

## Reality monitoring of physically similar and conceptually related objects

LINDA A. HENKEL and NANCY FRANKLIN  
State University of New York, Stony Brook, New York

Three studies showed that information used in determining a target memory's source may be derived not only from the target event itself, but also from other nontarget events or memories. Subjects were more likely to claim that an imagined object was perceived when it physically resembled or was conceptually related to another specific item that was actually perceived, relative to when there was no physical resemblance or semantic relation. Furthermore, error rates for imagined items increased with the number of perceived items that they resembled. However, subjects' orienting task at encoding (perceptually biased or perceptually plus conceptually biased) did not systematically affect error rates. The results indicate that reality monitoring decisions about a target object are influenced by similar physical and conceptual information that was derived from other objects.

Through the course of each day, we encode information into memory about both perceived and imagined events, and information surrounding the creation of these memories is often important for subsequent action and decision making. Suppose, for example, that you recall a phone call from abroad informing you that you have won the Nobel Prize. Successfully using this memory to guide behavior depends on your ability to determine whether the conversation took place or was imagined. But because you probably did not explicitly tag the memory with a label indicating its source (i.e., perception or imagination), you must engage in evaluative processes to make your assessment (Johnson, 1983; Johnson, Hashtroudi, & Lindsay, 1993). Given the importance of distinguishing externally derived memories from internally generated memories (i.e., *reality monitoring*), people's ability to determine a memory's source should be adaptable to the wide range of circumstances under which memories are encoded and to the high variability of their content.

According to the source monitoring model of Johnson and colleagues, strategic judgment processes, such as reasoning based on the plausibility of an event's occurrence, can help determine a memory's source (Johnson et al., 1993). In addition, they propose that a particularly useful mechanism of reality monitoring relies on average differences in the characteristics of memories derived from different sources (Foley & Johnson, 1985; Johnson,

Foley, Suengas, & Raye, 1988). Thus, the features or characteristics of a target memory serve as a cue as to the memory's source (see, e.g., Johnson et al., 1993; Johnson & Raye, 1981). For example, memories derived from imagination typically have more information about cognitive effort and generation processes than do memories derived from perception, and those derived from perception typically have more perceptual and contextual detail than do those derived from imagination. Because such differences are a matter of degree, subtle variations can lead to source errors. For example, reality monitoring errors can arise when the amount or quality of the features for a particular memory derived from imagination is more typical of memories from perception (e.g., very vivid imagery and/or reduced cognitive effort in image generation). A vivid memory arising from an imagined phone call about the Nobel Prize could lead to such an error. So could a memory stemming from images induced by a vividly written novel; certainly many of us have fallen prey to a mistaken belief that we have seen the movie after having read an especially evocative book.

The research to date supports Johnson's source monitoring model for both objects and events. People appear to determine the source of a memory by comparing its features to the average features of memories derived from perception and imagination (Johnson, Foley, Suengas, & Raye, 1988; Johnson et al., 1993). Specifically, when the features from an imagined memory are more similar to the average features for perceptually derived memories, source errors increase (Intraub & Hoffman, 1992; Johnson, Foley, & Leach, 1988; Johnson, Raye, Wang, & Taylor, 1979). Such work provides information about the features people use in reality monitoring and the values that are indicative of a particular source.

The studies discussed thus far examined source errors of a *specific* memory on the basis of how that memory's features compare to *average* characteristics. They show that the specific features used in evaluating a given mem-

---

The authors thank Marcia Johnson, Richard Gerrig, Arthur Samuel, Susan Brennan, Barbara Tversky, Suparna Rajaram, Ute Bayen, Kevin Murnane, Tom Nelson, and three anonymous reviewers for their help on earlier versions of this manuscript. Correspondence concerning this paper should be sent to L. A. Henkel, Department of Psychology, University of North Florida, 4567 St. Johns Bluff Road South, Jacksonville, FL 32224 (e-mail: lhenkel@unf.edu).

ory's source derive from the target itself. The present studies address a new but related question: Can the judgment of a memory's source be influenced by the features derived from other *specific* memories? That is, does greater similarity between specific imagery and perceptually based memories lead to a higher rate of source errors? If so, reality monitoring judgments can be influenced by information derived not only from the target itself but from other events or objects as well.

Research on *external source monitoring* suggests this may be the case. People's ability to determine which particular perceptual source gave rise to a target memory decreases as the different sources increase in the similarity of their perceptual features—for example, two female voices versus a male and a female voice (Bayen & Murnane, 1996; Bayen, Murnane, & Erdfelder, 1996; Ferguson, Hashtroudi, & Johnson, 1992; Johnson, DeLeonardis, Hashtroudi, & Ferguson, 1995). Likewise, greater semantic similarity between two external sources leads to higher error rates in judging source (Johnson, Raye, Foley, & Foley, 1981). For example, people are more likely to misattribute a statement from one speaker to another when the two speakers describe the same event than they are when the speakers describe different events (Lindsay, Johnson, & Kwon, 1991). The present studies address whether such types of similarity not only increase confusions between different externally derived memories but increase confusions between externally derived memories and internally generated ones.

According to Johnson's source monitoring model (Johnson et al., 1993), the extent to which a particular object was imagined vividly and with minimal cognitive processing influences the likelihood of its being mistakenly judged as having been perceived. But the "vividness" that gives rise to source errors may involve several kinds of information. People might think they have seen an imagined item because they have a particularly vivid memory for that *object itself*. This could arise, for example, because the memory itself has features more typical, on average, for perceptually derived memories. However, such an error could also arise if information from other events influences judgments of a target memory's origin. For example, people might think they have seen the imagined item if they have a particularly vivid memory of seeing an object with those *sensory features*, or if they have a vivid memory of seeing an object with those *semantic or conceptual features*. To illustrate, suppose that you imagine seeing a lollipop and that your memory has rich perceptual information (e.g., about its appearance) and has little information about the cognitive processes involved in generating the memory (e.g., because it was very easy to imagine). When you later mistakenly claim that you saw a lollipop, is your error based on your having a particularly perceptual-like memory for a lollipop itself, a particularly perceptual-like memory for something round with a stick at its bottom, or a particularly perceptual-like memory for a piece of candy?

The source monitoring model argues that sensory-perceptual information activated when one is remembering can be an important cue as to the origin of the information. Such information could, in fact, arise not only from the to-be-remembered target event but also from other memories. To determine whether reality monitoring is influenced by the similarity of a target memory's perceptual features to the features of other specific memories, we presented such perceptual information outside the context of the item itself. For example, suppose you imagined a lollipop and at another time saw something with similar structural features, such as a magnifying glass. You have a memory that can supply you with evidence that you saw a round object with a narrow stick at its bottom, but the evidence derives from an event other than that of having imagined the lollipop. If the viewed magnifying glass inflates error rates, it is because information outside the context of the lollipop affects your assessment. This would constitute evidence not only that perceptual information about a target affects reality monitoring but that perceptual information from other objects does so as well.

Circumstances of this sort may not be unusual. Memories vary in quality, age, and completeness, and features can be recovered independently of their corresponding objects (Brown & McNeill, 1966). The present studies test whether perceptual information is used in reality monitoring even when it is not integrated into target objects. Shape in particular as a basis of perceptual similarity was investigated, both because representation of shape in mental images appears to be isomorphic to that for percepts (Shepard & Chipman, 1970) and because shape is a reliable characteristic feature of object classes (Biederman, 1987).

As noted earlier, conceptual as well as perceptual information about a target memory may be a useful cue in determining its source. Research on eyewitness testimony provides indirect evidence for the use of conceptual information in source decisions (Lindsay, 1994). But because many of these studies use items that both fit the general semantic theme of the scene and physically resemble each other (e.g., different kinds of coffee cans), it is unclear whether errors arise from conceptual relation, physical similarity, or both. The present studies therefore examined whether conceptual information from nontarget events can influence judgments of a memory's source. For example, does having seen a candy bar (which belongs to the same functional class as an imagined lollipop but does not physically resemble it) increase people's belief that they have seen the lollipop? Such errors would indicate that reality monitoring makes use of conceptual information about target objects and that it is influenced by the presence of other, related memories.

To determine whether some of the information that enters into the evaluation of the origin of a particular event may be derived from other events, the present experiments therefore examined the contribution of similar percep-

tual and conceptual information from nontarget events to judgments of a target memory's source.

## EXPERIMENT 1

### Physical Similarity and Conceptual Relation

Experiment 1 tested whether reality monitoring errors can be affected by the similarity of the physical and conceptual features of specific imagined and perceived items. The subjects saw simple line drawings of objects and visually imagined others in the same style. Line drawings were used to minimize variability in the representation of structural information for imagined items and to allow for similarity between items to be defined in terms of basic component parts or geons (avoiding other, uncontrolled sources of similarity, such as color). Each imagined item (italicized) either physically resembled a perceived item (e.g., *lollipop*—magnifying glass), was conceptually related to a perceived item (e.g., *shirt*—pants), or had no resemblance or relation to perceived items (e.g., *book*—ax).

While seeing and imagining the objects, the subjects either had a perceptual orienting task, which required them to attend to sensory information such as shape and detail, or had a compound orienting task, which required them to process information about each object's appearance and function. These two orienting tasks were used to determine whether source errors differ for conceptually and perceptually similar pairs when people focus only on aspects of the items' appearance or when they divide attention between perceptual and conceptual features.

When making reality monitoring decisions, people evaluate memory for evidence that a target was seen or imagined. The present study examines whether the perceptual information entering into such judgments can derive not only from the target event but also from other experienced events. Using the above pairs as illustrations, if evaluation occurs only in the context of the target, error rates should be the same for the *lollipop* as for the *book*, since each was imagined and not seen. But if such evaluation is affected by evidence from other nontarget events, errors should be higher for the *lollipop*, which resembles a viewed object (magnifying glass), presumably because the subjects have evidence for their having perceived features that they would have encoded had they seen a *lollipop*. Similarly, if the subjects are affected by conceptual information activated by the target, then an imagined *shirt*, which is conceptually related to perceived *pants*, might yield higher error rates than an imagined *book*, which has no conceptual relation to any perceived item in the set.

These studies focus on reality monitoring for imagined rather than for perceived items. The use of common items and the simplicity of their depiction should encourage memories with vivid perceptual information and with little information about the cognitive operations that took place at encoding for both the imagined and the perceived items. As such, memories for imagined items should be likely to be mistakenly judged as perceived, because their features would be more typical for perceived memories,

on average (Johnson & Raye, 1981). Perceived items would be less likely to be mistaken as imagined, because they would be lacking in the features that, on average, are more typical for imagined memories (e.g., reduced perceptual information, rich information about cognitive operations). This asymmetry has been shown in other studies as well (see, e.g., Johnson et al., 1979; Johnson, Taylor, & Raye, 1977; Kahan & Johnson, 1990).

### Method

**Subjects.** Seventy-five native English speakers at the State University of New York (SUNY) at Stony Brook were tested individually in two sessions. Thirty-eight received the perceptual orienting task, and 37 received the compound orienting task. Subjects received course research credit for participating or were volunteers from a summer study program at the university.

**Materials and Design.** The stimuli were 78 slides, half of which were for perceived trials and half for imagined trials. Each slide for the perceived trials showed a simple black-and-white line drawing of a common object with its name below it (see Figure 1 for examples). For the imagined trials, just the name of an object appeared at the bottom of the slide. Equal numbers of three types of stimulus pairs were used. The members of a pair either resembled each other physically (with no conceptual relation), were conceptually related (with no physical resemblance), or were neither physically nor conceptually similar. Norms were established on the relative ease with which a given item could be imagined and on the item's physical complexity. Across the three types of stimulus pairs, items did not differ significantly on these dimensions.<sup>1</sup>

Two sets of stimuli, counterbalanced across subjects, each used the same 26 items for perceived trials. In Set 1, 13 perceived items each physically resembled 1 imagined item, and 13 were conceptually related to 1 imagined item. In Set 2, the 13 perceived items that had physically resembled an imagined item now were in conceptual pairs, and vice versa. Physical similarity was established by pilot subjects' ratings of resemblance for pairs of imagined items (1 = physically "not at all similar," 7 = "very similar"). The mean rating for pairs of physically similar items was 4.5 or higher, and for nonphysically similar pairs ratings were 2.0 or lower.

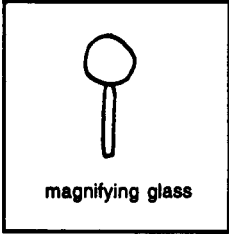
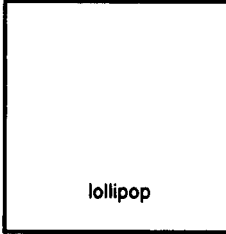
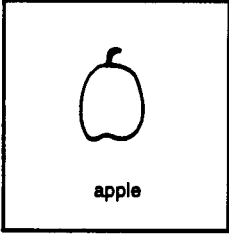
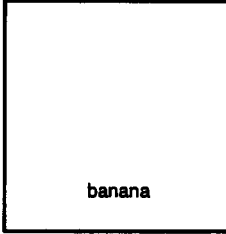
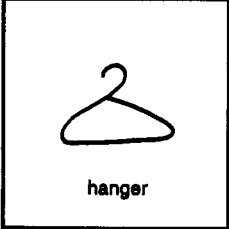
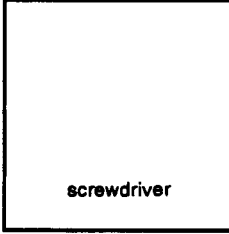
Conceptually related items were selected on the basis of several different kinds of relations: (1) category membership (e.g., *apple* and *banana* as instances of fruit, *knife* and *fork* as eating utensils), (2) co-occurrence in the world (e.g., *oyster* and *pearl*, *cigarette* and *ashtray*), or (3) second-order association (e.g., *leash* and *bone* as associates of dog). These were selected by the consensus of three trained judges.

Item order in the encoding phase was randomized, except that there was a minimum of 10 trials between the members of a pair. For half the pairs, the perceived member appeared first, and, for the other half, the imagined member appeared first. A single random ordering was used for all subjects.

For the source test, the names of 39 new items and the 78 old items were presented in random order. Of the new items, one-third resembled an object that had been perceived, one-third were conceptually related to a perceived item, and one-third were control items with no resemblance or relation.

**Procedure.** The subjects were told that one type of slide showed a simple black-and-white line drawing of an object with the object's name beneath it. Another type of slide gave the name of an object, and they were to imagine the item as a black-and-white line drawing similar in style to the pictures.

Subjects were randomly assigned to one of two conditions. For the perceptual orienting task, subjects estimated how long it took to draw each picture or how long it would take to draw each imagined item (see Durso & Johnson, 1980). For the compound orient-

<b>TYPE OF PAIR</b>	<b>PERCEIVED</b>	<b>IMAGINED</b>
<b>Physically Similar</b>	 <p>magnifying glass</p>	 <p>lollipop</p>
<b>Conceptually Related</b>	 <p>apple</p>	 <p>banana</p>
<b>Control</b>	 <p>hanger</p>	 <p>screwdriver</p>

**Figure 1.** Examples of perceptual and imagery trials for physically similar, conceptually related, and control pairs in Experiment 1.

ing task, they named a common function of the item. To induce those subjects performing the compound orienting task to generate visual images on the imagined trials, they first gave a rating of how well their image depicted the named object (e.g., fair, good, very good). For perceived trials in this condition, they rated the quality of the drawing before indicating the object's function.

For the encoding phase, 78 slides were presented for 7 sec each by a Kodak Carousel Projector, during which time the subjects stated aloud their judgment of drawing time or of quality plus function. The experimenter recorded all responses.

Subjects returned 2–3 days later and were given a surprise recognition/source monitoring test. The names of the items appeared one at a time on an IBM-PC computer, and the subjects indicated whether the item had been presented as a picture, had been imagined, or was new. They indicated their response by pressing the P, I, or N key (corresponding to *perceived*, *imagined*, and *new*). The test was self-paced.

### Results and Discussion

Six subjects (3 in each condition) failed to return for Session 2, leaving 35 in the perceptual orienting task condition and 34 in the compound orienting task condition. All analyses for the remaining 69 subjects used an alpha level of .05.

Overall recognition accuracy (hits plus correct rejections) was 81.4%, and overall reality monitoring accuracy (correct source attributions given old items were identified as old) was 75.9%. The two sources were ex-

pected to produce different rates and patterns of reality monitoring errors across the manipulated variables; therefore, for the analyses of primary interest, reality monitoring error rates were calculated for each similarity condition as the percentage of items in that condition for which subjects incorrectly attributed source (e.g., claimed an imagined item was perceived), given correct identification of the items in that condition as old (old items identified as either perceived or imagined). This measure conditionalizes source monitoring on correct recognition for each source.

Reality monitoring error rates were examined in a  $2 \times 2 \times 3$  analysis of variance (ANOVA), with orienting task (perceptual or compound), source (imagined or perceived), and pair type (physically similar, conceptually related, or control) as independent variables.<sup>2</sup> Mean error rates appear in Figure 2.

No significant difference in error rates was found between the perceptual ( $M = 29.4\%$ ) and compound ( $M = 26.4\%$ ) orienting tasks [ $F(1,67) = 1.33$ ,  $MS_e = 712.4$ ,  $p = .25$ ]. In addition, consistent with predictions and asymmetries typically found in the literature, the subjects were more likely to claim that an imagined item was perceived ( $M = 30.6\%$ ) than to claim that a perceived item was imagined [ $M = 25.2\%$ ;  $F(1,67) = 4.14$ ,  $MS_e = 755.3$ ,  $p < .05$ ].

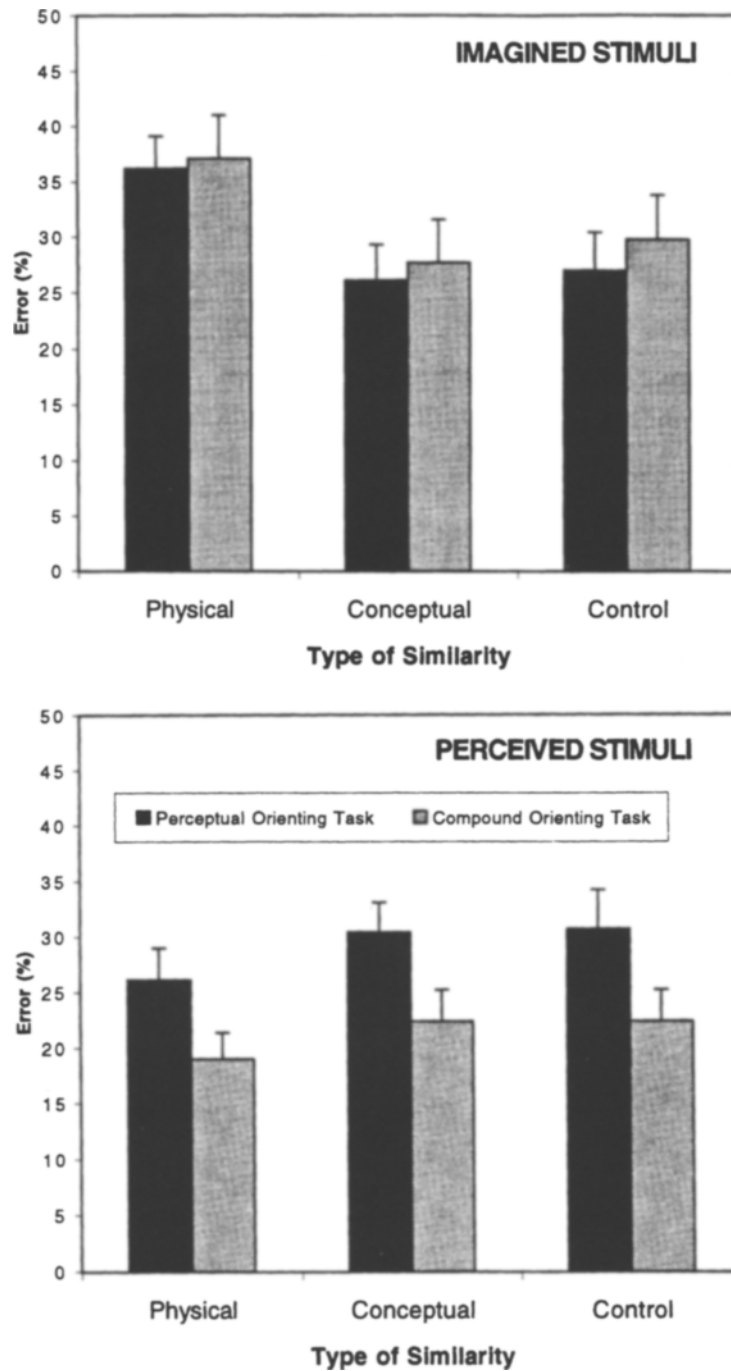


Figure 2. Percentage of reality monitoring errors for imagined and perceived stimuli in Experiment 1.

It was predicted that imagined items that resembled or were related to perceived items would have higher error rates than would control items. There was a significant pair type  $\times$  source interaction, where pair type affected reality monitoring for imagined but not for perceived items [ $F(2,134) = 9.79, MS_e = 198.9, p < .001$ ]. Post hoc Newman-Keuls  $t$  tests on imagined items showed more

errors when the target resembled a perceived item ( $M = 36.6\%$ ) than when it was conceptually related ( $M = 26.9\%$ ) or had no resemblance or relation ( $M = 28.4\%$ ) to a perceived item. However, conceptually related and control pairs did not differ significantly.

Taken together, the results indicate that when you have imagined an object, the likelihood of later judging that

you had perceived it increases if you have actually seen a different object with similar perceptual features. This happens even though the two items occurred several trials apart, even though they have no conceptual relation, and even though the experimenter never drew your attention to the fact that the items share physical features. Although you correctly remembered that *lollipop* was in the slide presentation and correctly remembered having perceived something round with a narrow stick beneath it, you incorrectly attributed those perceived features to the imagined object. There was evidence for having seen features possessed by the target, but the evidence came from a different item in memory.

Contrary to predictions, conceptual relatedness did not increase source errors, as might be expected on the basis of previous findings (see, e.g., Johnson et al., 1981). It is possible that reality monitoring is not influenced by conceptual information from other memories, but, because we used three different classes of conceptual relation, such a possibility was considered only tentatively, and the effect of conceptual similarity was investigated again in Experiment 2.

False alarms on new items (i.e., claims that a new item had been imagined or perceived) were higher to new items that resembled a perceived item ( $M = 15.7\%$ ) than to conceptually related ( $M = 11.2\%$ ) or control ( $M = 12.6\%$ ) new items [ $F(2,134) = 11.65$ ,  $MS_e = 29.3$ ,  $p < .001$ ]. Consistent with the main experimental results, these errors indicate that information about physical features from nontarget, perceived items can enter into judgments about a target item's source. Such errors point to the potential for similarity between recognition and source monitoring processes (Johnson et al., 1993).

Many measures of source monitoring involve both recognition memory for old items and memory for the source of the items, and this confounding of item detectability and source discriminability can be problematic. Several researchers have argued that the interpretation of source monitoring errors can be ambiguous when there are varying levels of recognition across the conditions in a study (Batchelder & Riefer, 1990; Murnane & Bayen, 1996). The reality monitoring errors found in the present study, however, cannot be solely based on differential recognition rates, because source monitoring and recognition were affected differently by the manipulated variables. For example, whereas recognition accuracy was higher for imagined than for perceived items and improved for the compound versus the perceptual orienting task,<sup>3</sup> reality monitoring performance was, in fact, worse for imagined items than for perceived ones and was not affected by orienting task. Furthermore, supplementary analyses that independently examine source discriminability and item detectability (reported in Appendix A) showed essentially equivalent results for the key experimental comparisons concerning reality monitoring. Taken together, these findings support the argument that the likelihood of people mistakenly believing they perceived an item that was actually imag-

ined increases as a function of the imagined memory's similarity to items that were in fact perceived. These errors appear to reflect differences in people's ability to identify the source of a memory and not just differences in people's ability to remember the items themselves.

## EXPERIMENT 2

### Conceptual Relation Due to Functional Category

One of the goals of Experiment 1 was to determine whether reality monitoring is affected by conceptual information consistent with a target but derived from a nontarget event. However, several kinds of relatedness were included in Experiment 1, and all types of conceptual information may not be equally likely to be activated at test from nontarget events or to be used in source judgments of target events. In Experiment 2, all conceptually related pairs, such as *pants* and *shirt*, belonged to the same functional (superordinate) category. We focused on category membership because function is a particularly important type of feature and appears to be one of the reasons why the basic level is privileged in memory (Tversky & Hemenway, 1984).

Experiment 2 examined the effect of physical resemblance and conceptual relation on reality monitoring in a factorial design, with two levels of physical resemblance (high or low) and two levels of conceptual relation (same or different functional category). Each pair of imagined and perceived items, then, had one of four relations: The items physically resembled and were conceptually related to each other (e.g., *bagel-doughnut*), resembled each other but were unrelated (e.g., *magnifying glass-lollipop*), had no resemblance but were related (e.g., *pants-shirt*), or had no resemblance or relation (e.g., *cane-feather*).

Subjects saw and imagined black-and-white line drawings and later indicated the source (perceived or imagined) of the items they recognized. If physical resemblance influences reality monitoring (as indicated by the outcome of Experiment 1), accuracy should be impaired for imagined items that physically resemble perceived items. If conceptual relation affects performance, then subjects should make more errors for items belonging to the same category than for unrelated items.

As in Experiment 1, physical resemblance and conceptual relation were expected to increase reality monitoring errors more for imagined than for perceived items, since the experimental situation was not designed to produce memories for perceived items that have characteristics typical of memories derived from imagination.

### Method

**Subjects.** Sixty-five undergraduates at SUNY at Stony Brook participated individually in two sessions. Of these, 33 were randomly assigned to the perceptual orienting task and 32 to the compound orienting task. All were native English speakers who received course research credit for participating.

**Materials and Design.** The stimuli were 104 slides, half for perceived and half for imagined trials. Each perceived trial consisted

of a black-and-white line drawing of an object with its name beneath. For imagined trials, an object's name appeared at the bottom of the slide. There were four types of pairs, crossing physical resemblance<sup>4</sup> with conceptual relation. Each cell included 13 pairs, and, within each pair, one member was perceived and one was imagined. In addition, to increase the face validity of the orienting tasks, 10 drawings of additional objects that depicted the specified objects poorly were randomly interspersed throughout the experimental trials. Conceptually related items were defined as belonging to the same functional category (e.g., *apple* and *banana* as instances of fruit). Slides were randomly ordered, with a minimum of 10 trials between pair members. For half of the pairs, the perceived member occurred first in the order of trials, and, in the other half, the imagined member occurred first.

**Procedure.** The general procedure was identical to that of Experiment 1. During the encoding phase, half of the subjects performed the perceptual orienting task while viewing 104 slides plus 10 filler slides showing poor depictions of objects, and half performed the compound orienting task. The experimental trials were each presented for 7 sec. All subjects returned after 2 days for the surprise recognition/source monitoring test, in which they judged 51 new and 104 old objects. The new items neither physically resembled nor were conceptually related to old items.

## Results and Discussion

Four subjects failed to return for the second session, resulting in 30 subjects in the perceptual orienting task condition and 31 in the compound orienting task condition.

Overall recognition accuracy was 84.2%, whereas overall reality monitoring accuracy was 73.5%. Reality monitoring errors were compared in a  $2 \times 2 \times 2 \times 2$  ANOVA, with orienting task (perceptual or compound), source (imagined or perceived), physical resemblance (high or low), and conceptual relation (high or low) as independent factors (see Figure 3).

The interaction of conceptual relation  $\times$  orienting task  $\times$  source was significant [ $F(1,59) = 5.92$ ,  $MS_e = 1,745$ ,  $p < .02$ ]. In light of this interaction, any simpler effects might be difficult to interpret. Because our primary interest was in the imagined items, the data were divided according to the source factor, and secondary analyses were conducted separately on imagined and perceived stimuli. Each was analyzed in a separate  $2 \times 2 \times 2$  ANOVA, using physical resemblance (high or low), conceptual relation (high or low), and orienting task (perceptual or compound) as independent variables. For each independent factor, data from the primary analysis are reported first, and, where results of these secondary analyses help to clarify the overall findings, they too are presented.

**Physical resemblance.** As predicted, subjects were more likely to err in their reality monitoring judgments for physically similar ( $M = 29.6\%$ ) than for dissimilar ( $M = 23.2\%$ ) pairs [ $F(1,59) = 34.63$ ,  $MS_e = 147.5$ ,  $p < .001$ ].

Confirming the prediction that physical similarity would affect performance for imagined items in particular, this physical resemblance effect interacted with source [ $F(1,59) = 6.47$ ,  $MS_e = 196.0$ ,  $p < .02$ ]. In the analysis on only imagined items, errors for targets that resembled perceived items ( $M = 36.0\%$ ) were higher, as predicted,

than errors for those with low resemblance [ $M = 26.4\%$ ;  $F(1,59) = 32.84$ ,  $MS_e = 171.2$ ,  $p < .001$ ]. In the analysis on only perceived items, physical resemblance was marginally significant [ $F(1,59) = 3.52$ ,  $MS_e = 172.3$ ,  $p < .07$ ], with physically similar pairs ( $M = 23.2\%$ ) producing more errors than control pairs ( $M = 20.0\%$ ).

**Conceptual relation.** With the definition of relatedness limited to functional category membership, reality monitoring error rates were higher for related ( $M = 29.2\%$ ) than for unrelated ( $M = 23.6\%$ ) pairs [ $F(1,59) = 22.41$ ,  $MS_e = 168.6$ ,  $p < .001$ ].

Again, the prediction regarding this effect was for imagined items in particular, and this prediction was supported by a marginally significant conceptual relation  $\times$  source interaction [ $F(1,59) = 3.57$ ,  $MS_e = 174.5$ ,  $p < .07$ ]. As predicted, imagined items that were related to perceived items ( $M = 35.1\%$ ) produced more errors than those that were not related [ $M = 27.3\%$ ;  $F(1,59) = 24.78$ ,  $MS_e = 150.7$ ,  $p < .001$ ]. Although a similar pattern occurred for perceived items (related  $M = 23.3\%$ , unrelated  $M = 20.0\%$ ), the difference was small and only marginal [ $F(1,59) = 3.46$ ,  $MS_e = 192.3$ ,  $p < .07$ ]. Apparently people use functional information to probe memory for evidence of perceived events. So they are liable to make errors for imagined items that are related to perceived items and are less likely to make the opposite error.

**Orienting task.** As in Experiment 1, orienting task did not reliably affect error rates (perceptual  $M = 26.5\%$ , compound  $M = 26.3\%$ ), nor did it interact with resemblance or conceptual relation.

**Source.** As expected, the subjects were more likely to claim that an imagined item was perceived ( $M = 31.2\%$ ) than to claim that a perceived item was imagined [ $M = 21.7\%$ ;  $F(1,59) = 9.97$ ,  $MS_e = 1,119.1$ ,  $p < .01$ ].

The possibility that source monitoring errors may be influenced by differential recognition rates between the various experimental conditions (see, e.g., Murnane & Bayen, 1996) cannot completely account for the present findings, because recognition accuracy was greater for imagined than for perceived items,<sup>5</sup> whereas source monitoring accuracy was lower for imagined than for perceived items. Furthermore, when the data were analyzed in a multinomial model, which examines the independent contributions of item recognition and source memory, the basic conclusions drawn here are upheld (see Appendix B).

Overall, the results replicate the findings of physical similarity in Experiment 1. In addition, they show that reality monitoring errors can be influenced by conceptual information derived from other specific memories, given a stricter definition of relatedness. Under the conditions in Experiment 2, physical resemblance and conceptual relation produced roughly the same rate of errors for imagined objects. More important, they seemed to do so in an additive way: The highest rate of errors was for imagined objects that both resembled and were conceptually related to perceived items. This supports the argu-

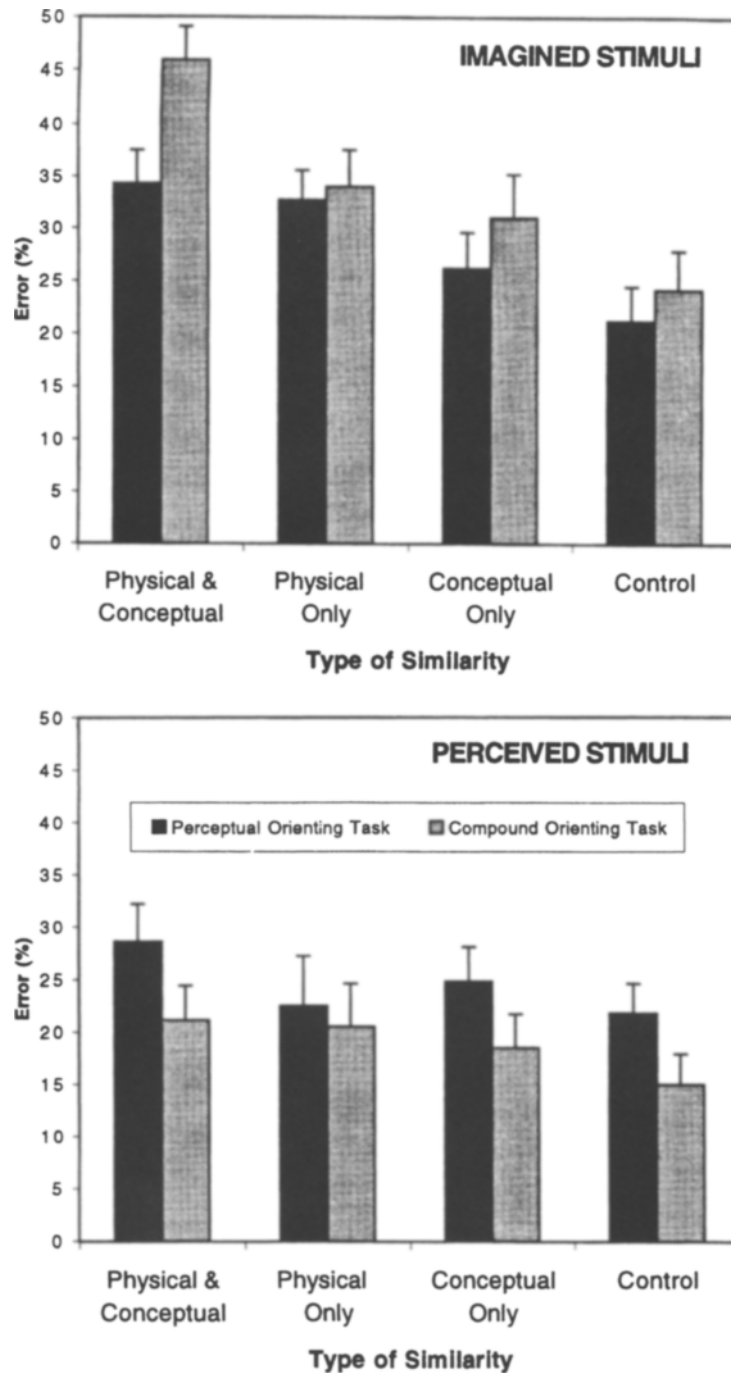


Figure 3. Percentage of reality monitoring errors for imagined and perceived stimuli in Experiment 2.

ment that reality monitoring makes use of several kinds of information in memory that may contribute separately to a decision (Johnson & Raye, 1981).

The finding that the presence of other, similar memories can affect reality monitoring for a target fits nicely with the findings in other literatures, as when subjects make schema-based errors in semantically rich settings (see, e.g., Brewer & Treyens, 1981) or association-based

errors in less rich settings (see, e.g., Roediger & McDermott, 1995). Often, such effects can be attributed to top-down reconstructive processes at test that generate information that results in source misattributions. For example, Brewer and Treyens' findings can be attributed to schema-based inference, and Roediger and McDermott's findings can be attributed to priming or inference based on semantic context. In contrast, at least some of the present



errors must be due to the evaluation of perceptual features, because some pairs had only a physical relation.

**EXPERIMENT 3**  
**Number of Confusable Memories**

The first two studies suggest that people may falsely believe that they have seen something they only imagined if they have perceived an object with similar perceptual or semantic features. Given that features belonging to other memories can influence reality monitoring, errors for imagined items should increase with the number of similar memories derived from perception.

Other findings provide indirect support for this prediction. Source misattribution errors in an eyewitness testimony paradigm increase with increased exposures to misleading postevent information (Zaragoza & Mitchell, 1996), and estimates of the number of times an item was physically presented increase not only with the number of actual presentations but also with the number of imagery trials (Johnson et al., 1979; Johnson et al., 1977; see also Rabinowitz, 1990). In the present study, however, the subjects do not experience the same item repeatedly. They experience the same sensory-structural features repeatedly by seeing and imagining different objects that resemble each other.

To keep the number of conditions and stimuli manageable, the study was designed to examine reality monitoring errors for imagined items in particular, and physical resemblance but not conceptual relation was manipulated. The subjects perceived and imagined black-and-white line drawings of objects and performed only the perceptual orienting task on each. For example, suppose that, over the course of the encoding phase, you imagined a *shovel* and perceived a *broom*, an *oar*, and a *flyswatter*. Are you more likely to think that you have seen the shovel than you would be if you had imagined the shovel and perceived only the broom?

To understand why errors may increase as evidence from structurally similar nontarget items accrues, reality monitoring errors that occurred when subjects imagined an item and perceived three perceptually similar items are compared with errors that occurred when they imagined all four items. Such multiple imagery trials in imagining several similarly shaped objects should also create memories with features similar to the target. However, the perceptual evidence should be richer or more detailed when several items with those features are actually perceived. Thus, errors in claiming that the target imagined item had been perceived should be higher when the three other items it resembles had been perceived rather than imagined.

**Method**

**Subjects.** Forty-seven undergraduates at SUNY at Stony Brook participated, either to fulfill a course requirement or for \$8. All were native English speakers.

**Materials and Design.** A pool was created consisting of sets of four physically similar objects.<sup>6</sup> Items were selected to be concep-

tually unrelated both within and across sets. Each set of items was randomly assigned to one of the experimental conditions.

The stimuli were 112 slides, 56 for imagined and 56 for perceived trials. Each critical imagined item belonged to one of four conditions ( $n = 8$  for each condition): (1) the imagined item resembled one perceived item, (2) the imagined item resembled three perceived items, (3) the imagined item resembled three other imagined items but did not resemble any perceived items, and (4) the imagined item resembled no other items. This allowed for the comparison of two nonzero levels of resemblance to perceived items, and it allowed for the comparison of many perceived items to many imagined items. These conditions satisfied the primary purposes of the study while maintaining as simple a design as possible (see Table 1). Eight sets of items were used in each of the imagined and perceived conditions. To provide equal numbers of imagined and perceived trials, 16 additional pictures were shown as fillers. They were selected according to the same criteria as the experimental items and were neither physically nor conceptually related to any other object. Stimulus order was determined as in Experiment 1. In the sets of one imagined and three perceived items, the imagined item occurred equally often in each of the four positions in the order of trials.

**Procedure.** The general procedure was identical to that in Experiments 1 and 2. All subjects received all experimental conditions, and they all performed the perceptual orienting task. The encoding phase consisted of 112 slides shown for 7 sec each, and testing occurred 2 days after encoding. For the reality monitoring task, all 112 experimental trials plus 48 new items were randomly ordered. The new items neither physically resembled nor were conceptually related to any old items.

**Results and Discussion**

Two subjects did not return for the second session, and data for 1 subject were destroyed through computer error. The data from the remaining 44 subjects were analyzed. Overall recognition accuracy was 77.0%, whereas overall reality monitoring accuracy was 73.8%.

Source errors for imagined items were examined in an ANOVA comparing resemblance to zero, one, and three perceived items and three other imagined items (see Fig-

**Table 1**  
**Design of Imagined and Perceived Conditions in Experiment 3**

	Examples	
	Imagined	Perceived
Imagery Conditions		
Resemblance of each imagined object to:		
Three perceived items	barbell	baton cotton swab dog bone wine bottle
One perceived item Three imagined items	bowling pin broom oar fly swatter shovel	—
No other items	tea kettle	—
Perceived Conditions		
Resemblance of each perceived object to:		
One imagined item and two perceived items	ice cream cone	funnel spinning top microphone ladder
One imagined item No other items	railroad track	belt

ure 4). Errors were predicted to increase over baseline when a physically similar item was perceived and to increase with the number of physically similar perceived items. In the first three imagery conditions, there was only one imagined item for each four-item set. Therefore, in the condition in which all four items were imagined, one item from each set was randomly selected as the target.

The main effect of condition was significant [ $F(3,129) = 15.81$ ,  $MS_e = 210.7$ ,  $p < .001$ ]. Post hoc  $t$  tests showed subjects to be more likely to claim that they saw an imagined item when it resembled three perceived items ( $M = 37.4\%$ ) than they were when it resembled one perceived item ( $M = 24.4\%$ ), three imagined items ( $M = 21.6\%$ ), or no other items ( $M = 17.2\%$ ). Further, they claimed to have seen an imagined item more often when it resembled one perceived item than they did when it resembled no other items. Both predictions thus held: The existence of physical similarity and the number of physically similar perceived items affected reality monitoring for imagined items. In contrast, imagining three physically similar items did not significantly affect performance relative to the no-other-item condition.

As in Experiments 1 and 2, perceiving a physically similar item increased reality monitoring errors for imagined items. But two new comparisons tell us more about the information used in reality monitoring decisions. First, the subjects were more likely to claim that they had seen an imagined item when it resembled several perceived items than they were when it resembled only one. That is, reality monitoring is sensitive to increases in evidence that sensory information was perceived, even when that evi-

dence derives from other events. Shape, not objects, varied in frequency of presentation, so we can say with confidence that reality monitoring utilizes sensory information.

Second, errors did not increase when several similar objects had been imagined. Although imagining four physically similar items provides the same number of *exposures* to the target shape as does imagining one and perceiving three, the quality of memories for that shape evidently differs. Thus, the difference in error rates between the imagine-four and the imagine-one/perceive-three conditions supports Johnson et al.'s (1993) source monitoring model. Memories derived from perception apparently possess more physical detail, contextual information, and other evidence of a perceived source than do those from imagery. The sensory perceptual information derived from *perceived* nontarget events is more likely to provide misleading cues as to the target memory's origin than is the information derived from *imagined* nontarget events.

Errors appear to arise when reality monitoring processes fail to identify what item was associated with particular sensory information. For example, suppose that a subject imagined a barbell and perceived a baton, a cotton swab, and a dog bone (as in Table 1) and was later probed at test with *barbell*. It is as if subjects recover evidence of having seen a barbell-shaped object but do not identify the item(s) that provided that evidence. This could lead to the mistaken belief that the barbell was seen.

These errors cannot be accounted for solely by differences in recognition rates, because recognition performance for imagined items did not differ across the experimental conditions.<sup>7</sup> An analysis comparing source

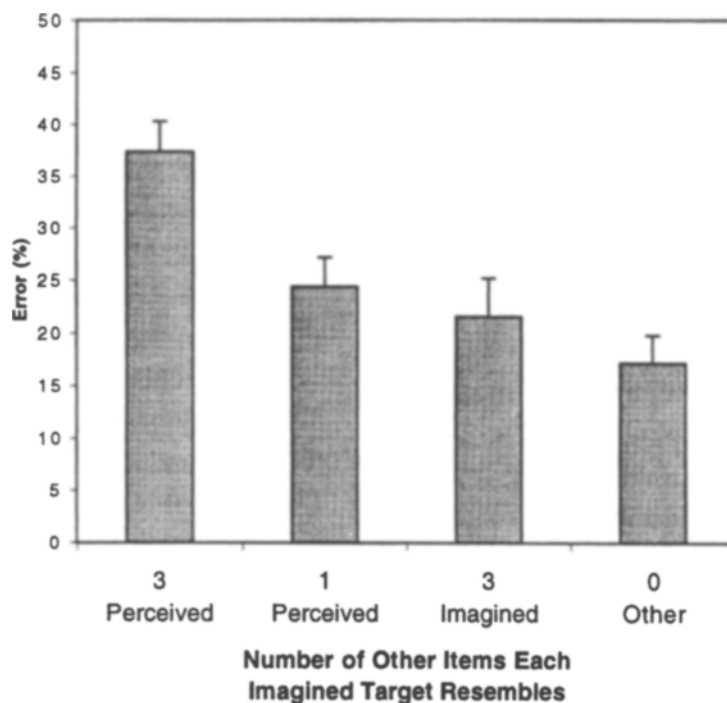


Figure 4. Percentage of reality monitoring errors for imagined stimuli in Experiment 3.

errors for imagined and perceived items is not appropriate in the ANOVA, because the experiment was designed to compare errors for imagined objects in several different conditions, and thus imagined and perceived items were allowed to differ on other dimensions as well. (See Appendix C for the results of multinomial analyses on these data, confirming these findings with independent parameters estimating recognition and source discriminability.)

This study leads to two new conclusions. First, evidence for perceptual experience can accumulate over memories for several objects possessing a given sensory feature (here, a given shape). Reality monitoring processes are sensitive to the strength of this evidence, and, because the information derives from different objects that are related only by virtue of their structural similarity, such evidence clearly can be based on structural information. Second, strength is not determined by simple frequency; reality monitoring processes distinguish between the cases of imagine one/perceive three and imagine four, even though both involve the same number of different objects possessing similar shapes. The two conditions differ in the quality of sensory information, and this information is evaluated and used in reality monitoring.

It is possible that the increase in errors that we attribute to *number* of items is actually due to the increased total viewing *time* for the target shape.<sup>8</sup> However, in either case, the perceived features come from different items, indicating that reality monitoring decisions are affected by physical features contributed by several different items.

## GENERAL DISCUSSION

Memories often provide rich and reliable information that guides us in making decisions as to their source. Johnson and colleagues have shown in numerous studies that we evaluate a target memory with respect to the average features of memories for imagined and perceived events (see, e.g., Johnson et al., 1993). This is generally a useful heuristic, but errors sometimes arise, as when people claim that they saw an object after imagining its shape, color, and texture especially vividly.

The present studies show that not only is reality monitoring influenced by the target memory's features in relation to the average features of perceived and imagined memories but that it can be influenced by the features derived from other *specific* memories. That is, reality monitoring processes act on information derived not only from the target event but from other events as well.

One factor affecting judgment of source is physical resemblance. The present studies show that other items with shapes similar to the target's affect people's ability to determine the target's source. This finding augments prior studies showing that reality monitoring operates on memory for sensory information (see, e.g., Johnson, Foley, & Leach, 1988) and that greater similarity between memories derived from different sources can lead to more source errors (see, e.g., Johnson et al., 1995). To return to the example at the beginning of the paper, it is as if, given the probe *lollipop*, people search memory for sensory infor-

mation associated with the object. If they find vivid instances of such features (and do not find other indications challenging the decision), they return a decision of *perceived*.

These findings, then, show not only that people use features of a target memory as cues for judging source but that such features need not be bound to the target object in order to be used. Rather, the features from other memories can provide information that is used in determining a target memory's source. Such similarity can create source errors, as shown here when people confuse having imagined an object with having seen a different object that looks like it. Further, the present results show that, as evidence increases for the perceptual experience of a feature (as when an imagined item resembles three perceived items), so does the likelihood of claiming that an object possessing that feature was perceived.

In addition to operating on structural information, reality monitoring appears to operate simultaneously on more conceptual information, which is consistent with other findings in the source monitoring literature (see, e.g., Johnson et al., 1981; Lindsay et al., 1991). If conceptual information associated with an imagined object is associated with another, perceived item, the source of the two memories may be confused. It is as if the probe *lollipop* activates a function (candy) and the memories of objects serving that function. Again, if the rememberer finds perceptually derived instances and no indications to the contrary, the likelihood that the rememberer will return a decision that the target was perceived increases.

It remains to be seen what sorts of conceptual information influence reality monitoring. The current work makes headway on this front and identifies function as a particularly important type of conceptual information (Experiment 2). The effect disappeared when co-occurrence and second-order association were also used to define relatedness (Experiment 1), although perhaps in a conceptually richer context, such as in an eyewitness testimony paradigm, such relations would affect performance.

Clearly, additional research can address these and other important issues emerging from the findings of the present studies. For example, it may be the case that the source errors found here arise primarily during retrieval and judgment processes at test rather than during encoding. However, this does not rule out source misattributions that arise from distortions or blends in features from specific perceived and imagined events during encoding (see, e.g., Metcalfe, 1990), which in certain settings should also lead to source misattributions. Two pieces of evidence lend support to the idea that the errors found in the present studies are likely arising at test, but this evidence is far from conclusive. First, the finding from Experiment 1 that people falsely claimed to have seen an item that was brand new when that item *physically resembled* something that they did in fact see suggests that such errors can indeed take place at the stage of evaluating a probe. Second, and perhaps more important, no differences in source errors were found as a function of the order of the members of a similar pair during encoding.

That is, errors were not influenced by whether the imagined item preceded or followed its corresponding similar perceived item in the order of slides. Thus, errors do not depend on the imagined member of a pair having been biased by the subjects having already seen the perceived member. Nonetheless, research aimed at dissociating encoding from retrieval processes is needed to adequately address this issue.

The findings regarding orienting task suggest that variations in the proportion of perceptual or conceptual processing at encoding do not affect errors for perceptually or conceptually similar items. But because both orienting tasks used here involved processing an item's appearance and had explicit imagery instructions, more specific manipulations of orienting task would be necessary before we can draw firm conclusions. Emphasis on the processing of structural and/or conceptual information affects performance in standard memory paradigms, as when perceptual processing boosts recall to structural retrieval cues (Morris, Bransford, & Franks, 1977) or boosts performance in an anticipated item recognition test (Frost, 1972). For reality monitoring, however, the set of features that are attended to at encoding may be a relatively less informative cue as to a memory's source than are more stable features of the item itself. Whether one attends to particular features of an item is more a function of one's goals than of the item or its source. In fact, other work has shown that orienting tasks that increase item memory do not necessarily increase source monitoring accuracy (Johnson, Nolde, & DeLeonardis, 1996; Lindsay & Johnson, 1989), which suggests that focusing attention on particular features of an event does not necessarily provide reliable cues as to the source of the event.

Reality monitoring is faced with the challenge of sorting through various aspects of memories and determining the source of bits and pieces of them. From what we know of how it works, it is well adapted for making use of available information, and, in many situations, that available information will guide correct judgments as to a memory's source. Misattributions of a memory's source can arise apparently when the available information provides misleading cues as to the memory's source, such as when a memory for an imagined item has features more typical, on the average, for perceived memories, or, as shown here, when the information for having perceived certain physical or conceptual features is actually derived from other nontarget memories.

#### REFERENCES

- BATCHELDER, W. H., & RIEFER, D. M. (1990). Multinomial models of source monitoring. *Psychological Review*, *97*, 548-564.
- BAYEN, U. J., & MURNANE, K. (1996). Aging and the use of perceptual and temporal information in source memory tasks. *Psychology & Aging*, *11*, 293-303.
- BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 197-215.
- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.
- BREWER, W. F., & TREYENS, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207-230.
- BROWN, R., & McNEILL, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning & Verbal Behavior*, *5*, 325-377.
- 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 & Verbal Behavior*, *19*, 416-429.
- FERGUSON, S. A., HASHTRoudI, S., & JOHNSON, M. K. (1992). Age differences in using source-relevant cues. *Psychology & Aging*, *7*, 443-452.
- FOLEY, M. A., & JOHNSON, M. K. (1985). Confusions between memories for performed and imagined actions. *Child Development*, *56*, 1145-1155.
- FROST, N. (1972). Encoding and retrieval in visual memory tasks. *Journal of Experimental Psychology*, *95*, 317-326.
- INTRAU, H., & HOFFMAN, J. E. (1992). Reading and visual memory: Remembering scenes that were never seen. *American Journal of Psychology*, *105*, 101-114.
- JOHNSON, M. K. (1983). A multiple-entry, modular memory system. *Psychology of Learning & Motivation*, *17*, 81-123.
- JOHNSON, M. K., DELEONARDIS, D. M., HASHTRoudI, S., & FERGUSON, S. A. (1995). Aging and single versus multiple cues in source monitoring. *Psychology & Aging*, *10*, 507-517.
- JOHNSON, M. K., FOLEY, M. A., & LEACH, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, *16*, 337-342.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*, 371-376.
- JOHNSON, M. K., HASHTRoudI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.
- JOHNSON, M. K., NOLDE, S. F., & DELEONARDIS, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory & Language*, *35*, 135-156.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67-85.
- JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, *94*, 37-64.
- JOHNSON, M. K., RAYE, C. L., WANG, A. Y., & TAYLOR, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 229-240.
- JOHNSON, M. K., TAYLOR, T. H., & RAYE, C. L. (1977). Fact and fantasy: The effects of internally generated events on the apparent frequency of externally generated events. *Memory & Cognition*, *5*, 116-122.
- KAHAN, T. L., & JOHNSON, M. K. (1990). Memory for seen and imagined rotations of alphanumeric characters. *Journal of Mental Imagery*, *14*, 119-130.
- LINDSAY, D. S. (1994). Memory source monitoring and eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 27-55). New York: Cambridge University Press.
- LINDSAY, D. S., & JOHNSON, M. K. (1989). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society*, *27*, 111-113.
- LINDSAY, D. S., JOHNSON, M. K., & KWON, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, *52*, 297-318.
- METCALFE, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General*, *119*, 145-160.
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, *16*, 519-533.
- MURNANE, K., & BAYEN, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition*, *24*, 417-428.
- RABINOWITZ, J. C. (1990). Effects of repetition of mental operations on memory for occurrence and origin. *Memory & Cognition*, *18*, 72-82.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Ex-*

perimental Psychology: Learning, Memory, & Cognition, 21, 803-814.

SHEPARD, R. N., & CHIPMAN, S. (1970). Second-order isomorphism of internal representation: Shapes of states. *Cognitive Psychology*, 1, 1-17.

TVERSKY, B., & HEMENWAY, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, 113, 169-193.

ZARAGOZA, M. S., & MITCHELL, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science*, 7, 294-300.

NOTES

1. To gather norms for a large number of objects, several forms with different subsets of items were used. The number of raters for various objects ranged from 30 to 40.
2. Errors did not differ significantly when the imagined member of a pair preceded or followed the perceived member of the pair in the order of slides during encoding, and, thus, data were collapsed across order of pair members for all analyses.
3. Recognition accuracy was higher for imagined ( $M = 77.8\%$ ) than for perceived ( $M = 75.4\%$ ) items [ $F(1,67) = 63.65, MS_e = 9.1$ ] and was higher for the compound ( $M = 81.8\%$ ) than for the perceptual ( $M = 71.4\%$ ) orienting task [ $F(1,67) = 8.99, MS_e = 1,247.3$ ]. Recognition accuracy varied as a function of pair type [ $F(2,134) = 3.92, MS_e = 6.57$ ], with physically similar and conceptually related items both having higher recognition than did control items. The only significant interaction for the recognition data, that between orienting task and source, is clarified by the multinomial analysis reported in Appendix A.
4. Resemblance was normed in the same manner as for Experiment 1. Mean ratings were 4.5 or higher for high-resemblance pairs and 3.0 or lower for low-resemblance pairs. Ratings did not differ significantly between the high resemblance/high conceptual relation and high resemblance/low conceptual relation conditions. The four conditions did not differ significantly in terms of physical complexity or ease of imaging of the respective items. Conceptually related items were selected by the consensus of three trained judges.
5. The independent variables again differentially affected recognition and reality monitoring performance. Certain recognition results are reported here to demonstrate this. Because of the complexity of the analysis, some interactions are not reported, because they do not bear on the issues at hand. As in Experiment 1, recognition accuracy was significantly higher for imagined ( $M = 80.5\%$ ) than for perceived items [ $M = 78.5; F(1,59) = 61.58, MS_e = 6.69, p < .001$ ]. In general, recognition for imagined items was not affected by orienting task, physical resemblance, or conceptual relatedness, whereas recognition for perceived items was affected by these variables. Consistent with the marginally significant interaction in Experiment 1, the compound orienting task increased recognition accuracy for perceived items ( $M = 82.5\%$ ), relative to the perceptual orienting task ( $M = 74.6\%$ ), whereas, for imagined items, no difference was found between the compound ( $M = 82.8\%$ ) and perceptual ( $M = 78.1\%$ ) orienting tasks.
6. Items were selected on the basis of pilot subjects' ratings of resemblance on a 7-point scale (7 = "very similar"). Mean ratings were 4.0 or higher overall for the pairs of each experimental set.
7. Recognition accuracy was higher for imagined items ( $M = 79.9\%$ ) than for perceived items ( $M = 75.9\%$ ). Across the four imagery conditions, recognition accuracy did not differ [ $F(3,129) < 1$ ], whereas for the perceived items, recognition was greater for the imagined one/perceive three condition than for the imagine one/perceive one condition [ $F(2,86) = 5.10, MS_e = 5.72$ ].
8. We thank Ute Bayen for this suggestion.

APPENDIX A

Multinomial Analysis of Experiment 1

Multinomial models of source monitoring estimate separate and independent parameters for item detectability (recognition), source discriminability (reality monitoring), and various kinds of response bias (see, e.g., Batchelder & Riefer, 1990;

Bayen, Murnane, & Erdfelder, 1996). The data from Experiment 1 were sorted into  $3 \times 3$  matrices containing the frequencies with which the subjects indicated whether an item was perceived, imagined, or new for each test item source (perceived, imagined, new). These frequency matrices are shown in Table A2.

The matrices were analyzed with Model 6c from Batchelder and Riefer's (1990) multinomial models, which provided a good fit for the data, with  $G^2 < 1$  for each  $3 \times 3$  table. This model was chosen because it was expected that recognition and source accuracy would differ for perceived and imagined items, as indeed was found. The parameter estimates of item and source memory for each condition are reported in Table A1. Higher values of  $D$  or  $d$  reflect higher accuracy in old/new recognition and source discriminability, respectively, whereas a value of zero on either parameter would indicate that discrimination was at chance levels.

Log-likelihood ratio tests ( $G^2$ ) on the parameters were consistent with the empirical findings from the ANOVAs for the key experimental comparisons. The critical value for comparisons with three matrices is 5.99 for significance at the .05 level, and, for comparisons between two matrices, 3.84. Source discriminability for imagined items ( $d2$ ) differed significantly across type of similarity ( $G^2 = 14.46$ ). Comparisons reveal that source discriminability was significantly lower when the imagined item physically resembled a perceived item than when it was conceptually related or had no similarity to a perceived item, and this pattern held for both orienting tasks (resemblance vs. conceptual relation,  $G^2 = 12.25$ ; resemblance vs. control,  $G^2 = 8.38$ ; conceptual relation vs. control,  $G^2 = 0.34$ ). Source discriminability for perceived items ( $d1$ ) was not influenced by type of similarity for either orienting task ( $G^2 = 2.48$ ).

Table A1  
Parameter Estimates for Experiment 1 Based on Model 6c

Orienting Task	Type of Similarity	Parameter Estimates				$G^2$
		$D1$	$D2$	$d1$	$d2$	
Perceptual	Physical	.66	.89	.60	.23	.00
	Conceptual	.67	.88	.51	.47	.00
	Control	.56	.87	.51	.45	.00
Compound	Physical	.82	.84	.72	.07	.00
	Conceptual	.84	.86	.67	.33	.00
	Control	.80	.84	.68	.27	.00

Note— $D1$ , detectability of perceived items;  $D2$ , detectability of imagined items;  $d1$ , source discriminability of perceived items;  $d2$ , source discriminability of imagined items;  $G^2$ , goodness-of-fit to Model 6c.

Table A2  
Response Frequencies for Perceived (P), Imagined (I), and New (N) Items in Experiment 1

Similarity	Source	Response Frequency					
		Perceptual Orienting Task			Compound Orienting Task		
		P	I	N	P	I	N
Physical	P	260	91	104	310	71	61
	I	153	270	32	143	244	55
	N	207	238	920	104	159	920
Conceptual	P	248	107	100	302	83	57
	I	106	312	37	106	287	49
	N	207	238	920	104	159	920
Control	P	219	101	135	292	80	70
	I	110	304	41	113	274	55
	N	207	238	920	104	159	920

The orienting task manipulation influenced source monitoring for imagined items, with lower *d2* parameters for the compound than for the perceptual orienting task ( $G^2 = 4.06$ ); but, for perceived items, source discriminability was higher for the compound than for the perceptual orienting task ( $G^2 = 8.39$ ). (This was only a marginal interaction [ $p < .09$ ] in the ANOVA.) The consistency of the findings from the multinomial analysis and the ANOVA lends credence to the argument that the effects we discuss are largely due to differences in discriminating the source of items rather than to differences in recognizing the items themselves.

Looking at recognition parameters, we see that recognition accuracy for imagined items (*D2*) was not influenced by type of similarity ( $G^2 = 0.73$ ) or by orienting task ( $G^2 = 3.52$ ). In contrast, recognition accuracy for perceived items (*D1*) was higher for the compound than for the perceptual orienting task ( $G^2 = 78.19$ ). In addition, recognition for perceived items differed across type of similarity for the perceptual orienting task ( $G^2 = 8.05$ ) but not for the compound orienting task ( $G^2 = 1.61$ ), with improved recognition for both physical and conceptual similarity over control items. These results are consistent with those from the ANOVA.

**APPENDIX B**  
**Multinomial Analysis of Experiment 2**

Batchelder and Riefer's (1990) Model 6c provided a good fit for all cells ( $G^2 < 1$ ), except for the items that were both physically and conceptually similar for the semantic orienting task. Parameter estimates and response frequencies are shown in Tables B1 and B2. Log-likelihood ratio tests comparing the four similarity conditions are significant at the .05 level when the critical value exceeds 7.81. Pairwise comparisons are significant when the critical value is greater than 3.84.

Source discriminability for imagined items (*d2*) replicates the pattern found in the ANOVA. The *d2* parameters differed significantly as a function of type of similarity ( $G^2 = 32.82$ ). Follow-up comparisons revealed that source discriminability decreased when the imagined item physically resembled or was conceptually related to a perceived item, relative to when there was no physical or conceptual similarity (resemblance vs. control,  $G^2 = 9.39$ ; conceptual relation vs. control,  $G^2 = 6.52$ ). Errors were higher when the imagined item both physically resembled and was conceptually related to a perceived item than when the imagined item had just one type of similarity to a perceived item (resemblance + conceptual relation vs. resemblance alone,  $G^2 = 7.13$ ; resemblance + conceptual relation vs. conceptual relation alone,  $G^2 = 9.83$ ; resemblance + conceptual relation vs. control,  $G^2 = 32.15$ ). Source discriminability for perceived items was not significantly influenced by type of similarity to imagined items ( $G^2 = 7.62$ ).

Recognition parameters for imagined items (*D2*) did not differ across the types of similarity ( $G^2 = 4.27$ ) or across the two orienting tasks ( $G^2 = 0.79$ ). Recognition for perceived items (*D1*) was higher for the compound than for the perceptual orienting task ( $G^2 = 131.91$ ). Although other differences were found for item detectability of perceived items (and these are largely consistent with the effects shown in the corresponding ANOVAs), these do not bear directly on the interpretation of the pattern of source errors for imagined items as a function of their similarity to perceived items.

**Table B1**  
**Parameter Estimates for Experiment 2 Based on Model 6c**

Orienting Task	Type of Similarity	<i>D1</i>	<i>D2</i>	<i>d1</i>	<i>d2</i>	$G^2$
	Physical	.55	.92	.69	.32	.00
	Conceptual	.70	.88	.62	.47	.00
	Control	.70	.89	.65	.57	.00
Compound	Physical and conceptual	.88	.88	.66	.00	8.45
	Physical	.86	.91	.71	.14	.00
	Conceptual	.89	.90	.74	.05	.00
	Control	.88	.88	.80	.32	.00

Note—*D1*, detectability of perceived items; *D2*, detectability of imagined items; *d1*, source discriminability of perceived items; *d2*, source discriminability of imagined items;  $G^2$ , goodness-of-fit to Model 6c.

**Table B2**  
**Response Frequencies for Perceived (P), Imagined (I), and New (N) Items in Experiment 2**

Similarity	Source	Response Frequency					
		Perceptual Orienting Task			Compound Orienting Task		
		P	I	N	P	I	N
Physical and conceptual	P	231	86	73	288	75	40
	I	126	239	25	170	195	38
	N	187	205	1,138	109	198	1,274
Physical	P	203	58	129	286	71	46
	I	120	248	22	114	260	29
	N	187	205	1,138	109	198	1,274
Conceptual	P	232	70	88	301	66	36
	I	93	262	35	125	244	34
	N	187	205	1,138	109	198	1,274
Control	P	238	65	87	314	51	38
	I	76	283	31	89	274	40
	N	187	205	1,138	109	198	1,274

**APPENDIX C**  
**Multinomial Analysis of Experiment 3**

Again, Batchelder and Riefer's (1990) Model 6c provided a good fit for all cells ( $G^2 < 1$ ). Parameter estimates and response frequencies are shown in Tables C1 and C2. Source discriminability for imagined items (*d2*) showed the same pattern as

**Table C1**  
**Parameter Estimates for Experiment 3 Based on Model 6c**

Condition	<i>D1</i>	<i>D2</i>	<i>d1</i>	<i>d2</i>	$G^2$
Imagine 1/perceive 1	.44	.89	.32	.54	.00
Imagine 1/perceive 0 (or imagine 0/perceive 1)	.52	.89	.48	.67	.00
Imagine 1/perceive 3	.58	.86	.22	.26	.00
Imagine 4/perceive 0*	–	.85	–	.59	.00

Note—*D1*, detectability of perceived items; *D2*, detectability of imagined items; *d1*, source discriminability of perceived items; *d2*, source discriminability of imagined items;  $G^2$ , goodness-of-fit to Model 6c. \*There were no items that were perceived in this condition; thus, parameter estimates for *D1* and *d1* are not meaningful.

**Table C2**  
**Response Frequencies for Each Condition in Experiment 3**

Condition	Source	Response Frequency		
		P	I	N
Imagine 1/perceive 1	P	125	73	154
	I	76	244	32
	N	233	227	1,652
Imagine 1/perceive 0 (or imagine 0/perceive 1)	P	156	65	131
	I	56	266	30
	N	233	227	1,652
Imagine 1/perceive 3	P	143	95	114
	I	118	196	38
	N	233	227	1,652
Imagine 4/perceive 0*	P	—	—	—
	I	67	244	41
	N	233	227	1,652

Note—P, perceived; I, imagined; N, new. \*There were no items that were perceived in this condition.

that found in the ANOVA ( $G^2 = 27.29$ ). Source discriminability decreased when the imagined item physically resembled three perceived items, relative to when it resembled one ( $G^2 = 10.87$ ) or no other perceived items ( $G^2 = 25.96$ ). When the imagined item resembled three other imagined items,  $d2$  did not differ from when the imagined item resembled no other items ( $G^2 = 1.25$ ). Source discriminability for perceived items ( $d1$ ) was not influenced by type of similarity to imagined items ( $G^2 = 4.83$ ).

Recognition parameters for imagined items ( $D2$ ) did not differ across the different conditions ( $G^2 = 1.12$ ). Recognition for perceived items ( $D1$ ) was higher for the imagine one/perceive three condition than for the imagine one/perceive one condition ( $G^2 = 9.40$ ). These findings are thus consistent with the ANOVA results.

(Manuscript received March 6, 1995;  
 revision accepted for publication June 24, 1996.)