The influence of linguistic labels on source-monitoring decisions

Lisa Geraci
Washington University, St. Louis, MO, USA

Nancy Franklin
State University of New York at Stony Brook, USA

Three studies explored the extent to which people use various object features, including linguistic label, shape, and category membership, to make decisions about the source of their memories. To isolate the influence of each feature, we used items that were related in the following four ways: as synonyms, as similar in shape and category membership, as homographs, or as unrelated. Participants read sentences and either saw or imagined a picture of the critical word's referent. Experiment 1 showed that participants committed more source errors for synonyms (e.g., *rabbit* and *bunny*) than for objects that were conceptually and perceptually similar (e.g., *doughnut* and *bagel*), which produced more errors than unrelated items. However, there was no effect of label, as people did not have more errors for homographs (e.g., baseball *bat* and flying *bat*) than unrelated items. In Experiment 2, presenting the critical word at study was not sufficient to lead people to use an item's label to make source decisions. However, Experiment 3 showed more source errors for homographs than unrelated pairs when semantic context was minimised at study, suggesting that people can use linguistic labels to make source decisions when other information is unavailable.

People rely on the qualities of their memories to determine whether they originated from perception or imagery, often by comparing the characteristics of the retrieved memory to memories generally associated with that source. For example, people may rely on the knowledge that, in general, memories of perceived or experienced events tend to be more vivid than memories of imagined events, which tend to be characterised by more information about cognitive processes associated with elaboration, organisation, retrieval, or identification of information (e.g., Finke, Johnson, & Shyi, 1988; Johnson, Foley, Suengas, & Raye, 1988; Johnson, Hashtroudi, & Lindsay, 1993; Johnson, Raye, Foley, & Foley, 1981). In this paper, we examine the features that people use to decide that an item was perceived or imagined. In particular, we consider whether an item's linguistic label constitutes an important feature for source monitoring. To do this, we examined source errors for imagined and perceived items that share the same label but not the same meaning (homographs), share the same meaning but not the same label (synonyms), share semantic and physical features but not label (similar items), or do not share any features (unrelated). We examined errors for these pairs of items to delineate the conditions under which labels might constitute an important feature for source monitoring.

Correspondence should be addressed to Lisa Geraci, Department of Psychology, Washington University, One Brookings Drive, Campus Box 1125, St. Louis, MO 63130-4899, USA. Email: lgeraci@artsci.wustl.edu

The first author was supported by NIA grant R01 AG17481-01 A1.

We thank Beth Marsh, Roddy Roediger, and Ayanna Thomas, as well as two anonymous reviewers for their very helpful comments on an earlier version of this manuscript. We also thank Beatrice Belizaire for her help with stimulus design and with running subjects.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210344000116

Recent work has attempted to define some of the object features that people rely on to make source decisions. This work shows that people use sensory features, such as modality (auditory or visual) (Henkel, Franklin, & Johnson, 2000; Markham, Howie, & Hlavacek, 1999) and voice quality (Bayen & Murnane, 1996; Bayen, Murnane, & Erdfelder, 1996; Ferguson, Hashtroudi, & Johnson, 1992; Johnson, DeLeonardis, Hashtroudi, & Ferguson, 1995; Johnson, Foley, & Leach, 1988). Results from these studies show that the more perceptually similar two memories are (e.g., both spoken in a female voice) the more likely people are to confuse their sources.

People also rely on higher-level semantic features to make source decisions. For example, recent work shows that people use stereotypical information to make source decisions. Sherman and Bessenoff (1999) had participants study a list of behaviours attributed to different people (e.g., a priest or a skinhead). Results showed that when source monitoring was difficult, participants incorrectly attributed behaviours to people based on their stereotype of that person; for example, they remembered that the priest (and not the skinhead) gave a stranger a quarter to make a phone call. This study suggests that people can also use semantic information as a heuristic for determining source.

Such heuristics are just that—simple devices that can be evoked to increase one's odds across situations of accurate source attribution. They are certainly not logic-based, and although they sometimes lead to error, they are on the whole quite useful. In the case of paradigms like that of Sherman and Bessenoff, people rely on semantic information, such as person schemas, for judging *a priori* behaviour probabilities. In the case of paradigms that manipulate sensory characteristics, the heuristics apply to the memory features themselves. If a target item is judged as being associated with qualities that are typical of a given source (for example, vividness, which is typical of a perceptual source), then the person is led to attribute the item to that source. This may also occur indirectly, when the features of a target are shared with another item. For example, a participant might be shown a lollipop and, elsewhere in the learning list, be asked to imagine a magnifying glass (Henkel & Franklin, 1998). Because the magnifying glass is associated with physical features (round top with stem) that the participant has a vivid memory for, people are more likely to make erroneous perceptual attributions to it than to control items. That is, people may be induced to incorrectly judge the quality of a memory's feature because of its association to another item for which the feature has quite different qualities (e.g., those associated with perception). We capitalise on this phenomenon in the current studies.

In these studies, we examine another feature that may be useful in source monitoring, an object's label. Label is an interesting feature to study because, in contrast to physical and conceptual features, its role in item identity is arbitrary and only established through convention. Thus, depending on how people use object features in source decisions, one could make different predictions regarding the role of labels. On the one hand, object labels might be irrelevant for source monitoring because they are not necessarily central to an item's identity in any meaningful way. Labels are almost always arbitrary with respect to other features of the object. At the same time, however, labels can be highly diagnostic of their referents because they are strongly associated with them. So, while the word *turtle* is not meaningfully related to its referent, a memory of seeing something labelled *turtle* constitutes good evidence for having actually encountered the item. Thus, because labels are strongly diagnostic of their referents, they may provide a potential source of information, and in some cases the only source of information.

There is evidence that under certain conditions linguistic labels contribute to people's decisions about whether a target memory was perceived or not. In some cases this is due to similarity in phonology. Research shows that people falsely recall words when they have been presented with other phonologically similar words (McDermott & Watson, 2001; Schacter, Koutstaal, Johnson, Gross, & Angell, 1997a; Schacter, Verfaellie, & Anes, 1997b; Sommers & Lewis, 1999; Wallace, Stewart, & Malone, 1995a; Wallace, Stewart, Sherman, & Mellor, 1995b). Sommers and Lewis's (1999) participants studied lists of similar-sounding words (e.g., *fat*, *that*, *sat*, *cab*, ...) and falsely recalled having seen the phonological associate, "*cat*". Interestingly, levels of false recall for phonological associates were similar to previously reported levels of false recall for semantic associates (Roediger & McDermott, 1995), suggesting that phonological word information is an important feature that can influence false recall and recognition judgements.

This work demonstrates two interesting points. First, it shows that linguistic features can lead to false memories. Second, it shows that features of surrounding items can influence memory for a separate target item. This is an important finding because memories are not stored in isolation and specific features from one memory can be confused with features from other surrounding or related memories. These breakdowns in the ability to associate features to their correct source can lead to what have been called ‘‘memory blends’’. In these instances, people incorrectly recognise a blend of two originally separate events (a greenish-blue hue after having seen a green car that was described as being blue; Loftus, 1977; see also Belli, 1988). Similar work shows that people falsely recognise a new object that contains previously seen individual features that are combined in the test item (Reinitz, Lammers, & Cochran, 1992). For example, people falsely recognise newly ‘‘merged’’ words, such as ‘‘insult’’ if they have studied ‘‘instruct’’ and ‘‘consult’’ (Underwood & Zimmerman, 1973). These conjunction errors also occurs with inherently cohesive stimuli, such as faces (Hannigan & Reinitz, 2000; Reinitz & Hannigan, 2001; Reinitz et al., 1992). Here, participants falsely recognise never-before-seen faces which were derived from previously seen features (e.g., eyes, mouth) that were present in two separate faces, particularly when those faces are presented close in time (Reinitz & Hannigan, 2001). Taken together, this work demonstrates that source-monitoring problems can cause people to falsely recognise new items that contain features from originally separate events.

Recent studies have begun to isolate which features cause source-monitoring errors across memories and, in turn, lead to new false memories. To do this, researchers have examined source-monitoring errors for items that share the particular feature of interest. In the Henkel and Franklin (1998) study mentioned earlier, participants studied items with the same general shape. They saw a picture of one item (e.g., a lollipop) and, elsewhere in the study set, they imaged how a picture of the other item (e.g., a magnifying glass) would look if it were perceived. In a subsequent source-monitoring task, participants were more likely to falsely claim to have seen the item that they had imagined (the magnifying glass) when they had seen a picture of a physically related item (the lollipop) than when they had seen a picture of a physically unrelated item. Similarly, when participants studied semantically related items (e.g., they perceived a dollar bill and imagined a penny), they were more likely to falsely claim to have seen the item that they had imagined (the penny) than if they had seen an unrelated item, presumably because they accurately remembered having seen an item from the same general category. These results demonstrate that features from surrounding memories can influence source judgements of target memories. Importantly, they also isolate two critical features (general shape and semantic category) that people use to make source decisions.

We follow this methodology to isolate and examine the use of linguistic labels for source monitoring. To do this, we examined two interesting cases of linguistic labels: synonyms (e.g., paddle and oar) and homographs (e.g., bat: sporting equipment or animal that lives in a cave). In the first case, the label differs between items but the meaning is the same and in the second case the label is the same but the meaning differs. As in Henkel and Franklin, participants in the present experiments studied a picture of one member of the pair and elsewhere in the study set imagined the other member. For instance, participants read sentences that biased one meaning of the homograph (‘‘The only animal who sleeps upside down in a cave is the b____’’) and saw a picture of the implied item (the animal). Elsewhere in the study list, participants read a sentence that biased the other meaning (‘‘The baseball player swung his b____ at the ball’’) and imagined a picture of the critical item (the piece of equipment).

We predicted that people rely on label to make source decisions, particularly when features of the target itself are lacking (as when it was imagined), and when physical and conceptual cues provided by other list items, are minimal. Under such circumstances, which are examined in Experiment 3, label should provide the most useful information for determining source. That is, people should be able to draw on the implication provided by qualities associated with the feature label, ‘‘bat’’, to decide that they had actually seen a bat. So, if an imagined item shares the same label as a perceived item, then at the time of retrieval, participants will have available perceptual information associated with that item (present in its homograph counterpart) to draw upon. If people use label to inform their source decisions, then in the current studies we should see greater source errors in the direction of judging that an imagined item had been perceived for homographs than for unrelated pairs. Although the use of this feature would lead to errors for the imagined items in our paradigm, we believe that it reflects a simple heuristic that would serve people well in the real world.

EXPERIMENT 1

Experiment 1 compared error rates for homographs to error rates from item pairs with other kinds of shared information. In total, this experiment included four pair types (synonyms, conceptually and perceptually similar pairs, homographs, and unrelated pairs). Items in unrelated pairs were completely unrelated to each other (e.g., ‘‘bread’’ and ‘‘plug’’). Homograph pairs consisted of two items with the same name (e.g., ‘‘bat’’) that were otherwise unrelated.[1] Synonym pairs consisted of two different labels that referred to the same object (e.g., ‘‘rabbit’’ and ‘‘bunny’’). Lastly, to replicate earlier work (Henkel & Franklin, 1998) and examine the relative importance of linguistic information to other ‘‘more meaningful’’ information, we also used conceptually and perceptually similar pairs (items with the same general shape and category, e.g., ‘‘banjo’’ and ‘‘guitar’’).

To isolate the role of each feature for source monitoring, we examined errors for imagined and perceived members of each pair. We were particularly interested in source errors in the direction of incorrectly judging that an imagined item had been perceived. This would reflect the standard finding in the literature (see Johnson et al., 1993, for a review), indicating that evidence for perception (or lack thereof) often dominates source decisions. An increase in these types of errors would indicate that people were using evidence of the specific features to determine that the item had been perceived.

If people use labels for source monitoring, then they should commit more source errors for homographs than for unrelated pairs. Further, if people use labels but they are less important than semantic and shape features, then source errors for synonyms and perceptually and conceptually similar items should be highest, followed by homographs, followed by unrelated pairs. Greater errors for synonyms than conceptually and perceptually similar items would indicate a relative disregard for labels in the presence of other more meaningful information. We note that, although the previously described false memory studies show effects of labels under certain circumstances, lexical information is not usually as salient in memory as semantic information. Most research shows that processing information for meaning as compared to more surface-level information, including linguistic features, leads to better memory performance (Craik, 1977; Craik & Lockhart, 1972; Craik & Tulving, 1975). Thus, we predict that while people may use the linguistic information to make source decisions, they will not use it to the same degree as other, more meaningful features such as category and shape. In Experiment 1, which compared all of these relations, this would be manifested by more source errors for synonyms and perceptually and conceptually similar items than for homographs.

Norming study

Two sets of undergraduates who did not participate in any other study provided norming data for our experimental materials. Synonyms and homographs were normed by a single set of 102 participants. For synonyms, participants were shown a simple line drawing and were asked which of two labels (e.g., ‘‘oar’’ and ‘‘paddle’’)[2] they would be more likely to use to refer to the object in the drawing. For homographs, they were given a word that has at least two distinct common interpretations, such as ‘‘lace’’, and were asked to provide the first meaning that came to mind. The two most common definitions for each homograph were used in creating the experimental stimuli. People’s preference for meaning (in homograph pairs) and for label (in synonym pairs) was generally asymmetric, and the norming data were used to select items and counterbalance dominance in the construction of critical stimuli in Experiment 1.

Because the sentences in Experiment 1 used missing words that participants had to supply, these sentences were normed to ensure that participants would produce the appropriate critical word. Each sentence, which included the missing word’s first letter followed by a blank of standard length, was given to 94 participants in the norming study (e.g., ‘‘The prisoner was confined to his

[1] For a few of our homograph pairs (e.g., horn), one meaning was originally derived from the other, although we doubt this semantic association still exists in the average person’s representation of these words.

[2] Participants were American speakers of English for whom these items are generally indistinguishable.

c____"). They were informed that the missing words always referred to a single concrete object, and they were asked to identify it in each sentence. Each study sentence used in Experiment 1 was completed correctly by at least 90% of the participants.

Method

Participants. A total of 40 native English-speaking undergraduates at the State University of New York at Stony Brook participated in Experiment 1 for research credit in their psychology course. All participants were tested individually in two sessions that occurred 4 days apart.

Design. We used a 2 × 4 within-subject design, in which Encoding Type (imagined or perceived) served as one variable and Pair Type (homograph, synonym, perceptually/conceptually similar, and unrelated) served as the other variable. Items were counterbalanced across participants so that for a given pair, participants in one condition perceived one member and imagined the other, while this was reversed for the other group. The order of imagined and perceived trials was also counterbalanced across pairs. Lastly, the order of presentation for dominant and non-dominant pair members for synonyms and homographs was counterbalanced within participants and imagined or perceived status was counterbalanced across participants.

Materials. The study list consisted of 20 critical items (10 pairs) for each stimulus type (homograph, synonym, conceptually/perceptually similar, and unrelated pairs), producing a total of 80 critical items. For each study list, half of the items were imagined and half were perceived. The test list also included 10 different nonstudied items (5 pairs) that were from a category of pairs used at study (synonyms, similar, homographs, unrelated), for a total of 40 nonstudied items. All items were concrete singular nouns. Items within each pair type were related on only the characteristic of interest and appeared in the study list at least 20 items apart.

The critical word in each sentence was presented only as its initial letter, followed by a blank of standard length. The following are example sentences for each pair type. For the synonyms, participants would read the following sentences: "The kayak was so light, he could move it with one p____." and "We rowed the canoe back to shore with an o____." to refer to "paddle" and "oar". They also read the sentences, "When he visited friends, he slept on the pull-out c____." and "In the living room you can relax on the s____." to refer to "couch" and "sofa". For conceptually/perceptually similar items, participants read sentences such as, "If in a hurry in the morning, grab a b____ with cream cheese." and "For breakfast, I only want a jelly d____ and some coffee." to refer to "bagel" and "doughnut". As another example, they read, "She wrote her letter with a ball point p____." and "Tomorrow bring a no. 2 p____ and an eraser." to refer to "pen" and "pencil". For the homograph, "bat", people read, "The only animal who sleeps upside down in a cave is the b____." and "The baseball player swung his b____ at the ball.". And for the homograph, "pipe", they read "The professor leaned back in his leather chair and took a drag from his p____." and "We called the plumber to fix the leaky p____.". For the unrelated pair, "book"/"flower", participants read, "Sara went to the library to get a b____ to read." and "On the way to school the boy picked a f____ for his teacher.".

For perceived trials, a line drawing of the referent was displayed in a standard orientation against a blank background and appeared in its most typical portrayal (regardless of the content of the sentence). Line drawings were presented under the study sentence on the computer screen. These stimuli were either taken from Snodgrass and Vanderwart (1980) or were drawn in the same style. For imagined trials, no pictures appeared under the sentences on the screen, and for these trials, participants were instructed to imagine what a line drawing would have looked like if it were provided.

All test sentences (corresponding to both studied and nonstudied critical items) were new. These sentences similarly biased a given interpretation of target words (e.g., "The vampire BAT preys on small animals.") but were new to ensure that people couldn't respond to critical items based only on their memory for the sentence in which they had been presented. For each test sentence, the critical word was presented in its entirety and in capital letters.

Procedure. Each study trial lasted 10 seconds, which pilot testing showed was enough time to read the sentences and perform the tasks. Trials consisted of the following. The participant read a sentence on the computer screen that contained a missing word, cued by its first letter, and was instructed to figure out the missing word. For perceived trials, a line drawing was provided beneath the sentence and participants were told to study the drawing for the entire time it appeared on the screen. For imagined trials, the sentence appeared without a picture beneath the sentence and participants were instructed to visually image what the object would look like if it were provided. Then the screen advanced and, depending on the previous screen, participants were asked to rate either the quality of their image or the quality of the drawing that was presented as good, fair, or poor. This rating was used to encourage participants to attend to the relevant task (studying or imaging the picture). Participants were instructed not to say the critical word aloud, since doing so would provide perceptual experience of the word form. We therefore cannot identify trials for which participants failed to determine the referent, although the norming data indicate that they would have been rare. The type of trial (imagined or perceived) was presented in a mixed fashion.

Participants were given five practice imagery trials with a separate set of items before the experiment began. For each practice trial, they read a concrete noun and formed a visual image of the referent. They were instructed that each image should be in the form of a simple black-and-white line drawing, depicting the object in its most common form, in standard orientation, and without background. Once they formed the appropriate image, they pressed a key to view a drawing of the item. Their goal was to create images that closely matched these line drawings. This process of "training" their imagery by showing them what the line drawing would have looked like occurred only for the pre-experimental practice trials. Participants began the experimental trials only when they were confident that they were able to follow the imagery instructions.

Participants returned 4 days after the first session and took a surprise source-monitoring test. For each test trial, they viewed a new sentence on the computer screen (for example, "The vampire BAT preys on small animals.") and pressed a key to indicate whether the item in capital letters had been imagined or perceived during the first session or whether it was new. Participants were told that none of the test sentences were identical to any they had read during the first session, but that some of the test sentences might provide a similar context to ones encountered in the first session. Participants were to make their judgements for the precise item given in capital letters and for the sense of the word instantiated by the current context. The instructions for the test phase provided several examples using items not presented in the experiment. Participants did not begin the test until the experimenter was confident that they understood the instructions.

At test, participants saw a total of 120 sentences. One-third (40) of the sentences contained a target word for which participants had formed an image, one-third contained a word for which they had seen a picture, and one-third contained a nonstudied item, where nonstudied items were drawn from the same stimulus categories as the critical items. The order of test sentences was pseudo-randomised such that the members of the critical pairs appeared at least 20 items apart in the list.

Results

Source monitoring. We were primarily interested in source-monitoring errors, and in particular, the frequency with which participants incorrectly judged that imagined items had been perceived. Although in principle source errors could occur in either direction, it is a well-established finding in the literature that source judgements for perceived items are relatively more accurate, with less systematic errors. This is believed to be because evidence for perception dominates source judgements. Thus, errors are more easily created by associating imagined items (indirectly) with evidence of prior perception, as we have done in this paradigm. We note that in all experiments, the perceived/imagined distinction concerns the referent (e.g., the flying animal that sleeps in caves), not the word itself (*bat*). Source errors were calculated as a proportion: for example, number of imagined items that were labelled as "perceived" divided by the number of imagined items correctly recognised as studied (hits). The alpha level was set at .05 for all analyses.

We first conducted a 2 $\times$ 4 repeated ANOVA using Encoding Type (originally imagined or originally perceived) and Pair Type (synonym, similar, homographs, and unrelated) as within-subject variables to examine their effect on source errors. We found a main effect of Pair Type, $F(3, 114) = 11.47$, $MSE = 0.06$, showing that source errors were affected by the relationships between items. There was a main effect of Encoding Type on source errors, $F(1, 38) = 9.27$, $MSE = 0.02$, indicating that people were more likely to make errors in the direction of judging that an imagined item was perceived (M = 0.32) than that a perceived item was imagined (M = 0.27). We obtained a significant interaction between Encoding and Pair Type on source errors, $F(3, 114)$ = 7.09, MSE = 0.04, such that false judgements of perception were affected by Pair Type more than false judgements of imagination.

A simple effects analysis examined this finding and showed that the proportion of these critical errors (falsely judging that an imagined item had been perceived) differed across Pair Type, $F(3, 117)$ = 23.88, MSE = 0.05. The mean proportions of these critical errors are shown in the first line of Table 1. Post-hoc Newman Keuls t-tests showed that people were more likely to incorrectly indicate that an imagined item had been perceived if they had also seen a conceptually and perceptually similar item than if they had seen an unrelated one, (M = 0.34 versus 0.20). This replicates earlier work showing additive effects of relatively low-level perceptual information and relatively high-level conceptual information on source-monitoring decisions (Henkel & Franklin, 1998). In addition, results showed that people were sensitive to the degree of semantic and physical overlap: Not only did synonyms produce greater source errors than unrelated items (M = 0.56 versus 0.20), but they also produced greater errors than conceptually and perceptually similar items (M = 0.56 versus 0.34). Lastly, we did not find evidence that people use labels to inform their source decisions because the difference in proportion of source errors for homographs and unrelated pairs was not significant (although we note that the means were in the predicted direction, unrelated pairs = 0.20; homograph pairs = 0.25). These post-hoc analyses showed no effect of Pair Type.

We predicted that these various shared features would influence source monitoring such that perceptual information from related perceived items would increase the likelihood that people would claim to have seen imagined items. Importantly, we did not expect the presence of these shared features to influence source judgements in the opposite direction, incorrectly judging that a perceived item had been imagined. A simple effects analysis supported this prediction and showed that there was no effect of Pair Type on these errors, $F(3, 117)$ = 0.54, MSE = 0.03, (unrelated = 0.29; homograph = 0.27; conceptually/perceptually similar = 0.33; synonym = 0.26).

TABLE 1
Mean proportion of critical source-monitoring errors (calling an imagined item ''perceived'') for each pair type

	Synonym	*Similar*	*Homograph*	*Unrelated*
Exp. 1	.56 (.26)	.34 (.24)	.25 (.22)	.20 (.21)
Exp. 2	.49 (.35)	.33 (.27)	.23 (.26)	.26 (.26)
Exp. 3	–	–	.36 (.24)	.28 (.21)

Standard deviations are shown in parentheses.

Consistent with previous work, when participants falsely recognised new items, they tended to identify them as having been imagined (M = 0.20) more often than perceived (M = 0.06), $F(1, 39)$ = 21.13, MSE = 0.01, presumably because these items had no perceptual information to indicate that they had been perceived (see Hoffman, 1997, for a similar argument). Post-hoc t-tests showed that this was true for each pair type: participants were more likely to call new items ''imagined'' than ''perceived'' for synonyms (M = 0.16 versus 0.06), conceptually/perceptually similar items (M = 0.19 versus 0.05), homographs (M = 0.18 versus 0.08), and unrelated items (M = 0.24 versus 0.07). These results demonstrate that participants did not simply have a tendency to call all synonyms and similar items ''perceived'' when they made source errors. Instead, this error seemed to depend selectively on having imagined the other pair member. An ANOVA showed that the overall tendency to call new items ''imagined'' was affected by Pair Type, $F(3, 117)$ = 3.67, MSE = 0.02, but this effect was largely driven by unrelated new items (unrelated = 0.24; homograph = 0.18; conceptually/perceptually similar = 0.19; synonym = 0.16).

Recognition. Mean proportions of hits and false alarms for each group are shown in the first lines of Table 2. An overall ANOVA on the proportion of hits showed that recognition differed by Pair Type, $F(3, 117)$ = 5.82, MSE = 0.01. Post-hoc Newman-Keuls t-tests showed that recognition was better for synonyms (M = 0.81) than conceptually/perceptually similar items (M = 0.76), homographs (M = 0.72) and unrelated items (M = 0.75), all of which did not differ from each other. Interestingly, recognition accuracy did not appear to be related to source errors, as similar and unrelated items had high recognition, but only similar items had high source errors. Thus, source-monitoring errors do not appear to result simply from recognition familiarity.

TABLE 2
Mean proportion hits and false alarms for each pair type

	Synonym	Similar	Homograph	Unrelated
Exp. 1				
Hits	.81 (.15)	.76 (.14)	.72 (.15)	.75 (.17)
FAs	.11 (.11)	.12 (.11)	.13 (.11)	.16 (.13)
Exp. 2				
Hits	.59 (.21)	.66 (.22)	.59 (.19)	.69 (.21)
FAs	.12 (.14)	.10 (.12)	.16 (.15)	.19 (14)
Exp. 3				
Hits	–	–	.69 (.15)	.71 (.15)
FAs	–	–	.22 (.21)	.22 (.20)

Standard deviations are shown in parentheses.

Discussion

The source-monitoring data replicate previous work and show that conceptual and perceptual information contribute to source decisions. Both synonyms and similar pairs led to greater source-monitoring errors for imagined items than did unrelated pairs.[3] The new finding, that people committed more source errors for synonyms than for similar pairs, indicates that source monitoring is also sensitive to the degree to which such information specifies the target. For similar items, people have evidence of having seen an object (e.g., a bagel) with similar perceptual and functional characteristics as the target item (e.g., an imagined doughnut). For these items, the degree of overlap is quite high. However, for synonym pairs, the degree of overlap is perfect (or near perfect), and this is reflected in the higher source errors for synonyms than for similar pairs.

Interestingly, the proportion of source errors was not greater for homographs than for unrelated pairs, despite the fact that, by necessity, the homographs were encountered twice as many times, and thus the word should have been activated twice as often as the other items. We note that the results did hint at greater source-monitoring errors for homographs than unrelated pairs, but we do not yet have clear evidence that labels are used in source-monitoring decisions. Experiments 2 and 3 examine the question further by considering two possible circumstances under which people might use labels to make source decisions. First, it is possible that, in the presence of other more meaningful information, people do not use labels to make source decisions but that they might under more impoverished conditions. Second, it is possible that people use linguistic labels only if they are physically present at study—that is, if the actual word form itself is presented so that the physical relation between members of a homograph pair is salient. Experiment 2 explores this second possibility.

EXPERIMENT 2

Method

Participants. A total of 36 undergraduates at SUNY Stony Brook participated in Experiment 2 for research credit in their psychology course. All were native English speakers. All were tested individually in two sessions, 5 days apart.

Design and materials. Experiment 2 used the same design as Experiment 1 except that the critical words appeared in the sentences at study. Unlike Experiment 1, the critical word in each sentence was presented in full. Because homograph pairs have the same form, we presented all other pair types twice to equate the number of exposures to the critical words across the four pair types. To do this, we presented each member of the other pair types twice, in two different sentences. However, participants were only required to either imagine or perceive the referent for one of the sentences and were not required to do anything for the other sentence, called ‘‘blank’’ trials. In these, the screen simply advanced to the next item. The order of blank vs critical trial was counterbalanced.

[3] Some readers might find the Imagine/Perceive distinction to be reminiscent of the Generate/Read distinction found in the recognition literature (e.g., Slamecka & Graf, 1978). Specifically, in the Perceive and Read conditions, the participant is presented with the stimulus in its entirety, whereas more cognitive effort is typically required in the Imagine and Generate conditions. The two types of paradigms focus on different phenomena, however. In the Generate/Read recognition paradigm, this greater cognitive effort generally leads to better performance for Generated items. In source-monitoring paradigms, on the contrary, Imagined items typically produce higher levels of source errors than do Perceived items because the quality of their features is more likely to be typical of perceived sources than vice versa. In any case, the results of the current study could not be attributed to a simple Generate/Read type of effect, since we are comparing four conditions that each use Perceived and Imagined items. We thank an anonymous reviewer for pointing out these potential misunderstandings.

Procedure. As in Experiment 1, participants were given several trials to practise the imagery task before the first session, and at study, participants read sentences presented one at a time on a computer screen. In this experiment, unlike Experiment 1, the sentence included the critical word, presented in capital letters. Participants had 5 seconds to read the sentence on the screen and (in non-blank trials) 5 seconds either to see a line drawing that was provided or to imagine how one would look. After each non-blank trial, participants judged the quality of their images or the drawings presented on the screen as good, fair, or poor.

In contrast to Experiment 1, the line drawings were not presented on the same screen as the sentences. Because Experiment 2 included blank trials that did not have pictures associated with them, the simple lack of a picture beneath the sentence could not be used as a cue for the participants to imagine the critical item, since blank trials would also not have a picture associated with them. Thus, three-screen critical trials were introduced in this study. Participants read the sentence and waited for the second screen. Next, the computer screen advanced, and one of three possibilities followed. For perceived trials, participants saw a line drawing of the critical item, which remained on the screen for 5 seconds. For imagined trials, the word ''Imagine'' was presented, and participants had 5 seconds to imagine what a line drawing would have looked like if it had been presented. Then the computer advanced to the third screen, and participants rated the pictures or their images. For blank trials, no second or third screen followed the presentation of the word pair. Instead, the screen simply advanced to the next word pair. As in Experiment 1, the type of trial (imagined, perceived, or blank) was presented in a mixed (not blocked) fashion. The order of the types of presentation (imagined, perceived, or blank) was also counterbalanced across study lists.

At test, participants made simultaneous recognition and source decisions on 120 items. One-third (40) of the sentences contained a target word for which participants had formed an image, one-third contained a word for which they had been given a picture, and one-third contained a nonstudied word.

Results

Source monitoring. As in Experiment 1, we first conducted a 2×4 repeated ANOVA using Encoding Type (originally imagined or originally perceived) and Pair Type (synonym, similar, homographs, and unrelated) as within-subject variables to examine their effect on source errors. We found a main effect of Pair Type, $F(3, 105) = 3.63$, $MSE = 0.02$, showing that source errors were affected by the relationships between items. This time there was no main effect of Encoding Type on source errors, $F(1, 35) = 0.63$, $MSE = 0.22$. Again, we obtained a significant interaction between Encoding and Pair Type on source errors, $F(3, 105) = 10.09$, $MSE = 0.04$, showing that false judgements of perception were affected by Pair Type more than false judgements of imagination.

Again, we were primarily interested in the frequency with which participants incorrectly judged imagined items as having been perceived depending on the relationship between the items in a pair. Mean error rates for each group are shown in the second line of Table 1. A simple effects analysis revealed that the overall proportion of critical errors (judging that an imagined item had been perceived) differed across Pair Types, $F(3, 105) = 10.27$, $MSE = 0.04$. Post-hoc tests replicated Experiment 1 and showed that people were more likely to make these errors if they had seen a picture of the imagined item's synonym ($M = 0.49$) than if they had seen an unrelated item ($M = 0.26$). While people were numerically more likely to incorrectly indicate that an imagined item had been perceived if they had also seen a conceptually and perceptually similar item ($M = 0.33$) than if they had seen an unrelated one ($M = 0.26$), this difference did not reach significance. Importantly, we replicated the finding from Experiment 1 that people were sensitive to identity above and beyond physical and conceptual similarity, comparing synonyms ($M = 0.49$) to similar pairs ($M = 0.33$). Again, however, presenting the word form did not lead people to rely on linguistic label, as critical source errors did not differ for homograph ($M = 0.23$) and unrelated ($M = 0.26$) pairs.

In this experiment, there was also a significant effect of Pair Type on people's tendency to call perceived items imagined, $F(3, 105) = 3.13$, $MSE = 0.03$, but this difference was due to lower errors for synonyms (as opposed to any systematic increase in errors for items with increasing overlap in features (unrelated = 0.31, homograph = 0.29, similar = 0.32, synonyms = 0.21).

Recognition. Mean proportions of hits and false alarms for each group are shown in the second lines of Table 2. An overall ANOVA showed that recognition hits differed for Pair Type, $F(3, 105) = 6.00$, $MSE = 0.02$. As in Experiment 1, post-hoc analyses showed that people recognised both similar ($M = 0.66$) and unrelated items ($M = 0.69$) better than synonyms ($M = 0.59$), which were equivalent to homographs ($M = 0.59$). However, as mentioned above, source errors were still highest for synonyms and similar items. Thus, again, superior recognition did not seem to drive the tendency to call imagined items perceived. In addition, when participants falsely recognised new items, they identified them as having been imagined ($M = 0.12$) significantly more often than perceived ($M = 0.02$), $t(35) = 4.47$, $SE = 0.02$, replicating Experiment 1 and consistent with the standard finding in the literature.

Discussion

These results suggest that providing the word in its entirety at study was not sufficient to lead people to use linguistic information for source monitoring. In Experiment 3, we explored a second factor that might affect use of labels in source monitoring. As discussed earlier, people may use linguistic labels only when other more meaningful information is sparse or not salient. It is possible that in Experiments 1 and 2, people did not rely on word form for source monitoring because the items were presented in the context of sentences. Having this context present would increase errors for synonyms and conceptually/perceptually similar items, because the sentence context would provide little discriminatory information. However, having this context present would decrease errors for homographs because it would provide very useful discriminatory information. In addition, it is possible that the mere presence of the other semantically related items reduced attention to or reliance on label information. So, it may be that when this semantic information is lacking, then people will rely more on label.

To test this hypothesis, in Experiment 3 we limited semantic information in two ways. First, we used only word pairs rather than entire sentences, minimising the semantic context provided by the longer and more detailed sentences. Second, we reduced the larger semantic context by eliminating the synonym and conceptually and perceptually similar pairs from the study set. Thus, participants were less likely to be cued to use semantic information in making source decisions.

EXPERIMENT 3

To establish semantically impoverished conditions, each target word was presented in as minimal a semantic context as possible that would still allow for its intended interpretation. Participants studied homographs, such as "umpire–BAT" and "cave–BAT", and unrelated words, such as "grain–BREAD" and "wall–PLUG". The critical word in each pair was always the second word, which was capitalised, and for each critical word the participant either saw a picture or imagined how it would look.

Method

Participants. A total of 46 native English speakers at the State University of New York at Stony Brook were tested in groups of two to six at a time in two sessions, 4 days apart. Participants received research credit in their undergraduate psychology course.

Design. Item Pair (homograph or unrelated) served as the within-subject variable. Items were counterbalanced so that half the participants imagined a given object and perceived the other member of the pair, whereas the other half did the opposite. As in Experiment 1, the order of the imagined and perceived trials was counterbalanced across participants, as was the order of dominant and subordinate meanings of the homographs. Homographs and unrelated pair members were always at least 20 trials apart in the study list.

Materials. A total of 20 homograph and 20 unrelated pairs were used. All were drawn from Experiment 1. In addition, 20 fillers (e.g., "push–OAR") were constructed from the synonym and similar pairs used in Experiments 1 and 2 (with no more than one member of each pair used in Experiment 3) and were interspersed throughout the study list. The purpose of the fillers was to lower the proportion (and thus the salience) of homograph words in the study set without changing the cue conditions for critical items at test. We took fillers from this set of already used items because these items had been selected to have no semantic or physical overlap with either the homographs or the unrelated pairs.

Targets were always the second word in the paired-associate (e.g., "cave–BAT" or "umpire–

BAT''). The first word was intended to provide a minimal semantic context within which to appropriately comprehend the referent of the critical word. This was particularly important for homograph pairs, whose meaning cannot be discerned in the absence of context. Context words were selected to be loosely associated with the critical items, to minimise the likelihood that the words in a pair would be encoded as a single familiar unit. As in Experiment 2, unrelated pairs appeared twice at study, with different word associates (e.g., ''cut–BREAD'' and ''grain–BREAD''), to equate the number of exposures to word form for each homograph and each unrelated critical item. One trial for each unrelated word was a blank trial, and we counterbalanced whether the blank trial was the first or second trial for the item.

At test, each critical word appeared with a new associate that instantiated the same meaning (e.g., ''flapping–BAT'' corresponding to the studied ''cave–BAT'', or ''shortstop–BAT'' corresponding to the studied ''umpire–BAT''). This was done to avoid the possibility that participants would make recognition and source-monitoring decisions based on the familiarity of the word pairs and not the critical word in the pair.

Procedure. At study, word pairs were presented one at a time on a computer screen. Following this presentation, participants were either shown a picture of the second word in the pair, or asked to image its referent (in its most simple and prototypical form, as in Experiments 1 and 2), or did nothing at all (blank trials). Before the study trials began, participants were told not to anticipate their task following the presentation of the word pair. Rather, they were to wait until the second screen appeared and indicated whether they were to imagine or perceive the item, or simply go on to the next word pair. To guard against the possibility that participants would begin to image all items as soon as they read the word pair, the word pairs were presented for a brief 3 seconds to allow enough time for participants to read the words and to understand their association but little time to form an image.

Next, the computer screen advanced, and one of three possibilities followed. For perceived trials, the second screen depicted a line drawing of the critical item for 5 seconds. For imagined trials, the word ''Imagine'' was presented on the second screen, and participants had 5 seconds to imagine what a line drawing would have looked like if it had been presented. For blank trials, no second screen followed the presentation of the word pair. Rather, the screen simply advanced to the next word pair. The order of the type of presentation (imagined, perceived, or blank) was counterbalanced across study lists. Following the perceived and imagined trials, the screen advanced and participants were asked to judge either the quality of the drawings presented or the quality of their images. They judged them as good, fair, or poor. Again, participants were given several trials to practise the imagery task before the experiment began.

Participants returned 4 days later and took a surprise source-monitoring test. They were told that they would see word pairs again but that the first word in the pair would be one they had not encountered in the earlier session. They were told that this word was there simply to denote the meaning of the second word. Thus, participants were told that they should answer only ''imagined'', ''perceived'', or ''new'' to the second capitalised word in the pair. They were explicitly warned to answer based on the meaning that was instantiated by the first word. There were 60 test pairs. One third of the tests pairs (20) contained a target that had been imagined, one-third contained a target that had been perceived, and one-third contained a new item. All new items were drawn from the same categories as the critical items (homographs and unrelated pairs).

Results

Source monitoring. Again, the critical comparison was the proportion of incorrect source-monitoring judgements in which people claimed that they had perceived an object that they had actually imagined. If people use label information to guide source attributions under these more semantically impoverished conditions, then source errors should be higher for the pairs of objects sharing the same name (i.e., the homographs) than for the unrelated pairs. A one-tailed *t*-test supported this prediction, showing significantly more of these critical source-monitoring errors for the homographs (M = 0.36) than the unrelated pairs (M = 0.28), $t(45) = 2.02$, $SE = 0.04$ (see Table 1). Unlike Experiment 2 but consistent with Experiment 1, source errors in the other direction (calling a perceived item imagined) did not differ for homographs (M = 0.29) and unrelated items (M = 0.31), $t(45) < 1.00$. Thus, these data suggest that the difference in source errors does not reflect a general inability to assess source in the homograph items. Here, it appears that the overlap in linguistic information selectively led people to claim that they had seen items that they had only imagined.

Recognition. The mean proportion of hits and false alarms are presented on the bottom lines of Table 2. Recognition hits did not differ between homographs ($M = 0.69$) and unrelated pairs ($M = 0.71$), $t(45) = 1.20$, $SE = 0.02$. Thus, as in Experiments 1 and 2, source-monitoring errors cannot be explained simply by differences in recognition familiarity. Rather, the source-monitoring differences reported above suggest that evidence of perceptual experience that was associated with perceived items was being applied to an event that had only been imagined, simply by virtue of sharing the same name. That is, under these conditions of decreased semantic context, labels appear to be used as separate cues for source monitoring even in the absence of recognition effects.

As in the first two experiments, and consistent with the literature, when participants falsely recognised new items, they were more likely to identify them as having been imagined ($M = 0.15$) than perceived ($M = 0.07$), $t(45) = 4.06$, $SE = 0.02$. Again, this resulted presumably from the fact that people did not have any perceptual detail associated with these items. When participants falsely recognised items, they were more likely to judge that these items were "imagined" as opposed to "perceived" both when the new item was a homograph, $t(45) = 3.14$, $SE = 0.19$, and when it was unrelated to the other items, $t(45) = 3.61$, $SE = 0.16$. Again, these errors suggest that participants do not have an overall preference to incorrectly call items "perceived" when the source is unknown. Instead, they appear only to do so in the case where they have seen a picture of the other homograph pair member.

One possible alternative account of these results is that when the participants were presented with a homograph, they occasionally brought to mind the other version of the word and created an image that incorporated characteristics of that other interpretation such that the encoding conditions for homograph pairs were quite different from those of control pairs. For example, participants could have read "flapping bat" during the learning phase and generated an image of a batter waving his bat ineffectually at a passing baseball.[4] We think this is unlikely, for the following reasons. First, we were very careful to select homographs that were not semantically related. If one tries hard enough, one can find relationships between any two list words (e.g., the control pair, "shoe and kite" both have dangly strings, and they seem to have more in common, for example, than the two interpretations of "hood" or of "calf"). Second, the data do not support this interpretation. If connections of this sort were driving the results, we would expect similar rates of source errors for the control items in this experiment, but errors were significantly greater for homographs than control items.

GENERAL DISCUSSION

To recap, these experiments explored whether an object's label influences source decisions. To isolate the use of this feature, we examined source-monitoring errors for perceived and imagined pairs of objects that either did or did not share the same label (synonyms, perceptually and conceptually similar items, homographs, and unrelated items). Overall results indicated that people rely on linguistic labels to make source decisions when other more meaningful information is sparse. Experiment 1 showed that when more meaningful information is available (either in the sentence or the overall experimental context), people do not rely on label information to make source decisions. Experiment 2 demonstrated that the use of labels did not depend on the presence of the visual word form. Lastly, Experiment 3 indicated that when the label was present and the semantic context was reduced, people did rely on label for source monitoring. Results from Experiment 3 showed that people made more source-monitoring errors for homographs than unrelated pairs. That is, people were more likely to judge that they had seen an originally imagined item if they had also seen an item with the same name. Because the errors were asymmetric, matching the pattern usually found in the literature of source errors greater for imagined than for perceived items, we can conclude that people were evaluating memory characteristics and sometimes misattributing them to other items with shared features (here, label).

These results add to the literature in two ways. First, they demonstrate that label can influence

[4] We thank J. Richard Hanley for this suggestion.

source decisions, particularly when other relevant features are not available. As such, these data isolate the influence of another object feature on source-monitoring decisions and suggest that this process flexibly depends on the availability of other, presumably more salient features. Second, these data demonstrate that source decisions prioritise highly diagnostic features over less diagnostic ones. We first discuss flexibility and prioritisation of features and then speculate on how label, in particular, might influence source decisions.

The source-monitoring error rates for the variously related pairs of objects provide evidence that people prioritise features based both on (1) how diagnostic the feature is of the item and (2) what other features are available in memory. Experiments 1 and 2 provided evidence that people are sensitive to the degree of perceptual and conceptual relatedness, as evidenced by higher error rates for synonyms (e.g., ''paddle'' and ''oar'') than for items that were only perceptually and conceptually similar in shape and category (e.g., ''bagel'' and ''doughnut''). It is, of course, not clear whether we obtained more errors for synonyms than similar items because people are sensitive to the amount of feature overlap (more feature overlap for synonyms than similar items) or because people are sensitive to types of characteristics that are the most central to the item's identity (i.e., it is the same object and not just physically and semantically similar). Regardless, these data indicate that source decisions flexibly draw on object features and prioritise certain features over others depending on their diagnosticity.

The homograph data provide further evidence that people's source decisions are flexible and depend on the availability of other features contained in surrounding items. In Experiments 1 and 2, when other contextual information was available (the critical items appeared in sentences, and the semantic context was rich), people did not rely on labels to make source decisions. Only when this information was sparse, as in Experiment 3 (which used word pairs and did not contain synonyms or semantically similar items), did people use labels to inform their source decisions. Thus, it appears that the availability of other more salient cues is relevant to whether people use labels to make source decisions. When semantic context (and therefore the availability of semantic cues) was minimised, labels became a valuable means by which to assess source. Although its actual value for source cues does not change with the availability of semantic information, it appears that label is treated as a secondary cue and may not be consulted unless more preferred cues are absent. Such devaluation under normal circumstances is sensible, given that cues provided by labels may be largely redundant with semantic cues. That is, when one encounters the word ''turtle'', one generally also encounters rich semantic (and often perceptual) information that can be later used in memory judgements, including source monitoring. Even when retrieving memories of verbal stimuli such as conversations, one is usually pressed to remember gist rather than precise wording, and so it would not be optimal for word form to be given primary status as a cue to source monitoring. Only when labels emerge as one of the best cues available does their covariance with the remembered item become valuable.

We turn next to the question of how labels may influence source decisions, and in particular, false judgements of perception. First, we note that in the case of the conceptually and perceptually similar items (e.g., ''bagel'' and ''doughnut''), people can base their judgements of perception on having remembered seeing something round and doughy with a hole in the middle. In turn, they may attribute this perceptual experience to either item, including the imagined doughnut. Here, the imagined item's referent is similar to that of the perceived item, and the source of the percept could be easily confused. In these situations, it seems that people are vividly remembering physical attributes that are indirectly associated with the imagined item. However, in the case of homographs, where there is no physical or semantic overlap between what people image and what they perceive, any increase in source errors must be due solely to the use of label. Indeed, results from Experiment 2 rule out the interpretation that an effect of an object's label is based on the visual similarity of word forms, because simply representing the word itself did not increase source-monitoring errors for homographs. This is somewhat different from previous research which has shown that the perceptual properties of the word form directly influence encoding and memory (Costa & Sebastian-Galles, 1998; Dijkstra, Grainger, & van Heuven, 1999). Therefore, the homograph source errors indicate that judgements of perception are not a simple function of orthographic or physical overlap.

This finding could be taken as support for a heuristic basis for judging that an item was perceived. Indeed, other research supports this hypothesis and shows that people commit source errors for items that have no physical similarity but are from the same semantic category (e.g., they think they saw an imagined sleeping bag because they saw a tent). Here, people are using the memory of seeing something from a similar category to judge that they perceived the target item that they only imagined. Remembering having encountered semantic/functional information about a target item may allow this information to be used indirectly to judge perception (Henkel & Franklin, 1998). And, as we have found in the current work, an item's linguistic label, which is even more abstract, can also be used indirectly to judge that something was perceived. We speculate that these errors may occur because people misattribute perceptual information from a related memory (e.g., of a perceived item labelled "bat") to an imagined target. However, we note that our data do not directly point to this process or definitively rule out others. Rather, our data simply establish conditions under which labels appear to influence source decisions.

Our data could be interpreted as consistent with other work showing that people claim to vividly remember items that they never saw. This work shows that people falsely recognise words that were not actually presented, simply because they saw semantically (Roediger & McDermott, 1995) or phonologically (Sommers & Lewis, 1999) related words. Moreover, people report consciously remembering specific supporting associative and perceptual details for these never-before-seen items (Norman & Schacter, 1997). According to one interpretation, people falsely remember seeing nonpresented words because, at the time of encoding, they imagined the associated word (Mather, Henkel, & Johnson, 1997). Although our data are consistent with this interpretation, the lack of semantic associations between homograph pair members and the lack of a homograph effect in Experiments 1 and 2 make this a less plausible explanation for our data.

The present studies address the general question regarding what features affect people's source decisions. We have suggested that these decisions are based on more than just physical similarity of items. In particular, our results suggest that judgements of perception are influenced not only by sensory information about having seen something similar to the target, but also by other associated features, such as an item's label. Label can be used as evidence for perception of a target via its abstract relation to the target and not through any direct overlap of physical or semantic characteristics. In fact, when physical overlap of homograph items is maximised, as with the full-word conditions of Experiment 2, the tendency towards source errors may nevertheless be low. As a secondary source of information about a memory's source, label appears to emerge as important only when other more salient cues are sparse. Taken together, our results fit with the source-monitoring literature and suggest that people's source decisions are impressively versatile, drawing on a range of information about the target itself, one's cognitive processes, the plausibility of remembered details and, importantly, related experiences.

Manuscript received 9 October 2002
Manuscript accepted 9 April 2003
PrEview proof published online 3 September 2003

REFERENCES

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, and Cognition*, *22*, 197–215.

Belli, R. F. (1988). Color blend retrievals: Compromise memories or deliberate compromise responses? *Memory & Cognition*, *16*, 314–326.

Costa, A., & Sebastian-Galles, N. (1998). Abstract phonological structure in language production: Evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 886–903.

Craik, F. I. M. (1977). Depth of processing in recall and recognition. In S. Dornic & P. M. A. Rabbit (Eds.), *Attention and performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

Dijkstra, T, Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 496–518.

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology & Aging*, *7*, 443–452.

Finke, R. A., Johnson, M. K., & Shyi, G. C. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, *16*, 133–137.

Hannigan, S. L., & Reinitz, M. T. (2000). Influences of temporal factors on memory conjunction errors. *Applied Cognitive Psychology*, *14*, 309–321.

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, and Cognition*, *26*, 321–335.

Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 371–383.

Johnson, M. K., De Leonardis, 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., 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.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition*, *5*, 696–699.

Markham, R., Howie, P., & Hlavacek, S. (1999). Reality monitoring in auditory and visual modalities: Developmental trends and effects of cross-modal imagery. *Journal of Experimental Child Psychology*, *72*, 51–70.

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

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*, 838–848.

Reinitz, M. T., & Hannigan, S. L. (2001). Effects of simultaneous stimulus presentation and attention switching on memory conjunction errors. *Journal of Memory & Language*, *44*, 206–219.

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, *20*, 1–11.

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.

Schacter, D. L., Koutstaal, W., Johnson, M. K., Gross, M. S., & Angell, K. E. (1997a). False recognition induced by photographs: A comparison of older and younger adults. *Psychology and Aging*, *12*, 203–215.

Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997b). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology*, *11*, 331–342.

Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*, 106–110.

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

Snodgrass, G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

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.

Underwood, B. J., & Zimmerman, J. (1973). The syllable as a source of error in multisyllable word recognition. *Journal of Verbal Learning & Verbal Behavior*, *12*, 701–706.

Wallace, W. P., Stewart, M. T., & Malone, C. P. (1995a). Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language*, *34*, 417–439.

Wallace, W. P., Stewart, M. T., Sherman, H. L., & Mellor, M. D. (1995b). False positives in recognition memory produced by cohort activation. *Cognition*, *55*, 85–113.