Psychology Faculty Publications   Psychology Department

2004

Erroneous memories arising from repeated attempts to remember

Linda A. Henkel
Fairfield University, lhenkel@fairfield.edu

Peer Reviewed

Repository Citation

Published Citation

Running Head: ERRONEOUS MEMORIES & REPEATED RECOLLECTION

Erroneous Memories Arising From Repeated Attempts to Remember

Linda A. Henkel

Fairfield University

Address correspondence to:
    Linda Henkel
    Department of Psychology
    Fairfield University
    North Benson Road
    Fairfield, CT 06430
    Email: lhenkel@mail.fairfield.edu
    Phone: 203-254-4000 x3269
    Fax: 203-254-4122

Abstract

The impact of repeated and prolonged attempts at remembering on false memory rates was assessed in 3 experiments. Participants saw and imagined pictures and then made repeated recall attempts before taking a source memory test. Although the number of items recalled increased with repeated tests, the net gains were associated with more source errors (i.e., calling imagined items "seen"). Source errors were particularly marked when earlier memory tasks involved remembering without consideration of the source of the remembered items, and occurred for both semantically related and physically similar items. Source misattributions occurred on both forced and free recall tests, thus were not merely due to guessing or changes in response criteria. Similar results were found when a prolonged recall period was used instead of separate memory tests. Hence, source confusions can increase when people reflect on their memories repeatedly or over prolonged periods of time without carefully considering their source.

KEYWORDS: False memory; False recall; Source monitoring; Hypermnesia; Repeated tests

Erroneous Memories Arising From Repeated Attempts to Remember

In many situations, people are called upon to report their memory of an event a number of times, such as an eyewitness to a crime reporting details to a police officer, to other investigators, lawyers, and friends and family. With repeated probing and prolonged searches of memory, as when someone repeatedly tries to add more information to his or her account, certain aspects of memory can be strengthened and reinforced and other aspects may be lost, degraded, or blended with other similar experiences. Thus although additional information may be recalled with repeated memory tests, such added information may not necessarily be accurate. Roediger, McDermott, and Goff (1997) refer to this as the "paradoxical effects of repeated testing," that is, that repeated retrieval or testing can enhance accurate retention while sometimes leading to an increase in intrusions and other memory errors.

The present studies examine the influence of repeated testing on people's ability to subsequently determine whether something they remember was really witnessed or was imagined, because remembering *how* something happened can often times be as important as remembering *that* it happened. Source memory is not perfect – the source or origin of a remembered event or piece of information is sometimes misattributed (e.g., Johnson, Hashtroudi, & Lindsay, 1993). For example, people confuse information that they heard with information that they read, confuse stories they were told and photos they have seen with their actual experiences, and confuse statements made by one person with statements made by another. In addition, people confuse internally generated events, such as those derived from their thoughts, feelings, inferences, and imagination, with externally experienced events, such as those involving perception. Source monitoring failures range from the extreme, as in cases in which people fabricate complex, bizarre, and implausible events that did not happen to them (see, e.g., Loftus & Ketcham, 1994), to the more commonplace,

as when people claim to have said something that they only thought of saying (e.g., "I'll be late for dinner").

The source monitoring framework of Johnson and colleagues offers a model of memory processes that accounts for such source confusions (Johnson & Raye, 1981; Johnson et al., 1993; Mitchell & Johnson, 2000). According to this model, determining the source of a remembered event is based on the nature of the memory representation as well as on the judgment processes involved in the act of remembering. These judgment processes include both strategic judgments and heuristic judgments. Strategic judgment processes rely on reasoning, such as considering the plausibility of an event in light of general knowledge, whereas heuristic judgment processes are more automatic and rely on general differences in memories derived from particular types of sources. Both types of judgment processes can be influenced by the various qualitative features of memories (e.g., the amount or quality of perceptual detail, contextual information, and affective reactions) in a way that promotes source errors. For instance, relying on a heuristic evaluation can lead to source errors when, for example, a rapid, relatively automatic assessment of the features of a memory derived from imagination indicates that the features are more characteristic for memories derived from perception (e.g., vivid sensory detail yet little information regarding the cognitive processes involved in the creation of the imagery). By relying on strategic judgment processes, one might erroneously reason that a remembered event had really occurred because it seems too detailed to have been fabricated.

According to the source monitoring framework, the qualitative features used in source judgments may have genuinely been part of the experience itself, but they can also arise from other, similar experiences (Henkel & Franklin, 1998) and from thinking about events in particular ways after they occurred (Johnson, Nolde, & De Leonardis, 1996; Koutstaal, Schacter, Johnson, &

Galluccio, 1999; Lane, Mather, & Villa, 2001). For instance, focusing on affective rather than

factual features of events while reminiscing alters the accessibility of potentially source-specifying

features and thus can impair one's ability to determine the source of those events (Hashtroudi,

Johnson, Vnek, & Ferguson, 1994).

In many source monitoring studies, substantial rates of source errors have been found when

memory is assessed and evaluated with just one memory test. What then occurs when memory is

repeatedly probed?   As noted, on the one hand, repeated memory efforts can in fact increase the

amount of information remembered. Research on the *testing effect* and *inoculation effect* shows

enhanced memory for studied items when people successfully remembered them on a prior memory

test (e.g., Brainerd, Reyna, Howe, & Kingma, 1990; Howe, 1972; McDermott, 1996), and numerous

studies have shown that repeated attempts at recall can increase the amount of information one

remembers, an effect known as *hypermnesia* (e.g., Payne, 1987). Several theories have been posited

to account for these effects, invoking mechanisms such as rehearsal, imaginal and elaborative

processing during retrieval efforts, and the development of and practice of organizational strategies

to cue additional memories (see, e.g., Erdelyi, 1982, 1996; Mandler, 1994; Payne, 1987; Roediger &

Payne, 1982).

On the other hand, the net gains in the amount of information recalled through repeated

attempts may come with a cost – namely greater intrusions of items or events that were not

presented at all and greater confusion about the source of what is remembered. Research with both

adults and children has demonstrated the detrimental effects that repeated testing and questioning

can have on memory accuracy. For example, repeatedly prodding young children with increasingly

suggestive questions can lead to more inaccurate responses (Cassel, Roebers, & Bjorklund, 1996;

Fivush & Schwarzmueller, 1995; Poole & White, 1995). With adults, repeated attempts to recall

details of a crime were associated with recall of more correct information as well as with a small but

reliable increase in memory intrusions (Bornstein, Liebel, & Scarberry, 1998; see also Eugenio,

Buckout, Kostes, & Ellison, 1982). In general, such memory errors induced through repeated

probing would be expected to reflect systematic biases that are a part of normal cognitive processes

(see, e.g., Schacter, 1999). For instance, people would be most likely to intrude items that were in

some way consistent with the context or their expectations, as seen when prior testing for words

from a thematically organized list is associated with subsequent increased recall for presented items

but also with increased false recall of thematically related words not on the list (McDermott, 1996;

Payne, Elie, Blackwell, & Neuschatz, 1996).

The present studies examine how repeated attempts at remembering what happened may

create a fertile environment for source misattributions. Semantic similarities and associations

certainly can evoke source confusions, as in the example cited above for intrusions of thematically

related material where items internally activated through list presentation are later misjudged as

having actually been heard. However, according to the source monitoring framework, source errors

can be induced through other types of similarity as well, such as phonological similarity

(McDermott & Watson, 2001; Sommers & Lewis, 1999) and physical similarity (Henkel &

Franklin, 1998; Henkel, Johnson, & De Leonardis, 1998), though much less research has focused on

these other types, particularly with regard to repeated testing. False memories for phonologically

related items (e.g., hearing "fat, that, cab, caught, and cot" and falsely remembering having heard

"cat") may occur in part due to inaccurate source assessments, when participants inappropriately

attribute the source of high activation for the critical nonpresented item to actual stimulus

presentation (Sommers & Lewis, 1999). False memories based on physical similarity (e.g., when

one mistakenly claims to have seen a lollipop that in fact had been imagined due to having seen a

physically similar object, such as a magnifying glass) can also be explained by the source monitoring framework, which argues that features from one experience can be inadvertently used in source judgments about other similar experiences (e.g., Henkel, Franklin, & Johnson, 2000). The present studies examine the influence of both semantic and physical similarity on source confusions when memory is repeatedly probed.

In order to examine the "paradoxical effects of repeated testing" on source confusions between seen and imagined items, the current studies directly manipulate internally generated events by asking participants to imagine items, instead of relying on word lists that participants hear or read to induce activation of associated concepts (e.g., Roediger & McDermott, 1995). The purpose of the present studies was threefold: (a) to examine whether repeated attempts at recollection may not only improve the overall net quantity of information remembered, but may simultaneously reduce people's accuracy in determining the source of that remembered information, (b) to explore the influence of different types of similarity on source confusions induced through repeated testing, and (c) to determine whether source misattributions arising from repeated probes of memory can be reduced by more careful evaluation and assessment of memory during recall.

In each of three experiments, the basic design involved participants viewing and imagining viewing various objects. Some of the imagined items physically resembled seen items, some were conceptually related to seen items, and some had no physical or conceptual relation to any seen items. The nature of the memory tasks varied within and across the three experiments. Because one potential problem in examining false memories across repeated tests is that participants may adopt a less stringent criteria when reporting items on the later tests, Experiment 1 employed a forced recall procedure often used in hypermnesia research to induce participants to remember progressively

more items on successive memory tests while controlling for shifts in response criteria for recall

across the repeated tests (Erdelyi & Becker, 1974). Experiment 2 utilized a free recall procedure in

which participants were encouraged to remember as much as they could but were not obligated to

remember a set number of items on the repeated memory tests, and Experiment 3 compared

repeated recall tests with one continuous, prolonged test period.   Following the last test in each

experiment, participants were asked to determine whether each remembered item was originally

imagined or perceived.

    It was hypothesized that source errors would be more prevalent when imagined items either

physically resembled or were conceptually related to seen items relative to control items in which

there was no physical or conceptual similarity to any seen items. This would replicate previous

findings and support claims from the source monitoring framework that judgments of a memory's

source can be influenced by the presence of vivid features from similar experiences (Henkel et al.,

2000; Henkel, Johnson, & De Leonardis, 1998). Expanding on these findings with the inclusion of

repeated memory tests, the present studies investigated whether in fact repeated recall tests would

systematically increase source confusions for both conceptually related and physically similar items.

 Although past research has shown that repeated memory tests can increase intrusions rates for

semantically related items (e.g., Payne et al., 1996), this was found in a paradigm in which the

falsely remembered item had never been presented at all, whereas in the current study, the critical

item is presented but is explicitly imagined rather than implictly triggered by associated words. In

addition, no research to date has examined this directly for physically similar items. It was

hypothesized that the "paradoxical effects of repeated testing" would be demonstrated – that is, as

the amount of items recalled increases systematically across the successive memory tests, so too will

source errors, in particular, source errors arising from confusions between the features of physically similar and conceptually related items.

Previous work has shown that in paradigms similar to that used here, participants are much more likely to make source errors by falsely claiming to have seen imagined items than to falsely claim to have imagined items that were in fact seen (e.g., Henkel & Franklin, 1998; Henkel et al., 2000). This asymmetry is in keeping with the source monitoring framework (Johnson et al., 1993): Source errors for imagined items would tend to occur more frequently than for perceived items because heuristic judgment processes would capitalize on misleading features from the vivid and relatively easy to envision imaginings inherent in the study's design. Thus these studies focus on errors in which people claim to have seen items that were in fact imagined.

In addition, to determine whether source errors for physically similar and conceptually related items could be differentially influenced over repeated tests by more careful scrutiny of the remembered event, the nature of the repeated memory tests was manipulated, with the two tasks varying in their specificity: One group indicated on each of the successive memory tests the source of each item they recalled, and the other group simply had to recall the items without explicit consideration as to whether they had been seen or imagined. The source monitoring framework claims that depending on the specifics of a given memory task, features of memories may be differentially accessed and evaluated, and these features in conjunction with flexible decision criteria in evaluating them can affect source accuracy (Johnson et al., 1993). Accordingly, several studies have shown that when people are induced to more carefully evaluate the features of their memories, to consider the specific potential sources for their memories, or use more strict criteria for their memory assessments, they in fact make fewer misattributions with a single probing of

memory (e.g., Dodson & Johnson, 1993; Hekkanen & McEvoy, 2002; Henkel et al., 2000; Koutstaal, Schacter, Galluccio, & Stofer, 1999; Lindsay & Johnson, 1989; Marsh & Hicks, 1998; Multhaup & Conner, 2002). Thus it was hypothesized that participants who evaluate the source of the remembered information while they repeatedly try to recall as much as they can from the original presentation will have more accurate source judgments on the final memory test than will those who recalled the items repeatedly without explicit consideration of their source. That is, memory tasks that require more specific consideration as to the source of the remembered information should result in recruitment of more useful source-specifying attributes associated with the memory and/or stricter judgment criteria in evaluating those features, and this effect likely would be magnified with multiple tests of memory.

EXPERIMENT 1

Method

Participants

Participants consisted of 121 undergraduates from the University of North Florida (89 females, 32 males; *M* age = 20.5 years, *SD* = 4.0) who participated for extra credit or a course requirement in their psychology courses.

Materials & Design

A three-way mixed factorial design was used. All participants took three repeated memory tests, thus test period (Test 1, 2, and 3) was a within-subjects variable. The similarity of the critical imagined items to seen items was manipulated within subjects (physical, conceptual, or no similarity). The specificity of the memory task during retrieval (forced recall or forced source recall) was manipulated as a between-subjects factor, with subjects randomly assigned to take either 3

forced recall memory tests (in which they recalled as many items as they could irrespective of the

item's source) or 3 forced source recall memory tests (in which they indicated the source of each

recalled item).

Stimuli were 60 slides, half of which were for perception trials and half for imagery trials.

For the perception trials each slide showed a simple black-and-white line drawing of a common

object with its name below it. For the imagery trials just the name of an object appeared at the

bottom of the slide. There were three types of stimulus pairs, and in each pair one member was

perceived and one was imagined. The members of a pair had one of three possible relations: The

members of the pair either physically resembled each other (e.g., imagine a lollipop, see a

magnifying glass), were conceptually related to each other (e.g., imagine a comb, see a hair brush),

or had no physical or conceptual similarity (control condition; e.g., imagine a sugar cube, see a

candle).   Equal numbers of each pair type were used. A single quasi random ordering of the slides

was used for all participants, with a minimum of 10 trials between pair members. For half the pairs,

the perceived member occurred first in the order of trials, and for half, the imagined member

occurred first. Assignment of members in a pair to source (seen or imagined) and to order

(perceived member first or second) was randomly determined, with a single set of stimuli used for

all subjects.

Physical similarity was established through earlier norms in which pairs of words were rated

with regard to the physical similarity between the referent objects using a 7-point scale (1 = no

physical similarity, 7 = high physical similarity). The mean rating for pairs of items for the physical

similarity condition in the present studies was 4.5 or higher, and for control and conceptually related

pairs was 2.0 or lower. Conceptually related items were selected by consensus of trained judges,

and these consisted of items with minimal physical similarity that belonged to the same functional

category.   Norms were also previously established for the relative ease with which a given item could be visually imagined and for its physical complexity. Items did not differ on these dimensions across the three types of stimulus pairs.

Procedure

Participants were tested individually and were told that the study involved how people make judgments about seen and imagined objects. They were told that on some trials they were to look at the drawings that were presented and on other trials just the name of an object would be shown and they were to visually imagine the named item as a simple black-and-white line drawing, similar in style to the drawings they viewed. For both seen and imagined items, they were to think of a typical function of each object (e.g., Durso & Johnson, 1980). Six practice trials were then given (half for perception trials, half for imagery) to familiarize participants with the task and to reduce primacy effects for the experimental stimuli. The 60 experimental trials were then presented one at a time for 5 seconds each. During the 5-second period, participants stated aloud a typical function of the seen or imagined object. Two filler trials appeared after the 60 experimental trials to reduce recency effects for the experimental stimuli.

Participants were randomly assigned to one of the two retrieval conditions in which the specificity of the memory test was manipulated. In both conditions they were given a response sheet consisting of 40 blank lines. Participants in the forced recall condition were told to list as many items as they could remember from the slide presentation and to make a response on each of the blank lines on the response sheet. They were instructed that if they got stuck before reaching the 40th line, they should stop, think, and try to recall more items and that if nothing else came to mind, they should proceed to make an educated guess for each of the remaining blank lines. Participants in the forced source recall condition were given the same instructions with the additional task of

indicating for each recalled item whether they had seen it as a picture or had visually imagined it. This was required even for items offered as guesses. For both the forced recall and source recall conditions, participants were told they would have 7 minutes to complete the memory task and were warned when 2 minutes, 1 minute, and 30 seconds remained. The response sheet was then removed and participants were told to sit quietly for the next several minutes and think about the objects from the presentation in an effort to improve their recall. After 5 minutes, the same instructions for the memory task for that condition (forced recall or forced source recall) were repeated, with the additional instruction that any items that had been recalled on the earlier memory test could be included. Seven minutes were given for the 40-line forced memory task, after which they were told to sit quietly for one more time and try to improve their memory. After the 5-minute think period, the final recall task was administered in which all participants were told to list as many items as they could from the slide presentation, filling all 40 lines and including items recalled on the earlier tests. After the 7 minutes, participants in both conditions were asked to indicate for each recalled item how certain they were of its source using the following scale: 1 = certain it was a picture, 2 = fairly certain it was a picture, 3 = not at all certain if it was a picture or imagined, 4 = fairly certain it was imagined, 5 = certain it was imagined, X = I made this one up and am certain it was not in the slide presentation at all. Response option X was given so that participants were not compelled to make source attributions for items that they had given for the sole purpose of filling up the lines on the response sheet on the final memory test.

Results & Discussion

An alpha level of .05 was used for all statistical analyses unless otherwise noted, and all post-hoc comparisons were Newman-Keuls. In the forced recall test condition, 6 of the 61 participants (9.8%) failed to follow instructions to record at least 40 items on each of the 3 recall

tests, and in the forced source recall condition 12 of the 59 participants failed to do so (20.3%). The total number of items recalled was significantly lower for this subset of noncompliant subjects ($M = 27.8$ items) than for the compliant subjects ($M = 40$), $t(118) = 14.19$, with no difference in recall rates for the noncompliant subjects in the forced recall condition ($M = 27.4$) or the forced source recall condition ($M = 28.1$). However, the primary analyses with and without these subjects did not differ, thus analyses reported below include all participants. As expected, source misattribution rates for perceived items (i.e., pictures falsely claimed as imagined) were extremely low ($< .03$) and thus data reported for the three experiments here focus on recall and source monitoring for imagined items.

Proportion of Items Recalled Across the Repeated Memory Tests

As seen in Table 1, the proportion of imagined items recalled increased from Test 1 to Test 2 and from Test 2 to Test 3, was higher overall for participants who took repeated forced recall tests compared to repeated forced source recall tests, and was higher for physically similar and conceptually related items than for control items. The proportion of imagined items that were recalled, regardless of what source they were attributed to, was examined in a 2 x 3 x 3 mixed factorial analysis of variance (ANOVA) for specificity of memory test, type of similarity, and test period. As expected, recall improved with repeated memory tests, $F(2, 236) = 61.51$, $MSe = .01$. Post hoc comparisons for this hypermnesia effect showed significantly higher recall for the third recall test (.49) than the second recall test (.44), which in turn were significantly higher than for the first (.41). Participants recalled more items on the forced recall test (.47) than on the forced source monitoring test (.43), though this effect was only marginally significant, $F(1, 118) = 3.53$, $MSe = .09$, $p = .06$. The interaction between specificity of memory test and test period was significant, $F(2, 236) = 3.65$, $MSe = .01$, which reflected the finding that although recall was higher on the initial

forced recall tests than on the source monitoring tests, retention did not differ for the two test types on the final recall test.

In addition, recall differed as a function of type of similarity, $F(2, 236) = 45.53$, $MSe = .05$, and the similarity x test period interaction was also significant, $F(4, 472) = 2.56$, $MSe = .01$. Follow up comparisons showed that overall recall was significantly higher for items that were either conceptually similar to seen items (.51) or were physically similar to seen items (.47) relative to control items with no physical or conceptual similarity to seen items (.36), although the relative magnitude of the differences between conditions varied across the test periods.

These findings show that people's ability to recall information improves with repeated memory tests: People remembered progressively more items with each additional memory attempt. Such results are consistent with previous work showing hypermnesia effects (e.g., Payne, 1987). The similarity of the imagined items to seen items also influenced their recallability, with participants recalling roughly half of the physically similar items and half of the conceptually similar items but only slightly more than one-third of the control items that had no physical or conceptual similarity to any seen items. People may have relied on associations between the items to enhance their recall, particularly as they effortfully tried to recall more items across the repeated tests.

Also associated with improved recall was the nature of the retrieval test. On the first memory test, the proportion of imagined items recalled was higher for participants who performed their memory task without consideration of the source of the remembered items (the forced recall test) than those who considered the source of each item as they were recalled (the forced source monitoring test). Thus recalling items with a memory test that requires more specificity reduced the number of studied items reported. This may have occurred because people initially required more time during the test period to reflect on and indicate the source of the items, thus had less time to

recall additional studied items and had to fill the remaining lines of the recall tests with items that

were guesses. In addition, it may be that with the further instruction of considering the source of

each recalled item, participants employed different criteria in retrieving and evaluating the items and

did not write down studied items they were uncertain of. In either case, by the third and final

memory test, participants apparently developed strategies to override these factors, because the

proportion of imagined items recalled did not differ in accordance with the specificity of the

memory test.

Source Errors Across the Repeated Memory Tests

As shown in Figure 1, source errors increased across the three test periods, and more errors

were made for physically and conceptually similar items than for control items with no physical or

conceptual similarity. The proportion of imagined items that were falsely claimed to have been seen

was examined in a 3 x 3 ANOVA (similarity x test period) for the participants given the repeated

source monitoring tests.[1] Participants given the repeated forced recall tests made source judgments

only after the third test, not on Tests 1 and 2, and thus cannot be included in this analysis. The

ANOVA showed that source error rates were significantly influenced by type of similarity, $F(2,$

$116) = 19.35$, $MSe = .01$, with more source errors for both physically similar (.12) and conceptually

similar items (.09) than for control items (.05). Source errors increased systematically with repeated

memory tests, $F(2, 116) = 27.10$, $MSe = .01$. Significantly more source errors were made on the

third memory test (.12) than on the second (.08), and significantly more errors were made on the

second test than the first (.05). Thus there were more than twice as many false claims of having

seen imagined items on the third recall test than on the first. The interaction between test period and

similarity was significant, $F(4, 232) = 5.19$, $MSe = .01$, with more source errors on the third recall

test than on the first for both physically similar and conceptually similar items, but no significant

increase in source errors for control items across the repeated tests, which may have been due to a floor effect. For all 3 test periods, more errors were made for physically similar and conceptually related items than control items.

Source Errors on the Final Memory Test

Figure 2 illustrates the effects of the different repeated retrieval tasks on source errors during the final memory test. Noticeably more source errors were made when the earlier memory tests did not require specification of the remembered items' source (forced recall) than when such specification was required (forced source monitoring), and errors were more frequent for physically and conceptually similar items than for control items. To confirm the significance of these patterns, the proportion of total imagined items that were falsely claimed as seen (i.e., judged as 1 "certain it was a picture" or 2 "fairly certain it was a picture") was examined in a 2 x 3 (specificity of memory test x type of similarity) ANOVA. Source error rates were significantly influenced by the similarity of the imagined items to perceived items, $F(2, 236) = 13.15$, $MSe = .01$: Error rates were higher for both physically similar (.19) and conceptually related items (.17) relative to control items (.12). The main effect for specificity of memory test on source errors was also significant, $F(1, 118) = 9.59$, $MSe = .06$. Participants who had taken repeated recall tests were more likely to falsely claim they had seen imagined items (.20) than those who had taken repeated source monitoring tests (.12). The interaction between specificity of memory test and similarity was also significant, $F(2, 236) = 6.01$, $MSe = .01$. Although source errors were always higher following the repeated forced recall tests than repeated source monitoring tests, the difference was less marked for physically similar items than for conceptually related or control items.

Separate analyses were also conducted examining participants' confidence for imagined items that they recalled but did not make accurate source judgments for on the final memory test. A

one-way ANOVA conducted on confidence ratings for misattributed imagined items yielded a significant main effect, $F(3, 354) = 56.91$, $MSe = .01$. The proportion of imagined items judged as "certain" they had been seen as pictures (.13) was significantly higher than either those judged as merely "fairly certain" they had been seen as pictures (.03), those judged as "unsure" whether they had been seen as pictures or imagined (.03), or those claimed as guesses to fulfill the requirements of the forced memory test (<.01). Thus participants exhibited high confidence that the recalled imagined items had actually been seen.

In addition, separate analyses were conducted to examine source misattributions of imagined items on the final memory test as a function or whether the items had been recalled on previous tests. Results showed that on the final memory test, people were much more likely to erroneously claim imagined items had been seen when the items had been previously recalled on one of the earlier memory tests. On the final forced recall test, the proportion of imagined items falsely claimed as seen was significantly higher for items that had been previously recalled (.18) than those not previously recalled (.02), $t(60) = 8.33$, $p < .001$. Likewise on the final forced source monitoring test, the proportion of imagined items misjudged as seen was significantly higher for items previously recalled (.09) compared to those not previously recalled (.03), $t(58) = 7.71$, $p < .001$. Of the items falsely claimed as seen on the final memory test, 85% had been recalled on one of the previous two memory tests, whereas 15% had not been recalled until the final test.

These findings support the general view that repeated testing can have contrary effects on memory: While repeatedly trying to remember successively more items does promote enhanced recall, such increases come with a cost, namely greater confusion in the source of those recalled items, particularly when memory is repeatedly assessed with less stringent criteria. On the whole, these results are consistent with predictions derived from the source monitoring framework, which

emphasizes that source assessments are based both on the qualitative features of memories as well as on flexible decision criteria employed in their evaluation. This study provides evidence that not only do similar semantic features and associations among items give rise to false memories (as also found in other paradigms using lists of thematically related words, e.g., Roediger & McDermott, 1995), but that physical similarity does as well (see also Henkel & Franklin, 1998). Furthermore, both types of false memories increase across repeated memory tests and decrease with more careful consideration as to the source of the remembered item. Source errors were much more likely when the items had been previously recalled on one of the earlier memory tests than when they hadn't been recalled until the final memory test, and in general the falsely attributed imagined items were rated with high confidence as having been seen. These memory errors therefore are far from negligible.

Experiment 1 involved a forced recall procedure in which participants generated responses for a relatively large fixed number of items on each memory test. This procedure is used in hypermnesia research to hold constant response criteria and guessing biases when recalling items across the multiple memory tests. However, studies have shown that when respondents are encouraged or forced to guess on earlier memory tests, the likelihood of later recalling the erroneous information and falsely attributing it to having been actually experienced increases (Ackil & Zaragoza, 1998; Zaragoza, Payment, Ackil, Drivdahl, & Beck, 2001). Obviously, the incidence of guessing would be quite high on the forced memory tasks where participants are required to guess items when they can no longer remember any more. Participants may include on their list of remembered items ones that otherwise would not have had strong enough features to merit inclusion, and they may alter the judgment criteria used in evaluating the source of these items and the items they did remember. The increase in source errors seen then in Experiment 1 may have

been driven by the guessing inherent in the forced memory procedure. Experiment 2 thus examines

source misattributions using a free recall procedure in which participants were not forced to guess to

fill a quota of items. This will allow us to determine whether repeated attempts at recollection and

the similarity between seen and imagined items will in fact increase source confusions when

participants are not required to provide guesses to fulfill the demands of the memory task. Past

research showing increases in false memories for nonpresented but thematically related items in a

list-learning paradigm across repeated recall tests shows that intrusions of nonpresented items occur

under both forced and free recall test situations (Payne et al., 1996), though that study addressed

only one particular type of similarity, did not explictly require participants to engage in imagery, and

did not directly ask for source discriminations during memory assessment (i.e., was the item heard

or only thought about?; see Multhaup & Conner, 2002).

<div align="center">**EXPERIMENT 2**</div>

Participants

Participants consisted of 128 undergraduates from the University of North Florida (96

females, 32 males; *M* age = 20.3 years, *SD* = 4.3) who participated for extra credit or a class

requirement in their psychology courses.

Materials & Design

Stimuli were the same as in Experiment 1. A 2 x 3 x 3 factorial design was used, with

specificity of memory task (free recall, source recall) manipulated between subjects, and type of

similarity (physical, conceptual, control) and test period (Test 1, 2, and 3) manipulated within

subjects.

Procedure

The procedure was the same as Experiment 1 with the following exception regarding the

memory test. All participants were given a response sheet consisting of 40 blank lines and were told

to list as many items as they could remember from the slide presentation and to continue trying

throughout the 7-minute recall period, but unlike with the forced memory procedure used in

Experiment 1, here they were not told that they must fill all 40 lines with guesses when they

couldn't remember any more items. In the free recall conditions, they were instructed to write down

as many items as they could remember, and in the source recall conditions, they were given the

additional instruction that as they recalled each item, they were to indicate if it had been one they

saw as a picture or was one they had imagined earlier. All participants took three 7-minute memory

tests interspersed by 5-minute think periods.

<div align="center">Results & Discussion</div>

Proportion of Items Recalled Across the Repeated Memory Tests

As can be seen in Table 2, the proportion of items recalled increased across successive

memory tests, with higher recall for free recall than source recall tests, and higher recall for physical

or conceptually similar items than for control items. A 2 x 3 x 3 ANOVA confirmed the

significance of these patterns: The main effects for specificity of memory test, type of similarity, and

test period were all significant. Participants recalled more items on average on the free recall tests

(.46) than on the source recall tests (.41), $F(1, 126) = 8.79$, $MSe = .10$, and their retention was

influenced by the similarity of the imagined items to seen items, $F(2, 252) = 32.61$, $MSe = .05$, with

higher recall for both physically similar (.46) and conceptually related items (.48) than for control

items with no physical or conceptual similarity (.36). Furthermore, recall improved systematically

with repeated memory tests, $F(2, 252) = 21.87$, $MSe = .02$. Post hoc comparisons for this

hypermnesia effect showed significantly higher recall for the third memory test (.47) than the second

memory test (.43), which in turn was significantly higher than for the first test (.40). No interactions

were significant.

These findings replicated the major findings from Experiment 1 with regards to recall rates: The number of items recalled increased across successive memory tests, recall was higher when people remembered items without regard to their source compared to when they had to indicate the source for each item as they were recalled, and people recalled more imagined items that were physically similar or conceptually related to a seen item relative to control items with no such similarity. Thus similar effects are obtained using both free and forced recall procedures.

Source Errors Across the Repeated Memory Tests

Figure 3 depicts the proportion of imagined items falsely claimed as seen and reveals that more source errors were made on the final memory test than on the earlier memory tests, and more source errors were made for items that either physically resembled or were conceptually related to seen items relative to items with no physical or conceptual similarity to seen items. A 3 x 3 ANOVA (similarity x test period) for the repeated source recall test conditions confirmed that source error rates were influenced by similarity, $F(2, 134) = 14.92$, $MSe = .01$, with significantly more errors for both physically similar items (.09) and conceptually similar items (.06) than for control items (.04). In addition, source errors increased systematically with repeated memory tests, $F(2, 134) = 10.42$, $MSe = .01$. Source errors were significantly higher on the third memory test (.08) than on the second (.06), and higher on the second than on the first (.04). Thus twice as many imagined items were falsely claimed as seen on the third memory test than on the first. The interaction was not significant.

Source Errors on the Final Memory Test

Figure 4 depicts the effects of the different retrieval tasks on source accuracy for the final memory test and shows more source errors when the earlier memory tests required participants to

recall items without explicit consideration of the source, and more source errors for items that either physically resembled or were conceptually related to seen items relative to control items with no relation to seen items. The proportion of total imagined items that were falsely claimed as seen was examined in a 2 x 3 (specificity of memory test x similarity) ANOVA.   The main effect for similarity was significant, $F(2, 252) = 19.49$, $MSe = .01$, with significantly more source errors for both physically similar (.15) and conceptually related items (.11) than control items (.07). The main effect for specificity of memory test on source errors was also significant, $F(1, 126) = 13.33$, $MSe = .03$. False claims that imagined items were seen occurred more frequently when participants previously took repeated free recall tests (.15) than when they took repeated source monitoring tests (.08). The interaction was not significant.

Closer examination based of the confidence ratings on the final memory test showed that imagined items that were falsely claimed as seen were most likely to be rated as "certain" they had been seen (.09) rather than "fairly certain" (.03) or "not at all certain" (.03). People rarely indicated that the items were guesses (i.e., completely made up and not part of the stimulus presentation at all, $M < .01$). A one-way ANOVA conducted on confidence level for misattributed imagined item yielded a significant main effect, $F(3,378) = 43.37$, $MSe = .01$.

A separate analysis to examine source misattributions on the final memory test showed that significantly more imagined items were falsely judged as seen when the items had been previously recalled on one of the earlier memory tests. For participants given repeated free recall tests, source errors were higher when items had been previously recalled (.13) than when not previously recalled (.01), $t(54) = 7.66$, $p < .001$. Likewise for participants given repeated source recall tests, source errors were significantly higher for items previously recalled (.07) compared to those not previously recalled (.01), $t(67) = 7.43$, $p < .001$. Of the items falsely claimed as seen on the final memory test,

87% had been recalled on one of the previous two memory tests, whereas 13% had not been recalled until the final test.

Thus results from Experiment 2 using a free recall (nonforced) testing procedure successfully replicated the major findings of Experiment 1, which employed a forced recall test procedure. Factors found to increase the amount of information recalled also served to decrease accuracy in judgments of the source of those remembered items. The number of items recalled increased across successive memory tests, as did the rate of source errors. Recall was higher and source errors more frequent when people recalled items without regard to their source than when they had to indicate the source for each item as they recalled them. Furthermore, both the amount of items recalled and the amount of source errors made increased as a function of similarity: That is, people recalled more imagined items that were physically similar or conceptually related to seen items and also made more source errors for them. Thus similar effects are obtained using both free and forced recall procedures, and in both cases people rate with high confidence that they saw items that in fact had only been imagined.

Comparisons Across Experiments 1 and 2

Because the same stimuli and procedures were used for Experiments 1 and 2, analyses across the two studies were performed to better understand the impact on recall and source performance of the forced memory procedure used in Experiment 1 but not used in Experiment 2. Tables 1 and 2 present the proportion of imagined items recalled across the three memory test periods separately for each experiment. Overall recall for imagined items did not differ between the forced memory tasks used in Experiment 1 (.45) and the nonforced memory tasks used in Experiment 2 (.43), $F(1, 246)$ $=1.22$, $MSe = .10$. Both the forced and free recall tasks yielded similar hypermnesia effects, that is, both showed increases in recall across the repeated tests (main effect for test period: $F(2, 492) =$

10.52, $MSe = .01$), with no significant interaction between test type (forced vs. not forced) and test period, $F(2, 492) < 1$, $MSe = .01$.

Analysis of source errors on the final memory test (illustrated separately for each study in Figures 2 and 4) showed that people were significantly more likely to erroneously claim an imagined item had been seen when they had taken repeated forced memory tests (.16) than nonforced memory tests (.11), $F(1, 246) = 9.04$, $MSe = .05$. Similarly, analyses across the two experiments for the conditions in which participants indicated the source of each recalled item during all 3 memory test periods showed greater source errors for imagined items in the forced source recall conditions (.09) than the nonforced source recall conditions (.06), $F(1, 125) = 4.68$, $MSe = .03$ (see Figures 1 and 3). Examination of the interaction between the nature of the memory test (forced vs. nonforced) and test period, $F(2, 250) = 3.62$, $MSe = .01$, showed that source errors increased across the three tests for both the forced and nonforced memory tests, though the rate of increase was more marked for the forced memory procedure (.06, .08, .12) than for the nonforced memory procedure (.04, .06, .08).

Thus these comparisons across the two experiments showed similar hypermnesia effects (i.e., increases in recall across successive memory tests) for both forced and nonforced memory procedures, with no difference in amount of imagined items recalled for the two tests, consistent with previous research (Erdelyi, Finks, & Feigin-Pfau, 1989; Roediger & Payne, 1985). However, source errors on the final memory test were more prevalent when participants had engaged in forced memory tests than nonforced tests, and the rate of increase in source errors was more marked across the repeated forced memory tests than the nonforced memory tests. The very nature of the forced memory procedure encourages participants to guess more, and they may in turn adopt a looser criterion and include vaguely remembered items or ones they thought were guesses that otherwise

would not have been listed. Accordingly, people listed more items overall (imagined items, seen items, filler items, and intrusions) for the forced recall procedure ($M = 37.34$ items) than the free recall procedure ($M = 28.02$), $F(1, 220) = 133.63$, $MSe = 35.71$. One possible explanation for the increased source errors then is that such forced guessing can lead people to later falsely believe that they experienced events that were not actually experienced (see Ackil & Zaragoza, 1998). In fact, in a study directly comparing forced and free recall procedures, false recognition of items that had been intruded by individual participants on a recall test one week prior was much higher when the earlier memory task was a forced recall task than a free recall task (Roediger, Wheeler, & Rajaram, 1993). Nonetheless, although the guessing inherent in the forced recall procedure used in Experiment 1 may have contributed to subsequent source misattributions, this cannot account for all such source errors because Experiment 2 clearly shows source misattributions using a memory task that does not encourage such guessing. Moreover, intrusion rates for nonpresented items were extremely low for the free recall tests (<2 items per subject on average), and it was rare for items intruded in the forced memory tasks to be judged as seen on the final memory test (<1 intruded item on average per participant was judged as seen). Consider also the fact that source judgments were made on a scale in which participants could have indicated that they were unsure whether a recalled item had been seen or imagined and could even have indicated that the item was completely a guess. When imagined items were falsely remembered as seen, people were much more likely to claim that they confidently remembered seeing those items than to say that they were only fairly certain, not certain at all, or were guessing. Taken together, this suggests that the illusion that imagined items had been seen is not just an artifact of the forced memory testing procedure but does in fact reflect genuine distortions in memory. This would be consistent with dozens of other studies showing that people can come to believe quite confidently and with supportive recollective detail that they perceived

events that were not in fact perceived (e.g., Mitchell & Johnson, 2000).

In sum, the findings from these two experiments clearly demonstrate that source confusions can be obtained both with forced recall and free recall test procedures, and therefore are not exclusively an artifact of guessing on repeated forced recall tests, and they illustrate important principles derived from the source monitoring framework: namely that the qualitative features of both physically similar and conceptually related items can be confused across imagination and perception, and that flexible decision criteria used in evaluating those features can lead to an inflation in errors when the source for the remembered item or features is not explicitly considered during repeated memory tests.

The use of a repeated memory test paradigm allows us to determine the effect of such repetition on subsequent retention rates and memory attributions, but just as there are many situations where people are called upon to repeatedly remember their experiences, such as an eyewitness or a crime victim telling police officers, lawyers, and friends about what they remember, it is also the case that sometimes circumstances may require very extended questioning and prolonged efforts to remember information.   To what extent then are the increases in source confusions observed in Experiments 1 and 2 due to the repetition of the memory tests themselves? That is, might similar increases in source errors occur when there is an extended effort to remember as much as one can? Experiment 3 thus addresses whether the factors found to produce increased recall and decreased source accuracy in Experiments 1 and 2 would similarly influence recall and source performance when memory is assessed in an extended test period rather than on multiple memory tests.

Previous research has shown that one lengthy and prolonged recall test can yield similar levels of recall to that found when separate, repeated recall tests are administered (Payne, 1986;

Roediger & Thorpe, 1978), and thus similar findings were expected in the present study. With

regard to increases in source errors, on the one hand more source errors might be made on the

repeated tests than on the extended test because there is greater opportunity for confusion in the

source of remembered items when participants reconstruct already recalled items on subsequent

memory tests. On the other hand, much as they do in repeated tests, people may reaccess

information during the extended test (e.g., mulling over what they have already recalled, using the

items they have recalled to trigger other items in memory), in which case one would not necessarily

expect a lower rate of source errors on the extended test.

However, because recall tasks by their nature are so demanding and require sufficient

retrieval cues and strong enough memory traces to pass participants' threshold to report on the tests,

this may mask possible differences between the different types of memory tests. For this reason,

Experiment 3 included a source recognition test following the repeated or continuous recall tasks in

which participants were presented with a list of old and new items and had to indicate which had

been seen, which had been imagined, and which were new. The relatively effortful recall tasks used

in these studies may encourage participants to rely on different qualitative features and different

decision criteria in making their source assessments than if memory was assessed with a recognition

test, and thus testing the generality of the influence of manipulated variables across different

measures of memory is important.

EXPERIMENT 3

Method

Participants

Participants consisted of 124 undergraduates from the University of North Florida (99

females, 25 males; *M* age = 23.6 years, *SD* = 5.1) who participated for extra credit or a course

requirement in their psychology courses.

Materials & Design

A 2 x 2 x 3 factorial design was used, with the number of separate recall attempts (3 repeated memory tests or one continuous one) and the specificity of the memory task (free recall or source recall) as between-subjects factors, and similarity (physical, conceptual, control) as a within-subjects factor. Stimuli consisted of 60 slides similar to those used in Experiments 1 and 2, half for perception trials and half for imagery trials. Ten of the imagined items physically resembled a seen item, 10 were conceptually related to a seen item, and 10 were control items in which there was no physical or conceptual similarity to any seen items. An additional 24 stimuli were used as fillers to reduce the possibility that participants may rely somewhat on strategic source decisions based on remembering something about one of the members of a pair (e.g., "if the apple was seen, then the banana must have been imagined"), which in theory would have served to reduce source errors in Experiments 1 and 2. These filler items consisted of 6 pairs of physically similar items and 6 pairs of conceptually similar items. Three pairs in each set had both members of a given pair presented as pictures and three pairs were imagined.

Procedure

The procedure for the stimulus presentation phase was the same as in Experiments 1 and 2, with the exception that in the present study, additional stimuli were used as fillers. The 60 experimental trials were presented one at a time for 5 seconds each in random order, interspersed with the 24 filler trials.

Participants were randomly assigned to one of the four retrieval conditions. The specificity of the memory task was manipulated by requiring participants to write down the names of all the items they could recall either while indicating for each recalled item whether it had been seen or

imagined (source recall) or without any instructions regarding consideration as to the source of the recalled items (free recall). The number of separate recall attempts was manipulated by having half the participants take a series of 3 separate memory tests, and half the participants take one continuous memory test for the equivalent amount of time.

For the separate memory test conditions, each recall test lasted 7 minutes. Participants were told to write the names of as many items as they could recall during the fixed amount of time (with or without instructions to specify the source of the recalled items, depending on the condition). Warnings were given when 3 minutes and 1 minute remained. The recall sheet was then taken away and they were instructed to try again to recall as many items as they could, writing any items they had on the first test as well as any additional ones they could recall. This was repeated for the third recall test as well. They thus took three 7-minute memory tests, which lasted 21 minutes total. For the continuous memory test conditions, they were not given separate memory tests but rather took one prolonged memory test that lasted 21 minutes. They were reminded at fixed times throughout the lengthy period to draw a line beneath the last item they had written and to continue trying to recall as many items as they could (instructions were given when 7 minutes, 3 minutes, and 1 minute remained). Participants in the free recall conditions were told after the third recall test or after the conclusion of the one continuous recall test to indicate the source of each recalled item (i.e., whether it had been seen as a picture or had been imagined), whereas participants in the source recall conditions made source assessments throughout the repeated or continuous memory test. The tests did not require confidence judgments in any conditions (in contrast to Experiments 1 and 2).

After the recall test phase, all participants were given a source recognition test. They were told that instead of trying to remember the items on their own, they would be presented with a list of items and were to determine whether each of the named items had been seen as a picture, had been

imagined, or was a new item that was not included in the slide presentation at all. The 124-item

source recognition test consisted of the 84 old items (64 experimental items + 24 fillers) presented

in random order, intermixed with 40 new items. No time limit was imposed for this source

recognition test.

<div align="center">Results & Discussion</div>

Proportion of Items Recalled on the Final or Continuous Recall Test

Table 3 shows the proportion of imagined items recalled across the three test periods and on

the continuous memory test. As seen, the proportion of items recalled increased across the test

periods and was higher for the free recall than the source recall conditions. A 2 x 2 x 3 (number of

separate recall attempts x specificity of memory test x similarity) ANOVA on the proportion of

imagined items recalled on the final memory test or on the one prolonged recall test showed a

significant main effect for similarity, $F(2, 240) = 36.43$, $MSe = .02$, with significantly higher recall

for both conceptually (.58) and physically similar items (.47) relative to control items (.43). The

proportion of items recalled was higher on the prolonged memory test (.55) than on the final

repeated memory test (.43), $F(1, 120) = 33.96$, $MSe = .04$, and was higher for the free recall (.52)

than the source recall (.47) conditions, $F(1, 120) = 4.98$, $MSe = .04$. A significant interaction

between specificity of memory test and similarity was found, $F(2, 240) = 5.35$, $MSe = .02$, and

follow up comparisons revealed higher recall on the free recall than on the source recall test for both

physically similar (.49 vs. .45) and conceptually similar items (.64 vs. .53) but no significant

difference for control items (.42 vs. .43).

The proportion of imagined items recalled at least once on one of the three recall tests was

also calculated to compare with performance on the prolonged memory test (see Roediger &

Thorpe, 1978). A 2 x 2 x 3 ANOVA examining this cumulative recall as a function of type of tests

(separate or continuous), specificity of memory test, and similarity yielded the same pattern of

results as in the analysis on final recall rates, and thus for the sake of brevity is not reported here

(see Table 3 for mean levels of cumulative recall).

Source Errors on the Final or Continuous Recall Test

To compare the effects of different retrieval tasks on source errors, the proportion of total

imagined items that were falsely claimed as seen on either the final recall test or the one prolonged

recall test was examined in a 2 x 2 x 3 (number of separate retrieval attempts x specificity of

memory test x similarity) ANOVA (see Figure 5). The main effect for number of retrieval attempts

on source errors was not significant, $F(1, 120) = 2.69$, $MSe = .03$. Thus people were just as likely

to mistakenly claim an imagined item had been seen after taking repeated memory tests (.13) or an

extended, continuous memory test (.15). However, the main effect for test specificity was

significant, $F(1, 120) = 19.38$, $MSe = .03$, with substantially more source errors for the free recall

condition (.18) than the source recall conditions (.10). The main effect for similarity was significant

as well, $F(2, 240) = 9.48$, $MSe = .01$, and follow up comparisons showed that source errors were

higher for both physically similar (.14) and conceptually related items (.17) than for control items

(.11). Analysis of the significant interaction between similarity and specificity, $F(2, 240) = 3.86$,

$MSe = .01$, showed that more errors were made following the repeated recall tests than the repeated

source tests but the relative magnitude differed across the similarity types.

Similar in concept to examining cumulative recall across the memory tests rather than just

total recall on the final memory test for participants given 3 separate tests (see Roediger & Thorpe,

1978), an additional analysis was run to compare cumulative rates of source errors (i.e., the

proportion of imagined items incorrectly claimed as seen on at least one of the three separate source

recall tests) to source errors on the continuous source recall test. As in the analysis using total

source errors for the final test, no difference was found in the amount of source errors when participants took 3 separate memory tests (.14) or one continuous memory test (.11), $F(1, 61) = 1.75$, $MSe = .01$.

These results replicate those from Experiments 1 and 2: Both recall rates and source error rates were higher when the repeated memory tasks did not require specification of the recalled item's source (free recall) than when they did (source recall). In addition, people were significantly more likely to claim to have seen imagined items when the imagined items either physically resembled or were conceptually related to seen items, in comparison to control items with no physical or conceptual similarity. The inclusion of a prolonged retrieval period to compare with separate, repeated memory tests spanning the same length of time shows that these contrary effects of improved recall but decreased source accuracy are not limited to situations where separate memory tests are used. Both effects occurred to the same extent when prolonged efforts were made to remember as much as possible.

Source Errors on the Source Recognition Test

After the recall tests, all participants took a source recognition test, in which a list of old and new items was presented and they indicated whether each had been seen, imagined, or was new. Old/new recognition accuracy was assessed as the proportion of old items correctly identified as old plus the proportion of new items correctly identified as new relative to the total number of old and new items. Performance was nearly at ceiling, with at least 97% accuracy in each condition.

The basic pattern of source misattributions depicted in Figure 6 suggests more source errors on the recognition test for physical and conceptually related items relative to control items, and more errors when the earlier recall tests did not require specification of the remembered items' source (free recall) than when such specification was required (source recall). To confirm this, the

proportion of imagined items erroneously claimed as seen was examined in a 2 (number of recall tests) x 2 (test specificity) x 3 (similarity) ANOVA. [2] The main effect for similarity was significant, $F(2, 232) = 33.08$, MSe = .01, and comparisons revealed significantly more errors for both physically similar (.30) and conceptually related items (.32) than for control items (.21). The main effect for test specificity was also significant, $F(1, 116) = 13.05$, $MSe = .09$, with higher errors when the earlier tests did not require specification of the remembered item's source (.33) than when they did require source discriminations (.22). The main effect for number of separate recall attempts was not significant, $F(1, 116) < 1$. Thus errors in claiming an imagined item had been seen on the source recognition test did not differ when people had earlier taken three recall tests (.26) or one prolonged one (.28). Although there was a significant three-way interaction, $F(2, 232) = 4.52$, $MSe = .01$, the patterns noted in the main effects were upheld, with physically similar and conceptually related items always yielding greater source errors than control items, though the relative magnitude of the differences varied across the different test conditions.

Additional analyses examined source judgments on the recognition test as a function of previous recall, that is, whether items that had been previously recalled were more likely to be misjudged as seen on the source recognition test than were items that had not been previously recalled, and this sheds some important light on the significant main effect of earlier test specificity noted in the above analysis. The proportion of originally imagined items that had been previously recalled and misjudged as seen on the recognition test (.18) was significantly higher than items not previously recalled but misjudged as seen on the recognition test (.10), $F(1, 116) = 67.82$, $MSe = 0.02$. A 2 (number of recall tests) x 2 (test specificity) x 3 (similarity) ANOVA was conducted on previously recalled items that were misjudged as seen on the source recognition test. The main effect for similarity was significant, $F(2, 232) = 20.69$, $MSe = 0.01$, with significantly more source

errors for conceptually related (.22) and physically similar items (.17) than for control items (.13).

No difference was found in source recognition errors when subjects had previously taken 3 separate recall tests (.16) or one continuous one (.19), $F(1, 116) = 2.09$, Mse = .04. However, subjects were significantly more likely to falsely recognize imagined items as seen on the recognition test when their earlier memory tests did not require specification of the remembered item's source (.23) than when the earlier tests required such specification (.13), $F(1, 116) = 23.51$, $MSe = .04$. An analogous 2 x 2 x 3 ANOVA was conducted on items not previously recalled but misjudged as seen on the source recognition test. Here again the main effect for similarity was significant, $F(2, 232) = 14.20$, $MSe = 0.01$, and the main effect for number of recall attempts was not significant, $F(1, 116) < 1$.

Thus there were more physically similar (.13) and conceptually related items (.10) misjudged as seen relative to control items (.06), with no more errors made for these unrecalled items when subjects had earlier attempted to recall them on three separate recall tests (.10) or on one continuous recall test (.08). More importantly was the absence of a main effect for specificity of memory test, $F(1, 166) < 1$. That is, subjects were no more likely to make source errors on the recognition test for items that they failed to recall on either the free recall or source recall test.   Thus although source recognition judgments for previously recalled items were influenced by the specificity of the earlier memory tests, source recognition errors for unrecalled items were not. These findings show that the act of successfully recalling items themselves leads to an increase in subsequent source recognition errors, and that earlier memory tests that encourage a more lax criterion in remembering items (i.e., recalling items without consideration as to whether they had been seen or imagined) subsequently increase the odds of misattributing the source of previously recalled items when they are encountered on a recognition test. The earlier free and source recall tests do not necessarily themselves alter the criteria one uses later when judging the source of items on a recognition test, as

evidenced by the null difference in source errors on the recognition test between these two

conditions for items that had not been previously recalled. Rather it appears that the act of recalling

items without consideration as to their source serves to strengthen the vividness of the remembered

item's features (thereby making it seem more perceptual like) and/or increase the ease with which

the imagined item comes to mind (thereby reducing the source-specifying cognitive operations

associated with the item that would correctly indicate its source as imagined), both of which would

lead to imagined events being misjudged as seen, according to the source monitoring framework.

The inclusion of the recognition test following the effortful recall tasks provided another

avenue with which to explore the effects of previous attempts at remembering on subsequent source

judgments. People's abilities to discriminate between old and new items was extremely good (near

ceiling), but their ability to discriminate between imagined and seen items was decidedly poorer,

with error rates of 30% and higher in some conditions. Results showed that the same factors that

influenced source errors during recall tasks also influenced source errors on the recognition task:

More source errors were made for physically similar and conceptually related items relative to

control items, and more source errors were made for items previously recalled without consideration

as to their source. In sum, these findings offer support for the predictions derived from the source

monitoring framework across several different types of memory tasks, including separate recall

tests, an equivalent-in-length continuous recall task, and a recognition test.

GENERAL DISCUSSION

The three experiments reported here assessed the influence of repeated and extended

attempts at remembering on people's ability to subsequently determine the source of the

remembered items, and they converged on the following basic findings: (a) Recall increased and

source accuracy decreased with repeated memory tests; (b) Imagined items were more likely to be

falsely judged as seen when they either physically resembled or were conceptually related to an item that was in fact seen; (c) When people repeatedly recalled items without consideration of their source, subsequent source errors were more frequent than when people considered the source of each recalled item during retrieval. These findings were obtained for both nonforced as well as for forced recall procedures, thus they were not simply due to lower evaluation criteria that may have been induced by forcing participants to guess on early memory tests in the forced recall conditions. These same effects were found not only following separate repeated memory tests, but also occurred when memory was assessed with one prolonged memory test period. In addition, source confusions made on an old/new and source recognition task showed similar effects as well. The findings across the three experiments support the general view that repeated testing can have contrary effects on memory: While repeatedly trying to remember successively more items does promote enhanced recall, such benefits come with a cost, namely greater confusion in the source of those remembered items, particularly when memory is repeatedly assessed with less stringent criteria. These falsely attributed memories are rated with high confidence as having been seen.

Taken as a whole, these findings illustrate several important principles derived from the source monitoring framework, which argues that source assessments are based both on the qualitative features of memories as well as on flexible decision criteria employed in evaluating features of memories. These studies show that source errors can arise from the features derived from other similar experiences one has had, and such similarity is not limited to semantic associations, but includes physical features as well. False memories induced by both conceptual and physical similarity increase across repeated memory tests and decrease with more careful consideration as to the source of the remembered item. The source errors resulting from repeated testing are explainable through the basic tenets of the source monitoring framework as well.

Repeated attempts to remember as much information as possible involve many different cognitive

processes, such as retrieving information that comes to mind readily, searching for additional

information, and evaluating that information, any of which may involve imagination,

embellishment, and reconstructive processes. The repetition that occurs across multiple retrieval

efforts may serve to strengthen the confidence and vividness with which one both veridically and

falsely recollects information. Supporting this point, other research has shown that the rates at which

people claimed they remember an event happening to them increase from when they are first

questioned to a later questioning if in the interim they had engaged in imagery involving that event

(Hyman & Billings, 1998; Hyman & Pentland, 1996). Likewise, repeated retrieval attempts using

guided imagery increased people's willingness to claim to remember events from their childhood

that probably did not occur (Porter, Yuille, & Lehman, 1999), and repeated imaginings of actions

led to an increase in people falsely claiming they had performed actions that they did not in fact

perform (Goff & Roediger, 1998; Thomas & Loftus, 2002). Thus, imagining events that did not

actually happen could lead one to mistakenly believe that they did happen (see also Nogrady,

McConkey, & Perry, 1985), and this effect can apparently be magnified by repeated imaginings.

Such repeated imaginings can serve to increase the vividness of imagined items and/or decrease the

cognitive operations associated with the remembered item (as imagery becomes less effortful with

repetition), both of which could provide misleading cues that the item was perceived.

Not only can postevent imagination sometimes inherent in repeated probings of memory

intrude upon the ability to correctly identify the source of one's experiences, repeated probings of

memory can in essence serve as repeated exposure to false information (Roediger, Jacoby, &

McDermott, 1996). Not surprisingly then, repeated exposure to misinformation has been shown to

increase false memory rates (Zaragoza & Mitchell, 1996), especially when the repeated exposure

involves imaginal processes that increase the vividness, detail, and fluency of the information

(Mitchell & Zaragoza, 1996; see also Reyna & Lloyd, 1997). Furthermore, other research has

shown that repeated exposure to similar items can increase false memory rates under some

circumstances (Seamon et al., 2002). Thus consistent with the source monitoring framework, as an

event becomes more vivid and easier to imagine on successive retrieval efforts, it becomes more

likely that the individual will mistakenly believe the imagined event really occurred.

Cognitive processes involved in encoding and extracting meaning may also contribute to

source misattributions. For example, when people falsely recall hearing words that were

thematically related to actually presented items, these intrusions may be due to incorrectly

attributing the relatively effortless yet vivid thoughts they had when listening to the lists to actually

hearing the words (see, e.g., Mather, Henkel, & Johnson, 1997; Multhaup & Conner, 2002). Both

constructive processes during encoding and reconstructive processes during retrieval could therefore

contribute to source confusions when people repeatedly try to remember their experiences. For

instance, as a consequence of multiple memory tests, people may note the relations between and/or

organize related items during retrieval, and they may trigger other items similar to the ones they do

recall.

In sum, one way in which multiple memory tests might create opportunity for the source

confusions seen in the present studies is because the repetition involves imaginal processes that

increase the vividness, detail, and fluency and decrease the cognitive operations associated with the

originally imagined item. As these features of imagined items become less discriminable from the

features of perceived items across the multiple memory tests, source errors increase. The increase in

source errors across the three memory tests, the substantially greater incidence of source errors for

items recalled previously on the earlier memory tests, and the elevated rates of source

misattributions for physically and conceptually similar items are all consistent with this

interpretation. In addition, source errors were more prevalent when people repeatedly recalled items

without regard to their source than when the task required them to indicate the source for each

recalled item. Whereas carefully considering the source of each recalled item may serve to

strengthen the binding between the features of the item (e.g., its shape, or its meaning) and its

specific occurrence, repeatedly trying to remember as much as possible without such careful

consideration is a recipe for source confusions because the features of similar items may not be

appropriately bound to their origin while increasing in their vividness.

Another possibility is that the repeated attempts at recollection do not so much influence the

vividness and accessibility of source specifying features, but that they bring about a change on later

memory tests in the criteria people use to determine which features are likely to represent an item

that was originally seen or imagined. For instance, as someone tries earnestly to remember more

items on the third test than the first, they may require less vividness or certainty to claim an item had

been seen. Several findings are inconsistent with this explanation, however. For one thing, source

errors were found using a forced recall test procedure, which holds constant response criteria for the

recall of items (as seen by equivalent rates of recall for imagined items for both the forced and free

recall tests; Erdelyi & Becker, 1974), though one could argue that that does not necessarily mean

that these tasks hold constant the criteria used for source assessments. However, confidence ratings

for falsely attributed items on the third test were relatively high, whereas one would expect less

confidence for misattributed items if participants simply lowered their criteria. In addition, source

errors were more prevalent for items that were recalled on earlier tests than items not recalled until

the third recall test, suggesting that the act of recalling the items influences subsequent source

attributions rather than merely inducing global changes in response criteria, which would

presumably influence source attributions for items irrespective of whether they had been previously

recalled. This interpretation is further strengthened by the finding from Experiment 3 showing

higher source error rates on the source recognition test as a function of specificity of the earlier

recall tests (free recall vs. source recall) only for items that had been previously recalled but not for

ones that had not been recalled. Thus the act itself of recalling items without considering their

origin produces memories in which potentially source-specifying features are misleading or

confused with other similar experiences. Taken together, these findings suggest that simple changes

in response criteria due to the demands of repeated memory tests are not the driving force behind the

increases in source errors. Instead, the paradoxical effects of repeated testing – increased recall and

decreased source accuracy -- can best be explained by the source monitoring framework's argument

that repeated memory tests can elicit misleading cues as to the originally imagined item's source,

especially for items that are physically or conceptually similar to actually seen items.

The source monitoring framework is a broad model that accounts for many aspects of correct

and incorrect remembering. However, other models have also been proposed to explain some of the

same phenomena (e.g., Gallo & Roediger, 2002; McDermott & Watson, 2001; Reyna & Llloyd,

1997; Schacter, Norman, & Koutstaal, 2000). These studies were not intended to provide rigorous

tests of different predictions derived from competing models. Nonetheless, it is hoped that this

paradigm can be applied in future research to differentiate among the models, for example, in

exploring potentially differential decay rates of various features of memories over time (see, e.g.,

Reyna & Brainerd, 1995; Suengas & Johnson, 1988) and examining directly conscious assessment

of the vividness of the features of remembered items. Directions for other research might include

consideration of strategic judgment processes that may be used in making source assessments. By

using larger sets of physically and conceptually similar items, one could examine organizational

strategies during recall, such as clustering of related items, and how this subsequently impacts on

source judgments.

In conclusion, these findings add to our understanding of the many factors that can influence

people's ability to determine whether something they remember was perceived or imagined.

Consistent with the source monitoring framework, both the features associated with one's

experiences and the criteria one uses to evaluate those features can affect the accuracy with which

one attributes memories to their sources. Repeated attempts to remember items has costs and

benefits: While they may improve the amount of information recalled, greater confusions in the

source of that information arises. Such source confusions are not random, but rather they follow

principled properties based on normal constructive and reconstructive memory processes.

References

Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34,* 1358-1372.

Bornstein, B., Liebel, L., & Scarberry, N. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12,* 119-131.

Brainerd, C. L., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 55,* (3-4, Whole No. 222).

Cassel, W., Roebers, C., & Bjorklund, D. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology, 61,* 116-133.

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106*, 541-557.

Durso, F. T., & Johnson, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior, 19,* 416-429.

Erdelyi, M. H. (1982). A note on the level of recall, level of processing, and imagery hypothesis of hypermnesia. *Journal of Verbal Learning and Verbal Behavior, 21,* 656-661.

Erdelyi, M. H. (1996). *The recovery of unconscious memories: Hypermnesia and reminiscence*. Chicago: University of Chicago Press.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures. *Cognitive Psychology, 6,* 159-171.

Erdelyi, M. H., Finks, J., & Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance,

with some observations on processing bias. *Journal of Experimental Psychology: General,* **118***,* 245-254.

Eugenio, P., Buckout, R., Kostes, S., & Ellison, K. W. (1982). Hypermnesia in the eyewitness to a crime. *Bulletin of the Psychonomic Society,* **19***,* 83-86.

Fivush, R., & Schwarzmueller, A. (1995). Say it once again: Effects of repeated questions on children's event recall. *Journal of Traumatic Stress,* **8***,* 555-580.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language,* **47,** 469-497.

Goff, L.M., & Roediger, H. L. (1998). Imagination inflation for actual events: Repeated imaginings lead to illusory recollections. *Memory & Cognition,* **26***,* 20-33.

Hashtroudi, S., Johnson, M. K., Vnek, N., & Ferguson, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology and Aging,* **9***,* 160-170.

Hekkanen, S. T., & McEvoy, C. (2002). False memories and source monitoring problems: Criterion differences. *Applied Cognitive Psychology,* **16**, 73-85.

Henkel, L. A., & Franklin, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition,* **26***,* 659-673.

Henkel, L. A., Franklin, N., & Johnson, M. K. (2000). Cross-modal source monitoring confusions between perceived and imagined events. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **26***,* 321-335.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General,* **127**,

251-268.

Howe, M. J. A. (1972). Repeated presentation and retention of meaningful information. *Psychological Reports, 31,* 840-842.

Hyman, I. E., & Billings, F. J. (1998). Individual differences and the creation of false childhood memories. *Memory, 6*, 1-20.

Hyman, I.E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language, 35,* 101-117.

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

Johnson, M. K., Nolde, S. F., & De Leonardis, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language, 35,* 135-156.

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

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14,* 220-237.

Koutstaal, W., Schacter, D. L., Johnson, M. K., & Galluccio, L. (1999). Facilitation and impairment of event memory produced by photograph review. *Memory & Cognition, 27,* 478-493.

Lane, S. M., Mather, M., & Villa, D. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition, 29,* 940-947.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17,* 349-358.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory: False memories and allegations of sexual abuse.* New York: St. Martin's Press.

Mandler, G. (1994). Hypermnesia, incubation, and mind popping: On remembering without really trying. *Attention and Perception X,* 4-33.

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **24***,* 1137-1151.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition,* **25***,* 826-837.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language,* **35**, 212-230.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language,* **45**, 160-176.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). New York: Oxford University Press.

Mitchell, K. J., & Zaragoza, M. S. (1996). Repeated exposure to suggestion and false memory: The role of contextual variability. *Journal of Memory and Language,* **35**, 246-260.

Multhaup, K. S., & Conner, C. A. (2002). The effects of considering non-list sources on the Deese-Roediger-McDermott memory illusion. *Journal of Memory and Language*, **47**, 214-228.

Nogrady, H., McConkey K. M., & Perry, C. (1985). Enhancing visual memory: Trying hypnosis, trying imagination, and trying again. *Journal of Abnormal Psychology,* **94***,* 195-204.

Payne, D. G. (1986). Hypermnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **12***,* 16-29.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin,* **101***,* 5-27.

Payne, D.G., Elie, C.J., Blackwell, H.M., & Neuschatz, J.S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language,* **35***,* 261-285.

Poole, D. A., & White, L. T. (1995). Tell me again and again: Stability and change in the repeated testimonies of children and adults. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y.S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 24-43). Thousand Oaks, CA: Sage.

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior,* **23***,* 517-535.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: Some foundational issues. *Learning and Individual Differences,* **7***,* 145-162.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences,* **9***,* 95-123.

Roediger, H. L., III, Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language,* **35**, 300-318.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not

presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, **21**,* 803-814.

Roediger, H. L., III, McDermott, K. B., & Goff, L. M. (1997). Recovery of true and false memories: Paradoxical effects of repeated testing. In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 118-149). New York: Oxford University Press.

Roediger, H. L., III, & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, **8**,* 66-72.

Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, **13**,* 1-7.

Roediger, H. L., III, & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, **6**,* 296-305.

Roediger, H. L., III, Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. *Psychology of Learning and Motivation*, **30,** 97-134.

Schacter, D. L. (1999). The seven sins of memory. *American Psychologist*, **54**, 182-203.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (2000). The cognitive neuroscience of constructive memory. In D. F. Bjorklund (Ed.), *False-memory creation in children and adults: Theory, research, and implications* (pp. 129-168). Mahwah, NJ: Erlbaum.

Seamon, J. G., Luo, C., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language, **46**,* 323-340.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, **40**,* 83-108.

Suengas, A. G., & Johnson, M. K. (1988).Qualitative effects of rehearsal on memories for perceived and

imagined complex events. *Journal of Experimental Psychology: General,* **117**, 377-389.

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory &*

Cognition, **30***,* 423-431.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false

memories. *Psychological Science,* **7***,* 294-300.

Zaragoza, M. S., Payment, K. E., Ackil, J. K., Drivdahl, S. B., & Beck, M. (2001). Interviewing

witnesses: Forced confabulation and confirmatory feedback increase false memories.

Psychological Science, **12***,* 473-477.

Author Note

Footnotes

1.   Because the proportion of items recalled differed across the different types of items, and thus the total number of items available for source errors differed as well, an additional analysis was performed on the proportion of imagined items falsely claimed as seen conditionalized on the total number of imagined items recalled (rather than based on the total number of imagined items presented). The analysis based on conditionalized scores yielded identical patterns to those found using unconditionalized scores, and this was the case across the 3 studies, thus only the data from the unconditionalized analyses are reported.

2. Similar results were obtained in an analogous analysis conducted on conditionalized source error rates, where the proportion of imagined items falsely claimed as seen was calculated based on the total number of imagined items correctly recognized as old.

APPENDIX A
CRITICAL STIMULI FOR EXPERIMENTS 1, 2, & 3

Physically Similar Items

Seen	Imagined
lunchbox	suitcase
watering can	tea kettle
magnifying glass	lollipop
oven mitt	boxing glove
flea collar	belt
cotton ball	cloud
railroad track	ladder
knitting needle	pencil
dynamite	candle
magnet	horseshoe

Conceptually Related Items

Seen	Imagined
butter knife	spoon
hairbrush	comb
dollar	nickel
telescope	binoculars
tulip	daisy
sailboat	canoe
table	chair
screwdriver	hammer
apple	banana
guitar	drum

Control Items

Seen	Imagined
hourglass	garbage can
bib	bowling pin
razor	barbell
cork	sugar cube
door	pearl
cane	pants
broom	peanut
swing	flashlight
flyswatter	kite
envelope	hanger

Table 1

Recall Accuracy for Imagined Items in Experiment 1

	Test 1		Test 2		Test 3	
	M	SD	M	SD	M	SD
Forced Recall Test						
Physical	.44	.15	.47	.19	.51	.17
Conceptual	.48	.19	.55	.20	.58	.19
Control	.37	.12	.39	.13	.42	.13
Overall	.43		.47		.50	
Forced Source Test						
Physical	.43	.15	.45	.16	.52	.17
Conceptual	.44	.18	.49	.19	.55	.19
Control	.31	.14	.31	.14	.39	.14
Overall	.39		.42		.49	

Table 2

Recall Accuracy for Imagined Items in Experiment 2

	Test 1		Test 2		Test 3	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Free Recall Test						
Physical	.45	.15	.50	.14	.53	.16
Conceptual	.48	.17	.51	.19	.56	.19
Control	.36	.14	.37	.16	.43	.17
Overall	.43		.46		.51	
Source Recall Test						
Physical	.41	.15	.43	.17	.45	.17
Conceptual	.42	.19	.45	.20	.49	.23
Control	.32	.15	.33	.17	.37	.18
Overall	.38		.40		.44	

Table 3

Recall Accuracy for Imagined Items in Experiment 3

	Three Separate Tests				One Continuous Test
	Test 1	Test 2	Test 3	Cum. Recall	Cum. Recall
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Free Recall Test					
Physical	.34 (.15)	.41 (.17)	.43 (.17)	.51 (.17)	.55 (.19)
Conceptual	.48 (.18)	.51 (.21)	.58 (.17)	.66 (.17)	.69 (.12)
Control	.29 (.13)	.32 (.12)	.35 (.13)	.41 (.11)	.50 (.20)
Overall	.37	.41	.45	.53	.58
Source Recall Test					
Physical	.34 (.16)	.35 (.19)	.41 (.15)	.48 (.16)	.49 (.18)
Conceptual	.36 (.16)	.41 (.17)	.46 (.18)	.52 (.17)	.59 (.15)
Control	.27 (.14)	.31 (.12)	.38 (.14)	.42 (.14)	.49 (.17)
Overall	.32	.36	.42	.47	.52

Note: "Cumulative recall" refers to the proportion of items recalled on at least one of the three separate memory tests or on the one continuous memory test.

Figure Captions

Figure 1. Source errors for imagined items for the forced source monitoring condition in Experiment 1.

Error bars indicate standard error.

Figure 2. Source errors for imagined items on final recall test in Experiment 1. Error bars indicate

standard error.

Figure 3. Source errors for imagined items for the nonforced source recall condition in Experiment 2.

Error bars indicate standard error.

Figure 4. Source errors for imagined items on final recall test in Experiment 2. Error bars indicate

standard error.

Figure 5. Source errors for imagined items on final recall test in Experiment 3. Error bars indicate

standard error.

Figure 6. Source errors for imagined items on source recognition test in Experiment 3. Error bars indicate

standard error.