Journal of Experimental Psychology:
Learning, Memory, and Cognition
2012, Vol. 38, No. 1, 1–15

0278-7393/11/$12.00 DOI: 10.1037/a0025053

Source-Constrained Recall: Front-End and Back-End Control of Retrieval Quality

Vered Halamish and Morris Goldsmith
University of Haifa

Larry L. Jacoby
Washington University in St. Louis

Research on the strategic regulation of memory accuracy has focused primarily on monitoring and control processes used to edit out incorrect information after it is retrieved (back-end control). Recent studies, however, suggest that rememberers also enhance accuracy by preventing the retrieval of incorrect information in the first place (front-end control). The present study put forward and examined a mechanism called *source-constrained recall* (cf. Jacoby, Shimizu, Velanova, & Rhodes, 2005) by which rememberers process and use recall cues in qualitatively different ways, depending on the manner of original encoding. Results of 2 experiments in which information about source encoding depth was made available at test showed that when possible, participants constrained recall to the solicited targets by reinstating the original encoding operations on the recall cues. This reinstatement improved the quality of the information that came to mind, which, together with improved postretrieval monitoring, enhanced actual recall performance.

Keywords: memory, metacognition, source-constrained retrieval, recall accuracy

Quality control in manufacturing can be achieved either by a postproduction monitoring process, which identifies and screens out defects at the "back end," or by investing in improved production techniques at the "front end," so that fewer defects are produced in the first place. Jacoby and colleagues (e.g., Jacoby, Shimizu, Daniels, & Rhodes, 2005; Shimizu & Jacoby, 2005) proposed this metaphor as a useful way of viewing the quality control of memory accuracy, emphasizing that both types of memory functions—pre- and postproduction—can be strategically regulated by the rememberer. Building on this idea, in the present study we examined the front-end and back-end control of memory recall quality.

Back-End Control of Memory Quality

Accuracy-oriented memory research, focusing on memory quality rather than quantity (Koriat & Goldsmith, 1996a, 1996b; Koriat, Goldsmith, & Pansky, 2000; Roediger, 1996), has emphasized the back-end control of memory quality: postretrieval monitoring and verification processes used to identify and screen out false memories. This emphasis is well illustrated in the highly influential source-monitoring framework (e.g., Johnson, Hashtroudi, & Lindsay, 1993), which assigns a primary role to postretrieval monitoring processes in attributing the information that comes to mind to a particular source. Accurate memory results from correct source-monitoring decisions, whereas false memories stem from source confusions or reality-monitoring errors. Other frameworks that emphasize postretrieval monitoring attributions and decisions are Jacoby, Kelley, and Dywan's (1989; Kelley & Jacoby 1998; Kelley & Rhodes, 2002) attributional framework, Whittlesea's (1997, 2002) SCAPE (production–evaluation) framework, and proposed memory editing mechanisms such as the distinctiveness heuristic (Dodson & Schacter, 2001, 2002; but see Gallo, Weiss, & Schacter, 2004) and recollection rejection (Brainerd, Reyna, Wright, & Mojardin, 2003).

Work by Koriat and Goldsmith (1994, 1996c; see Goldsmith & Koriat, 2008, for a review) has emphasized the role of postretrieval monitoring and control processes in free-report situations, in which rememberers regulate the accuracy and quantity of the information that they report from memory, either by withholding answers that are likely to be wrong (Koriat & Goldsmith, 1994, 1996c) or by choosing a level of precision or coarseness (grain size) at which the answers are likely to be correct (Goldsmith, Koriat, & Pansky, 2005; Goldsmith, Koriat, & Weinberg-Eliezer, 2002). This work is based on a model of the postretrieval metamemory processes underlying the strategic regulation of memory accuracy and quantity performance (Koriat & Goldsmith, 1996c): When attempting to recount past events, people first monitor the subjective likelihood that each item of information that comes to mind is correct and then apply a report criterion to the monitoring output in order to decide whether to volunteer the item. The setting of the report

This article was published Online First August 22, 2011.

Vered Halamish and Morris Goldsmith, Department of Psychology, University of Haifa, Haifa, Israel; Larry L. Jacoby, Department of Psychology, Washington University in St. Louis.

This research was supported by United States–Israel Binational Science Foundation Grant 2005356. Facilities for conducting the research were provided by the Institute of Information Processing and Decision Making, University of Haifa, and by the Max Wertheimer Minerva Center for Cognitive Processes and Human Performance. The article is based in part on a doctoral dissertation submitted to the University of Haifa by Vered Halamish. We thank Adi Goldberg, Rotem Kahalon, and Naama Mizrahi for their help in running the experiments. We also thank Phil Higham and Jeff Toth for helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Vered Halamish or Morris Goldsmith, Department of Psychology, University of Haifa, 31095 Haifa, Israel. E-mail: halamish@research.haifa.ac.il or mgold@research.haifa.ac.il

criterion depends on the relative utility of providing complete versus accurate information: The stronger the incentives for accuracy, the more selective people are in their reporting.

Koriat and Goldsmith's (1996c) model was supported in several experiments that examined the manner in which monitoring and control processes mediate between memory retrieval on the one hand and memory performance on the other. To enable this examination, a special *quantity-accuracy profile* (QAP) methodology was developed, based on a special procedure that combines both free- and forced reporting in conjunction with confidence judgments. This methodology or its adaptations have since been used in many other studies as well (see Goldsmith & Koriat, 2008).

In the QAP methodology, participants answer the same set of memory queries twice, once under forced-report instructions, which require them to respond to all queries (specific questions or cues), guessing if necessary, and again under free-report instructions, which allow them to withhold responses that they are unsure about (respond "don't know"). Confidence judgments (assessed probability correct) are also elicited for each forced-report response. In the free-report phase, a monetary or point bonus is gained for each correct response, a penalty is paid for each wrong response, and no penalty is paid (but also no bonus is gained) for withheld responses.

This design enables one to isolate and trace several different cognitive and metacognitive components that, according to the model, potentially mediate the effects of experimental manipulations or group differences on *output-bound accuracy* performance (the percentage of volunteered responses that are correct) and *input-bound quantity* performance (the percentage of correct responses out of the total number of queries). First, by treating each forced-report response as the person's best candidate answer to each question, the first component, *memory retrieval* (retention), can be assessed in terms of the percentage of these best candidates that are correct. Second, *monitoring effectiveness* can be evaluated in terms of both relative monitoring (i.e., the extent to which assessed probability correct discriminates between correct and incorrect best-candidate answers) and absolute monitoring (i.e., calibration bias or overconfidence, measured as the difference between mean assessed probability correct and actual proportion correct). Third, *report control policy* (criterion level) can be estimated as the cutoff on each participant's assessed-probability-correct ratings that best predicts his or her actual volunteering–withholding decisions in the free-report phase. Finally, *report control sensitivity* can be measured in terms of the strength of relationship between assessed probability correct and whether the best-candidate answer was volunteered. In general, these measures were found to be well behaved, responding to manipulations and correlating with the ultimate free-report performance measures (accuracy and quantity) in ways consistent with the model. A new version of this methodology was used in the present study (Experiment 2; see also Wahlheim & Jacoby, 2011, Experiment 3), adapting it to allow a more refined assessment of front-end memory control.

Front-End Control of Memory Quality

The monitoring and control processes addressed in Koriat and Goldsmith's (1996c) model, as well as the various proposed monitoring mechanisms discussed earlier, operate at the postretrieval stage, helping rememberers screen out false information after it has come to mind. When these mechanisms fail, the result is a commission error or false memory. Arguably, however, the entire burden of accurate remembering should not be placed solely on postretrieval monitoring and control. At least some of the blame for commission errors should be assigned to the faulty retrieval processes that brought the incorrect information to mind in the first place.

This idea has been promoted in several proposals stemming from the cognitive-neuropsychological study of memory confabulation. Moscovitch and Melo (1997), for example, distinguished between strategic processes taking place "at input" (or preretrieval) and "at output" (or postretrieval). At input, strategic processes are used to frame the memory query and to constrain memory search until local associative/cue-dependent processes produce a possible answer. At output, postecphoric strategic processes evaluate and verify the accuracy of the retrieved memory and place it in the proper temporal-spatial context in relation to other events. Along similar lines, Burgess and Shallice (1996) put forward a schematic model that includes a preretrieval *descriptor* process, which specifies attributes of the solicited trace (cf. Norman & Bobrow, 1979), as well as postretrieval editing and evaluation processes. Both Moscovitch and Melo (1997) and Burgess and Shallice (1996) concluded that both pre- and postretrieval deficits contribute to confabulation (see also Dab, Claes, Morais, & Shallice, 1999; Gilboa, 2010; Gilboa et al., 2006; Schacter, Norman, & Koutstaal, 1998; Shallice, 2001).

High rates of false memories can also be elicited in nonclinical populations. For example, research using the Deese–Roediger–McDermott paradigm (Roediger & McDermott, 1995) has demonstrated conditions in which the likelihood of recalling a nonpresented "critical lure" can sometimes be higher than the likelihood of recalling an item that was actually presented (Roediger & McDermott, 1995, Experiment 2; but see Koriat, Pansky, & Goldsmith, 2011). This work has engendered great interest both in the mechanisms that cause these lures to come to mind (e.g., Hege & Dodson, 2004; Nelson, McKinney, Gee, & Janczura, 1998) and in the ability or inability of postretrieval monitoring and control processes to edit out these false memories (e.g., Brainerd et al., 2003; Hicks & Marsh, 1999; McCabe & Smith, 2006; Schacter, Israel, & Racine, 1999).

The preceding examples emphasize the point that confabulations and other types of recall errors imply not only poor memory monitoring and control at the back end but also poor memory production at the front end; hence, the need to clarify the unique contribution of each to false remembering. The examples also raise the question of whether and how front-end control is exerted.

In this regard, we distinguish between the specific operation of memory production (e.g., the concept of *ecphory*; Tulving, 1983) and the more prolonged memory retrieval process in which that operation is embedded (see also Koriat, Goldsmith, & Halamish, 2008). We assume that rememberers cannot control the production operation itself, which we conceptualize as an automatic, ballistic operation (Guynn, 2003; Koriat et al., 2008; Moscovitch, 1994). Rememberers can nevertheless exert control over the production–retrieval process as a whole (for reviews, see Benjamin, 2008; Koriat et al., 2008). This can be done, for example, by generating additional internal cues to drive the production (ecphory) operation (cue specification and elaboration; e.g., Burgess & Shallice, 1996;

Cohn, Emrich, & Moscovitch, 2008; Norman & Bobrow, 1979; Schacter et al., 1998; Williams & Hollan, 1981) or by choosing an efficient production strategy (e.g., Higham & Tam, 2005, 2006).

In the present study, we focus on one particular way to control production at the front end, namely, source-constrained retrieval.

Source-Constrained Retrieval

Jacoby and colleagues (Jacoby, Kelley, & McElree, 1999; Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005; see also Marsh et al., 2009) proposed a front-end control mechanism called *source-constrained retrieval*—the focusing of retrieval such that only (mainly) information from the desired source comes to mind. This focusing is accomplished by mentally reinstating retrieval cues from the source context that specifically fit the way in which the target information was encoded.

Evidence for source-constrained retrieval in recognition memory has been obtained with a special "memory-for-foils" paradigm, introduced specifically for this purpose. In one study (Jacoby, Shimizu, Velanova, & Rhodes, 2005, Experiment 1), for example, participants studied words under either a deep or shallow encoding task and were given a recognition test requiring them to discriminate between the studied targets and new foil words. Following this, a final recognition test was administered to examine memory for the items appearing as foils on the preceding recognition test. Foil memory was superior when the participants had attempted to recognize deeply encoded rather than shallowly encoded study items on the preceding recognition test. This result counts against the view that recognition memory is a passive judgment based solely on trace strength or familiarity. Instead, it implies that the participants had invoked qualitatively different retrieval processes on the preceding recognition test, depending on the original manner of encoding: In attempting to constrain recognition to items that had been encoded in a particular way, participants apparently mentally reinstated the encoding task at test (e.g., pleasantness judgments or vowel counting), so that all recognition probes (both targets and foils) were processed deeply when the original encoding task had involved deep encoding (pleasantness judgments) and shallowly when the original task had involved shallow encoding (vowel counting).

The Present Study: Source-Constrained Recall

The present study was designed to demonstrate the operation of source-constrained retrieval in *cued recall* rather than in recognition memory, to examine its contribution to actual recall performance, and to isolate the specific front-end and back-end components that may underlie this contribution.

In Experiment 1, we adapted the memory-for-foils paradigm to reveal the operation of source-constrained retrieval in a cued-recall task. In Experiment 2, we manipulated the availability of information at retrieval regarding the specific operation used to encode each item, and hence the ability to use such source information to constrain retrieval. This, together with a special assessment procedure based on Koriat and Goldsmith's (1996c) QAP methodology, allowed us to examine the contribution of source-constrained retrieval to cued-recall accuracy and quantity measures and to isolate the underlying front-end and back-end components.

Experiment 1

In Experiment 1, Jacoby, Shimizu, Daniels, and Rhodes's (2005) memory-for-foils paradigm was adapted to a cued-recall situation. In the first phase of the experiment (*read-only* phase), participants were asked to read silently a list of unrelated word pairs, presented at a fast pace. The reason for this will be explained shortly. In the second, *processing* phase, the level of processing (Craik & Lockhart, 1972) of two additional lists of unrelated words pairs was manipulated within participants, by means of two typical incidental-encoding tasks: For the deeply processed list, participants were asked to indicate which of the two words in each pair was more pleasant, whereas for the shallowly processed list they indicated which word had more syllables. Next, in the third, *cued-recall* phase, memory for the pairs in each processed list was tested separately, by providing the first word and first letter of the second word of each pair as a cue and asking participants to recall the second word. Because the lists were tested separately, we hypothesized that participants would attempt to constrain their retrieval based on the knowledge that the source items had been encoded in a particular way—pleasantness comparisons or syllable comparisons. For example, participants might try to remember the second word to which the provided cue word had been compared earlier by once again judging the pleasantness or number of syllables of the cue word. Essentially, such source-constrained recall would constitute an attempt to home in on the target word by mentally reinstating the encoding task that had been performed on the cue–target pair during study (cf. Jacoby, Shimizu, Daniels, & Rhodes, 2005; Marsh et al., 2009; Polyn, Natu, Cohen, & Norman, 2005; Rajaram, Srinivas, & Roediger, 1998; Rugg, Allan, & Birch, 2000).

To reveal the participants' use of source-constrained recall in the cued-recall phase, a set of neutral cue words were added to the cued-recall tests, serving the same role as the foils in the original memory-for-foils paradigm. These cue words came from the read-only list that participants had simply read earlier in Phase 1. In the instructions to the Phase 3 cued-recall tests of both the deeply and shallowly processed Phase 2 lists, participants were told that some of the test cues might belong to the initial read-only list but were given no information as to which cues these might be. Believing that only a minority of the cue words on each test would relate to the read-only list, and without a reliable way to distinguish between these and the more numerous processed-list cues,[1] we expected that participants would attempt to constrain their retrieval by reinstating the Phase 2 processing task on all—or at least most—of the retrieval cues (both processed and read). Thus, although neither processing task (pleasantness or syllable comparison) had been used initially to encode the read-only cue words, source-constrained recall should now cause those words to be deeply processed when embedded in the cued-recall test of the

[1] Although no explicit information was provided, it was conceivable that participants might be able to discriminate the Phase 1 read-only cues from the Phase 2 processed cues on the basis of source memory alone. To make this more difficult, a fast presentation rate was used for the read-only pairs in Phase 1, and a filler task was introduced between the Phase 2 encoding tasks and the Phase 3 cued-recall tests, which, together with a relatively large number of word pairs, ultimately led to very low levels of memory for both types of items.

pleasantness-comparison list and to be shallowly processed when embedded in the cued-recall test of the syllable-comparison list. If so, the type of source-constrained recall invoked on the Phase 3 cued-recall test should be revealed by its influence on the memorability of the read-only cue words on a *subsequent recognition test* (Phase 4), in which participants had to discriminate these words from a set of new foil words. The critical prediction was that the read-only cue words that had been embedded in the Phase 3 cued-recall test of the pleasantness-comparison list would be better recognized on the Phase 4 recognition test than the read-only cue words that had been embedded in the Phase 3 cued-recall test of the syllable-comparison list.

Method

Participants. Sixteen Hebrew-speaking undergraduates at the University of Haifa participated in the experiment for course credit.

Materials. The Phase 1 read-only list consisted of 16 unrelated cue–target word pairs (e.g., *bee–stapler*), divided into two sublists of eight pairs each. The Phase 2 processed lists consisted of two additional lists, each with 16 pairs of similar composition. Another set of 16 individual words served as distractors in the recognition phase. All 112 words were selected on the basis of Hebrew word-association norms (Anaki & Henik, 2005; Bergerbest & Goshen-Gottstein, 2005; Faran, 2005; Rubinstein, Anaki, Henik, Drori, & Faran, 2005), with the restriction that no word was ever provided as the first association of any of the other words (i.e., the norm frequency of first association was zero).

Procedure. Participants were tested individually, with all instructions, stimulus materials, and data collection administered by computer. The session included four phases that are schematically presented in Figure 1.

Phase 1: Reading. Participants were told that the study involved pairs of unrelated words and that they were first going to be presented with a list of such pairs that would appear again later in the session, their present task being just to read those word pairs silently. These *read-only pairs* were presented in randomized order, with the word that would later be used as the cue word presented to the right of the target word (Hebrew is read from right to left). With the aim of preventing effective encoding of these words, each pair was presented for only 500 ms, with 1-s interpair interval.

Phase 2: Levels-of-processing manipulation. Immediately following the reading phase, participants were presented with the

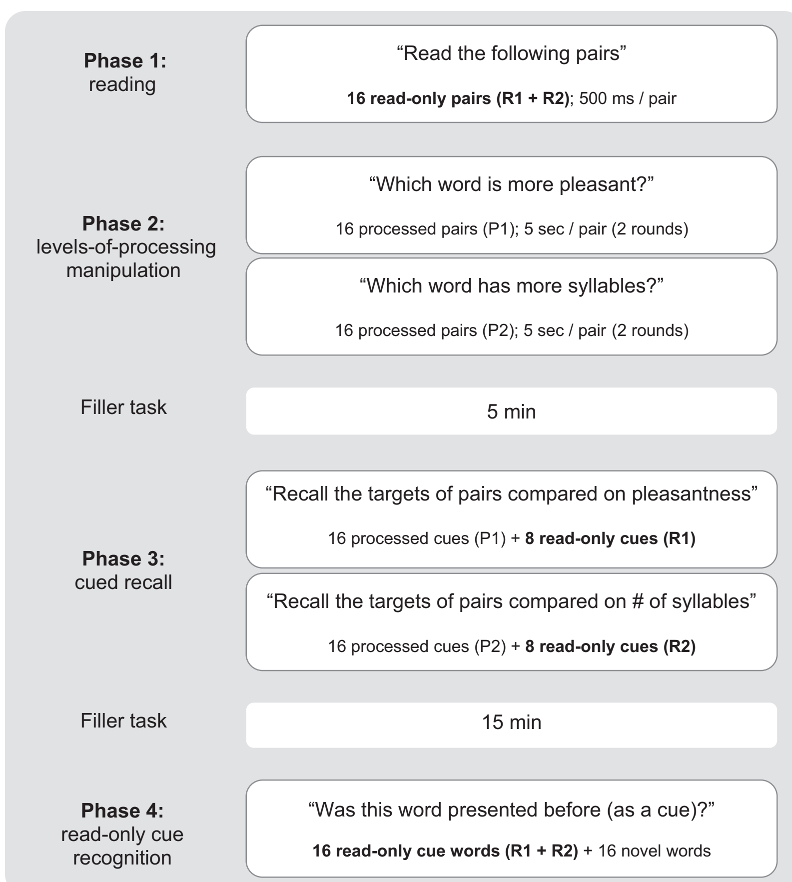

Figure 1. Schematic design and procedure of Experiment 1. The order of the processing tasks in Phase 2 (with matched test order in Phase 3) and the assignment of item lists (P1, P2, R1, R2) to conditions were counterbalanced across participants.

two word-pair lists for the levels-of-processing manipulation (*processed pairs*). The assignment of lists to processing conditions and the order of the two conditions were counterbalanced across participants. In the *pleasantness-comparison* (deep-processing) condition, participants were asked to indicate which of the two words in each pair was more pleasant, with the option of indicating that they were equally pleasant. In the *syllable-comparison* (shallow-processing) condition, participants indicated which of the two words in each pair had more syllables, with the option of indicating that they had the same number of syllables. Each word pair was presented for 5 s with 1-s interpair interval. Responses were made by pressing one of three keys. Participants were told that the study examined how practice affects the way people compare words, and so the same encoding task was repeated twice for each list (in the same random order) before moving on either to the next encoding task or to Phase 3.

Phase 3: Cued recall. Following a 5-min filler task (digit–symbol substitution), a self-paced cued-recall test was administered. Participants were instructed that the cued-recall task involved two stages and that in one stage they would be asked to recall the word pairs they had compared earlier for pleasantness and in the other stage they would be asked to recall the word pairs they had compared earlier for number of syllables. They were also informed that in both stages some of the retrieval cues might come from the list of word pairs that they had been asked to read in the first phase of the experiment.

In actuality, in each stage participants were presented with cues from the entire relevant list of 16 processed pairs (pleasantness comparison or syllable comparison) and from one sublist of eight read-only pairs, intermixed. The pairing of a specific sublist of read-only pairs with a specific list of processed pairs was counterbalanced across participants. The order of the two cued-recall test stages matched the order of the Phase 2 processing conditions for each participant. At the start of each stage, a title screen announced the type of processing task that had been used to encode the upcoming list of target items: “Word-Pair Recall: Pleasantness Judgments” or “Word-Pair Recall: Syllable Counting.” There was no reminder that some of the cues might point to read-only pairs.

On each test phase, the cue word and first letter of each target word were presented in a different randomized order for each participant, with the restriction that read-only or processed cues did not appear more than four times successively. Participants were asked to type in the remainder of the target word or to leave the item blank if they were unable to recall the target. As an explicit performance incentive, participants were told that they would gain 1 point for each correct answer, but would lose 1 point for each wrong answer, and that no points would be gained or lost for items left blank. Their goal was to maximize the number of points earned by their volunteered answers.

Phase 4: Read-only cue recognition. Following a 15-min filler task (nonverbal computer game), participants were given an old–new recognition test on the cue words from the read-only pairs (which had been read in Phase 1 and presented as recall cues in Phase 3). The 16 read-only cue words, intermixed with an equal number of new words, were presented in a different randomized order for each participant, with the restriction that old (read-only) or new words did not appear more than four times successively. Participants responded by pressing one of two keys labeled “previously presented” and “not previously presented.” Pretesting using a similar procedure indicated that participants tended to use a liberal response criterion that led to a ceiling effect on the hit rates. To induce a more conservative criterion, the instructions emphasized that it is important to respond “previously presented” only when one “really remembers” the word. Accordingly, participants were informed that when responding that a word was previously presented, they would gain 1 point if they are right but lose 3 points if they are wrong; when responding that a word was not previously presented, they would gain 1 point if they are right and lose 1 point if they are wrong.

Results and Discussion

Phase 3: Cued-recall performance. We first examined Phase 3 cued-recall performance for the Phase 1 read-only and Phase 2 processed (pleasantness or syllable comparison) pairs. As mentioned earlier (Footnote 1), poor retention for the items in both phases would serve to lessen the possibility that participants could discriminate the Phase 1 read-only cues from the Phase 2 processed cues, thereby increasing the likelihood that the same type of source-constrained recall strategy (i.e., mental reinstatement of the contextually relevant Phase 2 source task) would be applied to all cues. This goal appears to have been achieved: Recall for the processed pairs was at near floor levels for both the pleasantness-comparison and syllable-comparison encoding tasks (7% vs. 3%, respectively), $t(15) < 1$, with similarly low levels for the read-only pairs, regardless of whether they were tested in the context of pleasantness-compared or syllable-compared Phase 2 lists (6% in both cases), $t(15) < 1$. We assume that the lack of a levels-of-processing effect on recall of the Phase 2 processed pairs reflects a floor effect, which in turn reflects the difficulty of recalling 48 semantically unrelated word pairs after incidental encoding. As will be seen later in Experiment 2, these same processing tasks do yield the expected levels-of-processing effect on cued recall when a shorter word-pair list is used.[2]

Phase 4: Recognition of read-only cue words. The Phase 3 results indicate that cued-recall performance for both processed and read-only pairs was equivalent, regardless of the encoding task (for processed pairs) or test context (for read-only pairs). Equivalent recall, however, does not necessarily imply that the underlying recall processes were equivalent for both processing conditions during the cued-recall test. Qualitatively different retrieval processes might be involved, as will now be examined.

The critical prediction was that a qualitative difference in the depth of processing (*retrieval depth*; see Jacoby, Shimizu, Daniels, & Rhodes, 2005) of the retrieval cues on the Phase 3 cued-recall test would reveal itself in the Phase 4 recognition test of the read-only cue words, with better recognition performance for Phase 1 read-only cue words that had been embedded in the context of recalling the pleasantness-comparison list than for those embedded in the context of recalling the syllable-comparison list. The results supported this prediction, with hit rates of .83 and .70

[2] Also, in a pretest using the same incidental encoding tasks on similar lists of 16 unrelated word pairs in a between-subjects design, cued recall was significantly better following deep (pleasantness-comparison) than shallow (syllable-comparison) encoding (47% vs. 11%, respectively), $t(38) = 6.6$, $p < .001$.

for the pleasantness-embedded and syllable-embedded read-only targets, respectively, $t(15) = 2.83$, $p < .05$, $\eta^2 = .35$ (false-alarm rate = .27). Thus, although the difference in the encoding depth of the Phase 2 word pairs due to the different incidental encoding tasks did not exhibit itself in the Phase 3 cued-recall results (apparently because of a floor effect), the difference in source-constrained retrieval depth on the Phase 3 cued-recall test in the two source contexts was revealed by differences in the subsequent recognition of the Phase 3 read-only recall cues.

One might be concerned, however, that the observed difference in Phase 4 recognition of the read-only cue words may reflect a quantitative rather than qualitative difference in the Phase 3 cued-recall retrieval process. That is, perhaps participants invested more time in recalling the read-only items on the Phase 3 test when they were embedded in the pleasantness-comparison than in the syllable-comparison recall context. To examine this possibility, we calculated the retrieval time for the read-only items on the Phase 3 cued-recall test in terms of the time that lapsed between cue presentation and the participant's keypress that brought on the next cue. Mean retrieval time did not differ between the pleasantness-comparison and syllable-comparison recall contexts (4.8 s vs. 5.3 s, respectively), $t(15) < 1$. Thus, we found no support for the idea that a quantitative difference in Phase 3 retrieval time rather than a qualitative difference in retrieval depth (via mental reinstatement of the contextually relevant source encoding task) was responsible for the observed difference in Phase 4 recognition memory of the read-only cue words.

Experiment 2

Experiment 2 was designed to provide converging evidence for source-constrained cued recall and to shed additional light on its locus and performance consequences. The results of Experiment 1 suggested that in using the provided retrieval cues in the (Phase 3) cued-recall test, participants reinstated the earlier (Phase 2) encoding task to constrain their memory retrieval. As discussed earlier, we hypothesize that such reinstatement takes place at the front end, in the attempt to constrain and thereby enhance candidate production. However, it might also occur at the back end, during monitoring. For example, to monitor the correctness of a produced candidate, participants might compare the cue word and produced candidate in terms of their pleasantness or number of syllables. The familiarity or fluency of this comparison operation could then be used to decide whether the produced candidate is the sought-for target. The main purpose of Experiment 2, then, was to examine whether source-constrained recall does indeed constrain the production of candidate answers, thereby enhancing production quality. We also examined whether it might contribute to back-end monitoring as well and, ultimately, to cued-recall accuracy and quantity performance.

Toward this aim, the two incidental encoding tasks used in Experiment 1 were also used in the present experiment, and participants were subsequently given a cued-recall test. However, in this experiment the two encoding tasks were interleaved during the encoding phase, and the deeply and shallowly encoded pairs were intermixed at test, making it difficult—if not impossible—for participants to distinguish between deeply and shallowly encoded items on the subsequent cued-recall test. Hence, the critical manipulation in this experiment was the provision of explicit *source information* at test: In addition to the retrieval cue word (and target stem) for each item, participants in one group were informed about the task that had been used to encode that item (pleasantness comparison or syllable comparison), whereas no such source information was provided to participants in the other group. In this way, source information (and hence the opportunity for source-constrained recall) was manipulated between participants, independently of encoding depth, which was manipulated within participants.

Using a shorter word-pair list than in Experiment 1, we expected to observe the typical benefit of deep encoding over shallow encoding on production quality. The critical prediction, however, was that the participants who were provided with explicit source information would take advantage of that information to reinstate the encoding task and thereby constrain their retrieval, just as the participants in Experiment 1 presumably did. (Note that in Experiment 1, source information was also available because recall of the pleasantness-comparison and syllable-comparison word-pair lists was performed in two stages.) Thus, beyond the expected benefit of deep encoding, to the extent that rememberers can indeed use source information about encoding operations to strategically constrain retrieval, an additional benefit of source information on production quality should be observed.

To isolate the predicted benefit of source information on front-end production quality, and examine possible effects on back-end monitoring and report control, we adapted Koriat and Goldsmith's (1996c) QAP methodology, incorporating into it aspects of a procedure introduced previously by Weldon and Colston (1995; see also Guynn & McDaniel, 1999). The main modification was that in the initial forced-report stage, rather than ask participants only for their best-candidate response to each cue, we asked them to write down every candidate response that comes to mind when searching for the target, in the order that it comes to mind, *regardless of whether they believed it was the target* (cf. uninhibited report instructions in Bousfield & Rosner, 1970). Participants were instructed to stop searching (and recording produced candidates) when they believed they had produced the target word or when they no longer believed that they could produce a better candidate. If more than one candidate was produced for a given cue, they were instructed to mark one of them (not necessarily the last one) as their best candidate answer, that is, as the one they deemed most likely to be correct. As in the original QAP methodology, they were asked to assess the likelihood that each best-candidate answer was correct and to decide whether to report it for points (free-report stage): One point was gained for each correct reported answer, 1 point was lost for each wrong reported answer, and no points were gained or lost for withheld answers.

This procedure (see also Wahlheim & Jacoby, 2011, Experiment 3) enabled a relatively refined assessment of both front-end and back-end retrieval and reporting components. First, front-end production quality was assessed in terms of *first-candidate target percent*, that is, the percentage of items for which the actual target was the first candidate that came to mind. We chose to base the front-end measure on the percentage of first-candidate targets rather than on the percentage of targets ultimately produced, because the latter may be affected to a greater extent by back-end contributions, such as the decision to terminate or continue the search in light of the monitored correctness of the candidates produced so far. Thus, for example, wrongly identifying a produced nontarget as the target might sometimes prevent the search from continuing and (potentially) reaching the target. By contrast, the first candidate that comes to mind in searching for the target is, by definition, produced before the postretrieval monitoring and control processes exert their effects. Thus, to the extent that participants do in fact follow the instructions they were given—to write down all candidates that come to mind without any filtering—first-candidate target percent is less likely to be contaminated by the influence of those processes.[3]

In addition, we used several measures to assess back-end monitoring and control. First, to examine the effectiveness of monitoring the candidate words produced in response to each cue, *candidate monitoring* was assessed in terms of the ability to identify targets and reject nontargets from among all the candidate words that came to mind during the retrieval process. Second, the original QAP measures (Goldsmith & Koriat, 2008) were used to assess the effectiveness of *best-candidate monitoring* (i.e., monitoring the correctness of the best-candidate response to each cue) in terms of both calibration bias (over- or underconfidence) and resolution (relative monitoring) and to estimate the report criterion. Finally, the ultimate free-report accuracy and quantity performance was calculated, based on the correctness of the best-candidate responses that were volunteered for points. This allowed the contribution of source information, as mediated by the underlying front-end or back-end components, to be examined.

Method

Participants. Forty Hebrew-speaking undergraduates at the University of Haifa participated in the experiment, 10 for course credit and 30 for payment. They were randomly assigned to the *source-information* and *no-source-information* conditions, with the proportion of credit and payment participants equated in the two conditions.

Materials. Materials included 16 unrelated cue–target word pairs taken from those used in the levels-of-processing manipulation (Phase 2) of Experiment 1, divided into two lists of eight pairs each, and each list was further subdivided into two four-pair sublists.

Procedure. The experiment was run individually, in two phases. In the *levels-of-processing* phase, participants incidentally encoded the 16 word pairs by making the same types of comparison judgments that were used in Experiment 1: pleasantness (deep encoding) and number of syllables (shallow encoding). The assignment of word-pair list to processing task and the order of the two processing tasks were counterbalanced across participants. However, unlike in Experiment 1, here the two processing operations were interspersed by alternating the processing task every four items, such that Items 1–4 and 9–12 were processed with one comparison task and Items 5–8 and 13–16 were processed with the other comparison task. The same random order of item presentation for each list (i.e., the two interspersed four-item sublists) was used for all participants. Responses (left word greater, right word greater, or both words equal) were made by moving a joystick.

As in Experiment 1, participants were told that the study examined how practice affects the way people compare words, and so the incidental encoding procedure was performed twice in the same order. The word pairs were presented on a computer screen at a 5-s rate with 1-s interpair interval and a 5-s interblock interval. A symbol presented above each pair of words (☺ or #) indicated the type of judgment to be made (pleasantness or syllable counting, respectively). In addition, a pause screen was inserted before the beginning of each new four-item sublist to signal the change in processing operation. Upon completion, participants took a 5-min digit–symbol substitution test as a filler task.

In the *test phase*, participants were given a cued-recall test for the target words (the left word of each studied pair) from all 16 word pairs. Participants were given a booklet in which the cue word and target word stem (first letter) of each pair were provided as the test cue, with the target stem and adjacent blank space for the recording of candidate target words duplicated nine times for each cue. In addition, participants in the source-information condition were provided with information regarding the processing task that had been performed for each pair (using the same symbols as in the processing phase). No such information was provided to the no-source-information group. The same random order of test items was used for all participants.

Participants were instructed to use the cue word and target word stem to recall each studied target word. Importantly, they were instructed that while attempting to do so, they should write down *all* stem-compatible candidate words that came to mind, in the *order* that they came to mind, *without any screening*, in the provided spaces (using only as many spaces as necessary). The instructions stressed the importance of this aspect of the procedure (see Footnote 3), explaining that we were especially interested in "what goes through people's minds when they try to remember something." Nevertheless, participants were also instructed that if no stem-compatible candidate word came to mind in response to the cue, they should leave this item blank and move on to the next one (such cases were quite rare; see results below). Upon recognizing one of the candidate words as the target (not necessarily the last one produced), participants were instructed to mark it as their best-candidate answer and stop searching. If they decided to stop searching without recognizing any of the produced answers as the target, they were

[3] We emphasize that participants were not instructed to avoid engaging in postretrieval monitoring but rather to record each candidate that comes to mind regardless of that monitoring. Of course, we acknowledge the possibility that some participants might not entirely comply with the instruction to record all candidates that come to mind without any filtering (see General Discussion). Several steps were taken to minimize this potential problem. First, the instructions to participants strongly stressed the importance of compliance with this aspect of the procedure:

> We remind you that you are to write down every word that comes to mind while trying to remember the target word and which begins with the appropriate first letter. This point is very important to the success of the experiment, and we ask that you take care to comply with it as much as possible.

Second, to increase the ease of compliance, target stems were provided, in addition to the paired cue word, so that only produced candidates that are compatible with the stem would need to be recorded. Finally, because the production measure is based only on the first-candidate responses, subsequently produced candidates that fail to be recorded would only be of secondary concern.

nevertheless required to mark one best-candidate answer from among those they had produced (forced report). Next, participants were asked to provide a confidence judgment on a 0%–100% scale, reflecting the subjective assessed probability that the best-candidate answer was correct. Finally, under free-report instructions, participants indicated whether the best-candidate answer should be included in the calculation of their point score. To provide an explicit incentive for both quantity and accuracy, participants were told that they would gain 1 point for each volunteered correct answer, but would lose 1 point for each volunteered incorrect answer, with no points gained or lost for withheld answers. Their goal was to maximize the number of points earned. The test was self-paced, and participants were instructed to proceed item by item, in the order presented in the booklet.

Results and Discussion

Production, monitoring, control, and free-report performance measures were calculated for each participant in each of the experimental conditions. The results with respect to each of these measures are described in turn.

Front-end production quality. On average, participants produced at least one candidate word for 93.4% of all test items (only 6.6% of the items were left blank). This percentage was slightly higher for pleasantness-encoded pairs (95.6%) than for syllable-encoded pairs (91.2%), $F(1, 38) = 5.37$, $MSE = 71.34$, $p < .05$, $\eta_p^2 = .12$. The average number of produced word candidates per item (3.4) was also affected by the encoding task, with more candidates produced for syllable-encoded pairs (3.61) than for pleasantness-encoded pairs (3.26), $F(1, 38) = 5.78$, $MSE = 0.43$, $p < .05$, $\eta_p^2 = .13$. The percentage of items eliciting at least one candidate and the number of produced candidates per item were not affected by source information (both Fs < 1).

Our primary interest is in the quality of the production process. As described earlier, this was assessed in terms of first-candidate target percent: the percentage of items for which the target was produced as the first candidate that came to mind (see Figure 2). As expected, this measure was significantly higher for pleasantness-encoded pairs (26%) than for syllable-encoded pairs (5%), $F(1, 38) = 69.06$, $MSE = 126.95$, $p < .001$, $\eta_p^2 = .65$. More importantly, production quality was also higher in the source-information condition (21%) than in the no-source-information condition (11%), $F(1, 38) = 7.25$, $MSE = 258.94$, $p < .05$, $\eta_p^2 = .16$. This indicates that, as predicted, participants were able to use information regarding the source encoding operation to boost production quality, presumably by reinstating that operation on the provided retrieval cue, as suggested by the results of Experiment 1. The contribution of source information to production quality, however, was qualified by a significant interaction with encoding task, $F(1, 38) = 6.79$, $MSE = 126.95$, $p < .05$, $\eta_p^2 = .15$: For pleasantness-encoded pairs, production quality was higher with source information (34%) than without source information (18%), $F(1, 38) = 7.96$, $MSE = 331.83$, $p < .01$, $\eta_p^2 = .17$, whereas for syllable-encoded pairs the observed difference (7% vs. 4%, respectively) did not reach statistical significance, $F(1, 38) = 1.81$, $MSE = 54.07$, $p = .19$, $\eta_p^2 = .05$.

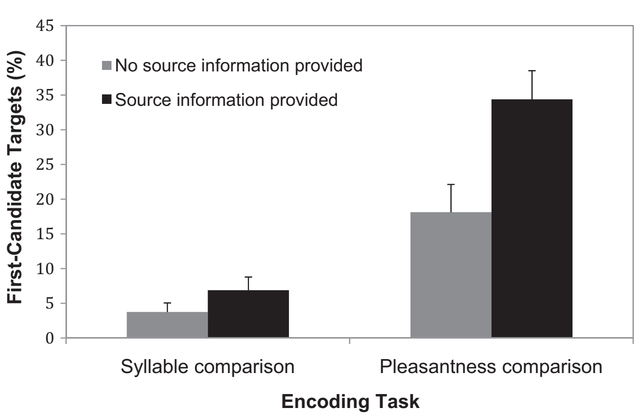

Figure 2. Mean production quality in Experiment 2, by encoding task and source-information condition. Error bars designate 1 standard error of the mean.

One potentially interesting explanation of this interaction is that source-constrained retrieval, implemented as the reinstatement of encoding operations, is an effective retrieval strategy when the encoding task involves pleasantness judgments but not when it involves syllable counting. That is, the internal cue “the word I compared to X in terms of pleasantness” might be effective in constraining retrieval, whereas “the word I compared to X in terms of number of syllables” might not. This idea is considered further in the General Discussion section.

Alternatively, the interaction may simply stem from the floor effect observed in the syllable-encoding condition, in which targets were produced as first-candidate responses for only 5% of the items and were produced in any position for only 7.5% of items. As in Experiment 1, this floor effect apparently reflects the difficulty of cued recall of unrelated word pairs after shallow incidental encoding, a problem that we tried to prevent by having participants encode the word pairs twice and by using a much shorter list ($n = 16$ items total; eight items for each encoding task) than in Experiment 1.

This low recall level for the syllable-encoded word pairs also yielded a large number of missing cases (about 40%) when calculating the back-end monitoring and control measures (when no target was produced or no best-candidate answer was volunteered for points). Nevertheless, the potential effects of source information on the back-end monitoring and control components, as well as on the ultimate levels of free-report accuracy and quantity performance, can be examined with respect to performance in the pleasantness-encoding condition. We therefore restrict the following analyses to the results from this condition.

Back-end monitoring and control. We now turn to examine the potential effects of source information on the back-end monitoring and report control of the pleasantness-encoded word pairs. The means of the relevant measures are presented in Table 1. First, candidate monitoring—defined as the ability to recognize produced targets and reject produced nontargets until the target is produced—was evaluated by calculating a corrected hit rate (hit rate minus false-alarm rate). Keeping in mind that participants were forced to choose a best-candidate answer for each item, target or nontarget recognitions were operationally defined as cases in which the target or a nontarget was identified as the best candidate with relatively high confidence (above 50%). Hit rate was then calculated as the proportion of targets that were correctly recog-

Table 1
Results of Experiment 2: Means and Standard Deviations of Production, Monitoring, Report Control, and Free-Report Performance Measures by Source Information Condition for Pleasantness-Encoded Items

Measure	Source information		No source information	
	M	*SD*	*M*	*SD*
Production quality (%)				
First candidate targets	34	19	18	17
Candidate monitoring				
Corrected hit rate	.93	.11	.84	.18
Best-candidate monitoring				
Resolution (gamma)	.98	.09	.98	.07
Corrected hit rate	.89	.17	.73	.21
Calibration bias (overconfidence)	.08	.11	.12	.14
Report control				
Report criterion	.53	.18	.47	.17
Fit rate	.99	.02	.98	.05
Free-report performance (%)				
Accuracy	83	24	63	36
Quantity	41	22	23	18

nized as targets, and false-alarm rate was calculated as the proportion of nontarget candidates falsely recognized as targets.[4]

Overall, participants were quite effective in their candidate monitoring, with a high hit rate (.93) and a low false-alarm rate (.04), for an average corrected hit rate of .89. This corrected hit rate was marginally higher when source information was provided (.93) than when it was not (.84), $t(33) = 1.90$, $p = .07$, $\eta^2 = .10$. Although only approaching statistical significance,[5] this difference suggests that participants may be using the provided source information to monitor the produced candidates for targets, in addition to using it to enhance the probability that the target will be produced in the first place (i.e., source-constrained recall). Alternatively, the enhanced quality of retrieved information produced by source-constrained recall at the front end may make it easier to identify targets and reject nontargets at the back end (see Rhodes & Kelley, 2005, for results consistent with this idea).

Whereas candidate monitoring refers to the ability to recognize the target when it is produced and to reject produced nontargets until then, best-candidate monitoring refers to the ability to distinguish best-candidate answers that are correct from those that are incorrect, for the purpose of deciding whether to report the answer. Thus, although these two types of monitoring may be based on partly or completely overlapping processes (see General Discussion), their separate analysis is motivated by the distinct control functions that they serve (continuing–terminating candidate production vs. volunteering–withholding the subjectively best produced candidate, respectively). One commonly used index of best-candidate monitoring, the within-participant Goodman–Kruskal gamma correlation between confidence and the correctness of each answer (Koriat & Goldsmith, 1996c; Nelson, 1984) was virtually at ceiling for both source-information conditions (.98) with no difference between them ($t < 1$). Because of the ceiling on gamma, we also examined best-candidate monitoring using a signal-detection analysis, treating correct answers held with greater than 50% confidence as hits and wrong answers held with greater than 50% confidence as false alarms (cf. Higham, 2002). Mean corrected hit rate[6] was significantly higher in the source-information condition (.89) than in the no-source-information condition (.73), $t(33) = 2.42$, $p < .05$, $\eta^2 = .15$. Thus, there is an indication that source information enhanced the effectiveness of monitoring the correctness of best-candidate answers.[7]

Two other aspects of back-end monitoring and report control were also examined. First, calibration bias, reflecting over- or underconfidence in the correctness of one's best-candidate answers, was calculated as the difference between mean assessed probability correct and the actual proportion of best candidates that were correct. Although participants were generally overconfident (mean bias = .10), the degree of overconfidence was unaffected by source information, $t(38) < 1$. Second, the report criterion set by each participant to guide the volunteer–withhold decisions was estimated with Koriat and Goldsmith's (1996c) procedure, which locates the cutoff point on each participant's confidence ratings that best predicts his or her actual report decisions (i.e., that maximizes the fit rate, the proportion of above-criterion items that were volunteered and below-criterion items that were withheld; see Goldsmith & Koriat, 2008, for details). Mean estimated report criterion was .50 (precisely normative under the operative payoff schedule), with an average fit rate of 98%. It too was unaffected by the source-information manipulation, $t(38) = 1.03$, $\eta^2 = .03$.

Free-report memory performance. Finally, after revealing the separate contributions of source information to the front-end production and back-end monitoring of the pleasantness-encoded items, we now examine whether these contributions translated into an enhancement of actual free-report quantity and accuracy performance (see Table 1).

Free-report quantity performance, calculated as the percentage of volunteered targets out of the total number of items, was rather low, averaging only 32%. In contrast, free-report accuracy performance, calculated as the percentage of volunteered targets out of

[4] Because recognizing one candidate as the target (correctly or incorrectly) necessarily entailed rejecting all other produced candidates for a specific item, an adjustment was needed in order to avoid undesirable dependencies: Each correctly rejected nontarget candidate was counted as an independent correct rejection only for items in which no candidate was recognized. In addition, even for such items, if the target was missed, only correctly rejected candidates that preceded the target in the production order were counted as independent correct rejections (to avoid giving credit for missing the target and continuing to produce and correctly reject nontargets).

[5] We chose to use the corrected hit rate as our measure of candidate monitoring because of its simple, straightforward interpretation. However, to ensure that the observed pattern of results did not depend on the particular distributional assumptions of this measure (see, e.g., Macmillan & Creelman, 2005), we also calculated candidate monitoring in terms of d' and A'. All three measures yielded an advantage of source information that either approached (corrected hit rate and A') or just reached (d') statistical significance.

[6] A statistically reliable effect of source information was also observed with alternative measures of best-candidate monitoring, d' and A'.

[7] Note that in the candidate and best-candidate monitoring analyses, four (out of 40) participants had to be omitted because they failed to produce any targets, thereby preventing the calculation of both corrected hit rate and gamma. A fifth participant had to be omitted because he provided only targets as best-candidate answers.

the total number of volunteered answers, was much higher, averaging 73%. This difference highlights the role of report option (back-end monitoring and report control) in allowing rememberers to achieve a high level of report reliability (output-bound accuracy), even though they may in fact remember very little. Importantly, with regard to our current concerns, both quantity and accuracy performance were enhanced significantly by source information: Free-report quantity was higher with source information (41%) than without source information (23%), $t(38) = 2.87$, $p < .01$, $\eta^2 = .18$, with a similarly large benefit for free-report accuracy (83% vs. 63%, respectively), $t(36) = 2.05$, $p < .05$, $\eta^2 = .10$. Thus, in addition to the typical overall benefit of deep versus shallow encoding, gained by making pleasantness comparisons rather than syllable comparisons,[8] source-constrained recall—based on explicit source information about the manner of encoding—allowed this benefit to be increased even further.

General Discussion

A growing amount of research in recent years is concerned with memory accuracy and distortion, in contrast to the traditional concern with memory quantity (Koriat et al., 2000). This shift brings to the fore questions about the mechanisms that underlie the accuracy and inaccuracy of memory reports. Much of the previous work on this topic has focused on back-end control of memory performance, examining postretrieval monitoring and control processes that are used to screen out wrong information after it has been produced (e.g., Brainerd et al., 2003; Dodson & Schacter, 2002; Goldsmith & Koriat, 2008; Lindsay, 2008). Recent studies, however, suggest that front-end control can also enhance memory accuracy, by increasing the quality of the retrieval process itself—homing in on target information and preventing nontarget information from being produced in the first place (e.g., Dab et al., 1999; Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005).

In this study, we examined one mechanism of front-end control—source-constrained recall: Rememberers attempt to constrain their retrieval to the source event by processing the available recall cues in qualitatively different ways (e.g., by judging their pleasantness or by counting their syllables), depending on the manner in which they were originally encoded. Cast more generally as a form of cue elaboration or specification (e.g., Burgess & Shallice, 1996; Norman & Bobrow, 1979), externally provided recall cues are refined and supplemented by additional, internally generated cues that enhance the efficiency of target retrieval. Experiment 1 provided evidence for the use of this mechanism in cued-recall situations. Experiment 2 showed that this mechanism indeed enhances production quality, with indications of improved back-end monitoring as well, ultimately enhancing free-report accuracy and quantity performance.

In discussing these results, we first address the evidence for source-constrained recall and the mechanisms underlying its enhancement of recall accuracy and quantity performance. We then address the more general idea of a division of labor between front-end (production) and back-end (monitoring) mechanisms of recall quality control, and discuss some theoretical and methodological issues that arose in the present attempt to examine this idea.

Source-Constrained Recall: Mechanisms and Performance Consequences

Building on previous findings from recognition memory (e.g., Jacoby, Shimizu, Daniels, & Rhodes, 2005), we hypothesized that rememberers can enhance the quality of the information that comes to mind during recall by internally reinstating the original encoding operation on the retrieval cue. The results supported this hypothesis. In Experiment 1, results from an adapted version of Jacoby et al.'s (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005) memory-for-foils paradigm indicated that participants reinstated the source encoding operation—pleasantness comparisons or syllable-count comparisons—on the recall cues, as revealed by differences in the ability to recognize the cues on a subsequent recognition test: Cue words of neutral word pairs from an initial read-only phase, embedded in the cued-recall test of the pleasantness-encoded pairs, were subsequently recognized better than the neutral cue words that had been embedded in the cued-recall test of the syllable-encoded pairs. By implication, the cue words on the cued-recall test were processed in qualitatively different ways—deeply, by reinstating the pleasantness-comparison operation, or shallowly, by reinstating the syllable-count operation—depending on the manner in which the majority of source items had originally been encoded.

The data from Experiment 1, however, did not allow a determination of whether the reinstatement of the source encoding operation was invoked during production or during postretrieval monitoring and whether in fact this reinstatement enhanced the ultimate recall performance. To achieve these aims, in Experiment 2, the encoding and test conditions were designed such that the participants could not know at test (without being explicitly told) which encoding operation had been used to encode each tested word pair. This allowed us to manipulate the availability of information about the source encoding operation at test. After restricting our analyses to the pleasantness-encoded word pairs, when source information was provided at test, so that participants could effectively reinstate the appropriate encoding operation for each retrieval cue, the percentage of targets produced as the first retrieved candidate response increased compared with when no source information was provided. This increase in front-end production quality, together with a somewhat less clearly indicated benefit of source information for back-end monitoring processes, led to increases in actual free-report memory performance, in terms of both memory quantity and accuracy.

In Experiment 2, the positive effect of source information on production quality, reflecting source-constrained retrieval, was observed for the pairs encoded by the pleasantness-comparison task but not for those encoded by the syllable-comparison task. Although, as discussed earlier, this could be due to the floor performance levels observed in the latter condition, it is interesting to consider a further, more theoretically interesting explanation: Perhaps the attempt to retrieve the target by reinstating the source encoding operation on the syllable-compared retrieval cues (e.g., "the word I compared to X in terms of number of syllables") was

[8] Free-report quantity and accuracy scores in the shallow-encoding (syllable-comparison) condition averaged 6% and 35%, respectively, with no effect of source information, consonant with the earlier results for production quality.

less effective in constraining retrieval than the parallel operation performed on the pleasantness-compared retrieval cues (e.g., "the word I compared to X in terms of pleasantness"). Such a difference in internally generated cue effectiveness might in fact be expected from the principle of cue distinctiveness or diagnosticity (Mäntylä & Nilsson, 1988; Nairne, 2002) or its converse, cue overload (Watkins & Watkins, 1975). In Experiment 2, there were many other words besides the target whose number of syllables was compared to a cue word, for example, having three syllables (e.g., when the cue word was *hospital*), but presumably only the target word was compared in terms of its pleasantness to a cue word involving "sickness, discomfort, and death" when the cue word *hospital* was processed in the pleasantness-comparison task. Thus, although source-constrained retrieval may be attempted in any case (Experiment 1), its effectiveness may depend on the diagnosticity of the retrieval cue that is generated on the basis of source information (Experiment 2; see also Nairne, 2002).

Of course, mental reinstatement of the encoding task is just one of many ways in which source information can potentially be used to constrain retrieval. In general, the mental reinstatement of any source feature based on remembered (or partly remembered) aspects of the source event could conceivably be used to constrain retrieval beyond what can be achieved based on externally provided retrieval cues alone (e.g., Reiser, Black, & Kalamarides, 1986; Smith, 1979; William & Hollan, 1981; for related neuroimaging evidence, see Polyn et al., 2005; Rugg & Wilding, 2000).

Front-End and Back-End Contributions to Recall Quality

Inspired by the manufacturing quality control metaphor (Jacoby, Shimizu, Velanova, & Rhodes, 2005), the present study was predicated on the assumption that front-end production and back-end monitoring and report processes make unique contributions to the quality of the final memory report, and that both types of processes are—to a greater or lesser degree—under the control of the rememberer.

Koriat et al. (2008) recently put forward a schematic conceptual framework for the control of remembering that assimilates these two types of control (see Figure 3). In that framework, preretrieval processes use heuristic cues to monitor the likeliness that sought-for information is available and accessible in memory, in deciding whether to initiate or forgo a memory search. If a search is initiated, implicit and explicit metacognitive knowledge is used to choose an appropriate search strategy and to generate an effective set of cues to guide the retrieval. After each retrieval (production) operation, postretrieval processes are used to assess the correctness of the produced information and to reject—or perhaps inhibit—information that is judged to be wrong or irrelevant. Postretrieval processes also decide (on the basis of assessed correctness, among other considerations) whether to continue the search and, if so, whether and how to update the retrieval (production) parameters. Finally, once the search is terminated, postretrieval report processes assess the likely correctness of the best-candidate answer and on that basis (among other considerations) decide whether to report the answer and at what level of precision or coarseness (grain size).

The present study focused on several components of this more encompassing framework. In particular, we focused on the preretrieval processes involved in setting up the search query by generating internal, source-constrained retrieval cues, on the postretrieval assessment of the correctness of the produced information, and on the decision whether to report the answer. Other components were considered implicitly: for example, the decision to focus on first-candidate target percent as the front-end measure, in order to avoid potential contamination by the (back-end) decision whether to continue or terminate the search. Indeed, along with the theoretical challenges, the study of controlled processes in remembering introduces some tough methodological challenges concerning how to allow participants control over their memory processes while still retaining experimental control in the examination and assessment of those processes (for related discussions, see Goldsmith & Koriat, 2008; Koriat & Goldsmith, 1996b, 1996c;

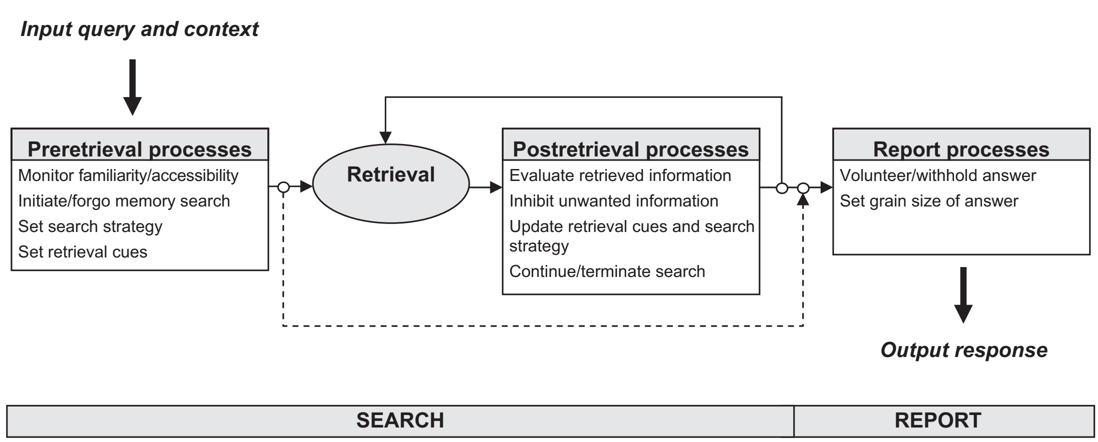

Figure 3. A schematic framework for the memory and metamemory processes involved in remembering (dashed line represents the decision to forgo a memory search). From "Controlled Processes in Voluntary Remembering" (p. 309), by A. Koriat, M. Goldsmith, and V. Halamish, 2008, in J. Byrne (Series Ed.) and H. L. Roediger III (Vol. Ed.), *Learning and Memory: A Comprehensive Reference: Vol. 2. Cognitive Psychology of Memory.* Oxford, England: Elsevier. Copyright 2008 by Elsevier. Reprinted with permission.

Nelson & Narens, 1994). Some of the challenges addressed in the present study are discussed next.

Assessing Front-End and Back-End Processes: Theoretical and Methodological Issues

Front-end production. As mentioned in the introduction, an advantage of Koriat and Goldsmith's (1996c) framework for investigating the strategic regulation of memory accuracy is that it is accompanied by a special assessment methodology (QAP) that combines both free and forced reporting in conjunction with confidence judgments to isolate and examine the separate contributions of retrieval, monitoring, and control to free-report memory accuracy and quantity performance. The QAP methodology is useful in identifying the unique contributions of monitoring and control processes at the reporting stage to memory performance, above and beyond the contribution of retrieval per se. However, its use of forced-report performance (proportion correct) as the index of retrieval does not allow one to distinguish the cognitive and metacognitive contributions to the retrieval process itself. This was noted, for example, by Rhodes and Kelley (2005) in using the QAP methodology to examine the mediation of age differences in memory accuracy by memory monitoring:

> The role of monitoring may, in fact, be underestimated in this paradigm if participants engage in some monitoring prior to choosing a response at the forced report stage (Koriat & Goldsmith, 1996[c]). Using protocol analysis in this paradigm with younger adults, we have found that participants occasionally engage in monitoring during the forced response stage, as they generate the deceptive (wrong) answer, reject it, and continue attempting to retrieve the correct answer. If there are individual differences in monitoring during forced report, that could produce an overestimation of the role of retrieval in determining memory accuracy in free report and an underestimation of the role of monitoring. More fine-grained methods are needed to capture the possible interplay of retrieval and monitoring processes. . . . (p. 591)

In the present study, we strived to achieve a relatively clean measure of retrieval–production quality to evaluate the claim that source-constrained retrieval operates as a mechanism for constraining candidate production in cued recall. Therefore, in Experiment 2, we adapted the original QAP methodology to include an uninhibited recall procedure, in which participants are instructed to record all candidate answers that come to mind while searching for the target (and that fit the target stem cue), in the order that they come to mind, without any filtering. This allowed us to base our measure of production quality on the percentage of cases in which the target appeared as the first produced candidate. In terms of the Koriat et al. (2008) scheme depicted in Figure 3, the first candidate produced in response to a memory query depends on metacognitive processes related to setting up the retrieval query, but its internal production is followed by, and is therefore unaffected by, postretrieval monitoring and control. Thus, although the correctness of the first produced candidate (like any subsequent candidate) is evaluated after it comes to mind, given the instruction to write down every candidate that comes to mind without any filtering, the data obtained regarding the correctness of the first-candidate response should be unaffected by the postretrieval evaluation. Also, another attractive aspect of examining production quality in terms of the correctness of first-candidate responses is that producing the target on the first retrieval attempt is generally preferable to producing it later (after producing one or more nontargets), because this places less burden on the back-end candidate-monitoring process to correctly evaluate and reject initially produced nontargets until the target is eventually produced.

Of course, this procedure (used also in Wahlheim & Jacoby, 2011, Experiment 3) has its shortcomings, stemming mainly from the assumption that participants can in fact faithfully record each candidate response that comes to mind, in the order that they come to mind. Thus, for example, perhaps in some cases multiple candidates might come to mind more or less simultaneously, so that the recorded order may be somewhat arbitrary or some candidates may be lost or edited out. We did our best to ensure that such cases would be infrequent and have minimal impact. As mentioned earlier (Footnote 3), beyond the very explicit and repeated instructions, target stems were added to the retrieval cues to constrain the range of viable candidates and thereby increase the ease of recording them. Also, basing the measure of production quality solely on the first-candidate responses serves to reduce the potential impact of misordering or omission of subsequent candidates. Thus, we believe that examining the correctness of the first-candidate responses can provide valuable information about production quality, beyond that obtained by the use of forced-report percent correct alone.

In this regard, we should note that in the present Experiment 2, the same pattern of effects observed with the first-candidate production measure was also observed with a measure similar to the standard forced-report measure—the percentage of items for which the target was produced in any position. In fact, the correlation between the two measures for the pleasantness-encoded pairs was .89. Thus, although the pattern of results would be the same, the use of the first-candidate target percent instead of the standard forced-report measure allows us to feel more comfortable that the production results were minimally affected by postproduction processes (for related procedures, see Guynn & McDaniel, 1999; Hege & Dodson, 2004; Higham & Tam, 2005, Experiment 3; Weldon & Colston, 1995).

Back-end monitoring. An additional potential benefit of the retrieve-and-report procedure relates to the measurement of postretrieval memory monitoring. In Koriat and Goldsmith's (1996c) original framework, as well as in related frameworks and work addressing the strategic regulation of memory reporting (e.g., Higham, 2002, 2007; Roebers, 2002), the effectiveness of postretrieval monitoring is assessed with respect to the ability to monitor the correctness of the best-candidate answer that comes to mind in response to each memory query. This approach stems in part from concern with the role of best-candidate monitoring in guiding the report control decision (i.e., volunteer or withhold) and the importance of best-candidate monitoring effectiveness in bounding the joint levels of free-report memory accuracy and quantity performance that can potentially be achieved (Goldsmith & Koriat, 2008; Koriat & Goldsmith, 1996c). However, as implied in Rhodes and Kelley's (2005) observation quoted earlier, in cued-recall situations in which there is only one correct target for each cue, two functionally distinct aspects of postretrieval monitoring can be distinguished: (a) candidate monitoring, reflecting the ability to recognize produced targets and reject produced nontargets over one or more retrieval attempts until the target is eventually produced (or until one deems that further retrieval is futile), and (b) best-candidate monitoring, reflecting the ability to monitor the likely correctness of the best-candidate response to a given cue, in deciding whether to report it (or in deciding whether to coarsen it; e.g., Goldsmith et al., 2002). Conceivably, both types of monitoring could be based on the same underlying processes but operate differently or give differential weight to different types of cues (e.g., heuristic vs. systematic, Mitchell & Johnson, 2000; experience based vs. information based, Koriat, Nussinson, Bless, & Shaked, 2008).

Regardless of whether the underlying processes are similar or different, as just explained, whenever more than one mutually exclusive candidate response is produced to a given query, candidate monitoring and best-candidate monitoring serve different functions, suggesting that one might want to make a separate evaluation of the effectiveness of each type of monitoring in fulfilling its function. The use of the uninhibited retrieve-and-report procedure allows one to do so.

There have been many discussions of issues relating to the measurement of memory monitoring, including analyses of the statistical properties of different measures (e.g., Benjamin & Diaz, 2008; Masson & Rotello, 2009; Nelson, 1984; Yaniv, Yates, & Smith, 1991) and conceptual distinctions between the various aspects of monitoring that may be captured by different measures (e.g., Nelson, 1996; Schraw, 2009; Yates, 1990). We believe that with regard to the monitoring of the products of retrieval, the distinction between candidate and best-candidate monitoring, although not at the focus of this article, may too be an interesting one to pursue, both theoretically and methodologically.

Conclusion

The approach and results of the present study add to those of a growing number of studies examining how rememberers exert control over the quality of the information that they retrieve and report (for reviews, see Benjamin, 2008; Koriat et al., 2008). By attempting to combine the literatures on front-end and back-end regulation of memory performance, this study takes another step toward understanding the mechanisms of self-regulated remembering—both front end and back end—that potentially contribute to memory quantity and accuracy performance.

References

Anaki, D., & Henik, A. (2005). Norms for semantic and associative relations in Hebrew. In A. Henik, O. Rubinstein, & D. Anaki (Eds.), *Norms for Hebrew words* (pp. 8–16). Beer-Sheva, Israel: Ben-Gurion University of the Negev.

Benjamin, A. S. (2008). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The Psychology of Learning and Motivation: Vol. 48. Skill and strategy in memory use* (pp. 175–223). London, England: Academic Press.

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York, NY: Psychology Press.

Bergerbest, D., & Goshen-Gottstein, Y. (2005). Measures of relation between words in Hebrew association norms and their implications for memory tests. In A. Henik, O. Rubinstein, & D. Anaki (Eds.), *Norms for Hebrew words* (pp. 83–96). Beer-Sheva, Israel: Ben-Gurion University of the Negev.

Bousfield, W. A., & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science, 20,* 75–76.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110,* 762–784. doi:10.1037/0033-295X.110.4.762

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory, 4,* 359–411. doi:10.1080/096582196388906

Cohn, M., Emrich, S. M., & Moscovitch, M. (2008). Age-related deficits in associative memory: The influence of impaired strategic retrieval. *Psychology and Aging, 23,* 93–103. doi:10.1037/0882-7974.23.1.93

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671–684. doi:10.1016/S0022-5371(72)80001-X

Dab, S., Claes, T., Morais, J., & Shallice, T. (1999). Confabulation with a selective descriptor process impairment. *Cognitive Neuropsychology, 16,* 215–242. doi:10.1080/026432999380771

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8,* 155–161. doi:10.3758/BF03196152

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46,* 782–803. doi:10.1006/jmla.2001.2822

Faran, I. (2005). Direction of relation of paired associate words. In A. Henik, O. Rubinstein, & D. Anaki (Eds.), *Norms for Hebrew words* (pp. 67–82). Beer-Sheva, Israel: Ben-Gurion University of the Negev.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51,* 473–493. doi:10.1016/j.jml.2004.06.002

Gilboa, A. (2010). Strategic retrieval, confabulations and delusions: Theory and data. *Cognitive Neuropsychiatry, 15,* 145–180. doi:10.1080/13546800903056965

Gilboa, A., Alain, C., Stuss, D. T., Melo, B., Miller, S., & Moscovitch, M. (2006). Mechanisms of spontaneous confabulations: A strategic retrieval account. *Brain, 129,* 1399–1414. doi:10.1093/brain/awl093

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. S. Benjamin & B. H. Ross (Eds.), *The Psychology of Learning and Motivation: Vol. 48. Skill and strategy in memory use* (pp. 1–60). London, England: Academic Press.

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52,* 505–525. doi:10.1016/j.jml.2005.01.010

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131,* 73–95. doi:10.1037/0096-3445.131.1.73

Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology, 38,* 245–256. doi:10.1080/00207590344000178

Guynn, M. J., & McDaniel, M. A. (1999). Generate—Sometimes recognize, sometimes not. *Journal of Memory and Language, 41,* 398–415. doi:10.1006/jmla.1999.2652

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 787–795. doi:10.1037/0278-7393.30.4.787

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychol-*

ogy: Learning, Memory, and Cognition, 25, 1195–1209. doi:10.1037/0278-7393.25.5.1195

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30,* 67–80. doi:10.3758/BF03195266

Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136,* 1–22. doi:10.1037/0096-3445.136.1.1

Higham, P. A., & Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language, 52,* 595–617. doi:10.1016/j.jml.2005.01.015

Higham, P. A., & Tam, H. (2006). Release from generation failure: The role of study-list structure. *Memory & Cognition, 34,* 148–157. doi:10.3758/BF03193394

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–400). New York, NY: Guilford Press.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12,* 852–857. doi:10.3758/BF03196776

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language, 52,* 493–504. doi:10.1016/j.jml.2005.01.007

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. doi:10.1037/0033-2909.114.1.3

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98,* 127–140. doi:10.1016/S0001-6918(97)00039-5

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 41, pp. 293–320). San Diego, CA: Academic Press.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123,* 297–315. doi:10.1037/0096-3445.123.3.297

Koriat, A., & Goldsmith, M. (1996a). Memory as something that can be counted versus memory as something that can be counted on. In D. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Vol. 2. Practical applications* (pp. 3–18). Hillsdale, NJ: Erlbaum.

Koriat, A., & Goldsmith, M. (1996b). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences, 19,* 167–188. doi:10.1017/S0140525X00042114

Koriat, A., & Goldsmith, M. (1996c). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517. doi:10.1037/0033-295X.103.3.490

Koriat, A., Goldsmith, M., & Halamish, V. (2008). Controlled processes in voluntary remembering. In J. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (pp. 307–324). Oxford, England: Elsevier.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51,* 481–537. doi:10.1146/annurev.psych.51.1.481

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 117–135). Mahwah, NJ: Erlbaum.

Koriat, A., Pansky, A., & Goldsmith, M. (2011). An output-bound perspective on false memories: The case of the Deese–Roediger–McDermott (DRM) paradigm. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 297–328). New York, NY: Psychology Press.

Lindsay, S. (2008). Source monitoring. In J. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (pp. 325–348). Oxford, England: Elsevier.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Mäntylä, T., & Nilsson, L.-G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 502–509. doi:10.1037/0278-7393.14.3.502

Marsh, R. L., Meeks, J. T., Cook, G. I., Clark-Foos, A., Hicks, J. L., & Brewer, G. A. (2009). Retrieval constraints on the front end create differences in recollection on a subsequent test. *Journal of Memory and Language, 61,* 470–479. doi:10.1016/j.jml.2009.06.005

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 509–527. doi:10.1037/a0014876

McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory, 14,* 570–583. doi:10.1080/09658210600624564

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

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. *Neuropsychology, 8,* 524–534. doi:10.1037/0894-4105.8.4.524

Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia, 35,* 1017–1034. doi:10.1016/S0028-3932(97)00028-6

Nairne, J. S. (2002). The myth of encoding–retrieval match. *Memory, 10,* 289–395. doi:10.1080/09658210244000216

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review, 105,* 299–324. doi:10.1037/0033-295X.105.2.299

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. Comments on Schraw (1995). *Applied Cognitive Psychology, 10,* 257–260. doi:10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology, 11,* 107–123. doi:10.1016/0010-0285(79)90006-9

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310,* 1963–1966. doi:10.1126/science.1117645

Rajaram, S., Srinivas, K., & Roediger, H. L., III. (1998). A transferappropriate processing account of context effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 993–1004. doi:10.1037/0278-7393.24.4.993

Reiser, B. J., Black, J. B., & Kalamarides, P. (1986). Strategic memory search processes. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 100–121). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511558313.011

Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis. *Journal of Memory and Language, 52,* 578–594. doi:10.1016/j.jml.2005.01.014

Roebers, C. M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology, 38,* 1052–1067. doi:10.1037/0012-1649.38.6.1052

Roediger, H. L., III. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100. doi:10.1006/jmla.1996.0005

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814. doi:10.1037/0278-7393.21.4.803

Rubinstein, O., Anaki, D., Henik, A., Drori, S., & Faran, I. (2005). Free associations norms in Hebrew. In A. Henik, O. Rubinstein, & D. Anaki (Eds.), *Norms for Hebrew words* (pp. 17–34). Beer-Sheva, Israel: Ben-Gurion University of the Negev.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience, 12,* 664–678. doi:10.1162/089892900562291

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4,* 108–115. doi:10.1016/S1364-6613(00)01445-5

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24. doi:10.1006/jmla.1998.2611

Schacter, D. L., Norman, D. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49,* 289–318. doi:10.1146/annurev.psych.49.1.289

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4,* 33–45. doi:10.1007/s11409-008-9031-3

Shallice, T. (2001). Deconstructing retrieval mode. In M. Naveh-Benjamin, M. Moscovitch, & H. L. Roediger III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik* (pp. 124–134). New York, NY: Psychological Press.

Shimizu, Y., & Jacoby, L. L. (2005). Similarity-guided depth of retrieval: Constraining at the front end. *Canadian Journal of Experimental Psychology, 59,* 17–21. doi:10.1037/h0087455

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460–471. doi:10.1037/0278-7393.5.5.460

Tulving, E. (1983). *Elements of episodic memory.* Oxford, England: Clarendon Press.

Wahlheim, C. N., & Jacoby, L. L. (2011). Experience with proactive interference diminishes its effects: Mechanisms of change. *Memory & Cognition, 39,* 185–195. doi:10.3758/s13421-010-0017-4

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 442–452. doi:10.1037/0278-7393.1.4.442

Weldon, M. S., & Colston, H. L. (1995). Dissociating the generation stage in implicit and explicit memory tests: Incidental production can differ from strategic access. *Psychonomic Bulletin & Review, 2,* 381–386. doi:10.3758/BF03210976

Whittlesea, B. W. A. (1997). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 37, pp. 211–264). New York, NY: Academic Press.

Whittlesea, B. W. A. (2002). Two routes to remembering (and another to remembering not). *Journal of Experimental Psychology: General, 131,* 325–348. doi:10.1037/0096-3445.131.3.325

Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very long-term memory. *Cognitive Science, 5,* 87–119. doi:10.1207/s15516709cog0502_1

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin, 110,* 611–617. doi:10.1037/0033-2909.110.3.611

Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice-Hall.

Received April 22, 2011
Revision received June 29, 2011
Accepted July 4, 2011 ■