memory for the solutions themselves by examining recognition or recall of solutions to easy versus difficult anagrams (e.g., Foley et al., 1989; Zacks et al., 1983). The Nairne and Widner (1987) finding suggests that if participants are directed to think about the operations involved in anagram solving, an advantage for difficult anagrams might also be observed. More generally, the absence of an advantage for letter transformation tasks leading to nonwords may have less to do with the unimportance of cognitive operations information *per se* and more to do with the possibility that memory tests vary in their sensitivity to the presence of cognitive operations information. Consistent with this point, the wording of source judgments has important consequences for accuracy (e.g., Foley et al., 2002; R. L. Marsh & Hicks, 1998).

In the present series of experiments, we examined the effects of anagram difficulty by way of two new source judgment tasks. Instructions for source tests explicitly directed attention to the cognitive operations activated

during encoding (by way of source judgments focused on solving vs. constructing anagrams) or to the difficulty level of the anagrams themselves. In Experiment 1, participants performed each of two anagram manipulation tasks, both solving and constructing anagrams. Participants were then surprised with a source-monitoring judgment task in which they were asked to remember the type of task performed on each item (solving anagrams or constructing anagrams from words). In Experiments 2–4, participants were surprised with a source-monitoring task in which they were asked to remember the kind of anagram generated or solved (easy or difficult). The rule for classifying letter strings as easy or difficult anagrams was based on the extent to which letter strings resembled the ordering of letters in words (transposing a pair of adjacent letters to create easy anagrams or scrambling the ordering as much as possible to create difficult anagrams).

EXPERIMENT 1

When making type of task judgments (solve or construct), we expected to find an advantage for solutions associated with difficult anagrams. Several lines of research informed this expectation. As we mentioned, an advantage for difficult anagrams is observed when participants are asked to remember who solved the anagrams, a judgment thought to draw on cognitive operations information (Foley et al., 2006). Along similar lines, there is some suggestion that individuals remember whether their generation attempts led to success or failure (Kinjo & Snodgrass, 2000). Moreover, in early source-monitoring studies (e.g., Foley, Johnson, & Raye, 1983; Johnson et al., 1981), when deciding whether or not a word was self-generated, participants reported considering cues related to the presence (e.g., "I remember coming up with that one") or absence (e.g., "I don't remember coming up with that one") of cognitive operations information (e.g., Johnson et al., 1981). Following the same logic, in Experiment 1, participants could base their decisions on the presence (or absence) of one kind of cognitive operation (e.g., "solved that one" vs. "didn't solve that one") or on the basis of memory for the type of operation itself (e.g., "I remember scrambling those letters" or "I remember finally discovering that one").

The extent to which an anagram letter arrangement resembles a word seems to influence the nature of the processing (data-driven vs. conceptually driven) evoked by an anagram (Srinivas & Roediger, 1990), also leading us to expect an effect of anagram difficulty on type of task judgments. Particularly for easy anagrams, because only two adjacent letters are transposed, the resemblance between the letter strings and their solutions is relatively high. In contrast, for difficult anagrams, in which the arrangement of letters is fairly random, the operations involved in rearranging the letters to "discover" the word solution may bear less resemblance to the operations involved in reading words. Similarly, when scrambling letters as much as possible to create difficult anagrams, participants may do more than simply read the words. Thus, the advantage for difficult anagrams was expected for both task conditions

(solve or construct). A nongenerate control condition (e.g., reading words) was not included in Experiment 1 because the benefit for generating solutions, in comparison with simply reading words, is well documented (e.g., Foley et al., 1989; Jacoby, 1991; Tacconat & Isingrini, 2004).

Method

Participants. Thirty-two undergraduates volunteered to participate, with each participant receiving \$3.00. Each participant contributed data to only one study in the series reported in this article.

Materials. Eighty high-frequency words were used—that is, all were A or AA from the Thorndike and Lorge (1944) norms, with an average length of seven letters. Easy and difficult anagrams were constructed using guidelines established by others (Mayzner & Tresselt, 1958; Zacks et al., 1983). The first letter of each anagram was the same as the first letter of its corresponding word solution. With this first-letter constraint, an easy anagram was defined as one in which the letter order was maximally similar to its word solution (two adjacent letters were switched), and a difficult anagram was defined as one in which the letter order was quite different from the ordering of the letters in the word. The specific materials (words and anagrams) used in the experiments reported in this series were among those used in our earlier studies of memory for anagram solving (Foley et al., 1989).

Independent confirmation of the anagram classifications was provided by normative data based on individual response times for solving the easy and difficult anagrams (Foley et al., 1989). With the exception of anagrams that were participant-generated, the anagrams used in this series were therefore representative of those used in our earlier studies in which baseline conditions were included (Foley et al., 1989). In comparison with baseline conditions reported in these previous studies (e.g., simply repeating words), these materials produce a positive generation effect in the context of anagram solving (i.e., generating words as solutions vs. repeating words).

Two sets of 40 items (one target set and one distractor set) were constructed. For the set of 40 targets, four lists were constructed, with an item assigned to easy solve, difficult solve, easy generate, and difficult generate conditions, across lists. For all of the experiments reported in this series, appropriate counterbalancing procedures were followed, with words rotated through anagram difficulty level when serving as target items. Two different, randomized presentation orders were used during both the encoding and test phases.

Procedure. Tested individually, undergraduates were told that the purpose of the study was to create a set of materials for a study of problem solving. On each encoding trial, before an anagram or word was presented, the participants were cued to solve or construct, respectively. The participants were given unlimited time either to solve the anagrams or to create anagrams, but they were told they should work as quickly and accurately as possible. The participants were further instructed about how to create the anagrams, following guidelines used for the experimenter-created ones (Mayzner & Tresselt, 1958; Zacks et al., 1983). The participants were told that: “An easy anagram is one in which a pair of letters is transposed, making the letter order maximally similar to its word solution. A difficult anagram is one in which the letter order is scrambled as much as possible. For example, an easy anagram for *ribbon* might be *rbion* and a difficult anagram might be *rbinob*.” The participants were also told that the first letter of an anagram should be the first letter of its word solution. These instructions were available throughout the generation phase of the experiment. For anagram-solving trials, all of the participants eventually discovered the correct solutions.

After a 3-min retention interval, during which time the participants were asked to count backward by 3s from 717, they were surprised by a source-monitoring test. The test consisted of 80 words (40 targets and 40 distractors). For each word, the participants were asked to make source decisions, classifying the word as one they generated as a solution for an anagram, as one from which they created an anagram, or as a new word. When classifying a test word as

“old,” the participants also used a 9-point scale to indicate the difficulty that they experienced in making each source decision, with higher ratings indicating greater perceived difficulty in rendering the source decisions.

Results and Discussion

Type of task judgment. The source-monitoring task was a judgment about the type of task associated with the test word. Source-monitoring accuracy was the proportion of correct task judgments to old items divided by the number of hits (responding “old” to old items). The proportions are presented in the top portion of Table 1. An ANOVA with type of task (solve or construct) and anagram difficulty (easy or difficult) as variables showed two main effects. Source-monitoring judgments were more accurate for words that had originally been in the solve condition ($M = .84$) than for those that had been in the construct condition ($M = .67$) [$F(1,31) = 16.93$, $MS_e = .05$, $p < .001$]. Source-monitoring judgments were more accurate for words that had originally been in the difficult anagram condition ($M = .79$) than for those in the easy condition ($M = .73$) [$F(1,31) = 4.45$, $MS_e = .03$, $p = .04$]. The interaction effect was not significant.

The source accuracy findings confirm our predictions that, as a result of directing participants to the operations involved in anagram manipulation by way of a type of task judgment, an advantage was observed for words in the difficult anagram condition in comparison with words in the easy anagram condition. Moreover, source accuracy was better for words generated as solutions than for words whose letters were rearranged to construct anagrams, suggesting that the kind of operation involved in generating materials has important consequences for memory.

Item memory. The proportion of hits (calling old items “old,” regardless of whether the type of task judgment was correct) is shown in the lower portion of Table 1. An ANOVA with type of task (solve or construct) and anagram difficulty (easy or difficult) as variables showed that item memory was better in the anagram solve ($M = .82$) than in the anagram construct condition ($M = .61$) [$F(1,31) = 62.17$, $MS_e = .02$, $p < .001$]. Item memory did not differ for easy and difficult anagrams. The interaction was not significant.

An advantage for difficult anagrams was not evident in item memory in Experiment 1, a pattern consistent with

Table 1
Proportion of Correct Type of Task Judgments (Source Accuracy) for Items Described As “Old,” and Proportion of Old Items Identified As “Old” in Experiment 1 As a Function of Type of Task and Anagram Rule Difficulty

Task	Anagram Rule Difficulty			
	Easy		Difficult	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Source Accuracy				
Solve	.80	.05	.87	.04
Construct	.65	.05	.70	.03
Item Memory				
Solve	.81	.03	.83	.02
Construct	.62	.02	.61	.03

previous studies of anagram solutions (e.g., Foley et al., 1989). Thus, an intriguing dissociation was observed between item and source memory. Although item and source memory are sometimes affected in similar ways by encoding and test manipulations, this pattern is not always found (e.g., Foley & Ratner, 1998; Hicks & R. L. Marsh, 2001; Lindsay & Johnson, 1991). Within the context of source judgments, item memory is typically assessed indirectly by estimating recognition performance from the source accuracy scores. In these cases, source-monitoring measures may confound source discrimination with item memory (Bayen, Murnane, & Erdfelder, 1996; Murnane & Bayen, 1996). To consider this possibility, we conducted another experiment, assessing item memory independent of source judgments. We replicated the item memory effects reported for Experiment 1.¹

Thus, the dissociation between source accuracy and item memory observed in Experiment 1 was not simply a result of assessing item memory within the context of a source test.

False positives. The proportion of new items called “old” (false positives) was computed as the proportion of new items to which an incorrect solve or construct response was made during the type of source judgment task. The mean proportions are reported in Table 2 as a function of incorrect source choice (solve or construct). As is clear from Table 2, the mean proportion of false positives was quite low. The ANOVA on these data revealed no significant effects.

Perceived difficulty of type of task judgment. The perceived difficulty ratings for type of task judgments were calculated for correct (type of task) source judgments. By way of reminder, perceived difficulty ratings were only collected for test words classified as “old,” because the ratings focused on the perceived difficulty of the source decisions. Because the incidence of false positives was so low, and perceived difficulty ratings were only collected for “yes” responses, there were an insufficient number of perceived difficulty ratings for new items to analyze. The perceived difficulty ratings for correct source judgments are reported in Table 3 as a function of type of task (solve or construct) and actual anagram difficulty (easy or

hard). An ANOVA revealed two significant effects. There was a main effect of type of task [$F(1,31) = 6.36$, $MS_e = 2.98$, $p = .02$], with decisions about words generated as solutions to anagrams perceived to be more difficult ($M = 4.61$) than decisions about words used to construct anagrams ($M = 3.82$). There was also a main effect of anagram difficulty [$F(1,31) = 11.66$, $MS_e = 1.33$, $p = .002$], with decisions about words associated with difficult anagrams perceived as more difficult ($M = 5.00$) than decisions about words associated with easy anagrams ($M = 3.42$). The interaction was not significant.

The effects of the anagram difficulty manipulation were similar for source accuracy and perceived difficulty ratings, in that source accuracy was better for words in the difficult anagram condition, and participants perceived their decisions about words associated with difficult anagrams to be more difficult than decisions about words associated with easy anagrams. We return to the importance of these aspects of our findings later in the article, after considering other metacognitive indices.

EXPERIMENT 2

In Experiment 2, we looked to see whether the source memory advantage for words associated with difficult anagrams is maintained when attention is directed to a property of the anagrams themselves. Rather than being asked to make type of task judgments, participants were asked to judge whether the word presented at test had originally been associated with an easy or difficult anagram rule. The participants were instructed to base their anagram difficulty judgments on the rules for defining anagrams as easy or difficult rather than on subjective impressions of difficulty. In Experiment 2, we also began to explore the basis for participants’ decisions, asking participants to describe what came to mind when making their anagram difficulty judgments.

Central tenets of the source-monitoring framework support the prediction that variations in the wording of source tests should influence source judgments because these variations in wording are thought to affect the characteristics of memory traces that are consulted, or the activation of beliefs about what affects one’s memory, or both (Foley et al., 2002; Johnson et al., 1981; R. L. Marsh & Hicks, 1998). In the present series of studies, shifting focus from the kind of operations initiated during encoding to properties of anagram materials may alter the effects of the anagram difficulty manipulation. Asking participants to make judgments about the kind of anagrams solved or generated may encourage thoughts about the pattern of letters, or thoughts about general features of cognitive operations information, or both, thus reducing the likelihood of observing an advantage for words associated with difficult anagrams. Along similar lines, directing attention to properties of the anagrams may be less likely to reveal an advantage for words associated with difficult anagrams because a finer level of analysis is required (e.g., number of letters switched) than that required by focusing on different types of tasks (e.g., switching letters in words or in anagrams).

Table 2
Proportion of New Items Identified As “Old” for Experiments 1–4 As a Function of Incorrect Source Judgment

	<i>M</i>	<i>SE</i>
Experiment 1: New item reported as		
Solved	.10	.01
Constructed	.11	.01
Experiment 2: New item reported as		
Easy anagram	.09	.01
Difficult anagram	.05	.01
Experiment 3: New item reported as		
Easy anagram	.18	.02
Difficult anagram	.14	.02
Experiment 4: New item reported as		
Easy anagram	.04	.01
Difficult anagram	.04	.01
Don’t know	.06	.01

Table 3
Average Perceived Difficulty Ratings for Correct Type of Task Judgment As a Function of Type of Task and Anagram Rule Difficulty in Experiment 1

Task	Anagram Rule Difficulty			
	Easy		Difficult	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Solve	3.66	.31	5.56	.28
Construct	3.19	.29	4.45	.32

Method

Participants. Seventy undergraduates participated in this experiment as part of their course requirement for a psychology course.

Materials and Procedure. Experiment 2 replicated Experiment 1 in that both type of task (solve or construct) and anagram difficulty (easy or difficult) were repeated measures variables. Participants solved and constructed easy or difficult anagrams (10 within each item category).² On each source test trial, the participants were asked to decide whether the test word was “new,” “easy” (i.e., associated with an easy anagram rule), or “difficult” (i.e., associated with a difficult anagram rule). After the anagram difficulty source test was completed, the participants were asked to comment on the way(s) in which they made these source decisions.³

Results and Discussion

Anagram difficulty judgment. The source-monitoring task was a judgment about the type of anagram (easy or difficult) associated with each test word. Source-monitoring accuracy was the proportion of correct anagram difficulty judgments to old items divided by the number of hits (responding “old” to old items). Table 4 reports source accuracy as a function of type of task (solve or construct) and anagram difficulty (easy or difficult). An ANOVA revealed a significant main effect of anagram difficulty [$F(1,69) = 62.85$, $MS_e = .07$, $p < .001$]. Anagram difficulty judgments were more accurate for words associated with easy anagrams ($M = .76$) than for words associated with difficult anagrams ($M = .51$). No other effects were significant. One possible explanation for the lower source accuracy for words associated with difficult anagrams is that these words were less memorable, perhaps because the greater focus on creating (or solving) the difficult anagrams led to a decrease in attention to the words themselves. However, as the next analysis of recognition performance indicates, this was not the case.

Item memory. The proportion of hits (calling old items “old” regardless of whether the anagram difficulty judgment was correct) is shown in Table 5. An ANOVA with type of task (solve or construct) and anagram difficulty (easy or difficult) as variables showed two significant main effects. There was a main effect of anagram difficulty [$F(1,69) = 36.41$, $MS_e = .03$, $p < .001$], such that item memory was better for words associated with difficult anagrams ($M = .69$) than for words associated with easy anagrams ($M = .56$). The anagram difficulty \times type of task interaction was significant [$F(1,69) = 7.60$, $MS_e = .03$, $p < .007$]. Post hoc tests confirmed that when solving anagrams, the participants recognized words associated with difficult anagrams better than they recog-

nized words associated with easy ones. However, when constructing anagrams, item memory did not differ for the easy and difficult anagrams.

False positives. The proportion of new items called “old” (false positives) was computed as the proportion of new items to which an incorrect “easy” or “difficult” response was made during the anagram difficulty judgment task. The mean proportions are reported in Table 2. Although the level of false positives was quite low, an ANOVA showed that when responding incorrectly to a new item, the participants were more likely to report “easy” than “difficult” [$F(1,69) = 16.15$, $MS_e = .003$, $p < .001$].

Metamemory remarks. After completing the source judgment task, the participants were asked to report on the way(s) in which they made their judgments. Drawing on coding schemes developed from the source-monitoring framework (Foley, Santini, & Sopasakis, 1989; Johnson, Foley, Suengas, & Raye, 1988), we created five categories for classifying responses: explicit references to properties of the materials (e.g., “remembering how the pattern [of letters] looked”), explicit references to emotional reactions (e.g., “remembering how frustrated I felt,” or “feeling proud I got it”), explicit references to cognitive operations (e.g., remembering how long it took to solve for a word, remembering strategies used to solve or construct an item, or references to strategies used at test), miscellaneous remarks, and “don’t know” responses. An example of a strategy used at test is deciding that the item “must have been an anagram because I don’t remember solving that one.” Two research assistants, neither of whom had knowledge about the purpose of the study, independently coded participants’ remarks, and the assistants’ classifications were quite similar (94% agreement).

Table 6 reports the percentage of participants who mentioned one or more of these kinds of remarks when commenting on the way in which they rendered their source decisions. As shown in the table, several partici-

Table 4
Proportion of Correct Anagram Difficulty Judgments (Source Accuracy) for Items Described As “Old” in Experiments 2, 3, and 4 As a Function of Type of Task and Anagram Rule Difficulty

Anagram Rule Difficulty	Type of Task			
	Solve		Construct	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Repeated Measure				
Experiment 2				
Easy	.80	.02	.72	.03
Difficult	.49	.04	.52	.03
Between-Groups Design				
Experiment 3				
Easy	.67	.05	.66	.05
Difficult	.44	.07	.55	.05
Experiment 4				
Easy	–	–	.59	.04
Difficult	–	–	.61	.04

Table 5
Proportion of Old Items Identified As “Old” in Experiments 2, 3, and 4 As a Function of Type of Task and Anagram Rule Difficulty

Anagram Rule Difficulty	Type of Task			
	Solve		Construct	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Repeated Measure				
Experiment 2				
Easy	.65	.02	.46	.03
Difficult	.84	.02	.54	.02
Between-Groups Design				
Experiment 3				
Easy	.74	.01	.75	.02
Difficult	.74	.02	.78	.02
Experiment 4				
Easy	–	–	.66	.05
Difficult	–	–	.76	.03

pants reported trying to remember the way the materials appeared in the encoding booklets or the kinds of operations initiated during encoding or test or both. Consistent with earlier source-monitoring studies showing that participants report thinking about cues associated with cognitive operations information when asked to remember who generated materials (e.g., Johnson et al., 1981), these metamemory remarks indicate that when source judgments focus on some aspect of anagram solving, strategies do indeed come to mind at test, and participants draw on these activations to render their decisions.

In Experiment 2, the source judgment instructions were intended to direct attention to the anagram materials themselves. However, the anagram difficulty judgments also could have been based on the type of task (solve or construct). When thinking about whether a word was associated with an easy or difficult anagram, for example, if a participant tried to remember how he or she worked on the item, perhaps remembering how the solution was “discovered,” then the participant could draw on cues about the cognitive tasks to arrive at the rule-based anagram classification. Indirect support for this possibility comes from the frequency with which explicit references were made to cognitive operations information in the metamemory responses.

The small but significant difference in the false positive rates, with more “easy” than “difficult” errors in response to new items, could reflect a similar strategy. If participants were trying to draw on cognitive operations cues when rendering decisions, if a new item seemed familiar but participants could not remember any information about the way they manipulated the letters in the test item, they may have reported “easy,” leading to the small but greater incidence of “easy” errors in comparison with “difficult” errors in the false positive data. To eliminate this cue to judgments, in Experiment 3 the cognitive task manipulation was a between-groups variable. Thus, as in Experiment 2, participants experienced two types of anagrams (easy or difficult ones) *but* they only initiated one type of task—solving or constructing anagrams.

EXPERIMENT 3

Method

Participants. Twenty-eight undergraduates participated in this study, each receiving \$3.00 for his or her participation.

Materials and Procedure. With type of task a between-groups variable, participants either solved *or* created anagrams, but experienced both types of anagrams (easy and difficult). Materials were counterbalanced, as in the other experiments reported in this series, so that each word appeared equally often across the anagram versions (e.g., easy anagram or difficult anagram). The number of participants receiving each of the counterbalanced sets was approximately equal ($n_s = 9, 9, \text{ and } 10$). On the surprise anagram difficulty judgment task, the participants were asked to remember whether each item was associated with an easy or difficult anagram.

Results and Discussion

Anagram difficulty judgment. The source-monitoring task was a judgment about the anagram (rule-based) difficulty associated with each test word. Source-monitoring accuracy was the proportion of correct anagram difficulty judgments to old items divided by the number of hits (responding “old” to old items). Table 4 reports source accuracy as a function of type of task (solve vs. construct) and anagram rule difficulty (easy or difficult). An ANOVA revealed a main effect of anagram difficulty [$F(1,26) = 4.41, MS_e = .17, p = .04$], with better source accuracy for words associated with easy anagrams ($M = .67$) than for words associated with difficult ones ($M = .49$). No other effects were significant.

Item memory. Item memory was computed as the proportion of old items to which an easy or difficult anagram choice was made. The means are reported in Table 5 as a function of type of task (anagram or solve) and anagram rule difficulty (easy or difficult). The ANOVA on proportion of hits revealed no significant effects.

False positives. The proportion of new items called “old” (false positives) was computed as the proportion of new items to which an incorrect “easy” or “difficult” response was made. The false positive rates, reported in Table 2, were quite low. The ANOVA including incorrect choice as a variable revealed no significant effects.

The results of the first three experiments in this series suggest that the advantage for solutions to difficult anagrams will be observed when source tasks direct attention to the cognitive operations activated during encoding. When the focus shifts to the anagram materials themselves, however, no such advantage is observed.

Table 6
Percentage of Participants Reporting Thinking About Properties of Anagrams, Cognitive Operations Cues, and/or Emotional Reactions When Rendering Type of Task Judgments in Experiments 2 and 4

Response Category	Percentage of Participants	
	Experiment 2	Experiment 4
Properties of materials	35	12
Cognitive operations cues	37	83
Emotional reactions	2	–
Miscellaneous remarks	1	1
Don't know	25	4

The higher source accuracy scores for words associated with easy anagrams in comparison with difficult ones was observed, regardless of whether the experiment had a within-subjects (Experiment 2) or between-subjects (Experiment 3) design. Before discussing further the importance of the differential effects of the anagram difficulty manipulation on the two source tasks investigated in Experiments 1–3, we consider why source accuracy was higher for words associated with easy anagrams than for words associated with difficult ones (Experiments 2 and 3). This advantage for easy anagrams might reflect a response bias to report “easy,” when in doubt, a tendency that could inflate accuracy scores for words associated with easy anagrams. Easy anagrams may essentially serve as repetition of target words because of their close resemblance to those words. In the last experiment reported in this series, we consider the possible role of a response bias to report “easy” by introducing at test the opportunity to report “don’t know.”

EXPERIMENT 4

With a three-choice source test like the ones included in Experiments 2 and 3 (i.e., “easy anagram,” “difficult anagram,” or “new” as the response options), disambiguating source accuracy from the operation of a general response bias is not possible (Murnane & Bayen, 1996). Unless participants are given the opportunity to report uncertainty by way of a fourth response option (i.e., “don’t know”), it is difficult to know whether the source advantage for easy items reflects better source accuracy or a bias to report “easy” when in doubt. This possibility is of greatest concern when the differences in item memory parallel those observed for source accuracy (unlike the patterns observed in Experiments 2 and 3). Nevertheless, in Experiment 4, a four-choice source test was introduced to address the possibility that a response of “easy” was a default response when participants were in doubt about anagram type. Participants were again asked to make anagram difficulty source judgments, but they were provided with a response option to express uncertainty. If they thought a test word *was* included in the encoding phase, but they could not decide whether the word was associated with an easy or difficult anagram, they were given the opportunity to report “don’t know.”

To discourage guessing, the option to report “don’t know” is sometimes included on other memory tests, including those used to study generation effects (e.g., Gardiner, 2000; Gardiner et al., 1999). In these other contexts, as well as in Experiment 4, participants were encouraged to use this option unless they felt their source decisions about particular words were based on experiences of remembering (e.g., some specific feature of the encoding encounter) or experiences of knowing (e.g., that they felt confident about the type of anagram associated with the test word). Thus, the “don’t know” response would serve as a default response when participants were unsure about how to classify words correctly recognized as part of the encoding series. If the source advantage for easy anagrams reported in Experiments 2 and 3 is an indication

of a tendency to report “easy” when in doubt about the anagram associated with words—perhaps because, under these conditions, participants are most likely to select the more word-like options—then when the option to report “don’t know” is provided, the source advantage for easy anagrams over difficult ones should be eliminated (or at least reduced).

Method

Participants. Twenty-four undergraduates volunteered to participate, each receiving credit toward their course requirement for an introductory psychology course. Males and females were represented proportionally.

Materials and Procedure. Participants only constructed anagrams. The materials and first phase of the procedure were identical to those followed for the anagram construction condition for previous experiments reported in this series. The participants were given the exact same instructions described earlier for creating easy and difficult anagrams. Two participants were replaced because they did not follow instructions, varying the way in which the anagrams were started.

After a 5-min retention interval, during which time participants were asked to count backward by 3s from 717, they were surprised by a source-monitoring test. The test consisted of 80 items (40 targets and 40 distractors). The source test, a one-stage test, now included four response options on each test trial. The participants were asked to decide whether each test item was “new,” seen during encoding and associated with an “easy anagram,” seen during encoding and associated with a “difficult anagram,” or seen during encoding but the participant could not remember whether easy or difficult (reporting “don’t know”). After completing the source test, the participants were asked to report on how they decided whether they encountered easy or difficult anagrams.

Results and Discussion

Anagram difficulty judgments. The source-monitoring task was a judgment about the anagram difficulty associated with each word. Source-monitoring accuracy was the proportion of correct anagram difficulty judgments to old items divided by the number of hits. When calculating source accuracy scores in Experiment 4, the denominator included three responses to old items—“easy,” “difficult,” and “don’t know.” Source accuracy scores are reported in Table 5 as a function of anagram difficulty. A one-way ANOVA showed no significant difference in the source accuracy scores for easy and difficult anagrams.

Item memory. The proportion of hits (calling old items “old” regardless of whether anagram difficulty judgment was correct) is shown in Table 5. An ANOVA including anagram difficulty as a variable revealed a main effect [$F(1,19) = 13.74, MS_e = .01, p = .001$], with better item memory for words associated with difficult anagrams than for words associated with easy ones.⁴

The frequency with which the “don’t know” option was used when responding to target words associated with easy and difficult anagrams was also analyzed. In a repeated measures ANOVA, there was no difference in the frequency with which this response option was selected. The mean proportions were .10 ($SE = .04$) and .12 ($SE = .05$) for words associated with easy and difficult anagrams, respectively [$F(1,19) = .89$].

False positives. The proportion of new items called “old” (false positives) was computed as the proportion of new items to which an incorrect “easy,” “difficult,”

or “don’t know” response was made. The mean proportions are reported in Table 3. As in the other experiments reported, the level of false positives was quite low. A repeated measures ANOVA including type of incorrect choice as a variable revealed no significant difference in the incorrect use of the response alternatives.

Metamemory responses. After the source test was completed, the participants were asked to describe the basis for their judgments. The same coding scheme reported for Experiment 2 was used. The participants’ metamemory remarks again suggested that when making anagram difficulty judgments, they drew on information about the appearance of the letter (or word) strings and/or information about the cognitive operations initiated when working on the materials (e.g., reporting thinking about ways they scrambled words, time spent deciding how to scramble words) when trying to decide whether they constructed an easy or difficult anagram for words recognized as targets. Table 6 reports the proportion of participants who made reference to properties of the materials, emotional reactions, cognitive operations cues, and/or miscellaneous remarks. Some participants reported thinking about features of the letter strings—for example, reporting thoughts about the number of letters in the words. Several reported thinking about the ways in which they constructed the anagrams (e.g., “if I remembered switching only one letter, I knew I made an easy one”) or the ways in which they rearranged letters while trying to solve anagrams. These themes were evident in the metamemory remarks reported for Experiment 2, and suggest that participants consider information about cognitive operations even when cues to different kinds of tasks (solve or construct) are not available (Experiment 4).

The absence of a source accuracy advantage for words in the easy anagram condition following the four-choice source judgments suggests that a bias to report “easy” when in doubt about items recognized as old may have contributed to the source advantage observed in Experiments 3 and 4. To determine the basis for this kind of bias in response to old items, future work might include anagrams at test to see if participants are more likely to pick easy anagrams over difficult ones, and if they even notice that the easy versions are anagrams and not words.

GENERAL DISCUSSION

Our source accuracy findings are consistent with theories that emphasize the role of cognitive operations in both item memory (the procedural reinstatement view, McNamara & Healy, 2000) and source memory (the source-monitoring framework, Foley et al., 2006; Johnson et al., 1993). As predicted, source-monitoring tests directing attention to the operations guiding anagram manipulations by asking participants to remember the way in which they operated on anagram materials (type of task judgment, Experiment 1) lead to an advantage in memory for words associated with difficult anagrams over those associated with easy ones. A similar advantage is observed when participants are asked to remember who generated

solutions to anagrams (self or partner, Foley et al., 2005). In contrast, when source tests ask participants to make rule-based anagram difficulty decisions, this advantage was not evident (Experiments 2–4).

These differential effects of the difficulty manipulation on source tests are consistent with other findings that point to the sensitivity of source judgments to the wording of test instructions. Within the context of remember/know judgments, an advantage is observed for difficult anagrams in comparison with easy ones (Dewhurst & Hitch, 1999). Similarly, R. L. Marsh and Hicks (1998) showed that source accuracy judgments vary depending on whether participants’ attention is directed to acts of generating (e.g., “did you generate this item or not?”) or reading (e.g., “did you read this item or not?”). Within the source-monitoring framework, variations in the wording of source tests are expected to influence accuracy by affecting the characteristics of memory traces that are consulted when rendering these judgments (e.g., Foley et al., 2002; Johnson et al., 1993; Johnson et al., 1981; R. L. Marsh & Hicks, 1998).

Consistent with this reasoning, intriguing patterns emerge when we integrate our findings with previous generation effect studies involving letter transformation tasks. As we mentioned earlier, when participants are asked to identify the letters that they transformed (rather than the nonwords they generated), a generation effect was observed for nonwords. This finding was originally reported as evidence that retention tests that are sensitive to what participants generate (the outcomes) will reveal the effects of generating (e.g., Nairne & Widner, 1987, Experiment 1). From our perspective, however, the report of an advantage for generating nonwords could be interpreted in another way. When retention tests focus attention on the operations themselves, by asking participants to remember the letters they switched (Nairne & Widner, 1987), to remember the kind of operation they performed (Experiment 1, solving or constructing anagrams), or to repeat the operations (Nairne & Widner, 1987, Experiment 2), an advantage for difficult cognitive tasks may indeed be observed. In contrast, when source tests require a finer level of discrimination (Experiments 2–4), perhaps based on relative amounts of one kind of operation associated with words, a similar advantage for difficult cognitive tasks may not be observed.

The new studies reported in this series, along with others recently reported (Foley et al., 2006; McNamara & Healy, 2000) invite further refinement, rather than rejection, of the cognitive operations explanation for the effects of difficulty manipulations. More often than not, previous tests of predictions about the effects of difficulty manipulations have focused on the possible role of effort or difficulty, independent of the kind of operations giving rise to that effort (e.g., generating related words, solving anagrams, transposing letters to create nonwords and anagrams, performing numerical operations). The likelihood of observing memory advantages for difficult versions of cognitive tasks relative to simpler ones seems to depend on the extent to which memory tests reinstate cognitive

operations initiated during encoding, or, at least, direct attention to thoughts about those cognitive operations. Future research might examine more closely the nature of this relationship so as to specify more fully the role of cognitive operations information in memory.

AUTHOR NOTE

This research was supported by NSF/REU Grant SBR-9619404. We thank Colleen Barber, Kerry Purtell, and Tracy Murrin-von Ebers for their help with data collection and analyses. Correspondence concerning this article should be addressed to M. A. Foley, Department of Psychology, Skidmore College, Saratoga Springs, NY 12866 (e-mail: mfoley@skidmore.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

REFERENCES

- BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 197-215.
- BLAXTON, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 657-668.
- DEWHURST, S. A., & HITCH, G. J. (1999). Cognitive effort and recollective experience in recognition memory. *Memory*, *7*, 129-146.
- DEWINSTANLEY, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review*, *2*, 538-541.
- FINKE, R. A., JOHNSON, M. K., & SHYI, G. C. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, *16*, 133-137.
- FOLEY, M. A., DURSO, F. T., WILDER, A., & FRIEDMAN, R. (1991). Developmental comparisons of explicit versus implicit imagery and reality monitoring. *Journal of Experimental Child Psychology*, *51*, 1-13.
- FOLEY, M. A., FOLEY, H. J., DURLEY, J., & MAITNER, A. (2006). Anticipating partners' responses: Examining item and source memory following interactive exchanges. *Memory & Cognition*, *34*, 1539-1547.
- FOLEY, M. A., FOLEY, H. J., & KORENMAN, L. M. (2002). Adapting a memory framework (source monitoring) to the study of closure processes. *Memory & Cognition*, *30*, 412-422.
- FOLEY, M. A., FOLEY, H. J., WILDER, A., & RUSCH, L. (1989). Anagram solving: Does effort have an effect? *Memory & Cognition*, *17*, 755-758.
- FOLEY, M. A., JOHNSON, M. K., & RAYE, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development*, *54*, 51-60.
- FOLEY, M. A., & RATNER, H. H. (1998). Children's recoding memory for collaboration: A way of learning from others. *Cognitive Development*, *13*, 91-108.
- FOLEY, M. A., SANTINI, C., & SOPASAKIS, M. (1989). Discriminating between memories: Evidence for children's spontaneous elaborations. *Journal of Experimental Child Psychology*, *48*, 146-169.
- GARDINER, J. M. (2000). On the objectivity of subjective experiences of autothetic and noetic consciousness. In E. Tulving (Ed.), *Memory, consciousness, and the brain: The Tallinn Conference* (pp. 159-172). Philadelphia: Psychology Press.
- GARDINER, J. M., RAMPONI, C., & RICHARDSON-KLAVEHN, A. (1999). Response deadline and subjective awareness in recognition memory. *Consciousness & Cognition*, *8*, 484-496.
- GLISKY, E. L., & RABINOWITZ, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 193-205.
- GREENE, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale, NJ: Erlbaum.
- HICKS, J. L., & MARSH, R. L. (2001). False recognition occurs more frequently during source identification than during old-new recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 375-383.
- HIRSHMAN, E., & BJORK, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 484-494.
- JACOBY, L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, *30*, 513-541.
- JOHNS, E. E., & SWANSON, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 180-190.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*, 371-376.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.
- JOHNSON, M. K., & RAYE, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 35-86). Cambridge, MA: Harvard University Press.
- JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, *94*, 37-64.
- KELLEY, M. R., & NAIRNE, J. S. (2001). von Restorff revisited: Isolation, generation, and memory for order. *Memory & Cognition*, *27*, 54-66.
- KINJO, H., & SNODGRASS, J. G. (2000). Does the generation effect occur for pictures? *American Journal of Psychology*, *113*, 95-121.
- KINOSHITA, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, *17*, 563-571.
- LINDSAY, D. S., & JOHNSON, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society*, *29*, 203-205.
- MARSH, E. J., EDELMAN, G., & BOWER, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition*, *29*, 798-805.
- MARSH, R. L., & BOWER, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 673-688.
- MARSH, R. L., & HICKS, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1137-1151.
- MAYZNER, M. S., & TRESSELT, M. E. (1958). Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology*, *56*, 376-379.
- MCDANIEL, M. A., EINSTEIN, G. O., & LOLLIS, T. (1988). Qualitative and quantitative considerations in encoding difficulty effects. *Memory & Cognition*, *16*, 8-14.
- MCLEROY, L. A., & SLAMECKA, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning & Verbal Behavior*, *21*, 249-259.
- MENAMARA, D. S., & HEALY, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory & Language*, *34*, 399-416.
- MENAMARA, D. S., & HEALY, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory & Language*, *43*, 652-679.
- MULLIGAN, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 436-450.
- MULLIGAN, N. W. (2002). The emergent generation effect and hypermnesia: Influences of semantic and nonsemantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 541-554.
- MULLIGAN, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 838-855.
- MURNANE, K., & BAYEN, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition*, *24*, 417-428.
- NAIRNE, J. S., PUSEN, C., & WIDNER, R. L. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, *13*, 183-191.
- NAIRNE, J. S., & WIDNER, R. L. (1987). Generation effects with non-

- words: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 164-171.
- RABINOWITZ, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology & Aging*, **4**, 259-268.
- ROSS, H. S., & BALZER, R. H. (1975). Determinants and consequences of children's questions. *Child Development*, **46**, 536-539.
- ROSS, H. S., & KILLEY, J. C. (1977). The effect of questioning on retention. *Child Development*, **48**, 312-314.
- SCHMIDT, S. R., & CHERRY, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition*, **17**, 359-369.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.
- SRINIVAS, K., & ROEDIGER, H. L., III (1990). Classifying implicit memory tests: Category association and anagram solution. *Journal of Memory & Language*, **29**, 389-412.
- STEFFENS, M. C., & ERDFELDER, E. (1998). Determinants of positive and negative generation effects in free recall. *Quarterly Journal of Experimental Psychology*, **51A**, 705-733.
- TACONNAT, L., & ISINGRINI, M. (2004). Cognitive operations in the generation effect on a recall test: Role of aging and divided attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 827-837.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College Press.
- TYLER, S. W., HERTEL, P. T., MCCALLUM, M. C., & ELLIS, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 607-617.
- ZACKS, R. T., HASHER, L., SANFT, H., & ROSE, K. C. (1983). Encoding effort and recall: A cautionary note. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 747-756.
- construct) and anagram difficulty (easy or difficult) as variables showed one main effect. Item memory was better in the solve condition ($M = .88$) than in the construct condition ($M = .64$) [$F(1,31) = 38.88$, $MS_e = .05$, $p < .001$]. The anagram difficulty effect was not significant; the means were .75 and .77 for words in the easy and difficult anagram conditions, respectively. Thus, the results fully replicated the item memory effects observed for Experiment 1.
2. To verify the classification of the anagrams created by the undergraduates as easy or difficult, we included two manipulation checks. Independent judges, who were not informed about the purpose of the study, were asked to classify the anagrams generated by the undergraduates as either easy or difficult to solve. These judges were given the same rules that the undergraduates followed when generating the anagrams. In every case, the judges' classifications (as easy vs. difficult) matched those intended by the participants when using the rules to create easy or difficult anagrams. An independent group of participants ($N = 14$) solved the anagrams created in Experiment 2, and the time to solve each anagram was measured. On average, the time to solve the difficult anagrams was significantly longer than the time to solve easy ones; the means were 12.26 sec and 5.67, respectively. Thus, both of these indices support the classifications of the anagrams as easy or difficult to solve.
3. Experiment 2 included a between-groups manipulation (instruction booklet) not reported in the text. The booklets used during the encoding phase varied in the detail of the prompts for the trials involving anagram solving. In one version, when participants were asked to solve an anagram, no reference was made to the designation of that anagram. In a second version, however, anagram type accompanied the solve prompts (i.e., solve easy anagram, solve difficult anagram). This wording had no effect on source or item memory. The analyses reported in the text therefore collapse across this variable.
4. Source accuracy and proportion of hits were computed in a second way, omitting the "don't know" responses from the denominators of each measure. Although source accuracy performance was relatively lower using these calculations (on average, 60%), the general pattern of findings reported for Experiment 4 was replicated when the correction factor was excluded.

NOTES

1. Experiment 1 was replicated using the same materials and general procedure and testing a new group of participants ($N = 32$). In this replication, item memory was assessed by use of a standard recognition test without a source judgment task. An ANOVA with type of task (solve or

(Manuscript received January 15, 2004;
revision accepted for publication October 22, 2005.)