Journal of Memory and Language **46,** 120–152 (2002)
doi:10.1006/jmla.2001.2796, available online at http://www.academicpress.com on IDEAL®

Dual-Retrieval Processes in Free and Associative Recall

C. J. Brainerd, Ron Wright, and V. F. Reyna

University of Arizona

and

D. G. Payne

Binghamton University

Recent dual-retrieval accounts of free recall postulate that a memory target can be recalled either by directly accessing its verbatim trace or by reconstructing it from semantic or other relational information. We introduce a simple paradigm, derived from the classic Estes RTT procedure, that separates direct access from reconstruction and that separates reconstruction from a metacognitive judgment process that authorizes reconstructed targets for output. Results are reported from four experiments, two that applied the paradigm to free recall and two that extended it to associative recall. The principal findings were that (a) direct access was enhanced by manipulations that made targets' surface forms easier to process or that focused recall on individual targets, (b) reconstruction was enhanced by manipulations that made targets' meaning content easier to process or that focused recall on groups of targets, and (c) such manipulations produced single dissociations, double dissociations, and reversed associations between direct access, reconstruction, and metacognitive judgment. We discuss how this paradigm might be exploited to unify dual-retrieval conceptions of recall and recognition.

Key Words: retrieval; free recall; fuzzy-trace theory; reconstruction; memory judgment.

For some years, the hypothesis that recognition tasks engage two independent retrieval operations, often called recollection and familiarity, has been the principal alternative to one-process theories in the signal-detection tradition. Although experimentation has traditionally centered on teasing apart the contributions of the two retrieval operations to hit rates (e.g., Gillund & Shiffrin, 1984; Jacoby, 1991; Mandler, 1980), the focus of research has lately been broadened to encompass false-alarm rates for distractors that share targets' meaning content (e.g., Payne & Elie, 1998; Reyna, 1996; Reyna & Brainerd, 1995; Rotello, 2000; Rotello & Heit, 1999). Our aim in this article is to broaden the focus still further to consider a dual-retrieval model for recall.

First, we review research on recent dual-retrieval conceptions of free recall, noting similarities and differences between the posited retrieval operations and the corresponding operations for recognition. Next, a paradigm is introduced, consisting of a simple experimental procedure and an associated modeling technology, that allows investigators to separate and quantify the retrieval operations that have been posited for recall. A series of experiments is then reported that made use of these tools, the main objective being to evaluate core hypotheses about the nature of the processes that are being measured.

This research was supported by a grant to the first and third authors from the National Science Foundation (SBR-9730143) and by grants to the third author from the National Institutes of Health, the Academic Medicine and Managed Care Forum, and the Department of Commerce. The authors express their appreciation to Debra McDonald, Adrian Monjes, and Keil Stockton for their assistance in conducting the research.

Correspondence and reprint requests should be addressed to C.J. Brainerd, Department of Special Education, Rehabilitation, and School Psychology, University of Arizona, Tucson, AZ 85721. E-mail: brainerd@u.arizona.edu.

Dual-Retrieval Explanations of Free Recall

Dual-retrieval accounts of recognition vary somewhat in the specifics of their assumptions (for reviews, see Brainerd, Reyna, & Mojardin, 1999; Clark & Gronlund, 1996). In fuzzy-trace

0749-596X/01 $35.00

theory's analysis (Brainerd et al., 1999), for instance, the representational content that retrieval operations access is stressed, and operational definitions are adopted that map different operations with distinct response patterns that are produced by three instructional conditions. These distinctions grew out of a prior literature, which suggested independent retrieval operations for different types of representational content (for a review, see Reyna & Brainerd, 1995). Other theories emphasize the phenomenologies that are induced by different retrieval operations (see Humphreys, Dennis, Chalmers, & Finnigan, 2001). Despite differences in the details, the modal dual-process account assumes that participants first evaluate a probe's level of global familiarity, where familiarity is viewed as a retrieval operation that accesses a continuous memory-strength scale like that in signal-detection theory. A probe is accepted (as old) if familiarity delivers a scale value that falls above some high criterion, and it is rejected (as new) if the value falls below some low criterion. The recollection operation is activated when the familiarity value falls between the two criteria. This slower back-up process searches participants' verbatim inventory of presentation events to determine whether the probe can be explicitly recollected as having been studied. The fact that the second operation is characterized as "recall" or as "recall like" (e.g., Horton, Pavlick, & Moulin-Julian, 1973; Mandler, 1980) suggests that, unlike recognition, recall tasks engage only the recollective operation.

Recently, however, dual-retrieval distinctions have been formulated to explain two sets of findings about free recall: cognitive triage effects (Brainerd & Reyna, in press; Reyna & Brainerd, 1995) and false recall of semantic associates of word lists (Lampinen, Neuschatz, & Payne, 1998; Payne & Elie, 1998; Payne, Elie, Blackwell, & Neuschatz, 1996; Reyna, 1992). Cognitive triage effects are counterintuitive relations between recall difficulty (as measured by total errors, trial number of the last error, etc.) and output position in a free-recall protocol. According to one-process accounts, output position should be a monotonic function of difficulty; the easier the item, the earlier the output position (for a review, see Brainerd, Reyna, Howe, & Kevershan, 1991). However, beginning with experiments by Battig (1965; Battig, Allen, & Jensen, 1965), when mean output position on Trial 2 of a free-recall experiment is computed for targets that were recalled on Trial 1 (easy) and for targets that were not (hard), output position is earlier, not later, for hard items. If output positions on criterion trials (errorless recall) are plotted against overall difficulty (total precriterion errors), a U-shaped relation is often observed, with harder items appearing at the beginning and the end of a protocol and easier items appearing in the middle (Brainerd et al., 1991; Brainerd, Reyna, & Howe, 1990).

Fuzzy-trace theory's distinction between direct access to verbatim traces versus reconstructive processing of gist traces has been used to explain such patterns (Brainerd & Reyna, in press; Reyna & Brainerd, 1995). It is assumed that one retrieval operation, the more accurate of the two, predominates at the start of output and provides direct access to verbatim traces of target presentations. When such traces are accessed, participants recall the targets by merely reading out surface information as it echoes in the mind's ear or flashes in the mind's eye, much as an actor would recite words as they are whispered by a prompter or seen on a script. As simple readout of surface forms that are present in consciousness is all that is required, direct access produces fast, confident, virtually errorless recall. Because direct access is susceptible to output interference that accumulates during recall, the number of targets that can be recalled in this way is maximized by retrieving verbatim traces in reverse order of their memory strength (weaker → stronger). The other retrieval operation, which is slower and less accurate, waxes as recall proceeds. This operation *reconstructs* targets by processing gist traces of the meaning content of studied lists. Meaning content is not unique to particular targets, of course, so reconstructive processing will sometimes generate candidates for output that are not part of studied lists (e.g., if COLLIE is studied, reconstruction may generate COLLIE, DOG, PET, and POODLE as output candidates). Therefore, a metamemorial judgment process is required to decide whether the participant is sufficiently confident that a candidate was present on the study list to authorize its output. Unlike verbatim traces, gist traces are accessed in the order of their relative memory strength (stronger → weaker), producing increasingly slower, more uncertain recall. Note that the familiar finding that output slows as recall proceeds (e.g., Payne, 1986) falls out as a predicted consequence of increasing reliance on reconstruction.

These distinctions explain triage effects as follows. First, when recalled targets are classified as hard (error on the previous trial) versus easy (correct on the previous trial), mean output positions are earlier for hard items because direct access (traces processed in a weaker → stronger order) dominates early in recall and is responsible for a larger proportion of total output than reconstruction (traces processed in a stronger → weaker order). Second, when targets are classified on criterion trials according to precriterion error rate, nonmonotonic relations with output position can appear because direct access (weaker → stronger) predominates initially, but reconstruction (stronger → weaker) predominates later.

Payne and associates (Payne et al., 1996; Payne & Elie, 1997, 1998) used similar distinctions to explain false recall and to generate additional predictions about it. A popular method for studying false recall is the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). Participants study lists of words that revolve around familiar themes (e.g, NURSE, SICK, LAWYER, MEDICINE, HEALTH, HOSPITAL, DENTIST, PHYSICIAN, ILL, PATIENT, OFFICE, STETHOSCOPE, SURGEON, CLINIC, CURE). All list words are semantic associates of an omitted critical word (DOCTOR in this instance). On free-recall tests, this critical unpresented word intrudes in a substantial proportion of protocols (Payne et al., 1996; Roediger & McDermott, 1995). On the basis of earlier work by Reyna and Kiernan (1994), Payne et al. proposed that studied items could be recalled by directly accessing verbatim traces or by reconstructively processing gist memories, but false recall of critical unpresented items (e.g., DOCTOR) should be due to the latter process. According to this proposal, critical unpresented items are falsely recalled because they are such excellent examples of lists' meanings that participants are likely to generate them as candidates, and once generated, participants are apt to be quite confident that they were present on study lists (i.e., metacognitive judgment authorizes output with high probability).

This analysis delivers some straightforward predictions that have been investigated by Payne and associates and by Toglia and associates (Toglia & Neuschatz, 1996, 1997). One is that intrusions of critical unpresented items should occur predominately at later output positions (where reconstruction is more likely). When Payne et al. (1996) and Roediger and McDermott (1995) plotted output positions of critical intrusions using Vincentized quintiles, approximately half of the intrusions occurred in the final quintile. Sommers and Lewis (1999) also reported that the mean output position of critical intrusions was in the second half of recall. In free recall of more traditional lists, intrusions of unpresented items also occur predominately in the second half of output (Brainerd, Reyna, Harnishfeger, & Howe, 1993). Two other predictions are that intrusion rates will increase under conditions of forced recall (requiring a number of items to be recalled that exceeds total output in standard recall) and under conditions of repeated testing. Forced recall should increase reconstructive retrieval because standard recall already exhausts items that can be output via direct access, thereby increasing the contribution of reconstruction to total output, and repeated testing should have the same effect because the interference that accumulates from earlier tests will differentially impair direct access on later tests.[1] Consistent with these predictions, Payne et al. found that forced recall increased intrusion rates by 76% and that intrusion rates increased by 30% after three free-recall tests. Forced recall is also known to increase intrusion rates for more traditional lists

[1] Payne et al. (1996) also proposed that repeated testing strengthens the gist memories that reconstructive retrieval processes rely on.

(Erdelyi, Finks, & Feigin-Pfau, 1989; Roediger & Payne, 1985).

Four other predictions whose effects bear on the dual-retrieval account of false recall have been investigated by Toglia and associates (Toglia & Neuschatz, 1996, 1997; Toglia, Neuschatz, & Goodwin, 1999). First, because targets' meanings can be identified before processing of their surface forms is completed (Stenberg, Lingren, Johansson, Olsson, & Rosen, 2000; Wallace, Stewart, Shaffer, & Barry, 1998), decreasing exposure duration at study ought to selectively impair verbatim traces, thereby increasing later reliance on reconstructive retrieval. Consistent with this hypothesis, Toglia and Neuschatz (1996) found that false recall of semantic associates increased as exposure duration decreased, a finding that has also been obtained for false recognition of semantic associates by Buchanan, Brown, and Westbury (1999) and by Seamon, Luo, and Gallo (1998). Second, instructions to process the meaning of studied targets, rather than surface form, should also increase later reliance on reconstructive retrieval (Reyna & Kiernan, 1994), thereby increasing false recall if reconstructive retrieval is responsible. Toglia et al. (1999) confirmed that meaning instructions increased false recall. Third, when participants study a single list composed of several DRM sublists, the core meanings of each sublist are cued more strongly when its targets are blocked together than when targets from different lists are intermixed. Hence, blocked presentation should increase reconstructive retrieval and false recall, and Toglia et al. and McDermott (1996) confirmed this prediction. Fourth, presenting DRM lists as pictures rather than words ought to impair direct access during *oral* recall, shifting retrieval toward reconstruction, because verbatim traces of pictures are not representations of phonetic information that can simply be read out of consciousness. Consistent with this prediction, Toglia and Neuschatz (1997) found higher levels of intrusions for written recall of picture lists than of word lists.

Summing up, there are key points of similarity and difference between this dual-retrieval conception of free recall and the modal dual-retrieval interpretation of recognition. Concerning similarities, one operation is thought to retrieve traces of specific targets (recollection in recognition and direct access in recall), whereas the other is thought to retrieve more global information about studied lists that could be shared by many targets (familiarity in recognition and reconstruction in recall). In addition, the global operation in both free recall and recognition is assumed to be responsible for false-memory reports (intrusion of semantically related items in free recall and false alarms to such items on recognition tests). However, there are also three fundamental differences between these dual-retrieval conceptions. First, the temporal and performance priorities of the target-specific and global operations are reversed. In recognition, the global operation is assumed to be the faster of the two and to be the default basis for accepting probes as old, with the target-specific operation acting as a back-up process when the global operation delivers an ambiguous result. In free recall, on the other hand, the target-specific operation is assumed to be faster and to be the preferred basis for output, with the global operation predominating when the target-specific operation cannot deliver output. Second, the global operation must generate *specific* items as candidates for output in free recall, whereas the global operation need only generate *nonspecific* feelings of familiarity in recognition. Third, the metacognitive judgment process that accompanies reconstructive retrieval has no clear counterpart in dual-retrieval accounts of recognition.

Finally, it is important to distinguish the present dual-retrieval conception, which emphasizes differences in representational content, from classic two-stage models of free recall, which do not. In the latter models, a single operation is posited that is executed in two steps. The standard example is generate/recognize theory (e.g., Bodner, Masson, & Caldwell, 2000), which assumes that participants access a target's trace (step 1) and then subject it to a familiarity check (step 2). The retrieved target is output only if its familiarity level exceeds some subjective criterion. The present dual-retrieval conception involves both a one-step operation (direct access) and a two-step operation (reconstruction followed by metamemorial judgment). The two-step operation differs from generate/recognize in two ways: (a) It is not item-specific traces that are being accessed. (b) Metamemorial judgment does not focus merely on global familiarity but also considers particularized sources of information that affect the chances that constructed items appeared on the study list, such as items' membership in semantic categories (see Experiment 1) or the phenomenologies that they induce (see Experiment 3).

Measuring Direct Access and Reconstruction

Quantifying the contributions of direct access and reconstruction to free recall, rather than merely invoking them to explain data patterns, requires a paradigm that separates their effects on output. A broadly applicable paradigm derives from Estes (1960). According to the dual-retrieval conception, a target can be successfully recalled in two ways: Direct access retrieves its verbatim trace, or it is reconstructed via semantic processing and the judgment operation authorizes output of the construction. Thus, successful direct access produces recall with probability one, but after successful reconstruction, the probability of recall is the probability of metacognitive authorization. A recall failure can also occur in two ways: Direct access and reconstruction both fail, or direct access fails and reconstruction succeeds but the constructed item is not authorized for output. Therefore, the probabilities of the two data events (c = target correctly recalled, n = target not recalled) may be expressed as

$$p(\mathrm{c}) = D + (1-D)RJ \tag{1}$$

and

$$p(\mathrm{n}) = (1-D)R(1-J) + (1-D)(1-R), \tag{2}$$

where D is the proportion of targets that are directly accessed, R is the proportion of targets that are reconstructed, and J is the proportion of reconstructed targets that metamemorial judgment authorizes for output.

These theoretical processes cannot yet be estimated, of course, because the number of processes (three) exceeds the number of free empirical probabilities (one). With memory models, this limitation, which is called nonidentifiability, is normally removed by enriching the data space to yield further empirical probabilities (for worked examples, see Brainerd, Howe, & Kingma, 1982). In this connection, consider the Estes (1960) RTT (R = "reinforce," T = "test") procedure in which, after the study trial, participants respond to two, independent recall tests. This procedure supplies four performance outcomes for each target, which yields three empirical probabilities for parameter estimation—namely, any three probabilities from the set [p(cc), p(cn), p(nc), p(nn)]. The expressions for the outcomes are

$$p(\mathrm{cc}) = D + (1-D)RJ^2, \tag{3}$$

$$p(\mathrm{cn}) = (1-D)RJ(1-J), \tag{4}$$

$$p(\mathrm{nc}) = (1-D)RJ(1-J), \tag{5}$$

$$p(\mathrm{nn}) = (1-D)R(1-J)^2 + (1-D)(1-R). \tag{6}$$

Relative to Eqs. (1) and (2), the only new feature concerns the metacognitive judgment process, J. The model assumes that a target that is directly accessed or reconstructed on the first test is likewise directly accessed or reconstructed on the second test, but that metacognitive judgment is variable over tests; authorizing a reconstructed target for output on the first test does not guarantee authorization on the second test (and conversely).[2]

Although the number of free empirical probabilities matches the number of processes to be estimated, this system is still not identifiable,

[2] Because Test 1 and Test 2 are independent experimental events, it might be thought that the expressions in Eqs. (3–6) should be derived via multiplication of Eqs. (1) and (2). That is, for instance, the expression for p(cc) in Eq. (3) should be derived by multiplying Eq. (1) by itself, yielding $p(\mathrm{cc}) = [D + (1 - D)RJ]\,[D + (1 - D)RJ] = D^2 + 2D(1 - D)RJ + [(1 - D)RJ]^2$. This procedure does not lead to the correct formulas, however, because although the recall tests are independent experimental events, the memory processes that control performance on the tests are not independent. Specifically, the model posits that a target that is directly accessed on Test 1 will also be directly accessed on Test 2 and that a target that is reconstructed on Test 1 will also be reconstructed on Test 2 (although it may not be authorized for output). Therefore, the probability of recall by direct access on both tests is D, not D^2, the probability of recall by direct access on one test but not the other is zero, and the probability of recall by reconstruction on both tests is $(1 - D)RJ^2$, not $[(1 - D)RJ]^2$. Similar statements apply to derivation of the expressions for p(cn), p(nc), and p(nn), as well as the corresponding derivations for the RTTT procedure that is described later (Eqs. (7–14)).

owing to an equality constraint that reduces the number of free probabilities to two. Eqs. (4) and (5) are identical, leaving $p(\text{cc}) + p(\text{nn}) + 2p(\text{cn}) = 1$, an expression in which only two probabilities are free because probabilities must sum to one. However, identifiability can be achieved with a simple extension of the RTT procedure that has been used in free-recall studies by Halff (1977) and others (for a review, see Batchelder & Riefer, 1999). Three independent recall tests are administered, which generate eight performance outcomes per target and seven free empirical probabilities (namely, any seven probabilities from the set [p(ccc), p(ccn), . . ., p(nnn)]). The expressions for the outcomes are

$$p(\text{ccc}) = D + (1-D)RJ^3, \quad (7)$$
$$p(\text{ccn}) = (1-D)RJ^2(1-J), \quad (8)$$
$$p(\text{cnc}) = (1-D)RJ^2(1-J), \quad (9)$$
$$p(\text{ncc}) = (1-D)RJ^2(1-J), \quad (10)$$
$$p(\text{cnn}) = (1-D)RJ(1-J)^2, \quad (11)$$
$$p(\text{ncn}) = (1-D)RJ(1-J)^2, \quad (12)$$
$$p(\text{nnc}) = (1-D)RJ(1-J)^2, \quad (13)$$

and

$$p(\text{nnn}) = (1-D)R(1-J)^3 + (1-D)(1-R). \quad (14)$$

These expressions involve no new assumptions; they merely extend Eqs. (3–6) from RTT to RTTT. Thus, correct recall is due to the same two causes (successful direct access or successful reconstruction coupled with metamemorial authorization), and nonrecall is due to the same two causes (failure of both direct access and reconstruction or successful reconstruction coupled with failure of metamemorial authorization).

To make these expressions transparent, the tree diagram in Fig. 1 shows how direct access, reconstruction, and metacognitive judgment deliver each of the eight response patterns. Eqs. (7–14) can be recovered by selecting each observable response pattern on the right side of Fig. 1 and tracing its path backward through the tree, to its point of origin. Although this system also contains equality constraints [e.g., $p(\text{ccn}) = p(\text{cnc}) = p(\text{ncc})$], there are enough free empirical probabilities to secure identifiable estimates of D, R, and J, to evaluate goodness of fit, and to test within- and between-condition hypotheses about parameter values. Such analyses are executed in the usual way by implementing Eqs. (7–14) in a likelihood function (for statistical machinery, see Riefer & Batchelder, 1988).

Because output interference impairs direct access during recall, it might seem that the model should posit that memory tests cause forgetting of direct access. This assumption is readily incorporated into the basic model by appending forgetting parameters that apply to Test 2 versus Test 1 and to Test 3 versus Test 2 (for examples, see Brainerd, Reyna, Howe, & Kingma, 1990). We have not included such parameters for two reasons. First, the forgetting assumption is better treated as an empirical question: If there is substantial forgetting, fit will fail for the basic model, and forgetting parameters can then be introduced to rectify the situation. Second, in previous RTTT experiments like the ones reported here (e.g., Halff, 1977; Payne, 1987), there was little or no evidence of forgetting between consecutive free recall tests (although forgetting could easily be induced by imposing long delays between consecutive tests).

On the usual a priori dimensions, the model in Fig. 1 supplies an attractive technology for measuring dual-retrieval processes in free recall: It is uncomplicated, mathematically tractable, and it can be applied to free recall paradigms without having to modify them. At this early stage of research, the key empirical question about the model is: Do D, R, and J behave in ways that are congruent with theoretical conceptions of the corresponding memory processes? The general objective of the experiments that we report was to accumulate preliminary data on this question by determining how the parameters react to manipulations that ought to divide along lines of direct access, reconstruction, or judgment. Although metacognitive judgment is of interest and is considered in our experiments, principal interest attaches to variables that should affect direct access or reconstruction. In Experiment 1, we manipulated the accessibility of well-understood semantic information during free recall, which should affect R but not D. In Experiment 2, we manipulated the difficulty of processing targets' surface content at study, which should affect D but not R. In Experiment 3, the model was extended from free

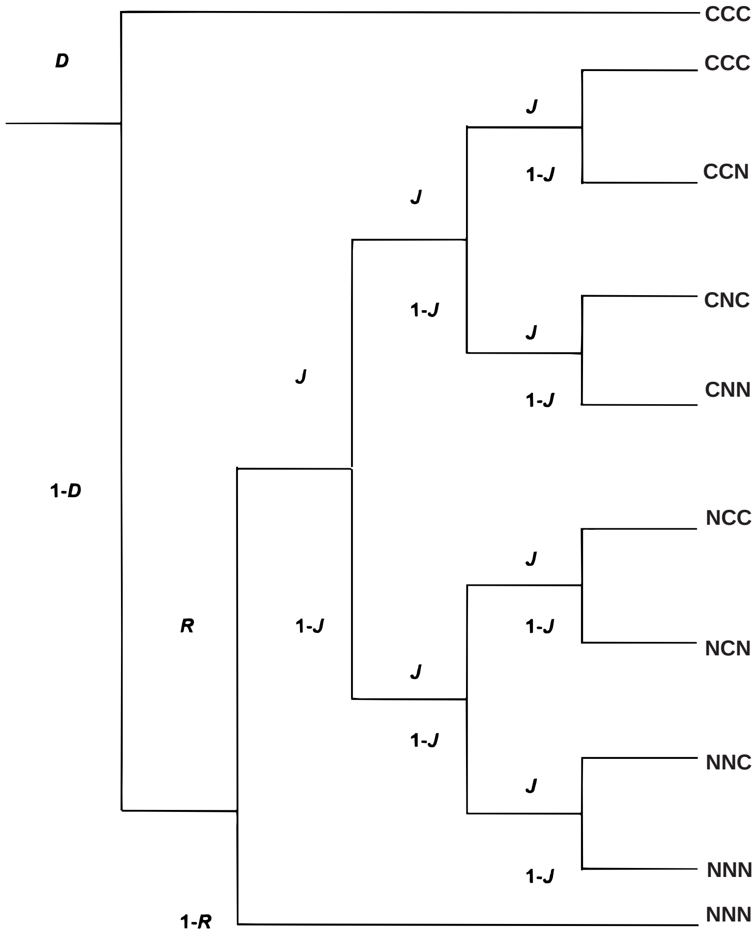

FIG. 1. Tree diagram for the dual-retrieval model of recall.

to associative recall, and a variable was manipulated (increasing target concreteness) that ought to have opposite effects on *D* and *R*. Finally, in Experiment 4, the model was applied concurrently to free and associative recall, on the hypothesis that relative to free recall, associative recall provides increased support for direct access and decreased support for reconstruction, thereby having opposite effects on *D* and *R*.

Experiments 3 and 4 also addressed a limitation of the dual-retrieval model as it is implemented in Eqs. (7–14). That implementation specifies that after the first trial, there is no retrieval learning, in the sense of increasing either the proportion of targets that can be directly accessed or the proportion that can be reconstructed. The probability of correct recall must therefore remain invariant across recall tests, so that the model is unable to fit data in which that probability increases over tests. For instance, recall probability obviously will increase if further study trials are interpolated between tests (i.e., if the design is $R_1TR_2TR_3T$, rather than RTTT). Even without further study trials, such increases will occur for the types of materials that produce hypermnesia effects (for a review, see Payne, 1987). Although the model, as it stands, will not fit such data, we show in Experiments 3 and 4 that this limitation is easily removed by relaxing the constraint on retrieval learning, allowing direct-access learning and reconstruction learning to occur after the first trial. Crucially, the model is still identifiable when this constraint is relaxed, and is therefore informative when applied to tasks in which recall probability increases over trials.

EXPERIMENT 1: MANIPULATING RECONSTRUCTION AND METAMEMORIAL JUDGMENT

Experiment 1 was a free-recall design that incorporated two transparent manipulations of semantic processing on recall tests. Both manipulations were designed to prolong reconstruction once it had begun, thereby increasing the total number of targets that would be reconstructively retrieved. Hence, it was expected that *R* should react to these manipulations, but *D* should not. One of the manipulations was included because it should also affect *J*, if this parameter measures participants' willingness to authorize output of reconstructed items.

The basic task was one that provided latitude for prolonging reconstructive retrieval via instructions or cues supplied during free-recall tests. Participants studied a list with an explicit and noticeable categorical structure—specifically, a list composed of four blocked sublists of exemplars of familiar semantic categories (e.g., items of furniture, articles of clothing). After studying the list, participants responded to three free-recall tests. Because the list's categorical structure was conspicuous, it was obvious to participants that semantic concepts could be processed to reconstruct targets that could not be directly accessed. By making such information prominent, we reasoned that it would be possible to prolong reconstructive retrieval via two test-phase manipulations, meaning instructions and category labels.

Free recall was performed under three test conditions, control, meaning instructions, and category labels. In the control condition, participants were given standard instructions to recall as many targets as they could remember in whatever order the targets came to mind. In the meaning condition, participants were given instructions that encouraged further reconstructive processing by informing them that in addition to recalling targets in any order they wished, they could recall unstudied words that were exemplars of the categories represented by the lists. If meaning instructions prolong reconstructive retrieval, then the total number of reconstructed targets should increase, thereby increasing the value of *R*, relative to the control condition. However, there is no reason to expect that meaning instructions would increase participants' metamemorial confidence in their reconstructions (parameter *J*) or that such instructions would increase the number of targets that can be directly accessed (parameter *D*). In the category labels condition, further reconstructive processing was encouraged by making the names of the four studied categories available as printed labels during free recall. Our assumption was that once reconstructive retrieval had begun, the availability of explicit labels would prolong reconstructive processing of the semantic categories, increasing target output and the value of *R*, relative to the control condition. If the effect of such labels on retrieval is simply to protract reconstructive processing of the categories, then there is no reason to expect that the number of directly accessed targets (parameter *D*) would increase. However, category labels should increase the value of *J*, if this parameter measures participants' degree of confidence that reconstructed items were present on study lists. A reconstruction should only be authorized for output if it is an exemplar of one of the studied categories. In the control and meaning-instruction conditions, there will be some uncertainty as to the *exact* scope of the categories that were studied because those categories are not explicitly named. That uncertainty is removed in the labeling condition: Participants can merely compare reconstructions to the available category labels, outputting those that match and suppressing those that do not.

Method

Participants. Participants were 90 undergraduate psychology students. They participated in the experiment to fulfill course requirements.

Materials and procedure. A pool of 192 words was formed by (a) selecting 24 familiar categories from the Battig and Montague

(1969) norms and then (b) selecting the first 8 exemplars from the production norms for each category, eliminating words that were very strong associates of selected words (e.g., TABLE–CHAIR). This pool was used to construct the lists presented to individual participants. Individual lists were constructed by randomly sampling 4 categories from the pool of 24. The list presented to the participant then consisted of these four sublists. The method of presentation made the categorical structure of the list obvious to the participant: The participant was informed that the upcoming study list would consist of groups of words that belonged to the same category of things, and targets were blocked by sublist during list presentation (i.e., all 8 exemplars of a given category were presented consecutively). Exemplar presentation order was random within sublists, and the order of presentation of sublists was random.

Thirty participants were assigned to each of three conditions: (a) control, (b) meaning instructions, (c) category labels. The participant first read a page of general memory instructions. The instructions stated that he or she would be listening to a list composed of four short sublists in which all the words belonged to the same category of things. Next, the participant listened to an audio recording of the list. The 8 targets on each sublist were presented at a 2.5-s rate, and there was a 10-s pause between consecutive sublists. After the four sublists had been presented, the participant was given a pencil and a blank sheet of paper, which were used to solve arithmetic problems for 1 min. To control time, the problems were presented orally by the experimenter. Next, the participant was given a form containing 80 spaces in which to write words, with the spaces being arranged in 4 vertical columns of 20 spaces each. Two minutes of written recall followed. The instructions for written recall stipulated that it was unnecessary to spell words accurately. After 2 min, the participant turned in the recall form, received a second blank sheet of paper, and solved arithmetic problems for 1 min. The participant was then provided with a second recall form, and another 2 min of written recall ensued. After 2 min, arithmetic problems were solved for 1 min, followed by a third recall test.

Each participant performed the recall tests under one of the three conditions. Instructions were provided orally by the experimenter. Participants in all conditions were reminded of the categorical structure of the list that they had just studied. The names of the categories were not mentioned, however. Participants in the control condition then received standard instructions to write as many of the studied words as possible within the next 2 min, in whatever order the words came to mind. Participants in the meaning condition received the same instructions, except they were told that they could also write unstudied words that belonged to the same categories as studied words. Finally, participants in the category labeling condition received the same instructions as participants in the control condition. However, on the written recall forms for this condition, the names of the four studied categories appeared as labels at the tops of the columns of spaces in which words were to be written.

Results and Discussion

The proportions of targets recalled in each condition on the three free-recall tests are shown in Table 1. The three testing conditions produced the expected ordering of recall accuracy. Across the three conditions, recall was poorer in the control condition than in the meaning condition, and it was poorer in the meaning condition than in the category labeling condition. A 3 (condition: control, meaning, category labeling) $\times$ 3 (recall test: 1, 2, 3) analysis of variance (ANOVA) revealed a main effect for condition, $F(2,116) = 27.73$, $MS_E = 0.06$. (All findings reported in this article, whether for ANOVAs or for analyses based on the dual-recall model, were reliable at the 0.05 level.) Post hoc tests (Tukey HSD) showed that the order of recall accuracy was control $<$ meaning $<$ category labeling. There was no main effect for test, indicating that there were no increases in recall probability across tests, and there was no Condition $\times$ Test interaction.

We estimated the parameters of the dual-retrieval model separately for the three conditions

TABLE 1

Observed and Predicted Proportions of Correct Recall in Experiment 1

Condition/test	Type of proportion: Observed	Predicted
Control:		
Test 1	0.47	0.49
Test 2	0.48	0.49
Test 3	0.46	0.49
Meaning:		
Test 1	0.55	0.52
Test 2	0.53	0.52
Test 3	0.55	0.52
Category cuing:		
Test 1	0.75	0.76
Test 2	0.77	0.76
Test 3	0.75	0.76

and evaluated goodness of fit in the usual way by implementing Eqs. (7–14) in a likelihood function and then maximizing the function for the sample data using numerical methods (for statistical details, see Riefer & Batchelder, 1988). These analyses, as well as all others involving the dual-retrieval model, were conducted with Hu's (1998) General Processing Tree software. Estimates of the three parameters for each condition appear in Table 2, along with their standard deviations. Values of the $X^2(4)$ test of goodness of fit for the conditions are also shown. This goodness-of-fit statistic evaluates the null hypothesis that the data were generated by the dual-retrieval model against the alternative hypothesis that the data were generated by some other process that is not isomorphic to the model. As can be seen, none of the values of the fit statistic exceeded the critical value for four degrees of freedom (9.49). In Table 1, the correct recall proportions that are predicted by the model for each condition and each test are presented, along with the corresponding observed values. Note that all of the predicted–observed correspondences were very close and that there was no tendency for the model to systematically underestimate or overestimate the actual data.

Inspection of the parameter values suggests that the two manipulations had three qualitative effects that we anticipated. First, the estimated value of R was larger with meaning instructions than with control instructions (0.51 versus 0.28), and it was larger with category labels than with meaning instructions (0.78 versus 0.51). Second, the estimate of J was larger with category labels than it was with either standard instructions or meaning instructions (0.81 versus 0.53 or 0.46). Third, the estimate of D was virtually the same in all conditions.

With models such as the present one, between-condition differences in parameter values are evaluated for statistical significance with a three-step procedure (e.g., see Brainerd et al., 1999). The first step is analogous to the omnibus F test in ANOVA: For all conditions of an experiment, an experimentwise likelihood ratio test, which is a X^2 statistic, is computed to evaluate the null hypothesis that parameter values do not vary between conditions. This null hypothesis was rejected in the present experiment, $X^2(6) = 83.16$. The second step is analogous to paired comparisons in ANOVA: For each pair of conditions, a likelihood ratio test is computed to evaluate the null hypothesis that parameter values do not differ between that specific pair of conditions. This null hypothesis was rejected for control versus meaning, $X^2(3) = 17.94$, for meaning versus category labels, $X^2(3) = 45.20$, and for control versus category labels, $X^2(3) = 42.26$. Hence, some of the parameters were affected by the manipulations. The third step, which has no analogue in ANOVA, determines which parameters were affected by these manipulations. For each pair of conditions, a likelihood ratio test is computed for each parameter that evaluates the null hypothesis that it has the same value in that pair of conditions. This final series of tests produced the following results.

Concerning reconstructive retrieval, R was smaller in the control condition than in either the meaning condition or the category labeling condition, and R was smaller in the meaning condition than in the category labeling condition. Thus, consistent with the notion that R measures semantic reconstruction of targets, its value increased as testing conditions made it progressively easier to process the correct category concepts during reconstructive retrieval. Concerning judgments as to

TABLE 2

Parameter Estimates (Standard Deviations in Parentheses) and Goodness-of-Fit Statistics for Experiment 1

	Statistic			
Condition	*D*	*R*	*J*	$X^2(4)$
Control	0.40(0.02)	0.28(0.02)	0.53(0.05)	8.85
Meaning instructions	0.37(0.02)	0.51(0.03)	0.46(0.03)	7.26
Category cues	0.37(0.06)	0.78(0.03)	0.81(0.03)	0.17

whether to authorize the output of reconstructed targets, the value of *J* was larger in the category labeling condition than in the control condition, but the values for the control and meaning conditions did not differ significantly. This pattern is consistent with the fact that category labels provide a reliable, objective basis for deciding whether a reconstructed target could have been a member of one of the study lists. Finally, concerning direct access, parameterwise tests revealed that *D* did not vary between conditions, which is consistent with the notion that simply making it easier to process correct category concepts during retrieval should not impair participants' ability to directly access verbatim traces of specific targets.

In summary, Experiment 1 produced straightforward outcomes with respect to the questions of central interest. First, the goodness-of-fit tests showed that for a simple RTTT design in which participants studied lists of category exemplars, the dual-retrieval model gave excellent accounts of the data of all conditions. The mean value of the fit statistic for this experiment was well below the critical value for rejection of the null hypothesis that the data were generated by the model. Second, consistent with the notion that *R* measures reconstructive retrieval, this parameter increased as a function of manipulations that should prolong reconstructive processing of semantic information during recall. Third, consistent with the notion that *J* measures participants' willingness to authorize the output of reconstructed targets, this parameter increased when a reliable, objective criterion of list membership was available that could be applied to reconstructed items.

EXPERIMENT 2: MANIPULATING DIRECT ACCESS

Experiment 1 yielded promising results on the reconstruction and judgment parameters: *R* reacted to two test-phase manipulations that should prolong reconstructive processing of correct semantic information; *J* reacted to a testing manipulation that supplied a decision criterion that should increase confidence in reconstructed targets. We conjectured earlier that this effect would occur because participants should be more confident about *any* reconstructed target if they can check it against the correct label before outputting it. On the other hand, the availability of correct category labels may cause reconstructive retrieval to generate targets that are better category exemplars, which should also increase confidence. This second hypothesis is inconsistent with other aspects of the data, however. Remember that total recall was highest in the category-labeling condition. The exemplars that formed the study lists varied considerably in their category representativeness (e.g., CHAIR is a better furniture exemplar than DRESSER). When we examined the category representativeness of the targets that were recalled in the three conditions, using available production norms (e.g., Battig & Montague, 1969), the average representativeness was *lowest* in the category-labeling condition (because more exemplars were recalled).

The fact that the manipulations in Experiment 1 had no effect on the parameter that measures direct access, *D*, is consistent with the hypothesis that it indexes some other form of retrieval that, unlike *R*, does not involve constructive semantic processing. However, this does not directly establish that *D* is dependent on processing the *surface* content of targets, nor does it rule out the possibility that R is dependent on processing surface as well as semantic content. The purpose of Experiment 2 was to investigate these latter possibilities.

To maximize comparability, the procedure in Experiment 2 was very similar to that of Experiment 1. Participants studied four lists of category exemplars, followed by three free-recall tests. In the dual-retrieval account, direct access is the preferred basis for output. Therefore, in contrast to Experiment 1, the focal manipulations were designed to disrupt direct access by disrupting participants' normal encoding of the surface forms of these familiar words by administering atypical visual study tasks. The tasks are illustrated in Fig. 2. For participants in the control condition (Fig. 2, left), targets on all lists were presented in a single, standard font. For participants in the unusual-fonts condition (Fig. 2, center), the targets on individual lists were presented in 10 atypical fonts. Obviously, these atypical fonts should interfere somewhat with participants' ordinary methods of encoding these words' orthographic forms. For participants in a consonant-counting condition (Fig. 2, right), words were presented in the same font as in the control condition, but participants were required to identify the consonants and to count them. This procedure ought to interfere more severely with ordinary methods of encoding words' orthographic forms. In the unusual fonts condition, participants could still follow normal encoding procedures with respect to differential processing of specific orthographic elements and the order in which different orthographic elements are processed (see also, Wallace et al., 1998). In the consonant-counting condition, participants were required to concentrate on arbitrary orthographic elements that had been selected by the experimenter.

Three predictions were of interest. The first is that if D measures retrieval of targets' verbatim traces, the observed ordering of its estimates should be control $>$ unusual fonts $>$ consonant counting. Second, because the lists' semantic content was made highly apparent to participants by presenting lists of same-category exemplars, the surface-difficulty manipulations should not impair memory for category concepts. Therefore, if R depends on memories of the latter sort, but not on memory for words' surface forms, R should not react to these manipulations. Third, J should not react to these manipulations either, because they do not provide information that would increase participants' confidence that reconstructed items appeared on study lists.

Method

Participants. Participants were 90 undergraduate communication and education students. They participated in the experiment for extra credit.

Materials and procedure. The pool of 192 words from Experiment 1 was expanded to 240 by selecting 2 further exemplars from the Battig and Montague (1969) norms for the 24 categories, again subject to the restriction that pairs of strong associates (e.g., TABLE–CHAIR) were not included. This pool was used to construct the lists presented to individual participants. Individual lists were constructed by randomly sampling 4 categories from the pool. Those four lists were then presented to a participant. To make the relevant category concepts obvious, list presentation was blocked by category (i.e., all 10 exemplars of a given category were presented consecutively). The order of list presentation was random.

Thirty participants were assigned to each of three conditions: (a) control, (b) unusual fonts, (c) consonant counting. The participant was provided with a pencil and a booklet consisting of 12 pages. Page 1 contained a brief descriptive title for the experiment. Page 2 contained instructions for the study portion of the experiments. Pages 3–6 contained the study lists. Sample study lists for the three conditions are shown in Fig. 2. Pages 7, 9, and 11 were blank pages for solving arithmetic problems, Pages 8, 10, and 12 contained four columns of blank spaces (20 spaces per column) for written recall. After completing the information on Page 1, participants were told to turn to Page 2 and read the study-phase instructions. These instructions indicated that the next four pages of the booklet contained printed word lists and that, to prepare

FIG. 2. Illustrative lists for the control condition (left), the unusual fonts condition (middle), and the consonant counting condition (right) in Experiment 2.

for later memory tests, the participant should read each word on each page. The instructions for the control and unusual fonts condition were the same: To ensure that all words were read, participants were told to cross out each word as it was read and to stop when the blank page following the fourth list had been reached. The instructions for the consonant-counting condition were similar to those for the control condition and told participants to count the number of consonants in each word as they read it and to write the number in a blank space beside the word. These participants were also told to stop when they reached the blank page following the fourth list. After reading the instructions, the participants proceeded through the study phase in a self-paced manner.

Next, the participants used the blank page to solve arithmetic problems for 1 min. The arithmetic problems were provided orally by the experimenter so that time could be controlled. The participants were then instructed to turn to the next page and to perform written recall for 2 min. The instructions, which were provided orally by the experimenter, stated that correct spelling was not essential. At the end of 2 min, the participants were instructed to turn to the next (blank) page, and another 1 min of arithmetic problem solving ensued. The participants then turned to the next page and performed a further 2 min of written recall under the same instructions as before. At the end of 2 min, there was another 1 min of arithmetic problem solving, followed by a final 2 min of written recall.

Results and Discussion

The proportions of targets recalled in the three conditions on each test are shown in Table 3. The study conditions produced the expected ordering of recall accuracy: Recall was better in the control condition than in either the unusual fonts condition or the consonant-counting condition, and it was better in the unusual fonts condition than in the consonant-counting condition. A 3 (condition: control, unusual fonts, consonant counting) × 3 (recall test: 1, 2, 3) ANOVA revealed a main effect for study condition, $F(2,116) = 24.81$, $MS_E = 0.04$. Post hoc tests (Tukey HSD) showed that the order of recall accuracy was consonant counting < unusual fonts < control. There was no main effect for test (i.e., recall did not increase across tests), and there was no Condition × Test interaction.

We estimated the parameters of the dual-retrieval model separately for the three conditions

TABLE 3

Observed and Predicted Proportions of Correct Recall in Experiment 2

	Type of proportion	
Condition/test	Observed	Predicted
Control:		
Test 1	0.54	0.54
Test 2	0.54	0.54
Test 3	0.54	0.54
Unusual fonts:		
Test 1	0.46	0.46
Test 2	0.46	0.46
Test 3	0.46	0.46
Consonant counting:		
Test 1	0.32	0.32
Test 2	0.31	0.32
Test 3	0.34	0.32

and evaluated goodness of fit in the same manner as in Experiment 1. Estimates of the three parameters, along with their standard deviations, appear by condition in Table 4. The values of the $X^2(4)$ test of goodness of fit for the conditions are also shown. As in Experiment 1, none of the values of the fit statistic exceeded the critical value for null hypothesis rejection (9.49), and the mean value of the three statistics (5.46) was slightly greater than half the critical value. The correct recall proportions that are predicted by the model for each condition are reported for each recall test, along with the corresponding observed values, in Table 3. As in Experiment 1, the predicted–observed correspondences were very close in all conditions, and there was no tendency for the model to systematically underestimate or overestimate the actual data. Once again, then, for a simple RTTT design in which participants studied lists of semantically related targets, the dual-retrieval model gave good accounts of the data.

Inspection of the parameter values suggests that these two surface-difficulty manipulations had the predicted qualitative effects. To begin with, *D* reacted to both manipulations. As predicted under the assumption that *D* measures retrieval of verbatim traces of individual targets, this parameter was larger in the control condition than in the unusual fonts condition (0.44 versus 0.32), and it was larger in the unusual fonts condition than in the consonant-counting condition (0.32 versus 0.17). In contrast, the parameter *R* did not respond reliably to these surface-difficulty manipulations. Its mean value, across the three conditions, was 0.30, and the average between-condition difference in its value was 0.04. Finally, the parameter *J* also was not affected by these manipulations. Its mean value was 0.65, and the average between-condition difference in its value was 0.04.

These patterns were confirmed by the appropriate significance tests. First, the experimentwise test showed that the null hypothesis that none of the three parameters varied reliably among conditions could be rejected, $X^2(6) = 144.31$. Next, the three conditionwise tests showed that the null hypothesis that none of the parameters varied reliably between specific pairs of conditions could be rejected for control versus unusual fonts, $X^2(3) = 22.85$, for control versus consonant counting, $X^2(3) = 68.79$, and for unusual fonts versus consonant counting, $X^2(3) = 51.37$. Finally, a series of nine parameterwise $X^2(1)$ tests were computed to determine which of the three parameters varied between which of the three pairs of conditions. These tests revealed that *R* and *J* did not vary between conditions but that *D* was larger in the control condition than in either the unusual fonts condition or the consonant-counting condition and that *D* was also larger in the unusual fonts condition than in the consonant-counting condition.

The findings of this experiment conform to the hypothesis that *D* indexes retrieval of traces of targets' surface forms because this parameter reacted to manipulations that should disrupt participants' normal methods of processing the surface content of familiar words. The fact that these same manipulations did not affect the other retrieval parameter adds important information to the results of Experiment 1. Although Experiment 1 established that *R* reacts to manipulations that increase the ease of semantic processing during retrieval, as would be expected if it measures meaning-driven reconstruction of

TABLE 4

Parameter Estimates (Standard Deviations in Parentheses) and Goodness-of-Fit Statistics for Experiment 2

	Statistic			
Condition	*D*	*R*	*J*	$X^2(4)$
Control	0.43(0.02)	0.32(0.02)	0.59(0.04)	9.16
Unusual fonts	0.32(0.02)	0.31(0.02)	0.67(0.04)	3.22
Consonant counting	0.19(0.01)	0.27(0.02)	0.66(0.04)	2.00

targets, it is possible that the success of reconstructive retrieval might also depend on the ease of processing targets' surface content. If so, *R* should react to manipulations of surface-processing difficulty, which it did not in the present experiment.

EXPERIMENT 3: SIMULTANEOUS MANIPULATION OF DIRECT ACCESS AND RECONSTRUCTION

Experiments 1 and 2 supplied evidence that *D* and *R* index different types of retrieval. In the familiar terminology of dual-process studies of recognition, Experiments 1 and 2 produced single dissociations: situations in which one parameter but not the other changed its value. Although any single dissociation is suggestive of different retrieval processes, those in Experiments 1 and 2 make a particularized case for this hypothesis because they were pursuant to manipulations that embody specific assumptions about the differences between the two retrieval operations. The semantic-processing manipulations of Experiment 1 dissociated the parameter that is thought to measure meaning-based reconstruction from the parameter that is thought to measure direct access, whereas the surface-processing manipulations of Experiment 2 dissociated the parameter that is thought to measure direct access from the parameter that is thought to measure meaning-based reconstruction.

Experiment 3 had three aims. The first was to extend the dual-retrieval model from free to associative recall. As the model has produced acceptable fits and parameter dissociations in free recall, it is a candidate for extension to other types of recall. Associative recall differs from free recall in that participants are provided with a specific, unique retrieval cue for each target. A related task in which unique retrieval cues are also supplied for targets, word stem completion, has already been analyzed using a different dual-retrieval approach, Jacoby's (1991) process-dissociation model of the recollection-familiarity distinction (Jacoby, 1998; Jacoby, Toth, & Yonelinas, 1993). We did not implement that approach in our research (but see the General Discussion, below) for four reasons. First, no procedural modifications are necessary to extend the direct access-reconstruction model to associative recall because the RTTT paradigm is equally applicable to associative recall (see Estes, 1960). In contrast, the boundary conditions that Jacoby (1998) has specified for the process-dissociation paradigm appear to preclude extension from stem completion to free recall. Second, we wished to study the relation between retrieval and metacognitive judgment in associative as well as free recall. As noted earlier, the recollection–familiarity distinction does not incorporate a separate metacognitive judgment operation. Third, Bodner et al. (2000) challenged applications of process dissociation to stem completion on the ground that performance was better accounted for by the generate/recognize distinction than by the recollection–familiarity distinction. Fourth, the manipulations used in Experiments 1 and 2, as well as in earlier research on fuzzy-trace theory, embody representational assumptions that differ from the assumptions of the process-dissociation model. In other words, the psychological interpretations of the parameters of the present model differ from those of the process-dissociation model.

The second aim of Experiment 3 was to determine whether it is possible to produce double dissociations with the present model, to identify situations in which manipulations that ought to have opposite effects on direct access and reconstruction cause *D* and *R* to change their values in opposite directions. Double dissociations are probative results in dual-process research because they mitigate the statistical criticism that some single dissociations may be artifacts of stochastic dependencies between conditional and unconditional parameters (Curran & Hintzman, 1995; Jacoby & Shrout, 1997) or artifacts of differential reliability of conditional and unconditional parameters (Brainerd, Wright, Reyna, & Mojardin, 2001).

To determine whether double dissociations could be induced, we manipulated a content variable that, based on prior experiments (Howe, Brainerd, & Kingma, 1985; Paivio, Walsh, & Bons, 1994), was expected to have opposite effects on participants' tendency to process surface versus semantic information on study trials: concreteness. Findings from those experiments suggest that in associative recall, there is both an increase in surface processing and a decrease in semantic processing as target concreteness is increased (e.g., when highly abstract nouns are replaced by highly concrete ones) and that the effect is most pronounced when the cue members of cue–target pairs are abstract. A likely basis for this effect is that when highly concrete targets (e.g., AMBULANCE, PARROT) are studied or retrieved, the accompanying phenomenology is laced with intense perceptual experiences (e.g., vivid auditory and visual images come to mind; Paivio, 1971; Paivio, Yuille, & Madigan, 1968), which usurps processing and focuses it on surface content. In contrast, highly abstract nouns (e.g., CONCEPT, THEORY) do not generate intense perceptual phenomenologies, which shifts processing toward words' semantic content (Paivio et al., 1994).

Although our principal motivation for studying concreteness was its potential for inducing double dissociations between *D* and *R*, the judgment process that authorizes output of reconstructed targets should also be affected by this variable. We saw in Experiment 1 that, consistent with the idea that *J* measures metamemorial confidence about whether reconstructed items are in fact studied targets, the value of this parameter increased when participants were given an objective criterion for deciding whether reconstructed items could possibly have appeared on study lists. Specifically, the value of *J* increased from roughly 0.5 to 0.8 when free recall tests supplied the names of the categories to which targets belonged. We reasoned that the vivid perceptual phenomenologies that are generated when concrete nouns are reconstructed would be interpreted as good subjective evidence that these items were encountered during the study phase and would therefore elevate metamemorial confidence. This conjecture follows from the simple fact that perceptual information is encoded as lists are studied, and, consequently, the presence of vivid perceptual content in retrieved memories is a reliable (though imperfect) guide as to whether items appeared on study lists (e.g., Mather, Henkle, & Johnson, 1997). Thus, it is rational, in the long run, for metamemorial judgment to rely on the presence–absence of perceptual phenomenology as a basis for deciding whether reconstructed items should be passed on for output. Note, further, that this implies a possible double dissociation between *R* and *J*, because prior research suggests that semantic processing during retrieval will decrease as target concreteness increases.

The third aim of this experiment was to extend the dual-retrieval model to situations in which the probability of correct recall increases across tests. The simplest way to ensure such increases in an RTTT design is to interpolate additional study trials following the first and second recall tests. This procedure was followed in the present experiment.

Method

Participants. Forty undergraduate communication students participated in the experiment for extra credit.

Materials and procedure. Using the Paivio et al. (1968) norms, we selected a pool of 90 familiar nouns (mean frequency value on the Thorndike–Lorge word count was 27). The pool was subdivided into 45 abstract nouns (mean concreteness rating on the Paivio et al. norms was 3.3) and 45 concrete nouns (mean concreteness rating was 6.2). The items from these subpools were randomly sampled to construct the paired-associate lists that were presented to individual participants. Each list was composed of 40 cue–target pairs. There were four types of pairs, with 10 pairs of each type: (a) concrete cues–concrete targets, (b) concrete cues–abstract targets, (c) abstract cues–concrete targets, and (d) abstract cues–abstract targets.

The procedure was similar to that of Experiments 1 and 2, the key differences being that paired-associate lists were studied, that a study trial preceded each of the memory tests, and that the memory tests involved associative rather than free recall. At the start of the experiment, after reading a page of instructions that described the experiment in general terms, the participant listened to an audio recording of a paired-associate list. The 40 pairs were presented in random order. Individual pairs were presented at a 2-s rate, and there was a 4-s pause between consecutive pairs. After list presentation was complete, the participant was given a pencil and a blank sheet of paper, and arithmetic problems were solved for 1 min using the same procedure as before. Next, the participant was given a form containing the 40 cue words from the study list, printed in random order, and was told to write the target words that went with the cue words in blank spaces that appeared next to the cues. Two minutes of written recall followed. After 2 min, the participant turned in the recall form and received a second study trial, which consisted of listening to another audio tape on which the pairs were presented in a new random order. The participant then received a second blank sheet of paper and solved arithmetic problems for 1 min. Next, the participant was provided with a second recall form, with the cue words printed in a new random order, and another 2 min of associative recall ensued. After 2 min, the participant turned in the recall form, received another study trial, which consisted of listening to a third audio tape on which the pairs were presented in a new random order, followed by another 1 min of arithmetic problem solving, followed by another 2 min of written recall.

TABLE 5

Observed and Predicted Proportions of Correct Recall in Experiment 3

	Type of proportion	
Condition/test	Observed	Predicted
Abstract cues/abstract targets:		
Test 1	0.14	0.14
Test 2	0.29	0.29
Test 3	0.47	0.47
Abstract cues/concrete targets:		
Test 1	0.31	0.29
Test 2	0.53	0.53
Test 3	0.65	0.65
Concrete cues/concrete targets:		
Test 1	0.24	0.24
Test 2	0.50	0.49
Test 3	0.61	0.62
Concrete cues/abstract targets:		
Test 1	0.23	0.26
Test 2	0.38	0.37
Test 3	0.50	0.50

Results and Discussion

The proportions of targets that were recalled in each of the four list conditions on each of the three recall tests are shown in Table 5. Because each recall test was preceded by a study trial, performance improved steadily across tests: Pooling over the four list conditions, the proportions of correct recall were 0.23 (Test 1), 0.43 (Test 2), and 0.58 (Test 3). The usual concreteness effects were also evident in that recall was more accurate when targets (or cues) were concrete than when they were abstract. These findings were confirmed in a 2 (cues: concrete versus abstract) × 2 (targets: concrete versus abstract) × 3 (recall test: 1, 2, 3) ANOVA that produced main effects for cue concreteness, $F(1,37) = 4.87$, $MS_E = 0.03$, for target concreteness, $F(1,37) = 97.75$, $MS_E = 0.03$, and for recall test, $F(2,74) = 100.79$, $MS_E = 0.04$. In addition, there was a Cue Concreteness × Recall Test interaction, $F(2,74) = 29.57$, $MS_E = 0.01$, and a Target Concreteness × Recall Test interaction, $F(2,74) = 8.89$, $MS_E = 0.01$. The reason for both interactions is that the effects of concreteness increased over recall tests. Post hoc tests showed (a) that recall was more accurate for pairs with concrete cues on Tests 2 and 3 but reliable differences were not observed on Test 1 and (b) that recall was more accurate for pairs with concrete targets on Test 3 but reliable differences were not observed on Tests 1 and 2. In contrast to the former result, other studies (e.g., Humphreys & Yullie, 1973) have obtained a cue concreteness effect after a single study trial. This discrepancy is presumably due to design differences, and in particular, it may be due to the fact that concreteness was a between-condition manipulation in most prior studies. Between-condition manipulations allow strategic factors to reinforce differences in the memory processes that are induced by concrete versus abstract cues, thereby enhancing performance differentials.

To evaluate the predicted double dissociation, we estimated the parameters of the dual-retrieval model in two ways: (a) for total abstract targets (pooled across concrete and abstract cues) versus total concrete targets (pooled across concrete and abstract cues) and (b) separately for each of the four list conditions (abstract–abstract, abstract–concrete, concrete–abstract, concrete–concrete). These estimates are reported in Table 6. To obtain them, an extended version of the dual-retrieval model was required. Remember, here, that the basic model (Eqs. 7–14) assumes that its parameters are invariant over recall tests and that the probability of correct recall therefore also remains invariant. However, in this experiment, by design, recall increased over tests. Let D_1, D_2, and D_3 be the probabilities of learning how to directly access targets on study trials 1, 2, and 3, respectively, and let R_1, R_2, and R_3 be the probabilities of learning how to reconstruct targets on study trials 1, 2, and 3, respectively. The basic model that was used in Experiments 1 and 2 assumes that $D_2 = D_3 = R_2 = R_3 = 0$, and it also assumes that J remains constant across tests (i.e., $J_1 = J_2 = J_3$, where the subscript indexes the recall test). However, by relaxing some of these assumptions, the model may be able to account for the data of the present experiment.

The procedure for finding an extended version of the model that would fit the data involved three steps. First, we explored the simplest possibility—namely, that there was no further retrieval learning on subsequent study trials (i.e., $D_2 = D_3 = R_2 = R_3 = 0$), but metamemorial confidence in reconstructed items increased (i.e., $J_1 < J_2 < J_3$). Although this is the simplest modification that might be able to account for the data, it is not plausible theoretically, and, therefore, it is not surprising that it was unsuccessful. When J was free to vary across recall tests, the relevant fit statistic, an $X^2(2)$ test, produced values that far exceeded the critical value of 5.99. Therefore, the second step was to relax the assumption that retrieval learning did not occur on recall tests and allow D_2, D_3, R_2, and R_3 to all be greater than zero. The revised model that implements this possibility appears in the Appendix. Note that because these expressions contain seven theoretical parameters (D_1, D_2, D_3, R_1, R_2, R_3, and J), all seven empirical probabilities are exhausted in parameter estimation, leaving no degrees of freedom for goodness-of-fit tests. Nevertheless, in the third step of evaluating the extended model, it was possible to establish goodness of fit for all data sets because, once the parameters had been estimated, at least two were always close to zero. Hence, fit could be evaluated in each data set with an $X^2(2)$ statistic that assumed that two of the parameters were zero, rather than with the $X^2(4)$ statistic used in Experiments 1 and 2. When these tests were performed, none produced a result that exceeded the critical value for null hypothesis rejection. The correct recall proportions that are predicted by the extended model for the four conditions are reported for each recall test, along with the corresponding observed values, in Table 5. Note that

TABLE 6

Parameter Estimates (Standard Deviations in Parentheses) for Experiment 3

	Parameter						
Condition	D_1	D_2	D_3	R_1	R_2	R_3	J
Total abstract targets	0.08(0.02)	0.13(0.03)	0	0	0.09(0.09)	0.44(0.19)	0.33(0.25)
Total concrete targets	0.05(0.03)	0.16(0.04)	0.12	0	0	0.26(0.28)	0.63(0.26)
Abstract cues							
Abstract targets	0.05(0.01)	0.06(0.02)	0	0.04(0.03)	0.06(0.05)	0.34(0.21)	0.33(0.22)
Concrete targets	0	0.11(0.04)	0.19(0.48)	0	0.06(0.05)	0.05(0.12)	0.54(0.33)
Concrete cues							
Abstract targets	0.11(0.02)	0.20(0.03)	0	0.05(0.04)	0.14(0.12)	0.61(0.17)	0.33(0.27)
Concrete targets	0.07(0.03)	0.21(0.03)	0	0	0.02(0.04)	0.49(0.43)	0.67(0.19)

Note. Zero entries for parameters mean that the estimation procedure assumed that those particular parameters had zero values.

the predicted–observed correspondences were very close in all conditions and that there was no tendency for the model to systematically underestimate or overestimate the actual data.

A conceptual limitation of the constraints that were used to evaluate fit is that although they produced acceptable fits (Table 5), they were a posteriori; none was specified in advance on theoretical grounds. Because these constraints involved one or more parameters having zero values, it might be thought that their success was due to unreliability of parameter estimation. We examined this possibility and concluded that it was not problematical. The reliability with which parameters are estimated depends primarily on the proportions of protocols that are used in estimation (e.g., Riefer & Batchelder, 1988). For the direct access parameters, all protocols are used to estimate D_1. The proportions that are used to estimate D_2 and D_3 are $(1 - D_1)$ for D_2 and $(1 - D_1)(1 - D_2)$ for D_3. In this experiment, those proportions were in the 0.7–0.9 range. For the reconstruction parameters, the proportions of protocols that are used in estimation are $(1 - D_1)$ for R_1, $(1 - D_1)(1 - D_2)(1 - R_1)$ for R_2, and $(1 - D_1)(1 - R_1)(1 - D_2)(1 - R_2)(1 - D_3)\ R_2$ for R_3. Those proportions were in the 0.6–0.9 range.

Turning to the parametric effects of the concreteness manipulation, three predicted relations were of interest, that relative to concrete targets, (1) abstract targets would be harder to directly access, (2) they would be easier to reconstruct, and (3) participants would be less willing to output them following reconstruction. As in Experiments 1 and 2, experimentwise tests followed by conditionwise tests followed by parameterwise tests were computed to evaluate these predictions. The experimentwise test produced a rejection of the null hypothesis that the model's parameters had the same values across the four conditions, $X^2(21) = 363.27$, and each of the six conditionwise tests produced a rejection of the null hypothesis that the model's parameters had the same values for that particular pair of conditions (mean $X^2(7) = 48.88$). Finally, the parameterwise $X^2(1)$ tests confirmed each of the predicted relations. Concerning direct access, when the three direct-access parameters (D_1, D_2, and D_3) were compared for abstract versus concrete targets, the value of D_3 was significantly larger for concrete targets. Concerning reconstruction, when the three reconstruction parameters (R_1, R_2, and R_3) were compared for abstract versus concrete targets, R_2 and R_3 were both significantly larger for abstract targets. Finally, concerning metamemorial judgment, there was a large difference between concrete and abstract targets, the probability that abstract targets would be authorized for output following reconstructive retrieval being only about half the corresponding probability for concrete targets.

An instructive way of summarizing the overall double dissociation between direct access and reconstruction is to compare the respective probabilities of targets being accessed via each retrieval operation *by the end of the experiment* (i.e., after three study trials). These probabilities are the most sensitive measures of the dissociation because they combine all of the retrieval learning that occurred on all of the study trials. At the end of the experiment, the probability of a target's being retrieved via direct access is given by the quantity $[D_1, + (1 - D_1)D_2 + (1 - D_1)(1 - D_2)D_3]$, whereas the probability of its being retrieved via reconstruction is given by the quantity $(1 - D_1)(1 - D_2(1 - D_3)[R_1 + (1 - R_1)R_2 + (1 - R_1)(1 - R_2)R_3]$. Calculation of these quantities for the data in the first row of Table 6 showed that the terminal probability of retrieving abstract targets via direct access was 0.20, and the terminal probability of retrieving them via reconstruction was 0.39. In sharp contrast, calculation of the same quantities for the data in the second row of Table 6 showed that the terminal probability of retrieving concrete targets via direct access (0.30) was larger than that for abstract targets, but the terminal probability of retrieving concrete targets via reconstruction (0.18) was smaller than that for abstract targets. Thus, after three study trials, the double dissociation between rates of direct access and reconstruction was quite marked. Note that in addition to doubly dissociating reconstruction from direct access, target concreteness doubly dissociated reconstruction from metamemorial judgment: The terminal probabilities of reconstructive retrieval were 0.39 and 0.18 for abstract and concrete targets, respectively, while the corresponding values of *J* for these same conditions were 0.33 and 0.63.

Although double dissociations mitigate the criticisms that single dissociations can be due to parameter correlations and differential reliability of parameter estimates, the hypothesis that different parameters measure the same process is not entirely ruled out. As Dunn and Kirsner (1988) pointed out, a manipulation can induce a double dissociation if parameters measure opposite aspects of a single process. To address this ambiguity, Dunn and Kirsner developed a further test, called reversed association, that can be conducted *if more than two paired values of the parameters are available*. If the same-process hypothesis is correct, plots of three or more values of one parameter against the corresponding values of the other parameter must yield monotonic-decreasing curves, such as the illustrative ones in panel A of Fig. 3. If such plots yield nonmonotonic curves (i.e., increases in one variable are sometimes associated with decreases in the other variable and are sometimes associated with increases), this is logically incompatible with the same-process hypothesis. Illustrative nonmonotonic curves are shown in Panel B of Fig. 3.

Dunn and Kirsner's (1988) reversed association test was used to pit the same-process and different-process interpretations of the present double dissociations against each other. There were two types of word pairs containing abstract targets (those with concrete cues and those with abstract cues) and two further types of word pairs with concrete targets (those with concrete cues and those with abstract cues), yielding four pairs of *D* and *R* values (and four pairs of *R* and *J* values). To perform the reversed association test, we plotted the terminal (Trial 3) probability of directly accessing targets against the terminal probability of reconstructing targets, and we also plotted *J* against the terminal probability of reconstructing targets. The resulting curves are shown in Fig. 4. Obviously, the same-process hypothesis can be rejected for *D* versus *R* and for *R* versus *J* because both curves are nonmonotonic.

Summing up, manipulating target concreteness in associative recall produced two double dissociations. One was between measures of the two forms of retrieval, with direct access being easier when targets were concrete but reconstruction being easier when targets were abstract. The other double dissociation was between reconstruction and metamemorial judgment, with targets being harder to reconstruct when they were concrete but, once reconstructed, being more likely to be authorized for output. Reversed association tests confirmed that neither double dissociation could be explained by a one-process hypothesis. These dissociations are of both technical and theoretical importance. On the technical side, the fact that

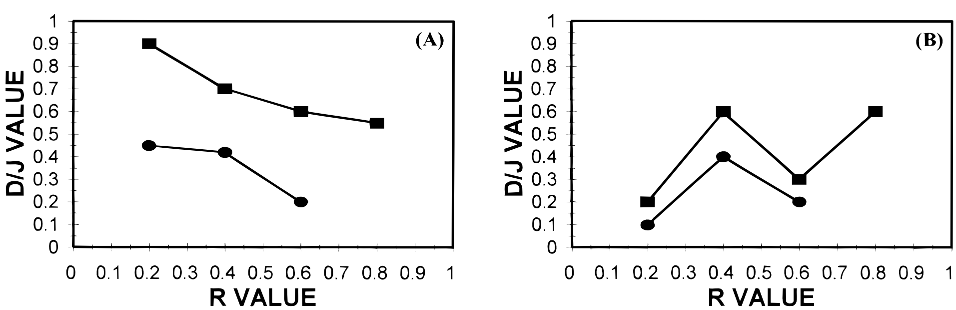

FIG. 3. Hypothetical monotonic and nonmonotonic relations between three or more paired estimates of *D* and *R*. Monotonic-decreasing relations (Panel A) are consistent with the hypothesis that *D* and *R* measure the same process and that double dissociations are due to a manipulation having opposite effects on that process. Nonmonotonic relations (Panel B) rule out this hypothesis, showing that double dissociations are due to *D* and *R* measuring separate processes.

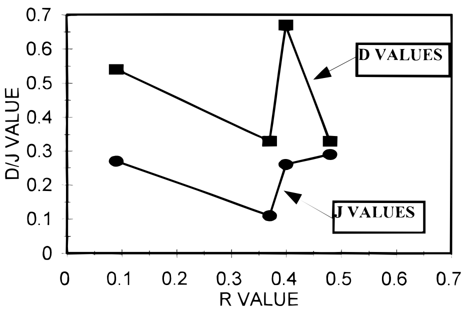

FIG. 4. Reversed association tests of the double dissociation between *D* and *R* and the double dissociation between *R* and *J*.

the direct access and reconstruction parameters were driven in opposite directions by a content manipulation provides evidence that is not subject to statistical caveats that have been raised in connection with single dissociations (see the General Discussion). On the theoretical side, both the direct access versus reconstruction dissociation and the reconstruction versus judgment dissociation are congruent with theoretical characterizations of these processes. Concerning the former dissociation, prior research (e.g., Paivio et al., 1994) suggests that, relative to abstract nouns, concrete nouns shift processing away from semantic content toward surface content, which should enhance direct access (if it involves processing surface content) and impair reconstruction (if it involves processing semantic content). Concerning the latter dissociation, although concrete targets may be more difficult to reconstruct than abstract targets, reconstructed concrete targets should be more apt to be authorized for output because their associated phenomenology is strongly perceptual.

EXPERIMENT 4: FURTHER SIMULTANEOUS MANIPULATION OF DIRECT ACCESS AND RECONSTRUCTION

In this last experiment, we applied the dual-recall model concurrently to free and associative recall, in order to seek additional evidence of double dissociation. Memory theorists have often treated associative recall as though it were a direct-access task (e.g., Greeno, James, Da Polito, & Polson, 1971; Underwood, 1953) and free recall as though it were a reconstruction task (e.g., Ackerman, 1985; Halff, 1977), a distinction that receives empirical support from the familiar finding that increasing the degree of semantic relatedness between targets (e.g., by presenting targets that are same-category exemplars) impairs associative recall but improves free recall (for a review, see Hunt & McDaniel, 1993). On its face, associative recall seems to slant retrieval toward direct access because it (a) provides a unique retrieval cue for each target (rather than global cues that span many targets) that is neither semantically or associatively related to the target, (b) provides retrieval cues that reinstate part of the surface information that is stored in verbatim traces (namely, the cue member of each pair), and (c) requires that targets *only* be recalled in response to the specific cues with which they were paired at study. The first feature focuses retrieval squarely at the level of individual items (rather than groups of items), the second focuses retrieval on the type of information that is ostensibly stored in item-level traces, and the third discourages the use of retrieval cues to reconstruct multiple output candidates.

Free recall, by comparison, seems on its face to be slanted more toward reconstruction. During the study phase, targets are not mapped with unique retrieval cues. On recall tests, only global, listwide retrieval cues are supplied to participants. Those cues do not reinstate the surface content of specific targets, and participants can output targets without regard to surface details of the earlier study context (e.g., without regard to the order in which they were studied). This is not to claim that participants only use global, listwide cues during retrieval. On the contrary, as specific targets are output, they may be used as retrieval cues for further targets. *Relative to associative recall*, however, where specific cue words are repeatedly mapped with targets on study trials and are presented on test trials, reliance on item-level cues should be reduced and reliance on global cues should be increased in free recall.

If *D* and *R* measure direct access and reconstruction, then, for a given set of *unrelated* targets, associative recall ought to produce larger values of D and smaller values of R than free recall: Taking free recall as the baseline task, associative recall provides stronger support for direct access *and* discourages reconstruction. (Remember, again, that cue members of paired associates are selected so as to be unrelated to their targets.) To evaluate this prediction, some participants in the experiment studied lists of unrelated words and performed free recall, whereas others studied lists of unrelated word pairs and performed associative recall. Two free-recall conditions were included to control for list-length differences between associative and free recall. For lists consisting of N targets, twice as many words are presented on paired-associate lists as on free-recall lists (because each target on a paired-associate list is studied with a unique cue word). To control for the possibility that differences in the numbers of studied words might contribute to differences in parameter estimates for associative versus free recall, participants in one free-recall condition studied lists whose length equaled the number of targets + cues on the corresponding paired-associate lists.

A final feature of the design is that, as in Experiment 3, each recall test was preceded by a study trial. In Experiment 3, it was found that an extended version of the dual-retrieval model that did not exhaust the available degrees of freedom could account for situations in which the probability of correct recall increased across tests. However, this result might be unique to the design of Experiment 3, or it might not generalize to either free recall or associative recall of other lists. To determine the result's generalizability, we used the same alternating study–test procedure in Experiment 4.

Method

Participants. One hundred and thirty-seven undergraduate communication students participated in the experiment for extra credit.

Materials and procedure. A pool of 200 words was formed by selecting words from the Toglia and Battig (1978) norms. This pool was used to construct the lists presented to individual participants. Four paired-associate lists were formed by selecting 80 words at random from the pool and then randomly pairing selected words to obtain 40 pairs. After the 40 pairs had been constructed for each list, they were inspected to determine if random pairing had produced any pairs in which there were obvious meaning relations between cues and targets. A few such pairs were detected. They were eliminated by exchanging the targets with the target members of other cue–target pairs. Four 40-item free-recall lists were then constructed using just the 40 targets from each of the paired-associate lists. Four 80-item free-recall lists were then constructed using both the 40 targets and the 40 cues from each of the paired-associate lists. The four lists were rotated through participants in the associative and free-recall conditions. As there were no performance differences in any of the conditions as a function of which list was studied, we do not discuss this factor further.

Participants were assigned randomly to three conditions: (a) associative recall, (b) 40-item free recall, and (c) 80-item free recall. At the start, participants in all conditions read a page of general memory instructions that described the experiment. Next, the participants listened to an audio recording of a list that was appropriate to their condition. To control presentation time, the words on individual lists were read at slightly different rates so that total presentation time was the same in all conditions. In the associative recall condition, pairs were presented at a rate of one pair every 6-s: A pair was presented during the first 2-s of each 6-s interval, followed by a 4-s pause. In the 40-item free-recall condition, words were presented at a constant 6-s rate. In the 80-item free-recall condition, words were presented at a constant 3-s rate. Following the study trial, participants engaged in the usual 1 min of arithmetic problem solving, followed by 2 min of written free recall. For the associative-recall condition, the instructions and procedure for the recall test were the same as in Experiment 3. For the two free-recall conditions, the instructions and procedure for the recall test were the same as in the control conditions of Experiments 1 and 2. For each condition, the first recall test was followed by a second study trial (new presentation order), another 1 min of arithmetic problem solving, and a second written recall test. For each condition, the second recall test was followed by a third study trial (new presentation order), another 1 min of arithmetic problem solving, and a final written recall test.

Results and Discussion

The proportions of targets recalled in the associative and free-recall conditions are shown in Table 7. The values for the associative and 40-item free-recall conditions are based on the actual number of studied targets, 40. For purposes of comparison, the values for the 80-item free-recall condition are based on a list length of 40 targets. (In order to compare correct recall proportions across all three conditions, it is necessary to calculate those proportions using the same baseline list length.) As would be expected, the proportions were consistently lower for 80-item than for 40-item free recall. It will be remembered that the reason for including both of these conditions was methodological, specifically, to control for the possibility that the larger number of words on paired-associate lists might be responsible for differences in the mix of retrieval processes that are used in associative versus free recall. Because recall was poorer in the 80-item condition that in the 40-item condition, conclusions about the effects of associative and free recall on direct access versus reconstruction must be based on comparisons of parameter estimates for associative recall to parameter estimates for both 40- and 80-item free recall.

It can be seen in Table 7 that the three conditions produced the following ordering of recall performance: 40-item free recall (mean proportion correct = 0.40) > 80-item free recall (mean proportion correct = 0.33) > associative recall (mean proportion correct = 0.26). It can also be seen that performance improved across the three recall tests, with the mean proportion correct rising from 0.15 for Test 1 to 0.30 for Test 2 to 0.43 for Test 3, which is expected because each recall test was preceded by a study trial. These two patterns were confirmed in a 3 (condition: associative recall, 40-item free recall, 80-item free recall) × 3 (recall test: 1, 2, 3) ANOVA of proportion correct, which produced a main effect for condition, $F(2,134) = 10.66$, $MS_E = 0.06$,

TABLE 7

Observed and Predicted Proportions of Correct Recall in Experiment 3

	Type of proportion	
Condition/test	Observed	Predicted
Associative recall		
Trial 1	0.10	0.10
Trial 2	0.27	0.26
Trial 3	0.44	0.45
Free recall, 40-item		
Trial 1	0.26	0.27
Trial 2	0.46	0.46
Trial 3	0.60	0.60
Free recall, 80-item		
Trial 1	0.10	0.11
Trial 2	0.18	0.17
Trial 3	0.25	0.25

and a main effect for recall test, $F(2,268) = 530.70$, $MS_E = 0.007$. Post hoc analyses (Tukey HSD) showed that the ordering of proportion correct by condition and by recall test were the same as indicated above. There was no Condition × Recall Test interaction.

To evaluate the effects of associative versus free recall on the two retrieval processes, we estimated the parameters of the dual-retrieval model for each of the three list conditions. Those estimates appear in Table 8, along with their standard deviations. (Estimates for the 80-item free-recall condition were based on a functional list length of 40.) It was again necessary to use an extended version of the dual-retrieval model that permitted retrieval learning on each of the three study trials (cf. the Appendix) because, as in Experiment 3, recall improved across tests in all three conditions. Also as in Experiment 3, a two-step procedure was used to locate a version of the dual-retrieval model that would provide acceptable fits to the data of all conditions. We first tried the simple expedient of preserving the assumption that retrieval learning did not occur after the first study trial, while allowing participants' willingness to authorize a reconstructed target for output (parameter *J*) to vary across trials. However, when only *J* was free to vary, the relevant fit statistic, a $X^2(2)$ test, produced values that far exceeded the critical value of 5.99. Therefore, second, we relaxed the assumption that retrieval learning did not occur after the first study trial, allowing D_1, D_2, and D_3 to be the probabilities of learning how to directly access targets on the respective study trials and allowing R_1, R_2, and R_3 to be the corresponding probabilities of learning how to reconstruct targets. Relaxing this assumption produced an acceptably fitting version of the dual-retrieval model for each of the three conditions. As can be seen in Table 8, the best-fitting version for the associative recall condition had values of D_1, D_2, D_3, R_1, and R_2 that were greater than zero but a zero value of R_3, whereas the best-fitting version for the two free-recall conditions had values of D_1, D_2, R_1, R_2, and R_3 that were greater than zero but a zero value of D_3. In addition, $R_1 = R_2$ in the associative condition and $D_1 = D_2$ in the 80-item free-recall condition. The correct recall proportions that are predicted by the model for each condition are reported for each recall test, along with the corresponding observed values in Table 7, where it can be seen that the predicted-observed correspondences were again very close, with no tendency for the model to systematically underestimate or overestimate the actual data.

As in Experiment 3, the constraints that produced acceptable fits were not specified a priori. Therefore, we again examined the possibility that unreliability of parameter estimation might be responsible for the success of those constraints. Reliability of parameter estimation did not appear to be problematical, however, because the proportion of protocols that was involved in estimating D_1, D_2, and D_3 and in estimating R_1, R_2, and R_3 always exceeded 0.50. This replicates findings from Experiment 3.

To determine how the type of recall test (associative versus free) affected direct access and reconstruction, we conducted between-condition parameter comparisons, using the same three-step procedure as before. That is, first, we computed an experimentwise likelihood ratio test to evaluate the null hypothesis that parameter values did not vary among the three conditions. This test produced a null hypothesis rejection, $X^2(14) = 273.58$. Second, for each of the three possible pairs of conditions, a likelihood ratio test was computed to evaluate the null hypothesis that parameter values did not differ between that specific pair of conditions. This null hypothesis was rejected for associative versus 40-item free recall, $X^2(7) = 82.88$, for associative versus 80-item free recall, $X^2(7) = 146.67$, and for 40- versus 80-item free recall, $X^2(7) = 18.91$ Thus, some of the parameters were affected by the associative versus free recall manipulation, regardless of whether the free-recall list was composed of 40 or 80 items.

Third, for each pair of conditions, we computed a parameterwise likelihood ratio test for each of the seven parameters (D_1, D_2, D_3, R_1, R_2, R_3, and J) to identify specific parameters that varied between those conditions. Taking the comparison of primary interest first, we noted earlier that theoretical considerations argue that, relative to free recall, associative recall is more of a direct access task and less of a reconstruction task, so that the associative versus free recall comparisons ought to produce double dissociations between parameters that measure these

TABLE 8

Parameter Estimates (Standard Deviations in Parentheses) for Experiment 4

	Parameter						
Condition	D_1	D_2	D_3	R_1	R_2	R_3	J
Associative recall	0.08(0.01)	0.16(0.01)	0.24(0.12)	0.07(0.02)	0.07(0.02)	0	0.32(0.10)
Free recall							
40 item	0.10(0.01)	0.22(0.01)	0	0.28(0.02)	0.10(0.04)	0.57(0.02)	0.51(0.03)
80 item	0.06(0.02)	0.06(0.02)	0	0.29(0.03)	0.11(0.04)	0.58(0.16)	0.48(0.05)

Note. Zero entries for parameters mean that the estimation procedure assumed that those particular parameters had zero values.

retrieval operations. They did. To begin, as in Experiment 3, it is instructive to compare the probabilities of targets being retrieved via direct access and reconstruction by the end of the experiment, because these probabilities combine all of the retrieval learning that occurred on the study trials. Calculation of these quantities for associative versus free recall (see Experiment 3 for formulas) revealed a robust double dissociation. In the associative condition, the estimated probabilities of direct access and reconstruction were 0.41 and 0.08, respectively. Thus, by the end of the experiment, associative recall was dominated by direct access. The picture was reversed in the two free-recall conditions. The respective estimates of direct access and reconstruction were 0.30 and 0.58 for 40-item free recall and 0.12 and 0.65 for 80-item free recall. The pattern of retrieval at the end of the experiment was therefore consistent with the notion that with homogeneous lists of unrelated targets, associative recall is more of a direct access task and free recall is more of a reconstruction task.

At the level of individual parameter comparisons, the parameterwise tests for associative recall versus 40-item free recall showed that D_3 was reliably larger in the associative-recall condition but that R_1 and R_3 were reliably larger in the free-recall condition. The parameterwise tests for associative versus 80-item free recall also showed that D_3 was reliably larger in the associative-recall condition but that R_1 and R_3 were reliably larger in the free-recall condition. In short, regardless of whether associative recall was compared to free recall of lists with the same number of targets or to free recall of lists with the same number of cues + targets, the same pattern of double dissociation between direct access and reconstruction was obtained. When values of the direct-access and reconstruction parameters were compared within each condition across the three recall tests, another finding emerged that reinforces the conclusion that associative recall was weighted toward direct access and free recall was weighted toward reconstruction. In the associative condition, direct-access learning occurred on all three study trials, and the amount of direct-access learning increased across study trials. However, reconstruction learning did not occur after the second study trial.[3] The pattern was reversed for free recall. With both 40- and 80-item free recall, direct-access learning did not occur after the second study trial, but reconstruction learning occurred on all study trials, and the amount of reconstruction learning was greater on the last study trial than on the first two.

In addition, an unpredicted difference was observed between associative and free recall. For both associative versus 40-item free recall and associative versus 80-item free recall, the judgment parameter, J, was reliably larger for free recall. Although this difference was not predicted, it is not unreasonable that metamemorial confidence in reconstructed targets would be higher in conditions in which reconstruction is the predominate mode of retrieval than in conditions in which other retrieval methods predominate.

What does the dual-retrieval model tell us about the reasons for the performance differential between 40- and 80-item free recall? The values reported above for the probabilities of targets being directly accessed versus reconstructed on the final recall test suggest that poorer performance in the 80-item condition was entirely attributable to reduced ability to directly access targets: The direct-access probabilities were 0.12 and 0.30 in the 40- and 80-item conditions, respectively, but the reconstruction probability was slightly higher in the 80-item condition than in the 40-item condition. At the level of individual parameter comparisons, the parameterwise tests for 40-item versus 80-item free recall revealed that the only parameter that differed reliably between the two conditions was one of the direct-access parameters, D_2.

A remaining concern about the dual-retrieval model is that because the goodness-of-fit tests in this experiment and Experiment 3 relied on a posteriori constraints to evaluate fit, no evidence

[3] In Experiment 3, which involved associative recall, this particular pattern was not present (see Table 6). In that experiment, reconstructive retrieval learning was confined to the third study trial, a difference that may be due to studying mixed lists composed of four types of pairs, rather than homogeneous lists. Regardless, the overall picture of associative recall favoring direct access was analogous in Experiments 3 and 4.

of fit has been obtained for associative recall using a priori tests, although such evidence has been obtained for free recall (Experiments 1 and 2). A priori fit results could be secured, however, by analyzing previous RTTT data for associative recall. Data of this sort were available from two previously reported experiments (Brainerd, Kingma, & Howe, 1985). When the a priori fit test in Experiments 1 and 2 was computed for those experiments, mean values were below the critical value of 9.49, providing a priori evidence of fit for associative recall. It might also be argued that support for the dual-retrieval model is limited by the fact that evidence of fit using a priori tests has not been obtained for situations in which retrieval learning occurs after the first study trial. However, it is easy to see that relevant data could be generated by extending the model from the present $R_1TR_2TR_3T$ design to a $R_1TR_2TR_3TR_4T$ design. In this extension, there are 16 empirical probabilities (15 of which are free) but only nine theoretical parameters to estimate (D_1, D_2, D_3, D_4, R_1, R_2, R_3, R_4, J). An a priori fit test with six degrees of freedom is therefore available. To acquire some pertinent findings, we computed this test for the conditions of four experiments reported by Brainerd et al. (1990), in which there were four or more study–test cycles in all conditions. This data set included both free and associative recall. Mean values of the $X^2(6)$ statistic were below the critical value of 12.59, providing a priori evidence of fit for both free and associative recall for situations in which further retrieval learning occurs after the first study trial.

GENERAL DISCUSSION

The present experiments were conducted with a view toward extending quantification of dual-retrieval processes to free and associative recall. The experiments were designed as an initial evaluation of a paradigm that allows two retrieval operations, direct access and reconstruction, plus a metacognitive judgment operation to be estimated. The paradigm consists of a simple experimental procedure, an Estes-type RTTT design, together with a retrieval model that is defined over that procedure. The model extracts estimates of: (a) the probability that a target can be directly accessed on a recall test; (b) the probability that a target can be reconstructed on a recall test; and (c) the probability that metamemorial confidence in a reconstructed target is sufficiently high to authorize its output. In the present section, we first summarize the main things that were learned about these processes from our experiments. Next, we discuss findings of parameter dissociation. Last, we sketch relations between our paradigm and paradigms that have been used to quantify dual-retrieval processes in recognition.

Direct Access, Reconstruction, and Metamemorial Judgment

A key preliminary finding is that the dual-retrieval model gave acceptable fits to data sets in which the probability of correct recall remained constant across the three free recall tests (Experiments 1 and 2). Further, an extension of the basic model, which permitted the probability of correct recall to vary over tests, gave acceptable fits to data sets in which this probability was made to increase by administering a study trial before each recall test (Experiments 3 and 4). Thus, quite apart from what research may ultimately show about theoretical interpretations of direct access, reconstruction, and metamemorial judgment, the fit analyses provided support for a model that assumes that these processes control recall. As with all mathematical models of memory, however, it is important to bear in mind that the mapping of posited retrieval processes onto model parameters is not necessarily one-to-one (Brainerd et al., 2001). More explicitly, although the model that we have used is implied by our dual-retrieval distinctions, other theoretical distinctions might imply the same model. Such contrasting process interpretations are traditionally adjudicated by determining how a model's parameters react to manipulations that embody those interpretations (Brainerd, Howe, & Desrochers, 1982).

In that connection, the selection of manipulations in the reported experiments was guided by a working conception of the two retrieval operations that was borrowed from fuzzy-trace theory (see Payne et al., 1996; Powell, Roberts, Ceci, & Hembrooke, 1999; Reyna & Brainerd, 1995). According to that conception, one operation retrieves verbatim traces of individual targets and reads out the stored surface information, and the other operation retrieves gist memories of lists' semantic content and processes them constructively to regenerate targets (with a judgment operation then deciding which constructions to output). These distinctions lead to the obvious expectation that direct access will be influenced by manipulations that affect the difficulty of processing targets' surface forms and by manipulations that orient retrieval toward individual targets (rather than toward relations among targets). The related expectation about reconstruction is that it will be influenced by manipulations that affect the processing of targets' semantic content and by manipulations that orient retrieval toward relations among targets (rather than toward individual targets). Beyond these general expectations, if the retrieval operations are truly distinct, it ought to be possible to identify manipulations that will dissociate direct access from reconstruction, and it ought to be possible to identify manipulations that will dissociate reconstruction from direct access.[4] We concentrated on candidate manipulations of this sort and were able to identify some that produced both types of dissociations.

Taking major findings for direct access first, in Experiment 2, to minimize variability in the processing of lists' semantic content, meaning relations among targets were made conspicuous to participants by presenting lists composed of blocks of same-category exemplars (e.g., 10 articles of furniture, followed by 10 animals, followed by 10 cities, followed by 10 metals). A manipulation was imposed that should moderately disrupt participants' normal methods of processing of words' surface forms (presenting words in several unusual fonts). Relative to a control condition in which words were presented in a single standard font, this manipulation decreased the probability of direct access (parameter *D*) by roughly 25%, but it did not affect the probability of reconstruction (parameter *R*). A further manipulation was imposed that should severely disrupt normal processing of words' surface forms (requiring participants to identify and count the consonants in each word). Relative to the same control condition, this manipulation decreased the probability of direct access by roughly 75%, but, again, the probability of reconstruction was not affected. In Experiment 3, another manipulation, concreteness of targets, was introduced that should influence the tendency to process surface information as targets are studied. It was found that cue–target pairs in which targets were very concrete (e.g., AMBULANCE) produced higher direct-access probabilities (and lower reconstruction probabilities) than pairs in which targets were very abstract (e.g., CONCEPT). Last, in Experiment 4, we compared free recall to a paradigm (associative recall) that should shift retrieval toward accessing traces of individual targets, rather than accessing semantic relations between them. Relative to free recall, associative recall increased direct-access probabilities and reduced reconstruction probabilities.

Turning to major findings for reconstruction, in Experiment 1, participants studied the same types of lists as in Experiment 2 (i.e., lists in which semantic relations were made conspicuous by presenting blocks of same-category exemplars). A manipulation was imposed that should moderately prolong constructive processing of category concepts during free recall (instructing participants to recall unpresented category exemplars as well as presented ones). Relative to a control condition involving standard free recall, instructions to recall both presented and unpresented exemplars increased *R*, but it did not affect *D*. A further manipulation was imposed that should more strongly affect constructive processing of category concepts

[4] The theoretical characterization of reconstruction must not be misinterpreted as predicting that all manipulations that affect the difficulty of processing targets' semantic content or that orient retrieval toward relations between targets will dissociate reconstruction from direct access. Similarly, the theoretical characterization of direct access must not be misinterpreted as predicting that all manipulations that affect the difficulty of processing targets' surface content or that orient retrieval toward individual targets will dissociate direct access from reconstruction. Tests of stronger, dissociative predictions demand that experimental conditions be carefully selected to avoid process contamination (e.g., Murphy & Shapiro, 1994).

during free recall (providing participants with labels for all the categories during recall tests). Relative to the same control condition, the availability of category labels during retrieval increased *R* more than the first manipulation but, again, *D* was not affected. In Experiment 3, it was anticipated that cue–target pairs containing very abstract targets would encourage semantic processing and discourage surface processing during the study phase, relative to pairs containing very concrete targets. Consistent with that expectation, pairs containing abstract targets produced higher levels of reconstruction (and lower levels of direct access) than pairs containing concrete targets. In Experiment 4, it was anticipated that free recall would produce higher levels of reconstruction and lower levels of direct access than associative recall. This pattern was observed.

Although we were chiefly interested in results for direct access and reconstruction, instructive findings were obtained for metacognitive judgment. In the dual-retrieval model, *J* is the probability that participants are sufficiently confident that a reconstructed item appeared on the study list to be willing to pass it on for output. If this interpretation is correct, *J* should increase as a result of manipulations that elevate such confidence, *independent of whether an item was actually studied*. An obvious manipulation of this sort was present in Experiment 1—providing category labels on recall tests. If participants know, as they did in that experiment, that all targets are exemplars of a few familiar categories and if participants are supplied with the names of those categories on recall tests, they possess objective grounds for having high confidence in reconstructed targets that match the names: A constructed item can only be a target if it is an exemplar of one of the specified categories, and all targets will necessarily be exemplars of one of the categories. Consistent with this reasoning, providing category labels during retrieval increased *J* by over 60%, relative to two conditions in which category labels were not provided. In Experiment 3, the fact that perceptual experiences are concomitants of word presentations on study lists suggests that concrete targets will produce higher values of *J* than abstract targets. Here, our reasoning was that because highly concrete words induce vivid perceptual phenomenologies when they are reconstructed, participants will be likely to interpret such experiences as evidence that such words were present on study lists. Consistent with this notion, the probability that a reconstructed concrete target would be authorized for output was roughly twice the probability that a reconstructed abstract target would be authorized for output.

Parameter Dissociations

Patterns of dissociation between *D*, *R*, and *J* accumulated across experiments that, taken together, support the notion that direct access, reconstruction, and metamemorial judgment are distinct processes. With respect to whether the two retrieval operations are distinct from each other, a total of five single dissociations and three double dissociations were obtained. Concerning single dissociations, both of the semantic-processing manipulations in Experiment 1 affected *R* without affecting *D*, both of the surface-processing manipulations in Experiment 2 affected *D* without affecting *R*, and the target concreteness manipulation in Experiment 3 affected *R* but not *D* when cues were concrete. Concerning double dissociations, the target concreteness manipulation in Experiment 3 drove the proportions of directly accessed targets and reconstructed targets in opposite directions when cues were abstract, and in Experiment 4, the proportions of directly accessed targets and reconstructed targets were driven in opposite directions when the associative-recall condition was compared to either free-recall condition.

With respect to whether metamemorial judgment is distinct from the two retrieval processes, because metamemorial judgment operates on constructed items, a natural working hypothesis would be that it is a slave operation that depends on the same factors that influence the ease of reconstruction. Under the slave hypothesis, *J* could be experimentally dissociated from *D* but not from *R*. This is not the pattern that fell out, however. On the one hand, there was evidence that metamemorial judgment can be singly and doubly dissociated from direct access: Both surface-processing manipulations in Experiment 2 affected *D* without affecting *J*, and in Experiment 4, associative recall decreased *J* but increased the proportion of directly accessed targets (relative to either free-recall condition). On the other hand, there also was evidence that metamemorial judgment can be singly and doubly dissociated from reconstruction: One of the semantic-processing manipulations in Experiment 1 (specifically, instructions to recall unpresented exemplars as well as targets) increased *R* without affecting *J*, the target concreteness manipulation in Experiment 3 drove the proportion of reconstructed targets and the value of *J* in *opposite* directions when cues were concrete, and the target concreteness manipulation produced the same double dissociation when cues were abstract.

Because parameter dissociations were prominent results in our experiments, some brief comments are in order regarding their interpretation as evidence of dissociations between the memory processes that they measure. It has long been customary to interpret parameter dissociations in this manner (see Brainerd, Howe, & Desrochers, 1982), but it has recently been noted that stochastic dependencies between parameters can contribute to certain types of dissociations—specifically, to single dissociations involving conditional as opposed to unconditional parameters (e.g., Curran & Hintzman, 1995; Jacoby & Shrout, 1997). Further, even a double dissociation is not logically incompatible with a one-process hypothesis because such a result can occur if parameters measure opposite aspects of a single process (Dunn & Kirsner, 1988).

This is not the venue for a full mathematical treatment of this problem. Nevertheless, the problem can be addressed with Dunn and Kirsner's (1988) reversed association test. This test can rule out one-process explanations of double dissociations for manipulations that can be broken down into more than two levels (which creates more than two possible pairings of the values of dissociated parameters). Under a one-process hypothesis, monotonic-decreasing curves must result when values of one parameter are plotted against values of the other. A one-process hypothesis can be rejected if nonmonotonic (reversed association) curves are obtained. The latter pattern was present in Experiment 3, where four pairings of direct access with reconstruction and four pairings of reconstruction with metamemorial judgment were possible. Both the plots of direct access against reconstruction and the plots of metamemorial judgment against reconstruction were nonmonotonic rather than monotonic-decreasing.

Comparison to Other Dual-Retrieval Paradigms

Two paradigms are in current use that allow investigators to quantify the contributions of dual-retrieval operations to recognition, process dissociation (e.g., Jacoby, 1991), and conjoint recognition (e.g., Brainerd et al., 2001). In each instance, the experimental procedure involves responding to recognition probes under differing instructions as to which to accept and which to reject. With process dissociation, there are two methods of target presentation (e.g., List 1 versus List 2) and two instructional conditions, inclusion (accept all targets) and exclusion (accept only a subset of targets, such as ones that were presented in List 1 rather than List 2). With conjoint recognition, there is one method of target presentation and three instructional conditions: verbatim (accept only targets), gist (accept only distractors that are semantically related to targets), and verbatim + gist (accept both targets and distractors that are semantically related to targets).

Because both of these models measure retrieval operations that figure in recognition, it might seem inappropriate to use them to measure dual-retrieval processes in recall (for relevant arguments, see Bodner et al., 2000). As noted, however, Jacoby and associates (Jacoby, 1998; Jacoby et al., 1993) have estimated the process-dissociation model's parameters using a procedure that resembles associative recall: Participants respond to word stems that can be completed with either of two words, one of which appeared on the study list, under two sets of instructions (complete stems with studied words or the first word that comes to mind versus complete stems with unstudied words only).

Similarly, it would be very easy to extend the present model to recognition tasks because the RTTT paradigm is just as applicable to recognition as to free and associative recall. Suppose that participants study a list and then respond to three recognition tests on which probes can be either targets or distractors. As before, there are eight distinct performance patterns for targets, although correct responses and errors now refer to accept–reject decisions about test probes. The respective expressions for these performance patterns for *targets* are presented in the Appendix. In these expressions, the parameters *D*, *R*, and *J* have the same meanings as in the corresponding expressions for recall. Thus, these expressions implement a direct access–reconstruction account of recognition, not a familiarity–recollection account.

In sum, the current situation is that (a) different dual-retrieval interpretations have been proposed for recognition and free recall, (b) the specific dual-retrieval operations that have been posited for recognition (recollection and familiarity) have also been measured for a type of cued recall (stem completion) using the process-dissociation procedure (Jacoby, 1998; Jacoby et al., 1993), (c) the specific dual-retrieval operations that have been posited for free recall (direct access and reconstruction) have been extended to another type of cued recall (associative recall), and (d) the latter operations can easily be measured for recognition by implementing the RTTT paradigm with recognition tasks. In factorial design terminology, the recollection–familiarity distinction and the direct access–reconstruction distinction are fully crossed with recognition and recall.

This situation creates an exciting opportunity to move in the direction of unification of dual-process theories via experimentation in which the respective paradigms that have been devised for these theories are applied in tandem to recognition and recall. A question that should occupy center stage is whether it is necessary to continue with separate dual-process accounts of recognition and recall, or whether, instead, it is possible to find convincing support for a unified theory. The type of design that could make progress toward an answer is one in which participants respond to both recognition and recall tests (a) under conditions that allow the direct access–reconstruction model to be applied to both types of data and (b) under conditions that allow the recollection–familiarity models to be applied to both types of data. Such a design would yield separate estimates of direct access, reconstruction, recollection, and familiarity for both recall and recognition. With separate estimates available for both recognition and recall, investigators could proceed to test hypotheses about relations between them by manipulating variables that embody different assumptions about these retrieval processes.

APPENDIX

Dual-Retrieval Model with Variability of Direct Access and Reconstruction over Recall Tests

Let D_1, D_2, and D_3 be the probabilities of learning how to directly access targets on Trials 1, 2, and 3, respectively, and let R_1, R_2, and R_3 be the probabilities of learning how to reconstruct targets on Trials, 1, 2, and 3, respectively, and let J be the probability that a reconstructed target is authorized for output. When the assumption of the basic model (Eqs. (7–14)) that $D_2 = D_3 = R_2 = R_3 = 0$ is relaxed, so that retrieval learning may occur on any trial, the new dual-retrieval model is

$$p(\text{ccc}) = D_1 + (1-D_1)R_1J[D_2 + (1-D_2)J\{D_3 + (1-D_3)J\}], \quad \text{(A1)}$$

$$p(\text{ccn}) = (1-D_1)R_1J^2(1-D_2)(1-D_3)(1-J), \quad \text{(A2)}$$

$$p(\text{cnc}) = (1-D_1)R_1J(1-D_2)(1-J)[D_3 + (1-D_3)J], \quad \text{(A3)}$$

$$p(\text{ncc}) = (1-D_1)R_1(1-J)[D_2 + (1-D_2)J\{D_3 + (1-D_3)J\}] + (1-D_1)(1-R_1)[D_2 + (1-D_2)R_2J\{D_3 + (1-D_3)J\}], \quad \text{(A4)}$$

$$p(\text{cnn}) = (1-D_1)R_1J(1-J)^2(1-D_2)(1-D_3), \quad \text{(A5)}$$

$$p(\text{ncn}) = (1-D_1)R_1(1-J)(1-D_2)J(1-D_3)(1-J) + (1-D_1)(1-R_1)(1-D_2)R_2J(1-D_3)(1-J), \quad \text{(A6)}$$

$$p(\text{nnc}) = (1-D_1)(1-R_1)(1-D_2)(1-R_2)[D_3 + (1-D_3R_3J + (1-D_1)(1-R_1)(1-D_2)R_2(1-J)[D_3 + (1-D_3)J] + (1-D_1)R_1 \times (1-J)(1-D_2)(1-J)[D_3 + (1-D_3)J], \quad \text{(A7)}$$

and

$$p(\text{nnn}) = (1-D_1)(1-R_1)(1-D_2)(1-R_2)(1-D_3)(1-R_3) + (1-D_1)(1-R_1)(1-D_2)(1-R_2)(1-D_3)R_3(1-J) + (1-D_1)(1-R_1)(1-D_2)R_2(1-J)^2(1-D_3) + (1-D_1)(1-D_2)(1-D_3)R_1(1-J)^3. \quad \text{(A8)}$$

Extension of the Dual-Retrieval Model to Recognition

Let *D*, *R*, and *J* have the same meanings as in Eqs. (7–14), and let β be the usual bias parameter for recognition.

(Estimates of β are obtained in the usual way from false-alarm rates for unrelated distractors (e.g., Buchner, Erdfelder, & Vaterrodt-Plunecke, 1995)). The parallel dual-retrieval model for recognition tasks is

$$p(\text{ccc}) = D + (1-D)RJ^3 + (1-D+-)RJ^2(1-J)\beta + (1-D)RJ(1-J)^2\beta^2 + (1-D)R(1-J)^3\beta^3 + (1-D)(1-R)\beta^3, \quad (A25)$$

$$p(\text{ccn}) = (1-D)RJ^2(1-J)(1-\beta) + 2(1-D)RJ(1-J)^2\beta(1-\beta) + (1-D)R(1-J)^3\beta^2(1-\beta) + (1-D)(1-R)\beta^2(1-\beta), \quad (A26)$$

$$p(\text{cnc}) = (1-D)RJ^2(1-J)(1-\beta) + 2(1-D)RJ(1-J)^2\beta(1-\beta) + (1-D)R(1-J)^3\beta^2(1-\beta) + (1-D)(1-R)\beta^2(1-\beta), \quad (A27)$$

$$p(\text{ncc}) = (1-D)RJ^2(1-J)(1-\beta) + 2(1-D)RJ(1-J)^2\beta(1-\beta) + (1-D)R(1-J)^3\beta^2(1-\beta) + (1-D)(1-R)\beta^2(1-\beta), \quad (A28)$$

$$p(\text{cnn}) = (1-D)RJ(1-J)^2(1-\beta)^2 + (1-D)R(1-J)^3\beta(1-\beta)^2 + (1-D)(1-R)\beta(1-\beta)^2, \quad (A29)$$

$$p(\text{ncn}) = (1-D)RJ(1-J)^2(1-\beta)^2 + (1-D)R(1-J)^3\beta(1-\beta)^2 + (1-D)(1-R)\beta(1-\beta)^2, \quad (A30)$$

$$p(\text{nnc}) = (1-D)RJ(1-J)^2(1-\beta)^2 + (1-D)R(1-J)^3\beta(1-\beta)^2 + (1-D)(1-R)\beta(1-\beta)^2, \quad (A31)$$

and

$$p(\text{nnn}) = (1-D)R(1-J)^3(1-\beta)^3 + (1-D)(1-R)(1-\beta)^3. \quad (A32)$$

REFERENCES

Ackerman, B. P. (1985). Children's retrieval deficit. In C. J. Brainerd and M. Pressley (Eds.), *Basic processes in memory development* (pp. 1–46). New York: Springer-Verlag.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review*, **6,** 57–86.

Battig, W. F. (1965). Further evidence that strongest free-recall items are not recalled first. *Psychological Reports*, **17,** 745–746.

Battig, W. F., Allen, M., & Jensen, A. R. (1965). Priority of free recall of newly learned items. *Journal of Verbal Learning and Verbal Behavior*, **4,** 175–179.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, **80** (3, Pt. 2).

Bodner, G. E., Masson, M. E. J., & Caldwell, J. I. (2000). Evidence for a generate–recognize model of episodic influences on word-stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26,** 267–293.

Brainerd, C. J., Howe, M. L., & Desrochers, A. (1982). The general theory of two-stage learning: A mathematical review with illustrations from memory development. *Psychological Bulletin*, **91,** 634–665.

Brainerd, C. J., Howe, M. L., & Kingma, J. (1982). An identifiable model of two-stage learning. *Journal of Mathematical Psychology*, **26,** 263–293.

Brainerd, C. J., Kingma, J., & Howe, M. L. (1985). On the development of forgetting. *Child Development*, **56,** 1103–1119.

Brainerd, C. J., & Reyna, V. F. (in press). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior*.

Brainerd, C. J., Reyna, V. F., Harnishfeger, K. K., & Howe, M. L. (1993). Is retrievability grouping good for recall? *Journal of Experimental Psychology: General*, **122,** 249–268.

Brainerd, C. J., Reyna, V. F., & Howe, M. L. (1990). Children's cognitive triage: Optimal retrieval or effortful processing? *Journal of Experimental Child Psychology*, **49,** 428–447.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kevershan, J. (1991). Fuzzy-trace theory and cognitive triage in memory development. *Developmental Psychology*, **27,** 351–369.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development*, **53,** 2–3 (whole no. 222).

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, **106,** 160–179.

Brainerd, C. J., Stein, L. M., & Reyna, V. F. (1998). On the development of conscious and unconscious memory. *Developmental Psychology*, **34,** 342–357.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (in press). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **27,** 307–327.

Buchanan, L., Brown, N., & Westbury, C. (Nov.1999). *Generating false memories: Conscious or nonconscious spread of activation during study?* Paper presented at the Psychonomic Society meeting, Los Angeles, CA, Nov.

Buchner, A., Erdfelder, E., & Vaterrodt-Plunecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes. *Journal of Experimental Psychology: General*, **124,** 137–160.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3,** 37–60.

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21,** 531–547.

Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology*, **58,** 17–22.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental process: The principle of reversed association. *Psychological Review*, **95,** 91–101.

Erdelyi, M. H., Finks, J., & Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General*, **118,** 245–254.

Estes, W. K. (1960). Learning theory and the new “mental chemistry.” *Psychological Review*, **67,** 207–233.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91,** 1–67.

Greeno, J. G., James, C. T., Da Polito, F. J., & Polson, P. (1971). A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior*, **10,** 331–345.

Halff, H. M. (1977). The role of opportunities for recall in learning to retrieve. *American Journal of Psychology*, **90,** 383–406.

Horton, D. L., Pavlick, T. J., & Moulin-Julian, M. W. (1993). Retrieval-based and familiarity-based recognition and the quality of information in episodic memory. *Journal of Memory and Language*, **32,** 39–55.

Howe, M. L., Brainerd, C. J., & Kingma, J. (1985). Storage-retrieval processes of learning-disabled children: A stages-of-learning analysis of picture-word effects. *Child Development*, **56,** 1120–1133.

Hu, X. (1998). General processing tree (Version 1.0) [Computer software]. Memphis, TN: University of Memphis.

Humphreys, M. S., Dennis, S., Chalmers, K. A., & Finnigan, S. (in press). Dual processes in recognition: Does a focus on measurement operations provide a sufficient foundation. *Psychonomic Bulletin & Review*, **7,** 593–603.

Humphreys, M. S., & Yuille, J. C. (1973). Errors as a function of noun concreteness. *Canadian Journal of Psychology*, **27,** 83–94.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, **32,** 421–445.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, **30,** 513–541.

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user’s guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 3–26.

Jacoby, L. L., & Shrout, P. E. (1997). Toward a psychometric analysis of violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23,** 505–510.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, **122,** 139–154.

Lampinen, J. M., Neuschatz, & Payne, D. G. (1998). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology: Development, Learning, Personality, Social*, **16,** 181–224.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87,** 252–271.

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

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

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, **22,** 85–94.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and Why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20,** 1196–1204.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, **76,** (No. 1, Pt. 2).

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

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

Payne, D. G., & Elie, C. J. (1997). Comparing false memories for pictures and words. Paper presented at the Psychonomic Society meeting, Philadelphia, PA, Nov.

Payne, D. G., & Elie, C. J. (1998). Repeated list presentation reduces false memories for pictures and words. Paper presented at Psychonomic Society, Dallas, TX, Nov.

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

Powell, M. B., Roberts, K. P., Ceci, S. J., & Hembrooke, H. (1999). The effects of repeated experience on children’s suggestibility. *Developmental Psychology*, **35,** 1462–1477.

Reyna, V. F. (1992). Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 103–127). New York: Springer-Verlag.

Reyna, V. F. (1996). Conceptions of memory development, with implications for reasoning and decision making. *Annals of Child Development*, **12,** 87–118.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, **7,** 1–75.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, **30,** 178–191.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, **95,** 318–339.

Robinson, K. J., & Roediger, H. L., III, (1997). Associative processes in false recall and false recognition. *Psychological Science*, **8,** 231–237.

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

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

Rotello, C. M. (2001). Recall processes in recognition memory. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 40, pp. 183–221). San Diego: Academic Press.

Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language*, **40,** 432–453.

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, **9,** 20–26.

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

Stenberg, G., Lingren, M., Johansson, M., Olsson, A., & Rosen, I. (2000). Semantic processing without conscious identification: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26,** 973–1004.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Toglia, M. P., & Neuschatz, J. S. (1996). False memories: Where does encoding opportunity fit into the equation? Paper presented at the Psychonomic Society meeting, Chicago, IL, Nov.

Toglia, M. P., & Neuschatz, J. S. (1997). False memories: Are pictures really worth a thousand words? Paper presented at the Psychonomic Society meeting, Philadelphia, PA, Nov.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, **7,** 233–256.

Underwood, B. J. (1953). Studies of distributed practice: IX. Learning and retention of paired associates as a function of intralist similarity. *Journal of Experimental Psychology*, **45,** 143–149.

Wallace, W. P., Stewart, M. T., Shaffer, T. R., & Barry, C. R. (1998). Are false recognitions influenced by prerecognition processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 284–298.

(Received October 5, 2000)
(Revision received January 18, 2001)
(Published online September 17, 2001)