Published in final edited form as:

J Mem Lang. 2010 October ; 63(3): 425–445. doi:10.1016/j.jml.2010.05.002.

Recollective and Nonrecollective Recall

C. J. Brainerd and **V. F. Reyna**

Department of Human Development, Cornell University

Abstract

The study of recollective and nonrecollective retrieval has become controversial, owing to several critiques of traditional recognition-based measurement (e.g., remember/know, ROC, process dissociation). We present a new methodology in which subjects merely study and recall lists, using any standard paradigm (associative, cued, free, or serial recall), the data are analyzed with a Markov model whose parameters measure recollective and nonrecollective retrieval, and the model's fit is compared to that of one-process models. The power of this approach is illustrated in some experiments that dealt with two classic questions: (a) What are the process-level differences between associative and free recall, and (b) why does taxonomic organization improve free recall but impair associative recall? Fit results showed that a dual-retrieval model is both necessary and sufficient to account for associative and free recall data, in contrast to the sufficient-but-not-necessary pattern that prevails in the recognition literature. Key substantive findings were that associative recall is more reliant on recollective retrieval and less reliant on nonrecollective retrieval than free recall, that taxonomic organization impairs recollective retrieval in both paradigms, and that taxonomic organization enhances the reconstruction component of nonrecollective retrieval in free recall.

Keywords

recollection; reconstruction; familiarity; taxonomic organization; Markov models

The question of how to separate and quantify the contributions of recollective and nonrecollective retrieval to memory performance has become a matter of contention. It was Mandler (1980) who, in a seminal paper, stimulated widespread interest in distinguishing these two forms of retrieval, although similar ideas can be found in an early article by Strong (1913), and dual-process distinctions between recognition and recall appeared in an early article by Muller (1913). Both Strong and Mandler discussed the distinction between recollective and nonrecollective retrieval in connection with old/new item recognition and, hence, that is how it has customarily been studied. The core methodological tactic is to enrich item recognition with further tasks that are intended to disentangle its recollective and nonrecollective components. Remember/know judgments (Tulving, 1985) are far and away the most extensively used procedure. In addition, confidence judgments, which produce the receiver operating characteristic (ROC), have often been used (e.g., Yonelinas, 1994), as

Corresponding author: C. J. Brainerd, B-43 MVR Hall, Cornell University, Ithaca, NY 14853, Telephone: 607-254-1163, FAX: 607-255-9856, cb299@cornell.edu.

Correspondence concerning this article should be addressed to C. J. Brainerd, Department of Human Development, Cornell University, Ithaca, NY 14853. cb299@cornell.edu

have inclusion versus exclusion instructions, which yield the recollection and familiarity parameters of Jacoby's (1991) process-dissociation model.

Such methodologies have produced an extensive literature on how recollective and nonrecollective retrieval (usually called familiarity) are ostensibly affected by experimental manipulations that embody their respective process definitions, by neurological impairments, and by activity in specific brain regions (for a review, see Yonelinas, 2002). However, much of the theoretical ground has been cut from under such work by a recent series of critiques of recognition-based methodologies. Although different critiques have focused on different methodologies, they converge on the following conclusion: Dual-retrieval processes are unnecessary to account for the fine-grained structure of item recognition, and instead, acceptable accounts are supplied by one-process models in the signal detection tradition, such as the unequal variance model (UVSD) of Murdock and Lockhart (1970) and, more recently, of Glanzer, Kim, Hilford, and Adams (1999). In such models, targets are random variables that are sampled from a single memory strength distribution whose mean and variance are μ_T and $\sigma^2{}_T$, and distractors are random variables that are sampled from another strength distribution whose mean and variance are $\mu_D \leq \mu_T$ and $\sigma^2{}_D \leq \sigma^2{}_T$.

Turning to objections to specific techniques for separating recollective from nonrecollective retrieval, no less than four major critiques of the most frequently used methodology, remember/know, have appeared (Donaldson, 1996; Dunn, 2004, 2008; Rotello, Macmillan, Reeder, 2004). Likewise, major critiques of the ROC and process dissociation techniques have been published by Wixted (2007) and by Ratcliff, Van Zandt, and McKoon (1995), respectively. Most recently, Malmberg (2008) surveyed findings generated by all of these procedures and concluded that there is overwhelming evidence that item recognition is governed by a single strength process, and hence, these procedures are unreliable methods of separating recollective from nonrecollective retrieval. Crucially, Malmberg and the authors of some of the other critiques stress that the recollective-nonrecollective distinction as an important conception that ought to be a continuing focus of research, which leaves us with the task of finding new methods of separating and quantifying these retrieval operations. Two approaches have been proposed, one of which was implemented in the experiments that we report below.

The first was discussed by Malmberg (2008), who proposed shifting the focus of research from item recognition to associative recognition and plurality discrimination on the ground that in the latter paradigms, subjects activate a recollection rejection process to weed out distractors that resemble studied targets (e.g., *bagpipe-cobalt*, when *bagpipe-turtle* was studied, or *frogs*, when *frog* was studied).1 Malmberg showed that this solution has much to recommend it, but he also noted that one-process models have been formulated and evaluated for both associative recognition (e.g., Wixted, 2007) and plurality discrimination (e.g., McClelland & Chappell, 1998; Park, Reder, & Dickison, 2005). Although Malmberg pointed out that some of these models make unrealistic assumptions about levels of overlap between targets and distractors, their existence means that the controversy over recognition-based measurement is unlikely to abate by switching to associative recognition and plurality discrimination.

[1]This process is commonly called recall-to-reject. However, we have pointed out elsewhere that recall-to-reject is a misleading label, whereas recollection rejection captures the process exactly (Brainerd, Reyna, Wright, & Mojardin, 2003). Remember that the underlying idea is that target recollective phenomenology is used to reject related distractors (e.g., "No, I didn't study *bagpipe-turtle* because *bagpipe-cobalt* is flashing in my mind's eye"). It follows that recall-to-reject would be an appropriate label only if recall were always recollective. As we discuss in this paper, however, recall is sometimes recollective and sometimes nonrecollective (see also, Barnhardt, Choi, Gerkens, & Smith, 2006).

The other approach, which was proposed by the present writers (Brainerd, Reyna, & Howe, 2009), is to shift from recognition to recall. The advantage of this solution is that there is a long tradition of memory research in which learning to recall the items on a list has been found to consist of two distinct processes (e.g., Bower & Theois, 1964; Kintsch & Morris, 1965; Greeno, 1968; Waugh & Smith, 1962). Using recall data, then, it is possible to achieve separation by identifying recollective retrieval with one of them and nonrecollective retrieval with the other. We show how that is done in the next section, and then, to illustrate the power of this technique, we use recall-based separation to answer two questions: (a) What are the relative contributions of recollective and nonrecollective retrieval to associative versus free recall (Experiment 1), and (b) why does semantic organization have opposite effects on associative versus free recall (Experiments 2 and 3)?

Dual-Retrieval Processes in Recall

As mentioned, there is a long research tradition whose results support the hypothesis that learning to recall the items on a list involves two distinct processes. The contemporary memory literature exhibits little awareness of this fact, perhaps because the pertinent research stretches back a half-century to the heyday of the all-or-none learning controversy and the origins of Markov modeling. Before that era, the dominant conception, which was inspired by theories of conditioning, had been that learning to recall items is governed by a single graded strength process (see Estes, 1960). The research that challenged that one-process conception consists of a substantial literature in which two-stage Markov models were successfully fit to the data of standard recall paradigms and one-stage models were rejected (for reviews, see Brainerd, Howe, & Desrochers, 1982; Greeno, 1974; Greeno, James, Dapolito, & Polson, 1978). Three findings from that literature are especially relevant to the use of recall data to measure dual-retrieval processes.

The first is that the hypothesis that a single graded strength process underlies recall was rejected with converging data from several paradigms—such as the replacement design (Rock, 1957), the miniature-experiment design (Estes, 1960), and the selective reminding design (Halff, 1977). The replacement design, which involves associative recall, provides a particularly incisive demonstration. Subjects participate in a standard series of study-test trials, except for one modification. At the end of each test cycle, items that were not recalled are removed from the study list and replaced by new word pairs. The list for the next study cycle consists of previously-studied pairs (which were recalled on the prior test) and pairs that have not been previously studied. If recall involves a graded strength process, then, naturally, it must be more accurate over trials for lists in which all pairs preserved than for lists in which new pairs are added after each test cycle. However, accuracy levels were found to be virtually identical for the two types of lists. Another compelling demonstration comes from selective reminding experiments. Subjects participate in a standard series of free recall trials, except that after Trial 1, the study list is restricted (a) to items that were recalled on the previous test cycle for some subjects or (b) to items that were not recalled on the previous test cycle for other subjects. When the list is restricted to previously recalled items, subsequent study cycles do not improve recall, which again violates the predictions of a graded strength process.

The second result that bears on present concerns is that improvements in recall over trials were found to involve two discrete changes, rather than one. Here, there are two pertinent groups of experiments. In the first, models were fit to associative and free recall data that posit a *single* discrete change (one-stage Markov models; see Greeno & Steiner, 1964) and that posit two discrete changes (two-stage Markov models; see Greeno, 1968). Comparative fit analyses consistently favored the latter (for a review, see Brainerd et al., 1982). In the second group of experiments, the question of whether the two discrete changes were due to a single underlying process (i.e., a double-threshold process) or two distinct processes was examined. If distinct processes are involved, it should be possible to dissociate the two discrete changes with selected experimental manipulations. Such dissociations were reported in several articles. For instance, dissociations were produced by manipulations such as picture versus word lists, concrete versus abstract noun lists, high versus low intralist similarity, immediate versus delayed recall tests, and different types of transfer tests (Brainerd, Desrochers, & Howe, 1981; Humphreys & Yuille, 1970; Greeno, James, and DaPolito, 1971; Pagel, 1973).

The third and most recent finding consists of evidence that the two recall processes can be interpreted as recollective and nonrecollective retrieval. Because discussion of that evidence requires defining the processes and formulating them in a measurement model, that discussion is postponed until those matters are attended to. Before doing that, we return to an issue that was mentioned above—namely, that although there are extensive data running counter to the hypothesis that learning to recall involves a single strength process, there is limited awareness of this fact in contemporary memory research. This is well illustrated by some influential recall models that are predicated on one-process retrieval assumptions, with Howard and Kahana's (2002) model of free recall, Brown, Neath, and Chater's (2008) model of free and serial recall, and Klein, Addis, Kahana's (2005) model of serial recall being cases in point. These models account for many findings in their respective domains, such as how recall is affected by presentation order on study trials. Our point about them is not that they should be dismissed because one-process assumptions are unrealistic for recall. As Gronlund and Ratcliff (1989) noted in connection with global memory models, the fact that a model can explain many effects in its target domain counts in its favor and argues for its continue use, notwithstanding that focused tests contradict some of its core assumptions. Rather, our point about contemporary one-process models of recall is that their sheer existence is apt to be interpreted as somehow challenging the use of recall data to measure dual-retrieval processes. That would be a misinterpretation: The aforementioned models are not grounded in the classical work on one- versus two-process conceptions of recall that we have summarized, they do not implement statistical tests of the comparative fit of one- versus two-process models, and hence, they do not bear on the use of recall data to measure dual-retrieval processes. It may seem puzzling that those models account for several important effects if such a fundamental assumption as whether recall involves a single retrieval process is unrealistic, but actually, this exemplifies a familiar lesson about memory models. That lesson is that models whose core assumptions proved to be nonviable have sometimes achieved high levels of success in accounting for empirical effects, simply because the assumptions did not carry much weight in those effects. Indeed, which empirical predictions deliver tests of core assumptions may not be obvious, and close analysis of a model's mathematical machinery may be required to isolate such predictions (e.g., Gronlund & Ratcliff, 1989; Hintzman & Curran, 1994).

Dual-Recall Theory

Now, we turn to the tasks of defining dual-retrieval processes for recall and of formulating a measurement model that quantifies those processes, both of which are predicated on an abstract characterization of the structure of recall data. On a recall test, list items must be emitted by subjects, rather than simply being classified as old or new. This means that the probability of successful recall before the first study cycle is effectively zero. By design, then, the items in a recall experiment usually start in a no-recall state U, in which the correct response probability is zero. Once study cycles have begun, learning to recall seems to consist of two discrete changes, as just noted: Items may enter a partial-recall state P from U, in which the average probability of successful recall is $0 < p < 1$, or they may enter a perfect-recall state L from U, in which the probability of successful recall is 1. (For items that enter state *P*, the second stage of learning [entry into state *L*] can be completed on a later trial.) Brainerd et al. (2009) proposed dual-retrieval interpretations of these two stages, formulated a mathematical model that delivers multiple measures of recollective and nonrecollective retrieval, and applied that model to a large corpus of recall data to validate the dual-retrieval interpretation of the two stages. We summarize key points in connection with each of these three topics before describing our experiments.

Brainerd et al. (2009) proposed that nonrecollective retrieval results from the first type of learning (entry into state *P*) and recollective retrieval results from the second (entry into state *L*). The two forms of retrieval are assumed to involve different types of episodic traces, following contemporary dual-trace conceptions of representation, such as fuzzy-trace theory (FTT; Brainerd & Reyna, 2005; Reyna & Brainerd, 1995), the processing implicit and explicit representations (PIER) model of Nelson and associates (Nelson, McGivney, Gee, & Janczura, 1998; Nelson, Schreiber, & McEvoy, 1992), and the retrieving efficiently from memory (REM) model of Shiffrin and associates (e.g., Shiffrin, 2003; Shiffrin & Steyvers, 1997). In such conceptions, subjects store episodic traces of individual list items and episodic traces of some of their features (especially semantic features). The former are variously termed verbatim traces (FTT), explicit traces (PIER), and episodic traces (REM), while the later are variously termed gist traces (FTT), implicit traces (PIER), and lexical/ semantic traces (REM). Traces of the former sort fully identity *individual* list items, whereas traces of the latter sort partially identify list items and delimit *sets* of potential list items (Brainerd et al., 2009).

In dual-recall theory, recollective retrieval is defined as direct access of verbatim traces of prior presentations of individual items from the study list. The term direct access is used in the sense in which it was used by Kintsch (1970) and Clark and Gronlund (1996), as an operation that retrieves an item’s trace without first activating, searching through, and comparing traces of other items. Direct access supports the errorless performance of state *L* because verbatim traces symbolically reinstate items’ surface forms, which allows them to simply be read out of consciousness as they echo in the mind’s ear or flash in the mind’s eye. The latter is the sort of vivid, realistic phenomenology that is the hallmark of traditional conceptions of recollection, and hence, direct access is the recollective form of recall.

The nonrecollective form of retrieval, on the other hand, is defined as a reconstruction operation plus a slave familiarity judgment operation. Reconstructive retrieval is a classic delimited search operation (see Crowder, 1976) that uses episodic traces of some of the targets’ features (e.g., “household pet”) to constrain the generation of candidate sets to ones that are restricted enough to be rapidly searched (e.g., *canary, cat, dog, parakeet*). Thus, reconstruction *regenerates* list items from the second type of episodic trace. Because such traces are not item-specific (e.g., “household pet” is not a specific animal, “Italian seasoning” is not a specific herb, “stringed instrument” is not a specific musical instrument), a mechanism is required to get from such traces to items that could actually be output. A delimited search accomplishes that by constructing of sets of candidate items that are small enough to be rapidly searched within the time constraints of a recall test (Brainerd et al., 2009). Thus, for a given list item, learning how to reconstruct it is chiefly a matter of learning to sample target features that delimit such search sets, with smaller sets producing better performance because they can be more fully processed within the time constraints of a recall test. A further consideration, which will become important later, is that, other things being equal (e.g., set size), features that delimit search sets that contain *more than one list item* ought to enhance reconstruction under certain types of recall conditions (e.g., “household pet” is an especially good feature if *canary, cat, dog,* and *parakeet* are all on the list).

Because reconstructive retrieval generates *sets* of candidate items, rather than directly accessing individual items, a further operation is needed to decide which members of a set to output, and this is where familiarity enters the picture. Brainerd et al. (2009) proposed that output is controlled by post-construction familiarity judgments; that is, constructed items generate familiarity signals. Constructions that generate strong familiarity signals are likely to be output, while those that generate weak signals are likely to be withheld. The nonrecollective form of retrieval supports the imperfect recall of state *P*, rather than the errorless recall of state *L*, because, of course, familiarity judgment may fail to output constructions that are in fact list items, and it may output constructions that are not list items (intrusions).

Quantifying Recollective and Nonrecollective Retrieval

Direct access, reconstruction, and familiarity judgment can all be measured by fitting a three-state absorbing Markov chain to the error-success data of recall experiments—where the model's states are *U*, *P*, and *L*, with *L* being an absorbing state. This is a two-process model because, as noted, there is a pair of learning events—namely, entering *P* and entering *L*—which are identified with nonrecollective and recollective retrieval, respectively. Mathematically, the model, which is shown in Equation 1, consists of a starting vector (*W*) and an interstate transition matrix (*M*). The parameters of *W* are the probabilities of an item being in each possible state on the first recall test, and the parameters of *M* are probabilities of being in each possible pair of states on consecutive recall tests. Thus, the parameters of *W* are unconditional probabilities, while the parameters of *M* are conditional probabilities.

$$W=[L(1),P_E(1),P_C(1),U(1)]=[D_1,(1-D_1)R_1(1-J_1),(1-D_1)R_1J_1,(1-D_1)(1-R_1)];$$

$$M=\begin{array}{lllll} & L(n+1) & P_E(n+1) & P_C(n+1) & U(n+1)\\ L(n) & 1 & 0 & 0 & 0\\ P_E(n) & D_{3E} & (1-D_{3E})(1-J_{3E}) & (1-D_{3E})J_{3E} & 0\\ P_C(n) & D_{3C} & (1-D_{3C})(1-J_{3C}) & (1-D_{3C})J_{3C} & 0\\ U(n) & D_2 & R_2(1-D_2)(1-J_2) & R_2(1-D_2)J_2 & (1-D_2)(1-R_2)\end{array}$$

(1)

Importantly, in light of the on-going dispute over whether recognition involves one retrieval process or two, this model supplies a simple statistical test of the null hypothesis that recall does not involve more than a single nonrecollective process (by assessing fit when state *U* is deleted) or more than a single recollective process (by assessing fit when state *P* is deleted), along with a simple statistical test of the null hypothesis that recall does not involve more than a pair of recollective and nonrecollective processes (by assessing fit when states *U* and *P* are present). (The statistical machinery for these tests is presented in the Appendix.) Brainerd et al. (2009) provided an identifiability proof that demonstrated that if this model is defined over recall designs of the form $S_1T_1T_2$, S_2T_3, S_3T_4, …, it yields four identifiable parameters that measure recollective retrieval, two identifiable parameters that measure the reconstruction component of nonrecollective retrieval, and four identifiable parameters that measure the familiarity component of nonrecollective retrieval. That is, the model delivers multiple measurements of all of these processes for any recall experiment, if the experiment has the following structure: Subjects respond to two separate recall tests following the first study cycle, and after that, trials consisting of a study cycle followed by a single recall test continue until performance is errorless.

The specific parameters that measure recollective and nonrecollective retrieval are displayed and defined in Table 1. Three points should be mentioned that will become relevant later. First, there are twice as many direct access parameters as reconstruction parameters. That is because an item can become directly accessible if it is in state *U* (when it is neither directly accessible nor reconstructable) or if it is in state *P* (when it is reconstructable but not directly accessible), whereas it can only become reconstructable if it is in state *U*. Second, there are measures of "early" and "late" occurrences of these processes (e.g., D_1 and R_1 measure recollective and nonrecollective retrieval, respectively, on Trial 1, whereas D_2 and R_2 measure these processes after Trial 1). Thus, one can determine how prior learning affects the difficulty of learning to retrieve recollectively versus nonrecollectively, and specific predictions about this follow from theoretical conceptions of direct access and reconstruction. Third, with respect to recollective retrieval, there are separate parameters that measure the difficulty of learning how to directly access an item before it becomes reconstructable (D_1 and D_2) and while it is reconstructable (D_{3C} and D_{3E}). (Hence, with respect to D_{3C} and D_{3E}, although retrieval is nonrecollective in state *P*, a process that will eventually make items recollectable is operating in that state.) Consequently, it is possible to determine whether nonrecollective retrieval aids or impairs recollective retrieval.

Evidence for Dual-Recall Theory

We return now to the topic that was postponed earlier--namely, findings that support the hypothesis that the two recall processes can be interpreted as recollective and nonrecollective retrieval, respectively. Evidence of that sort was provided by Brainerd et al. (2009), using a corpus of 207 data sets in which subjects of various ages (children, adolescents, young adults, elderly adults) learned to recall lists to stringent performance criteria (one or two errorless tests) under associative, cued, free, or serial recall conditions, using the $S_1T_1T_2$, S_2T_3, S_3T_4, … design. A series of initial fit analyses were conducted for these data sets, which revealed that: (a) The hypothesis that learning to recall involves only a single retrieval process was consistently rejected at high levels of confidence, (b) there was no statistical support for the hypothesis that learning to recall involves more than two retrieval processes, and (c) the two-process model in Equation 1 gave good accounts of the observed distributions of fine-grained statistics of these data sets (e.g., errors before first success, trial number of the last error, successes before the last error, total errors). Thus, unlike the dual-process recognition literature, recall could not be reduced to a one-process model (finding a), and there was consistent support for a dual-retrieval conception (findings b and c), which brings us to findings that support the recollective and nonrecollective interpretations of the two retrieval operations.

We mention six such findings, by way of illustration. The first concerns developmental changes between early and late adulthood. A standard assumption about recollective retrieval is that it exhibits more pronounced aging declines than nonrecollective retrieval (see Yonelinas, 2002), which leads to the prediction that the *D* parameters in Table 1 should decline more with age than the *R* or *J* parameters. That pattern was observed when parameter estimates were compared for several sets of younger versus older adult data. Second, another standard assumption about aging is that whereas nonrecollective retrieval is largely spared during healthy aging, it declines noticeably in older adults with certain neurocognitive impairments (e.g., Alzheimer's dementia or Parkinson's disease), which leads to the prediction that there should be large declines in *R* parameters for cognitively impaired subjects. That pattern was observed when parameter estimates were compared for healthy older adults versus older adults with Alzheimer's dementia. Third, list length is a standard experimental manipulation that has often been proposed (e.g., Yonelinas, 1994) as one that should selectively affect recollective retrieval, which leads to the prediction that longer lists ought to decrease estimates of the *D* parameters but not the *R* or *J* parameters.

That pattern was observed when parameter estimates were compared for data sets in which subjects learned to recall shorter versus longer lists under otherwise identical conditions. Fourth, repetition is another standard manipulation that has often been proposed (e.g., Jacoby, 1991) as one that should affect recollective retrieval more than nonrecollective retrieval, which leads to the prediction that D ought to increase more as study cycles accumulate than R or J parameters. That pattern was observed when changes in parameter estimates were compared over trials. The fifth result concerns how difficult it is for an item to become recollectable before versus after it has become reconstructable. It ought to be easier after than before because once an item can be reconstructed, this provides additional covert "presentations" of the item beyond its study cycle presentations. That leads to the prediction that if the D parameters measure recollective retrieval, estimates of the parameters in the D_1/D_2 pair should be smaller than estimates of the parameters in the D_{3C}/D_{3E} pair. That pattern was observed when estimates of the D_1/D_2 and D_{3C}/D_{3E} pairs were compared over data sets. The last result concerns how difficult it is for an item that is in state U to become reconstructable on the first trial of an experiment (parameter R_1) versus on later trials. Brainerd et al. (2009) noted that it ought to be easier on later trials because each trial provides additional opportunities for subjects to sample the types of features that best-support reconstruction (i.e., features such as "household pet," rather than "animal," which generate candidate sets that are small enough to be rapidly searched). That leads to the prediction that $R_1 < R_2$, a pattern that was observed when these parameters were compared over data sets.

Such results notwithstanding, Brainerd et al.'s (2009) evaluation of the dual-retrieval model has two salient limitations. The first is that the evaluation was retrospective. The data sets in their corpus had been previously reported, over a period of several years, so that the experiments that generated those data had originally been conducted for other purposes. Consequently, none incorporated manipulations that had been designed, prospectively, to test predictions that follow from the process definitions of recollective and nonrecollective retrieval. The other limitation is that all of the data sets were ones in which the most stringent possible performance criterion, errorless recall, was imposed. This has three disadvantages. First, with the young adult subjects of mainstream memory research, an errorless criterion constrains the lengths of the lists that can be administered, with lists longer than about 20 items requiring prohibitively large numbers of study-test cycles to reach criterion. Second, even with much shorter lists, an errorless criterion is impractical for use with children, or with adult populations whose recall is compromised, such as adults with neurocognitive impairments. Such subjects are typically unable to achieve errorless recall. Third and most worrisome of all, the errorless criterion may be responsible for the fact that one-process models of recall were consistently rejected. In other words, reliance on dual-retrieval processes may depend on having to meet a stringent performance criterion, which usually requires a protracted sequence of study-test trials. Here, it can reasonably be argued that subjects may rely entirely on the error-prone nonrecollective form of retrieval when they are not under criterion pressure, but they are forced to resort to recollective retrieval as well in order to meet a stringent criterion.

These limitations were removed in the experiments that we report below. The experiments were designed prospectively to test specific predictions that follow from the process definitions of recollective and nonrecollective retrieval. Also, the experiments implemented simplified designs in which (a) all subjects participated in a short sequence of study-test trials, and (b) no performance criterion was imposed.

Dual-Retrieval Processes in Associative and Free Recall

In the remainder of this article, we show how dual-recall methodology can be used to explain variations in the process demands of different recall paradigms and to explain why some manipulations have opposite effects on different types of recall. With respect to the first point, we consider potential variations in the contributions of recollective and nonrecollective retrieval to associative versus free recall. With respect to the second point, we consider how such variations can explain why a classic content manipulation, taxonomic organization of lists, improves free recall while simultaneously impairing associative recall (Schwenn, & Underwood, 1968).

Associative Versus Free Recall

In associative recall, subjects study cue-target pairs and attempt to remember the targets that map with specific cues when those cues are presented on test cycles. Thus, targets must be recalled at specific times in response to specific cues and not at other times in response to other cues, which are very recollective types of demands. Retrieval is far more open ended in free recall. Subjects study individual target items and merely attempt to remember as many of them as possible on test cycles, given only the generic cue to recall items from the just-studied list. Thus, targets can be recalled at any time (and may be recalled multiple times), in any order, and there is no forced mapping of targets with specific retrieval cues, all of which allows increased latitude for reconstruction to operate.

Dual-retrieval ideas about process-level differences between associative and free recall echo proposals about associative versus item recognition in the recognition literature (see Malmberg, 2008; Yonelinas, 2002). According to those ideas, associative recall is more recollectively slanted than free recall because its high baseline level of retrieval cue specificity promotes recollective retrieval, relative to free recall: If recollective retrieval involves direct access of verbatim traces of individual target items, such a process ought to be facilitated by the item-specific cues that associative recall supplies. In contrast, if nonrecollective recall involves reconstructing items from partial-identifying information (e.g., semantic features) that delimits sets of candidate items, such a process ought to be impeded by the availability of item-specific cues. These predictions were investigated in Experiment 1.

Taxonomic Organization

A hoary finding about free recall is that subjects remember more items when they study lists of words that are exemplars of a few familiar taxonomic categories than when they study lists of unrelated words that are matched on other difficulty factors (Mandler, 1967). This finding has, of course, stimulated a vast literature on free recall of categorized lists (e.g., Brown, Flores, & Goodman, 1991; Gollin & Sharps, 1988). It has also been the subject of numerous developmental studies (e.g., Bjorklund, 1987; Bjorklund & Jacobs, 1985; Bjorklund & Muir, 1988), which have been motivated by the fact that categorized lists do not improve free recall before age 10 or 11. This pattern has been used to track the emergence of the ability to connect semantic features across list items (Brainerd, Reyna, & Ceci, 2008).

A less well-known and infrequently studied finding about taxonomic organization is that in associative recall, subjects remember *fewer* items with categorized than with unrelated lists (Schwenn & Underwood, 1968). This result was first reported by Underwood, Ekstrand, and Keppel (1963), who compared associative recall of lists on which the cue members of pairs were unrelated to each other and target members were also unrelated to each other (UU) to recall of lists on which cue members were exemplars of a few taxonomic categories but target members were unrelated (CU) and to recall of lists on which target members were exemplars of a few taxonomic categories but cue members were unrelated (UC). They reported that recall was poorer for CU lists than for UU lists and that recall did not differ for UC versus UU lists. Subsequent research, in which there was full factorial manipulation of taxonomic organization (UU, CU, UC, and CC lists), revealed that categorization suppresses recall on both the cue and target sides of pairs (Brainerd, Kingma, & Howe, 1985).

Dual-retrieval theory explains the opposite effects of categorization on associative versus free recall via (a) differences in the paradigms' relative reliance on recollective and nonrecollective retrieval and (b) differences in categorization's effects on recollective versus nonrecollective retrieval. Concerning a, in the dual-retrieval account, associative recall is more recollective than free recall because its high baseline level of cue-target specificity ought to favor direct access of verbatim traces of individual targets(see Brainerd, Wright, Reyna, & Payne, 2002). Consequently, associative recall should be especially sensitive to manipulations that either increase or dilute cue-target specificity. Concerning b, categorizing either the cue or the target side of pairs reduces such specificity and, hence, ought to impair recollective retrieval. Also in the dual-recall account, free recall is more reconstructive than associative recall because it provides only generic retrieval cues that do not map with individual list items (Brainerd et al., 2002). Therefore, free recall should be especially sensitive to manipulations that increase or decrease the availability of features that make items easier to reconstruct—such as features that can be used to construct search sets that contain multiple rather than single list items. Categorization is transparently a manipulation of this sort. Taxonomic features such as "clothing," "food," and "furniture" yield search sets that contain multiple list items when lists are categorized, but not when lists are unrelated. Categorization should also have a positive effect on the familiarity component of nonrecollective retrieval. In adults, a key property of free recall of categorized lists is that items are output in bursts of same-category exemplars (clustering). Reconstructions of *deer* or *steak*, for instance, ought to be more likely to be authorized for output if they are part of bursts of other animals or foods (categorized lists) than if they are not (unrelated lists).

In sum, although the quality of retrieval cues affects both associative and free recall, the same cues have opposite effects on accuracy in the two paradigms owing to the fact that (a) associative recall has a recollective slant and categorization should interfere with that process by reducing cue-target specificity, whereas (b) free recall has a reconstructive slant and categorization should support that process by increasing the number of list items in search sets and by enhancing their perceived familiarity. In parametric terms, categorization should decrease the D parameters in associative recall (by diluting the cue-target specificity that supports this paradigm's dominant recollection process). In addition, categorization should increase the R and J parameters in free recall (by enhancing this paradigm's dominant reconstruction and familiarity processes). These are obvious predictions of dual-retrieval theory. Some less obvious ones are described in Experiments 2 and 3, below.

Experiment 1

This experiment had two aims, one methodological and the other theoretical. The methodological objective was to resolve the above question as to whether the strong support that Brainerd et al. (2009) obtained for the dual-retrieval model was due to the fact that they analyzed data sets in which subjects were required to meet a perfect-recall criterion. As mentioned, it is conceivable that subjects only use (error-prone) nonrecollective retrieval when no performance criterion is imposed, but they will be forced to resort to recollective retrieval as well in order to meet a stringent criterion. It would be highly desirable, therefore, to determine whether the dual-retrieval model fares as well when subjects participate in only a small, fixed number of study-test trials and no performance criterion is imposed. In anticipation of the need to do this, Brainerd et al. developed a no-criterion, fixed-trails implementation of the dual-retrieval model that provides identifiable estimates of all of the parameters in Table 1. This procedure was used in all of our experiments.

The theoretical objective of this experiment was to evaluate the dual-retrieval model's conception of process-level differences between associative and free recall. It will be remembered that the model posits that associative recall is more reliant on recollection and less reliant on reconstruction than free recall because associative recall provides a unique retrieval cue for each target and requires that a target only be recalled when its corresponding cue is presented. This is analogous to proposals that have been made in the dual-process recognition literature about associative versus item recognition (e.g., Yonelinas, 2002). The dual-retrieval account's specific predictions are that *D* parameters will be larger and *R* parameters will be smaller in associative than in free recall and that a larger proportion of associative than free recall will therefore be due to recollective retrieval.

Method

Subjects

The subjects were 48 undergraduates who participated in the experiment to fulfill a course requirement. Each subject was randomly assigned to one of two conditions: (a) associative recall of word pairs or (b) free recall of individual words.

Materials

The subjects learned to recall lists of words that were drawn from the Toglia and Battig (1979) norms. First, a pool of words was formed by randomly sampling 250 items from the 2,852 words on these norms. All words on the Toglia-Battig norms are rated for six semantic properties (concreteness, imagery, meaningfulness, familiarity, number of semantic attributes, and pleasantness). The mean value of the words in our pool on each of these properties did not differ significantly from the overall mean for the full sample of 2,852 words. This pool was then used to generate the lists that were administered to individual subjects in each of two conditions.

For subjects in the associative recall condition, the lists that were administered to individual subjects were generated as follows: A total of 80 words was sampled from the pool and then randomly paired to generate the 40 cue-target pairs that would be presented on study trials. Once the pairs were generated, they were examined to determine if there were obvious surface or semantic relations between any of them. In the few instances in which such relations were detected, they were eliminated by repairing cues and targets. For the subjects in the free recall condition, the lists that were administered to individual subjects were generated by randomly selecting one of the cue-target lists that had been generated for one of the subjects in the associative recall condition and simply using the 40 target members of that list as the free recall list. That is, each subject in the free recall condition was yoked to a (randomly selected) subject in the associative recall condition in the sense that they recalled the same words. In this way, the to-be-recalled targets were the same in the two conditions.

Procedure

Associative recall—The subjects learned to recall their respective lists using a $S_1T_1T_2$, S_2T_3, S_3T_4 procedure that is required to estimate the parameters of the dual-retrieval model (Brainerd et al., 2009). That is, the procedure consisted of three study-test trials, with the first trial consisting of a study cycle that was followed by two separate recall cycles and the other trials consisting of a study cycle followed by a single recall cycle. Following general memory instructions, the 40 word pairs were presented visually (computer screen), in random order, at a 3 sec rate. To eliminate short-term memory effects, the subject then participated in a 30-sec buffer activity (letter shadowing), which was followed by the first recall test. The cue words of the 40 pairs were presented in random order, at a 3 sec rate, and the subject attempted to recall the target word that had been presented with each cue. After the first recall test, the subject participated in another 30-sec of letter shadowing, followed by a second recall test that used the same procedure as the first. That was followed by two further trials, each of which consisted of a study cycle on the list of word pairs, 30-sec of letter shadowing, and an associative-recall test.

Free recall—The subjects in this condition also learned to recall their respective lists using the $S_1T_1T_2$, S_2T_3, S_3T_4 procedure. Following general memory instructions, the 40 words were presented visually (computer screen), in random order, at a 3 sec rate. The subject then participated in 30-sec of letter shadowing, followed by the first free recall test. The subject was simply instructed to recall as many words from the study list as possible, with recall continuing until 15 sec had elapsed without the emission of a word. After the first recall test, subjects participated in another 30-sec of letter shadowing, followed by a second free recall test that followed the same procedure as the first. This was followed by two further trials, each of which consisted of a study cycle on the list of words, 30-sec of letter shadowing, and a free recall test.

Results and Discussion

With respect to overall recall performance, the mean total errors per item were 2.52 (associative) and 1.92 (free), and the mean trial of last error for each item was 2.62 (associative) and 2.29 (free). Thus, free recall was easier than associative recall, a standard outcome when subjects recall the same targets via the two procedures (e.g., Halff, 1977). Subjecting these statistics to a 2 (paradigm: associative or free) X 2 (statistic: total errors or trial of last error) analysis of variance (ANOVA) merely confirms what is obvious from these means—namely, a main effect for paradigm such that free recall was easier than associative recall.

The important findings, which bear on the process loci of the differences between associative and free recall, appear in Table 2. The parameter estimates are organized so that estimates of the recollection parameters appear at the top, estimates of the reconstruction component of nonrecollective retrieval appear in the middle, and estimates of the familiarity judgment component of nonrecollective retrieval appear at the bottom. Although the statistical machinery in which the dual-retrieval model is implemented generates parameter estimates and goodness-of-fit tests at the same time, we discuss the two types of results separately, presenting the fit results first.

Model Fit

The first test evaluated the null hypothesis that retrieval is simpler than the model supposes and that, in particular, associative and free recall involve only a single nonrecollective process, which corresponds to the model in Equation A19. This test (see Equation A21) produces a G^2 statistic with 9 degrees of freedom for each experimental condition. For purposes of significance testing, G^2 statistics are asymptotically distributed as χ^2 (see Riefer & Batchleder, 1988), so that the critical value of this particular statistic to reject the null hypothesis at the .05 level is 16.92. In both conditions, the value of this statistic exceeded the critical value by a wide margin, the observed values being 153.77 (p = .0001; associative recall) and 203.17 (p = .0001; free recall). Thus, the data could not be accommodated by a model that posits only a single nonrecollective retrieval process. It also could not be accommodated by a model that posits only a single *recollective* retrieval process: As shown in the Appendix (see Equations A22 and A24), when the test statistic in Equation A21 produces a null hypothesis rejection, this automatically rules out a model that posits only a single recollective retrieval process.

The second fit test evaluated the null hypothesis that retrieval is not more complex than the model supposes and that, in particular, associative and free recall do not involve more than the posited pair of recollective and nonrecollective processes. This test (see Equation A18) produces a G^2 statistic with 4 degrees of freedom for each experimental condition. Thus, the critical value of this statistic to reject the null hypothesis at the .05 level is 9.49. In both conditions, the observed value of this statistic did not exceed the critical value to reject the null hypothesis, those values being 8.35 (p = .08; associative recall) and 0.65 (p = .96; free recall). The results of the first two fit tests support the conclusion that associative recall and free recall data involve more than a single retrieval process (one-process models were rejected at high levels of confidence) but that it is unnecessary to posit more two processes (the two-process model could not be rejected at conventional levels of confidence).

The third fit test, the results of which are displayed at the top of Table 3, evaluated the ability of the dual-retrieval model to account for the observed distributions of two fine-grained statistics of the data: the total number of errors per item and the trial number of the last error for each item. (Errors are omissions rather than intrusions.) Table 3 provides comparisons of the observed versus predicted distributions of these two statistics. For both conditions, the observed probability that the total number of errors per item had each of the five possible values (0, 1, 2, 3, 4) is reported along with the corresponding probabilities predicted by the dual-retrieval model, and the observed probability that the trial of last error had each of the five possible values (0, 1, 2, 3, 4) is reported along with the corresponding probabilities predicted by the dual-retrieval model. Visual inspection reveals close correspondence between observed and predicted distributions of these statistics in both conditions. Fit was evaluated for each observed-predicted pairing (2 paradigms X 2 statistics) with a Kolmogorov-Smirnoff test. The null hypothesis that predicted and observed distributions were equivalent could not be rejected in any of these tests.

These fit results provide an answer to the first question that motivated this experiment. Unlike the corpus of recall data that Brainerd et al. (2009) analyzed, the subjects in this experiment did not have to meet any performance criterion, and they participated in only a short sequence of study-test cycles. Nevertheless, fit tests revealed that the data could not be fit by a one-process model but were well-fit by a two-process model.

Process-Level Differences between Associative and Free Recall

Returning to Table 2, we consider how the observed values of the parameters bear on process-level predictions about associative versus free recall—specifically, whether associative recall produces larger values of direct access parameters and smaller values of reconstruction parameters than free recall. Visual inspection of the observed values of the D and R parameters shows that they fell out in accordance with prediction. Estimates of the individual parameters are reported for each condition, and the means of these pairs are also presented (in boldface).

First, with respect to the D parameters, although the mean values for the "early" recollective parameters (D_1/D_2 pair) did not differ reliably for associative versus free recall, the mean value of the "late" recollective parameters (D_{3C}/D_{3E} pair) for associative recall was nearly three times the corresponding value for free recall (.52 versus .18). To test these differences for significance, we used the appropriate likelihood ratio statistic (Brainerd et al., 2009, Equation A20) to evaluate the null hypotheses that D_{3C} and D_{3E} were equal in the associative and free recall conditions. Both null hypotheses were rejected, $G^2(1)$ = 12.17 (p

= .0005; D_{3C}) and 6.89 (p = .009; D_{3E}). (In this experiment and those that follow, the .05 level of confidence is used for all likelihood ratio tests of hypotheses about parameter values.)[2]

Next, with respect to the R parameters, the effects of associative versus free recall were reversed, relative to their effects on direct access, as expected: The mean value of the R_1/R_2 pair for free recall was more than twice the corresponding value for associative recall (.54 versus .26*).* We tested the between-condition difference in the values of each R parameter for statistical significance, and each $G^2(1)$ test produced a value that exceeded the critical value to reject the null hypothesis of between-condition equivalence of parameter values, $G^2(1)$ = 7.38 (p = .007; R_1) and 14.54 (p = .0001; R_2). Summing up, associative recall enhanced recollection and impaired reconstruction, relative to free recall.

Further inspection of Table 2 suggests that the familiarity judgment component of nonrecollective recall also differed for the two recall paradigms: The observed values of the four J parameters were all larger for free than for associative recall. We tested the between-condition difference in the values of each J parameter for statistical significance, and three of the four $G^2(1)$ tests produced a value that exceeded the critical value to reject the null hypothesis of equivalence of parameter values. Those parameters were J_1, $G^2(1)$ = 10.72 (p = .001), J_2, $G^2(1)$ = 8.22 (p = .004), and J_{3E},$G^2(1)$ = 14.54 (p = .0001). Thus, beyond making it easier to reconstruct items, free recall also makes it more likely that reconstructed items will seem familiar enough to output.

The overall difference in the relative reliance of associative and free recall on recollective versus nonrecollective retrieval can be summarized by using the dual-retrieval model and the parameter values in Table 2 to estimate the amount of recall that is due to each type of retrieval by the last trial of the experiment. The dual-retrieval model can be analyzed to derive an expression for the probability of correct recall on any recall test (see Brainerd et al., 2009). As these expressions involve only the direct access, reconstruction, and familiarity judgment parameters in Table 1, they can be partitioned into recollective and nonrecollective components, from which the probability of each type of recall can be estimated. These values can then be compared to the observed probability of correct recall on a given test to determine the portion that is due to recollective versus nonrecollective retrieval. We derived the expression for Test 4 in order to estimate the proportions of recollective and nonrecollective retrieval on the last test for associative versus free recall. Inserting the parameter values in Table 2 for the associative condition, the estimated probability of recollection-based recall was .48, and the estimated probability of nonrecollection-based recall was .11. Thus, the proportion of correct associative recall on the last test that was due to recollective retrieval was .48 ÷ (.48 + .11) = .81. Inserting the parameter values in Table 2 for the free recall condition in the same expression, the estimated probability of recollection-based recall was .37, and the estimated probability of nonrecollection-based recall was .37. Thus, the proportion of correct free recall on the last trial that was due to recollective retrieval was .37 ÷ (.37 + .37) = .50. In short, by the last recall test, associative recall was heavily recollective, whereas free recall was evenly split between the two modes of retrieval.

[2]It might be thought that the fact that the D_1/D_2 pair did not differ reliably between the associative and free recall disconfirms the prediction that recollective retrieval is more difficult in free recall. That interpretation would be incorrect, for two reasons. First, note that the mean value of this pair was smaller for free recall, as predicted. Second, note, too, that the fact that this difference was not reliable was most likely a floor effect. The values of D_1 and D_2 were not very far above zero in either condition, so little parametric variance was available with which to detect between-condition differences. In contrast, the values of D_{3C} and D_{3E} were far above zero in both conditions, and as reported, those parameters produced highly reliable differences in the predicted direction.

Experiment 2

The first experiment provided statistical support for the dual-retrieval model when subjects participated in a short sequence of trials in which no performance criterion was imposed. The experiment also revealed process-level differences between associative and free recall that are similar to the pattern that has often been hypothesized for associative versus item recognition (e.g., Malmberg, 2008; Yonelinas, 2002). Explicitly, recollective retrieval parameters had larger values for associative than for free recall, whereas the opposite was true for nonrecollective parameters.

In the second experiment, we move on to the other major question, Why does categorization drive the difficulty of free and associative recall in opposite directions? This experiment was concerned with the first part of this pattern; that the accuracy of free recall increases when lists are taxonomically organized. Three points will be remembered from the earlier discussion of the dual-retrieval model's explanation of this effect. First, the model expects that reconstruction will be enhanced by taxonomic organization because subjects will be able to store episodic traces of certain semantic features (e.g., "clothing," "furniture") that will allow them to generate search sets that contain multiple targets. Second, the model also expects that taxonomic organization will elevate familiarity judgment because reconstructed items will seem more familiar when they are part of clusters of same-category exemplars. Third, the model expects that although the net effect of categorization on accuracy is positive in free recall, it could have a negative effect on reconstruction in this paradigm. Any such effect should be small, however, owing to the paradigm's reconstructive slant.

Although the basic manipulation in this experiment was free recall of categorized versus unrelated lists, we included a second variable, the number of exemplars per category in categorized lists, to generate tests of further dual-retrieval predictions. Each subject recalled one of three types of lists: an unrelated list or a categorized list whose targets were exemplars of three taxonomic categories or a categorized list whose targets were exemplars of six taxonomic categories. As all lists were the same length, the first of the two categorized lists contained twice as many exemplars per category as the second of the two categorized lists. How should this affect the processes that are responsible for categorization effects? Here, Brainerd et al. (2009) proposed that reconstruction is influenced by two properties of the partial-identifying features that subjects use to generate search sets—namely, whether features generate sets that contain multiple targets and whether the features generate smaller versus larger search sets. Other things being equal, reconstruction is enhanced by features that generate sets with multiple targets and, other things being equal, by features that generate smaller rather than larger sets (because smaller sets can be more rapidly processed). The latter property differentiates the two types of categorized lists in this experiment, so that in the dual-retrieval account reconstruction will be better with six-category lists than with three-category lists.

Method

Subjects

The subjects were 120 undergraduates who participated to fulfill a course requirement. Each subject was randomly assigned to one of three conditions: (a) unrelated lists (U), (b) categorized lists on which each item was an exemplar of one of three familiar taxonomic categories (C3), and (c) categorized lists on which each item was an exemplar of one of six familiar taxonomic categories (C6).

Materials

The subjects learned to recall lists of words drawn from the Battig and Montague (1969) norms. First, a pool of 320 words was formed by sampling the 8 most frequent exemplars from 40 of the 56 Battig-Montague categories. This pool was then used to generate the lists that were administered to individual subjects in each of the three conditions, with each list consisting of 24 words. The lists that were administered in each of the three conditions were constructed as follows.

In the U condition, the 24 words were unrelated. The lists for individual subjects were generated, first, by randomly sampling 24 of the 40 taxonomic categories in the word pool and, second, by randomly sampling one exemplar per category. In the C3 condition, the 24 words were taxonomically related. The lists for individual subjects were generated by randomly sampling 3 of the 40 taxonomic categories in the word pool and placing all 8 exemplars of each category on the list. In the C6 condition, the 24 words were also taxonomically related. The lists for individual subjects were generated by, first, randomly sampling 6 of the 40 taxonomic categories in the word pool and by, second, randomly sampling 4 of the 8 exemplars of each category and placing them on the list.

Procedure

The procedure in all three conditions was the same as in the free recall condition of Experiment 1, except that all lists consisted of 24 rather than 40 words and the lists in conditions C3 and C6 were taxonomically organized.

Results and Discussion

With respect to overall performance statistics, the mean total errors per item were 1.97 (U), 1.80 (C3), and 1.60 (C6), and the mean trial of last error for each item was 2.20 (U), 2.17 (C3), and 1.98 (C6). One-way ANVOAs followed by trend analyses showed what is obvious from these values: (a) With the total errors measure, performance was poorer in the U condition than in either categorized condition and that performance in the C3 condition was poorer than in the C6 condition, and (b) with the trial of last error measure, performance in the C6 condition was better than in either of the other conditions. That C6 performance was more accurate than C3 performance conforms to the dual-retrieval prediction.

Results that bear on the process loci of these effects appear in Table 4. The parameter estimates are organized as in Experiment 1, with estimates of the recollective parameters appearing at the top, estimates of the reconstruction component of nonrecollective retrieval appearing in the middle, and estimates of the familiarity judgment component of nonrecollective retrieval appearing at the bottom. Before considering these findings, we report the same goodness-of-fit analyses that were conducted in Experiment 1.

Model Fit

The first series of fit tests evaluated the null hypothesis that retrieval is simpler than the dual-retrieval model assumes and that, in particular, learning to recall these lists involved only a single nonrecollective process. As before, this test produces a G^2 statistic with 9 degrees of freedom for each condition, which is asymptotically distributed as χ^2. In all three conditions, the value of this statistic exceeded the critical value of 16.92 by a wide margin, the observed values being 242.99 (U), 244.97 (C3), and 281.14 (C6). These results also rule out an alternative model that assumes that recall involves only a single recollective retrieval process (see above).

As in Experiment 1, the second series of fit tests evaluated the null hypothesis that learning to recall these lists did not involve more than two retrieval processes. These tests generated the same $G^2(4)$ statistic as before. In each condition, the observed value of this statistic did not exceed the critical value (9.49) to reject the null hypothesis, the values being 0.17 (U), 1.97 (C3), and 7.41 (C6).

Also as in Experiment 1, the third series of fit tests, the results of which are displayed in the middle of Table 3, evaluated the ability of the dual-retrieval model to account for the observed distributions of the total number of errors per item and the trial number of the last error for each item. For each condition, the observed probability that the total number of errors had each of the five possible values (0, 1, 2, 3, 4) is reported along with the corresponding probabilities predicted by the model, and the observed probability that the trial of last error had each of the five possible values (0, 1, 2, 3, 4) is reported along with the corresponding probabilities predicted by the model. Fit was evaluated with Kolmogorov-Smirnoff tests, and none of the 6 tests (3 conditions X 2 statistics) yielded a rejection of the null hypothesis that predicted and observed distributions were equivalent.

Process-Level Effects of Taxonomic Organization

Returning to Table 4, we consider the main dual-retrieval prediction first—that taxonomic organization improves free recall because it enhances reconstruction. The values of the R_1/R_2 pair confirm that prediction. The average value of this pair for the C3 and C6 conditions (.56) was considerably larger than the corresponding average for the U condition (.32). To test this difference for statistical reliability, we used the same likelihood ratio tests as in Experiment 1. That is, we tested (a) the null hypothesis that R_1 and R_2 were equal in the U and C3 conditions and (b) the null hypothesis that R_1 and R_2 were equal in the U and C6 conditions. Both hypotheses were rejected, $G^2(2)$ = 16.56 and 15.79. In addition, the values of the reconstruction parameters for the C3 and C6 conditions provide support for the dual-retrieval prediction that reconstruction should be easier when taxonomic categories are smaller: The value of R_1 (but not R_2) was noticeably larger in the C6 condition than in the C3 condition. When we tested the null hypothesis that values of R_1 were the same in these two conditions, the test statistic, $G^2(1)$ = 13.29, exceeded the critical value that was required to reject this null hypothesis (3.84). Thus, beyond the fact that taxonomic organization enhanced reconstruction, the enhancement was greater when lists contained fewer exemplars per category. It should be noted, however, that this advantage should decrease and, perhaps, disappear as the number of exemplars per category becomes very small (say, two or three): As the number of exemplars becomes very small, categorized lists become increasing like unrelated lists and less like categorized lists.

Next, we consider findings for the familiarity judgment component of nonrecollective retrieval, which the dual-retrieval account predicts will be easier with categorized lists. The values of the J parameters in Table 4 reveal support for this prediction: (a) For U versus C3, the value of one of the four J parameters (J_{3C}) was reliably larger [by a $G^2(1)$ test] in the C3 condition, and (b) the values of three of the four J parameters (J_1, J_{3C}, and J_{3E}) were reliably larger [by $G^2(1)$ tests] in the C6 condition than in the U condition. In addition, familiarity judgment was easier with six categories than with three. $G^2(1)$ tests showed that the null hypothesis of between-condition equivalence of parameter values could be rejected for J_2 and J_{3C} in the C3/C6 comparison.[3]

[3]A comparison of parameter estimates for the U condition to parameter estimates for the free recall condition of Experiment 1 (which also involved unrelated lists) discloses that recollection parameters were smaller and reconstruction parameters were larger in Experiment 1. This is most probably a list length effect. Longer lists have been found to make it more difficult to rely on recollective retrieval during recall (Brainerd et al., 2009), and the lists Experiment 1 were nearly twice as long as those in the present experiment.

Last, we consider categorization effects on recollective retrieval. Here, remember that the model predicts that categorization generally interferes with recollective retrieval. Although the net effect of categorization on free recall was *positive*, categorization *interfered* with recollective retrieval in free recall in the manner that was predicted: The average value of the D_1/D_2 pair was larger for unrelated lists (.22) than it was for either C3 (.08) or C6 (.06). When we conducted between-condition tests for differences in parameter values, we found that D_1 was significantly larger in the U condition than in the C6 condition, $G^2(1) = 24.72$, and that D_2 was significantly larger in the U condition than in either the C3 or the C6 condition, $G^2(1) = 12.67$ and 17.14.

These results for the D parameters illustrate two important points, a general one about model-based measurement and a specific one about theoretically-predicted process loci of categorization effects. The general point is that model-based measurement can detect an informative class of situations that cannot be detected by global performance measures—namely, situations in which manipulations affect some memory processes in opposite ways. A global performance measure, such as the mean number of errors or the mean trial of last error, can only tell us whether a manipulation's *net effect* on accuracy is positive or negative, whereas model-based measurement can show that a manipulation whose net effect is positive (or negative) can nevertheless have negative (or positive) effects on particular processes. The other point is about the theoretical issue that motivated this experiment, why categorization has opposite effects on associative versus free recall. According to the dual-retrieval model, categorization ought to impair associative recall because (a) it suppresses recollective retrieval by diverting retrieval away from direct access of verbatim traces and toward reconstruction, and (b) associative recall should be especially sensitive to such suppression because it is slanted toward recollective retrieval. However, the first part of this prediction is not paradigm-specific. Although the recollective slant of associative recall should make it particularly sensitive to recollective interference from categorization, such interference should also occur in other paradigms, as long as recollection makes some contribution to performance, as it does in free recall. This result was obtained in the present experiment.

Experiment 3

We turn to associative recall in the final experiment and report a study that is an associative recall analogue of Experiment 2. In other words, subjects again learned to recall unrelated and taxonomically organized word lists, using the no-criterion $S_1T_1T_2$, S_2T_3, S_3T_4 procedure, except that study lists consisted of word pairs and recall tests involved associative rather than free recall. The taxonomic organization variable was factorially manipulated over the cue and target sides of word pairs.

Method

Subjects

The subjects were 140 undergraduate students who participated in the study to fulfill a course requirement. Each subject was randomly assigned to one of four associative recall conditions: (a) unrelated cues and unrelated targets (UU), (b) categorized cues and unrelated targets (CU), (c) unrelated cues and categorized targets (UC), and (d) categorized cues and categorized targets (CC).

Materials

The materials were the same as in Experiment 2 in that the same pool of 320 words was used to generate the lists that were administered to individual subjects. The lists were different, of course, as the procedure involved associative rather than free recall. Each list consisted of 32 word pairs, with the lists that were administered in each of the four conditions being constructed as follows. For the UU lists, the 32 cue words were chosen by randomly sampling one exemplar apiece from 32 randomly selected categories, and the 32 target words were chosen by randomly sampling one exemplar apiece from 32 randomly selected categories. The study list was then constructed by randomly pairing the 32 cues with the 32 targets, subject to the constraint that exemplars from the same category were never paired. For the UC lists, the cue were words were chosen in the same way, but the target words were chosen by randomly selecting 4 exemplars from each of the 8 categories from which cue words had not been selected. The study list was then generated by randomly pairing each of the 32 cues with one of the 32 targets (8 categories x 4 exemplars per category). For the CU lists, the target words were selected in the same manner as in the UU condition, while the cue words were chosen via the procedure that was used to choose target words in the UC condition. The study list was then constructed by randomly pairing each of the 32 cues (8 categories x 4 exemplars per category) with one of the 32 targets. For the CC lists, the cue words were chosen by randomly selecting 4 exemplars from each of 8 randomly selected categories, and the target words were chosen by randomly selecting 4 exemplars from another 8 randomly selected categories. The study list was then generated by randomly pairing each of the 32 cues with one of the 32 targets, subject to the constraint that for each of the 8 cue categories, each of the 4 exemplars had to be paired with a cue word from a different cue category.

Procedure

The procedure in all four conditions was that same as in the associative recall condition of Experiment 1, except that the lists consisted of 32 rather than 40 word pairs, and in three of the conditions, the words on one or both sides of the pairs were taxonomically organized.

Results and Discussion

In an experiment such as this, the focus is normally on how manipulating taxonomic organization of lists affects overall performance statistics, such as total errors or (less often) the trial number the last error. With respect to such statistics, the mean total errors per item were 1.48 (UU), 1.74 (UC), 2.22 (CU), and 2.86 (CC), and the mean trial of last error for each item was 1.72 (UU), 2.21 (UC), 2.37 (CU), and 3.10 (CC). Subjecting these statistics to a 2 (cues: unrelated or related) X 2 (targets: unrelated or related) analysis of variance (ANOVA) merely confirms what is obvious from these means—namely, main effects for categorization (changing from unrelated to categorized items impairs recall on both the cue and target sides) and an interaction (the impairment is greater on the cue side than on the target side). Such findings reveal nothing about whether recollective or nonrecollective retrieval is at the bottom of these effects.

Those findings appear in Table 5, where estimates of the recollection parameters are reported by condition at the top, estimates of the parameters that measure the reconstruction component of nonrecollective recall are reported in the middle, and estimates of the parameters that measure the familiarity judgment component of nonrecollective recall are reported at the bottom. Before considering these results, we again report goodness-of-fit tests for the dual-retrieval model.

Model Fit

As before, the first series of fit tests evaluated the null hypothesis that retrieval is less complex than the dual-retrieval model supposes and that, in particular, associative recall involves only a single nonrecollective retrieval process. This is the same fit test as in Experiments 1 and 2, which produces a G^2 statistic with 9 degrees of freedom and a critical value to reject the null hypothesis of 16.92 for each of the four conditions. The observed value for each of the four conditions exceeded the critical value by a wide margin, those values being 94.58 (UU), 127.59 (UC), 153.56 (CU), and 198.87 (CC).

Also as before, the second series of fit tests evaluated the null hypothesis that associative recall does not involve more than the posited pair of recollective and nonrecollective retrieval processes. This test produces a G^2 statistic with 4 degrees of freedom for each experimental condition and a critical value to reject the null hypothesis of 9.49. The observed value for each of the four conditions was below the critical value, those values being 5.82 (UU), 0.60 (UC), 2.73 (CU), and 9.44 (CC).

The third series of fit tests, the results of which are displayed at the bottom of Table 3, evaluated the dual-retrieval model's ability to account for the observed distributions of two fine-grained statistics of the data, the total number of errors per item and the trial number of the last error for each item. For each of the four conditions, the observed probability that the total number of errors per item had each of the five possible values (0, 1, 2, 3, 4) is reported along with the corresponding probabilities predicted by the dual-retrieval model, and the observed probability that the trial of last error had each of the five possible values (0, 1, 2, 3, 4) is reported along with the corresponding probabilities predicted by the dual-retrieval model. Visual inspection reveals close predicted-observed correspondence, and 8 Kolmogorov-Smirnoff tests (4 list conditions X 2 statistics) showed that the null hypothesis that predicted and observed distributions were equivalent could not be rejected in any of these tests.

Process-Level Effects of Taxonomic Organization

Returning to Table 5, we consider whether taxonomic organization suppresses associative recall because it interferes with recollective retrieval, as the dual-retrieval explanation anticipates. As in earlier experiments, the recollection parameters are presented in two groups in Table 5—namely, those that measure the difficulty of learning to recollect items when items are still in state U (the D_1/D_2 pair) versus those that measure the difficulty of learning to recollect items after they have escaped U by becoming reconstructable (the D_{3C}/D_{3E} pair).

The D values in Table 5 are obviously consistent with prediction. With respect to the D_1/D_2 pair, taking UU lists as a baseline, the mean value of this pair declines from .32 to .22 when targets are categorized (UC lists) and from .32 to .19 when cues are categorized (CU lists). To test these differences for significance, we used the appropriate likelihood ratio statistic (Brainerd et al., 2009, Equation A20) to evaluate the null hypothesis that D_1 and D_2 were equal in the UU and UC conditions and the null hypothesis that these parameters were equal in the UU and CU conditions. Both null hypotheses were rejected, $G^2(2) = 10.74$ and 16.08. There are further declines in the mean value of the D_1/D_2 pair when the targets of UC pairs are categorized (.22 to .05) or when the cues of CU pairs are categorized (.19 to .05). Both declines were reliable, $G^2(2) = 9.47$ and 16.09.

Moving on to the D_{3C}/D_{3E} pair and again taking UU lists as a baseline, categorization also impaired recollection after items had become reconstructable. However, the effect was more limited in that it was confined to the target side of pairs. The mean value of D_{3C} declined (from .36 to .11) when UU was changed to UC but not when it was changed to CU. When the null hypothesis that the D_{3C} values were the same in the UU and UC conditions was tested for significance, it was rejected, $G^2(1) = 56.73$. Also, the mean value of D_{3C}/D_{3E} pair declined (from .37 to .16) when CU was changed to CC. When the null hypothesis that both parameters' values were the same in the two conditions was tested for significance, it was rejected, $G^2(2) = 10.86$.

Summing up the findings for recollective retrieval, as the dual-retrieval model predicts, recollection parameters were suppressed when taxonomic organization was imposed on either the cue or the target side of word pairs. The effect was more pronounced before items became reconstructable than after, so the ability to rely on nonrecollective recall compensated somewhat for this manipulation's deleterious effects on recollection.

Turning to categorization's effects on nonrecollective retrieval, we consider its effect on both the reconstruction and familiarity judgment components. Although associative recall should be far more sensitive to the recollective effects of categorization, nonrecollective effects might be observed, too. As we know, categorization allows subjects to store episodic traces of features (e.g., "clothing," "furniture") that will generate search sets that contain multiple targets, and this will be helpful in free recall (because subjects' task is to recall as many items as possible--in any order and at any time during a test cycle). However, this effect of categorization would be harmful in associative recall: Subjects' task is more constrained, in that they must recall a specific target when a specific cue is presented and not recall other targets when that cue is presented. Obviously, that will be more difficult whenever reconstruction processes features that generate search sets containing multiple targets. Just as obviously, reconstruction will be more likely to process such features if subjects study pairs in which cue words trigger salient semantic features that map with multiple targets (as when cues are categorized) than if subjects do not study such pairs (as when cues are unrelated).

The values of the reconstruction parameters in Table 5 provide some support for this prediction. If we take UU as the baseline, the mean value of the R_1/R_2 pair decreases when taxonomic organization is imposed on cues (from .74 to .31), but not when targets are categorized. On the other hand, if we take CC as the baseline, the mean value of this pair increases when taxonomic organization is removed from the cue side (from .38 to .71), but not when it is removed from the target side. We tested both the UU → CU decline and the CC → UC increase for significance by evaluating the null hypothesis that R_1 and R_2 were equal in the indicated pairs of conditions. Both null hypotheses were rejected, $G^2(2) = 17.78$ and 7.35.

Finally, inspection of the estimates of the *J* parameters in Table 5 reveals no systematic tendency for the "early" pair (J_1/J_2) to be suppressed or enhanced by categorizing the cue side of pairs or by categorizing the target side. Inspection of the "late" pair (J_{3C}/J_{3E}) suggests that familiarity judgment may have been suppressed by categorizing the target side of pairs, but the likelihood ratio tests revealed no reliable differences between the UU and UC conditions or between the CU and CC conditions.

General Discussion

The chief aims of this paper were to introduce a basic recall methodology for studying recollective and nonrecollective retrieval and to exemplify its use in testing prospective predictions about these processes. This technique, the dual-recall procedure, has two tactical advantages, relative to traditional recognition-based procedures for gaining leverage on the two forms of retrieval. First, it is simpler. Whereas recognition-based procedures require that subjects make meta-cognitive judgments about their old/new responses (e.g., remember/know, confidence, inclusion-exclusion), the present procedure relies on low-burden recall tasks. Subjects merely study and recall target materials, using any standard method of recall (e.g., associative, cued, free, serial) and without having to meet any performance criterion. Unlike meta-cognitive judgment tasks, this procedure is easily used with children and with adults whose episodic memories are compromised by neurocognitive impairments. The resulting data are analyzed with an absorbing Markov chain (Equation 1) whose parameters

NIH-PA Author Manuscript

NIH-PA Author Manuscript

supply multiple measurements of recollective retrieval and of the two components of nonrecollective retrieval, reconstruction and familiarity judgment. This model allows for rigorous evaluations of fit, and for any set of target data, one is able to: (a) test the hypothesis that recall involves *less than* the two processes that are required by the distinction between recollective and nonrecollective retrieval; (b) test the hypothesis that recall involves *more than* these two processes; and (c) assess the degree of correspondence between observed distributions of fine-grained statistics of recall data and the corresponding distributions that are predicted by the model.

The second tactical advantage of the dual-recall procedure is that it avoids the on-going controversy over whether recognition involves both recollective and nonrecollective components or only the latter (e.g., Donaldson, 1996; Dunn, 2004, 2008; Gillund & Shiffrin, 1984; Hirshman & Master, 1997; Malmberg, 2008; Rotello et al., 2004; Wixted, 2007). In contrast to recognition, there is a long history of Markov modeling studies that converge on the conclusion that recall improves over trials in accordance with two distinct threshold processes, rather than a single strength process (for reviews, see Brainerd et al., 1982; Greeno, 1974; Greeno et al., 1978).

Turning to the other aim of this paper, to exemplify how the dual-recall procedure is implemented to test theoretical predictions, we made three choices that we thought would illustrate the procedure to good advantage. First, although Brainerd et al. (2009) reported extensive statistical support for the dual-retrieval model, all of the data sets that they analyzed came from experiments that imposed perfect-recall criteria on subjects. It is conceivable that subjects use only a single retrieval process—especially, error-prone nonrecollective retrieval—if no performance criterion is imposed, and they participate in only a short sequence of trials. Therefore, we conducted experiments of that sort in order to determine whether there is strong support for the dual-retrieval model under less taxing conditions. Second, we chose to study associative versus free recall. That allowed us to investigate the recall analogue of an issue that has figured centrally in the dual-process recognition literature (see Malmberg, 2008; Yonelinas, 2002). The standard view is that associative recognition is a more recollectively oriented task than item recognition because, in the former, subjects activate a recollection rejection process in order to suppress false alarms to distractors whose cue members appeared on the study list. Therefore, it is of immediate theoretical interest to investigate the parallel prediction that associative recall is more recollectively slanted than free recall. There is an independent theoretical rationale for that prediction from the recall literature because the theoretical distinctions that underlie the dual-recall procedure (Brainerd et al., 2009) posit that the availability of a unique cue for each target, together with the requirement that a target can only be output when its corresponding cue is present, forces associative recall to be more recollective and less reconstructive than free recall.

The third choice was to use taxonomic organization as a focal manipulation. That choice was motivated by the fact that taxonomic organization has opposite effects on accuracy in associative versus free recall and by the fact that different process-level explanations can be evaluated with the dual-recall procedure. As is well known and widely studied, free recall is more accurate when lists consist of items that are exemplars of a few taxonomic categories than when lists consist of unrelated items. As is less well known and infrequently studied, associative recall is impaired when taxonomic organization is imposed on word pairs. Why? According to the dual-retrieval explanation, the baseline situation is that associative recall is more recollective and less reconstructive than free recall. Hence, imposing taxonomic organization impairs associative recall because it interferes with the dominant recollective mode of retrieval, but it enhances free recall because it supports the dominant reconstructive mode of retrieval. We showed that these explanations can be rigorously tested by varying taxonomic organization on associative and free recall tasks and then estimating the recollective and nonrecollective parameters of the dual-retrieval model.

The experiments themselves produced some instructive findings, which for convenience we will divide into background data that bear on the potential usefulness of the present approach to dual-retrieval processes and findings that bear on substantive questions about associative versus free recall and about the contrasting effects of taxonomic organization on different recall tasks. With respect to background results, we conducted three-step fit evaluations of the dual-retrieval model in each experiment. Those evaluations converged on the important conclusion that statistical support for the dual-retrieval model is strong when subjects participate in a short sequence of trials on which no performance criterion is imposed. Here, there were two groups of background results, one bearing on the inadequacy of one-process models and the other bearing on the adequacy of the dual-retrieval model.

With respect to the former group of results, echoing the on-going debate over whether recognition involves only a nonrecollective process (e.g., Dunn, 2004, 2008), we tested the hypothesis that a model that involves only nonrecollective retrieval (or recollective retrieval) could account for associative or free recall data. Across the experiments, this hypothesis was tested for nine distinct experimental conditions and was always rejected at a high level of confidence. For instance, even the smallest value of the G^2 statistic for any of these tests (for the UU condition of Experiment 1) was more than five times the critical value that is to required reject a one-process model. Thus, in contrast to the weight of fit results in the recognition literature, which favors a one-process model (see Malmberg, 2008), the corresponding results for recall ran consistently against it.

With respect to the other group of background results, it does not follow from rejection of a one-process model that the dual-retrieval conception is adequate. Therefore, we also tested the hypothesis that a model that involves both recollective and nonrecollective retrieval is sufficient to account for recall data. Across the experiments, this hypothesis was never rejected. As these analyses relied on global fit tests, a remaining question was whether the two-process model could give good accounts of fine-grained features of recall data. It did.

Turning to substantive findings, perhaps the most informative results were concerned with the process-level differences between associative and free recall because those results are congruent with the hypothesis that associative recall is more recollective and less reconstructive free recall. Consistent with that notion, the mean value of the late pair of recollection parameters, D_{3C} and D_{3E}, for associative recall was nearly three times the corresponding value for free recall. Also, the mean value of the two reconstruction parameters, R_1 and R_2, was only half as large for associative recall as for free recall, suggesting that associative recall is both more reliant on recollective retrieval and less reliant on nonrecollective retrieval. Further analysis of these parameter differences pointed to the conclusion that associative recall is strongly recollective, whereas free recall is more balanced between recollective and nonrecollective retrieval. Specifically, when parameter values were used to estimate the percentages of recollective and nonrecollective retrieval on the final recall test, associative recall was found to be overwhelmingly recollective (81%), but free recall was found to be evenly divided between recollective and nonrecollective retrieval.

Other instructive results were obtained for the question of why taxonomic organization has opposite effects on associative versus free recall. The overall picture for associative recall was consistent with the predictions of the dual-retrieval framework. More explicitly, in line with the notion that introducing salient semantic overlap on either the cue or target side of pairs should interfere with the dominant recollective model of retrieval, mean values of the

D parameters were smaller when either cues or targets belonged to a few taxonomic categories than when they were unrelated. The corresponding picture for free recall was congruent with prediction, too, inasmuch as taxonomic organization was expected to enhance familiarity, and the mean values of the *J* parameters were larger for categorized than for unrelated lists. The dual-retrieval model made two further predictions about free recall that received support: (a) *R* parameters will have larger values for categorized than for unrelated lists, and (b) for categorized lists of the same length, *R* parameters will have larger values for lists with fewer exemplars per category.

To conclude, we have shown that recall-based measurement can provide a productive approach to the controversial problems of separation and quantification of dual-retrieval processes, we have shown that recall-based measurement can be used to investigate theoretical predictions that have figured centrally in the dual-process recognition literature, and we have shown how those things can be accomplished with simple no-criterion designs that are appropriate for children and neuro-cognitively impaired adults. Nevertheless, it might be argued that a disadvantage of our approach is that there may be little or no overlap between these recall processes and the corresponding recognition processes; that the *D* parameters may measure something quite different than what traditional recognition indexes of recollection measure and the *J* parameters may measure something quite different than what traditional recognition indexes of familiarity measure. We note three points in response to this argument.

First, the argument loses much of its force when one considers the on-going controversy over whether recognition actually involves only a single retrieval process. While it is true that there is a large literature in which data on recollective and nonrecollective retrieval have been reported for recognition, recent critiques of recognition-based methodologies seriously challenge the reliability of those data (Malmberg, 2008). Ironically, the strong statistical support for the dual-retrieval model of recall means that a more convincing case can currently be made for the reliability of recall-based measurement than for the reliability of recognition-based measurement. Second, the above argument is an empirical question, not a proof of logical incompatibility. The answer will ultimately turn on the extent to which recall- and recognition-based measurements of dual-retrieval processes react similarly to experimental manipulations. In that connection, Brainerd et al. (2009) noted several such similarities when they applied the dual-retrieval model to a large corpus of recall data (e.g., that life-span trends in direct access and familiarity judgment parameters were the same as reported patterns for recollection and familiarity measurements in the recognition literature). Additional similarities were identified in our experiments (e.g., that associative recall is more recollectively slanted than free recall and that free recall is more nonrecollectively slanted than associative recall). Third, to prepare the ground for careful comparison of recall- and recognition-based measurements, Brainerd et al. developed a recognition version of Equation 1, which provides parameters that measure the traditional recognition conceptions of recollection and familiarity (see their Equation 5). They then showed how to evaluate whether recall- and recognition-based measurement are tapping the same retrieval processes by using these two models in tandem, within individual experiments. A core experimental design was outlined in which all subjects study the same list and are exposed to the same manipulations, but some subjects perform recall during test cycles while others perform recognition. Recollection and familiarity parameters are estimated for recall data with the recall model and for recognition data with the recognition model, and then, the two sets of estimates are compared to determine the extent to which they react similarly to experimental manipulations.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Acknowledgments

Portions of the research were supported by grants from the Natural Sciences and Engineering Research Council to the first author and by National Institutes of Health Grant 1RC1AG036915-01 to both authors. The PC-based software that was used to conduct the analyses of the dual-retrieval model that are reported in this paper (goodness-of-fit tests, parameter estimation, and significance tests of parameter values) is available from the first author upon request. We are indebted to J. C. Dunn, K. J. Malmberg, S. Lewandowsky, and A. P. Yonelinas for their comments on earlier drafts of this article.

References

Barnhardt TM, Choi H, Gerkens DR, Smith SM. Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. Journal of Memory and Language. 2006; 55:213–231.

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

Bjorklund DF. How changes in knowledge base contribute to the development of children's memory. Developmental Review. 1987; 7:93–130.

Bjorklund DF, Jacobs JW. Associative and categorical processes in children's memory: The role of automaticity in the development of free recall. Journal of Experimental Child Psychology. 1985; 39:599–617.

Bjorklund DF, Muir JE. Children's development of free recall memory: Remembering on their own. Annals of Child Development. 1988; 5:79–123.

Bower, GH.; Theios, JA. A learning model for discrete performance levels. In: Atkinson, RC., editor. Studies in mathematical psychology. Stanford, CA: Stanford University Press; 1964.

Brainerd CJ, Desrochers A, Howe ML. Stages-of-learning analysis of picture-word effects in associative memory. Journal of Experimental Psychology: Human Learning and Memory. 1981; 7:1–14.

Brainerd CJ, Howe ML, Desrochers A. The general theory of two-stage learning: A mathematical review with illustrations from memory development. Psychological Bulletin. 91:634–665.

Brainerd CJ, Kingma J, Howe ML. On the development of forgetting. Child Development. 1985; 56:1103–1119.

Brainerd, CJ.; Reyna, VF. The science of false memory. New York: Oxford University Press; 2005.

Brainerd CJ, Reyna VF, Ceci SJ. Developmental reversals in false memory: A review of data and theory. Psychological Bulletin. 2008; 134:343–382. [PubMed: 18444700]

Brainerd CJ, Reyna VF, Howe ML. Trichotomous processes in early memory development, aging, and neurocognitive impairment: A unified theory. Psychological Review. 2009; 116:783–732. [PubMed: 19839684]

Brainerd CJ, Reyna VF, Wright R, Mojardin AH. Recollection rejection: False-memory editing in children and adults. Psychological Review. 2003; 110:762–784. [PubMed: 14599242]

Brainerd CJ, Wright R, Reyna VF, Payne DG. Dual-retrieval processes in free and associative recall. Journal of Memory and Language. 2002; 46:120–152.

Brown GDA, Neath I, Chater N. Serial and free recall: Common effects and common mechanisms? A reply to Murdock (2008). Psychological Review. 2008; 115:781–785.

Brown SC, Flores LM, Goodman KM, Conover JN. Clustering and recall: Do high clusterers recall more than low clusterers because of clustering? Journal of Experimental Psychology: Learning, Memory, and Cognition. 1991; 17:710–721.

Clark SE, Gronlund SD. Global matching models of recognition memory: How the models match the data. Psychonomic Bulletin & Review. 1996; 3:37–60.

Crowder, RG. Principles of learning and memory. Hillsdale, NJ: Erlbaum; 1976.

Dunn JC. Remember-know: A matter of confidence. Psychological Review. 2004; 111:524–542. [PubMed: 15065921]

Dunn JC. The dimensionality of the remember–know task: A state-trace analysis. Psychological Review. 2008; 115:426–446. [PubMed: 18426296]

Estes WK. Learning theory and the new mental chemistry. Psychological Review. 1960; 67:207–223. [PubMed: 13820869]

Gillund G, Shiffrin RM. A retrieval model for both recognition and recall. Psychological Review. 1984; 91:1–67. [PubMed: 6571421]

Glanzer M, Kim K, Hilford A, Adams JK. Slope of the receiver-operating characteristic in recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:500–513.

Gollin ES, Sharps MJ. Facilitation of free recall by categorical blocking depends on stimulus type. Memory & Cognition. 1988; 16:539–544.

Greeno JG. Identifiability and statistical properties of two-stage learning with no successes in the initial stage. Psychometrika. 1968; 33:173–215. [PubMed: 5242170]

Greeno, JG. Representation of learning as discrete transition in a finite state space. In: Krantz, DH.; Atkinson, RC.; Luce, RD.; Suppes, P., editors. Contemporary developments in mathematical psychology. San Francisco, Calif: Freeman; 1974.

Greeno JG, James CT, DaPolito FJ. A cognitive interpretation of negative transfer and forgetting of paired associates. Journal of Verbal Learning and Verbal Behavior. 1971; 10:331–345.

Greeno, JG.; James, CT.; DaPolito, FJ.; Polson, PG. Associative learning: A cognitive analysis. Englewood Cliffs, N.J: Prentice-Hall; 1978.

Greeno JG, Steiner TE. Markovian processes with identifiable states: General considerations and applications to all-or-none learning. Psychometrika. 1964; 29:309–333.

Gronlund SD, Ratcliff R. Time course of item and associative information: Implications for global memory models. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1989; 20:1355–1369.

Halff HM. The role of opportunities for recall in learning to retrieve. American Journal of Psychology. 1977; 90:383–406.

Hirshman E, Master S. Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. Memory & Cognition. 1997; 25:345–351.

Howard MW, Kahana MJ. A distributed representation of temporal context. Journal of Mathematical Psychology. 2002; 46:269–299.

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

Humphreys MS, Yuille JC. Errors as a function of noun concreteness. Canadian Journal of Psychology. 1973; 27:83–94.

Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. Journal of Memory and Language. 1991; 30:513–541.

Kintsch, W. Models for free recall and recognition. In: Norman, DA., editor. Models of human memory. New York: Academic Press; 1970.

Kintsch W, Morris CJ. Application of a Markov model to free recall and recognition. Journal of Experimental Psychology. 1965; 69:200–206. [PubMed: 14279778]

Klein KA, Addis KM, Kahana MJ. A comparative analysis of serial and free recall. Memory & Cognition. 2005; 33:833–839.

Malmberg KJ. Recognition memory: A review of the critical findings and an integrated theory for relating them. Cognitive Psychology. 2008; 57:335–384. [PubMed: 18485339]

Mandler, G. Organization and memory. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation. Vol. 1. San Diego, CA: Academic Press; 1967. p. 327-372.

Mandler G. Recognizing: The judgment of previous occurrence. Psychological Review. 1980; 87:252–271.

McClelland JL, Chappell M. Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. Psychological Review. 1998; 105:724–760. [PubMed: 9830377]

Murdock BB, Lockhart RS. Memory and theory of signal detection. Psychological Bulletin. 1970; 74:100–109.

Nelson DL, McGivney VM, Gee NR, Janczura GA. Interpreting the influence of implicitly activated memories on recall and recognition. Psychological Review. 1998; 105:299–324. [PubMed: 9577240]

Nelson DL, Schreiber TA, McEvoy CL. Processing implicit and explicit representations. Psychological Review. 1992; 99:322–348. [PubMed: 1594728]

Pagel JC. A Markov analysis of transfer in paired-associate learning with high intralist similarity. Journal of Verbal Learning and Verbal Behavior. 1973; 12:456–470.

Park H, Reder LM, Dickison D. The effects of word frequency and similarity on recognition judgments: The role of recollection. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:568–578.

Reyna VF, Brainerd CJ. Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences. 1995; 7:1–75.

Reyna, VF.; Mills, BA. Interference processes in fuzzy-trace theory: Aging, Alzheimer's disease, and development. In: MacLeod, C.; Gorfein, D., editors. Inhibition in cognition. Washington, DC: American Psycholoical Association; 2007. p. 185-210.

Riefer DM, Batchelder WH. Multinomial modeling and the measurement of cognitive processes. Psychological Review. 1988; 95:318–339.

Rock I. The role of repetition in associative learning. American Journal of Psychology. 1957; 70:186–193. [PubMed: 13424758]

Rotello CM, Macmillan NA, Reeder JA. Sum-difference theory of remembering and knowing: A two-dimensional signal detection model. Psychological Review. 2004; 111:588–616. [PubMed: 15250777]

Schwenn EA, Underwood BJ. Effect of formal and associative similarity on paired-associate and free-recall learning. Journal of Verbal Learning and Verbal Behavior. 1968; 7:817–827.

Shiffrin RM. Modeling memory and perception. Cognitive Science. 2003; 27:341–378.

Shiffrin RM, Steyvers M. Model for recognition memory: REM: Retrieving effectively from memory. Psychonomic Bulletin & Review. 1997; 4:145–166. [PubMed: 21331823]

Strong EK. The effect of time-interval upon recognition memory. Psychological Review. 1913; 20:339–371.

Toglia, MP.; Battig, WF. Handbook of semantic word norms. Hillsdale, NJ: Erlbaum; 1978.

Tulving E. Memory and consciousness. Canadian Psychologist. 1985; 26:1–12.

Underwood BJ, Ekstrand BR, Keppel G. An analaysis of intralist similarity in verbal-learning with experiments on conceptual similarity. Journal of Verbal Learnin and Verbal Behavior. 1964; 4:447–462.

Waugh NC, Smith JA. A stochastic model for free recall. Psychometrika. 1962; 27:141–145.

Wixted JT. Dual-process theory and signal-detection theory of recognition memory. Psychological Review. 2007; 114:152–176. [PubMed: 17227185]

Yonelinas AP. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1994; 20:1341–1354.

Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language. 2002; 46:441–518.

Appendix

Consider a fixed-trials, no-criterion recall experiment of the form $S_1T_1T_2$, S_2T_3, S_3T_4; that is, there are three study cycles on the target list, with two separate test cycles following the first study cycle and one test cycle following each of the remaining study cycles. Because there are four separate recall tests, each item on the list can exhibit one of 16 different patterns of successes and errors over these tests: $C_1C_2C_3C_4$, $C_1C_2C_3E_4$, $C_1C_2E_3C_4$, …, $E_1E_2E_3E_4$, where C_i indicates that the item was recalled on the *i*th test and E_i indicates that it was not recalled. All of the dual-retrieval model's parameters (Table 1) can be estimated for any set of such data by (a) expressing the observed probability of each of the 16 error-success patterns as a function of these parameters and then (b) solving the set of 16 equations for those data. Concerning Step a, each of the 16 equations is just the expression that the Markov model in Equation 1 gives for one of the possible patterns of errors and success over the four tests. For instance, Equations A1 and A2 are merely the ways that the Markov model expresses the probability of successful recall on all tests (A1) and the probability of successful recall on the first three tests but not the fourth test (A2). In these expressions, it will be noted that six variables appear that are not parameters of the Markov model, as that model's parameter set appears in Equation 1 and Table 1—namely, the variables α, β, μ, λ, π, and θ. These variables are merely abbreviations for certain combinations of the parameters in Equation 1 that allow Equations A1–A16 to be expressed in more tractable form. The definitions of these variables, in terms of the model's parameter set, are $\alpha = R_1J_1(1 - f)J_{3C}$, $\beta = D_{3c} + (1 - D_{3c})J_{3C}D_{3c}$, $\mu = R_1(1 - J_1)$, $\lambda = R_1J_1(1 - f)(1 - J_{3C})$, $\pi = R_1J_1f$, and $\theta = D_{3c} + (1 - D_{3c})J_{3C}$.

Continuing with Step a, the 16 expressions are:

$$p(C_1C_2C_3C_4,)=D_1+\alpha\beta, \tag{A1}$$

$$p(C_1C_2C_3E_4,)=\alpha J_{3C}(1-D_{3c})^2(1-J_{3C}), \tag{A2}$$

$$p(C_1C_2E_3C_4,)=\alpha\theta(1-D_{3c})(1-J_{3C}), \tag{A3}$$

$$p(C_1E_2C_3C_4,)=\pi[D_2+R_2J_2D_{3c}+R_2J_2(1-D_{3c})J_{3C}]+\lambda[D_{3E}+(1-D_{3E})J_{3E}D_{3C}+(1-D_{3E})J_{3E}(1-D_{3c})J_{3C}], \tag{A4}$$

$$p(E_1C_2E_3C_4,)=\mu\beta(1-f)J_{3E}, \tag{A5}$$

$$p(C_1C_2E_3E_4,)=\alpha(1-D_{3c})(1-J_{3C})(1-D_{3E})(1-J_{3E}), \tag{A6}$$

$$p(C_1E_2C_3E_4,)=\pi R_2J_2(1-D_{3E})(1-J_{3E})+\lambda(1-J_{3C})(1-D_{3E})J_{3E}(1-D_{3c}), \tag{A7}$$

$$p(E_1C_2C_3E_4,)=\mu(1-f)(J_{3C})^2(1-D_{3c})^2(1-J_{3C}), \tag{A8}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$p(\mathrm{C}_1\mathrm{E}_2\mathrm{E}_3\mathrm{C}_4,)=\pi\{(1-D_2)(1-R_2)[D_2+R_2J_2)]+R_2(1-J_2)\theta\}+\lambda\theta(1-D_{3E})(1-J_{3E}), \quad (A9)$$

$$p(\mathrm{E}_1\mathrm{C}_2\mathrm{E}_3\mathrm{C}_4,)=\alpha\theta(1-f)J_{3E}(1-D_{3E})(1-J_{3E}), \quad (A10)$$

$$p(\mathrm{E}_1\mathrm{E}_2\mathrm{C}_3\mathrm{C}_4,)=[(1-D_1)(1-R_1)\{D_2+R_2J_2[D_{3c}+(1-D_{3c})J_{3C}]\}+\mu f\{D_2+R_2J_2[D_{3c}+(1-D_{3c})J_{3C}]\}+\mu(1-f)(1-J_{3E})[D_{3E}+(1-D_{3E})J_{3E}D_{3E}+(1-D_{3E})^2(J_{3E})^2] \quad (A11)$$

$$p(\mathrm{C}_1\mathrm{E}_2\mathrm{E}_3\mathrm{E}_4,)=\pi[((1-D_2)(1-R_2))^2+(1-D_2)(1-R_2)R_2(1-J_2)+R_2(1-J_2)(1-D_{3E})(1-J_{3E})]+R_1J_1(1-f)(1-D_{3E})(1-J \quad (A12)$$

$$p(\mathrm{E}_1\mathrm{C}_2\mathrm{E}_3\mathrm{E}_4,)=\mu(1-f)J_{3E}(1-D_{3c})(1-J_{3C})(1-D_{3E})(1-J_{3E}), \quad (A13)$$

$$p(\mathrm{E}_1\mathrm{E}_2\mathrm{C}_3\mathrm{E}_4,)=\mu(1-D_{3c})(1-J_{3C})[fR_2J_2+(1-f)(1-J_{3E})(1-D_{3E})J_{3E}]+(1-D_1)(1-R_1)R_2J_2(1-D_{3c})(1-J_{3C}), \quad (A14)$$

$$p(\mathrm{E}_1\mathrm{E}_2\mathrm{E}_3\mathrm{C}_4,)=(1-D_1)(1-R_1)(1-D_2)(1-R_2)[D_2+R_2J_2]+\theta(1-D_1)(1-R_1)R_2(1-J_2)+\mu f\{(1-D_2)(1-R_2)[D_2+R_2J_2]+R_2(1-J_2)\theta\}+\mu\theta(1-f)(1-D_{3E})(1-J_{3E})^2, \tag{A15}$$

$$p(\mathrm{E}_1\mathrm{E}_2\mathrm{E}_3\mathrm{E}_4,)=(1-D_1)(1-R_1)[((1-D_2)(1-R_2))^2+(1-D_2)(1-R_2)R_2(1-J_2)+R_2(1-J_2)(1-D_{3E})(1-J_{3E})]++\mu f[((1-D_2)(1-R_2))^2+(1-D_2)(1-R_2)R_2(1-J_2)+R_2(1-J_2)(1-D_{3E})(1-J_{3E})]+\mu(1-f)(1-D_{3E})(1-J_{3E})^3. \tag{A16}$$

The parameter *f*, which does not appear in Table 1, is a forgetting parameter that measures the probability that a item that has escaped from state *U* to state *P* on Test 1 falls back to state *U* on Test 2. Brainerd et al. (2009) found that in their large corpus of data sets, the value of *f* was always zero or very close to it. Hence, we ignore this parameter in the experiments that we report.

Concerning Step b, the likelihood of any sample of data (and estimates of the parameters in Table 1) is easily obtained by maximizing the following likelihood function, using a computer search program, such as GPT (Hu, 1998):

$$L_{11}=\Pi(p_i)^{N(i)}. \tag{A17}$$

NIH-PA Author Manuscript
NIH-PA Author Manuscript

The p_i are the 16 expressions on the right sides of Equations A1 – A16. The exponent of each p_i is an empirical data count that corresponds to one of the 16 possible error-success sequences. Specifically, each exponent is just the total number of times that the indicated error-success sequence was observed in sample data. Because 11 parameters are estimated, the likelihood value in A17 is computed with 11 degrees of freedom. A goodness-of-fit test that evaluates the null hypothesis that no more than two retrieval processes are required to account for the data is then obtained in the usual way by computing a likelihood ratio statistic that compares the likelihood in A17 to the likelihood of the same data when all 15 observable probabilities are free to vary. That test statistic, which is asymptotically distributed as $\chi^2(4)$, is

$$G^2 = -2\ln[L_{11}/L_{15}], \qquad \text{(A18)}$$

where L_{15} is the likelihood of the data when all 15 observable probabilities are free to vary.

A second goodness-of-fit test can also be computed that evaluates the null hypothesis that no more than one retrieval process is required to account for the data—specifically, that learning to recall involves only a single nonrecollective process. This one-process model is obtained from the two-process model by merely eliminating one of the Markov states, as follows:

$$W' = [L(1), P_E(1), P_C(1)] = [D'_1, (1-D'_1)(1-J'_1), (1-D'_1)J'_1];$$

$$M = \begin{array}{llll} & L(n+1) & P_E(n+1) & P_C(n+1) \\ L(n) & 1 & 0 & 0 \\ P_E(n) & D'_{3E} & (1-D'_{3E})(1-J'_{3E}) & (1-D'_{3E})J'_{3E} \\ P_E(n) & D'_{3C} & (1-D'_{3C})(1-J'_{3C}) & (1-D'_{3C})J'_{3C} \end{array}. \qquad \text{(A19)}$$

The likelihood of any set of data over which A17 can be defined (i.e., the data of a no-criterion experiment of the form $S_1T_1T_2$, S_2T_3, S_3T_4) can also be estimated for the one-process model in A19 by maximizing the following likelihood function:

$$
\begin{aligned}
L_6 = & [D'_1 + (1 - D'_1)(1 - J'_1)J'_{3C}D'_{3C}/(1 - (1 - D'_{3C})J'_{3C})]^{N[Q(1)Q(2)]} \\
& \times [(1 - D'_1)J'_1J'_{3E}(D'_{3C}/(1 - (1 - D'_{3C})J'_{3C}))]^{N[R(1)Q(2)]} \\
& \times [(1 - D'_1)J'_1(1 - J'_{3E})]^{N[R(1)R(2)]} \\
& \times [(1 - D'_1)J'_1J'_{3E}(1 - D'_{3C})(1 - J'_{3C})/(1 - (1 - D'_{3C})J'_{3C})]^{N[R(1)S(2)]} \\
& \times [(1 - D'_1)(1 - J'_1)(1 - J'_{3C})]^{N[S(1)R(2)]} \\
& \times [(1 - D'_1)(1 - J'_1)J'_{3C}(1 - D'_{3C})(1 - J'_{3C})/(1 - (1 - D'_{3C})J'_{3C})]^{N[S(1)S(2)]} \\
& \times [D'_{3E} + (1 - D'_{3E})J'_{3E}D'_{3C}/(1 - (1 - D'_{3C})J'_{3C})]^{N(R,Q)} \\
& \times [1 - D'_{3E} - (1 - D'_{3E})J'_{3E}D'_{3C}/(1 - (1 - D'_{3C})J'_{3C})]^{N[R,R]+N[R,S]} \\
& \times [(1 - J'_{3E})(1 - (1 - D'_{3C})J'_{3C})/((1 - J'_{3E})(1 - (1 - D'_{3C})J'_{3C}) + (1 - D'_{3C})(1 - J'_{3C})J'_{3E})]^{N[R,R]} \\
& \times [1 - (1 - J'_{3E})(1 - (1 - D'_{3C})J'_{3C})/((1 - J'_{3E})(1 - (1 - D'_{3C})J'_{3C}) + (1 - D'_{3C})(1 - J'_{3C})J'_{3E})]^{N[R,S]} \\
& \times [1 - (1 - D'_{3C})J'_{3C}]^{N[S,R]} \\
& \times [(1 - D'_{3C})J'_{3C}]^{N[S,S]}.
\end{aligned}
\tag{A20}
$$

The exponent of each term of A20 is an empirical data count that corresponds to possible pairings on consecutive test cycles of three observable states: Q = the state on all recall tests after the last error for items with one or more errors and the state on all recall tests for items with no errors; R = the state on all recall tests for an item when the item is not recalled; S = the state on all recall tests for an item when the item is recalled on that test but is not recalled on at least one later recall test. The first six terms in A20 range over the starting vector, and the remaining terms range over the transition matrix. The fit statistic for the one-process model is

$$G^2 = -2\ln[L_\theta/L_{15}], \tag{A21}$$

which is asymptotically distributed as $\chi^2(9)$. L_{15} is the likelihood in the numerator of A18 (i.e., the likelihood when all 15 observable probabilities are free to vary), and L_6 is the likelihood of the same data under the assumption that learning to recall involves only a nonrecollective process (i.e., the model in A19).

A final goodness-of-fit test can be computed, which also the evaluates the null hypothesis that no more than one retrieval process is required to account for the data but, unlike A21, assumes that the process is recollective rather than nonrecollective retrieval. This recollective one-process model, like the nonrecollective one-process model, is obtained from the two-process model by merely eliminating one of the Markov states, as follows:

$$W'' = [L(1), U(1)] = [D'_1, (1 - D'_1)];$$
$$M'' = \begin{array}{c|cc} & L(n+1) & U(n+1) \\ \hline L(n) & 1 & 0 \\ U(n) & D'_2 & 1 - D'_2 \end{array}. \tag{A22}$$

The likelihood of any set of data over which A17 can be defined can also be estimated for the one-process model in A22 by maximizing the following likelihood function:

$$L_2 = (D'_1)^{N[C(1)]} \times (1 - D'_1)^{N[E(1)]} \times (D'_2)^{N[C,C]} \times (1 - D'_2)^{N[E,E]}. \tag{A23}$$

The exponent of each term of A23 is an empirical data count that corresponds to possible pairings on consecutive test cycles of two observable states: C = the state on all recall tests when an item is recalled and E = the state on all recall tests when an item is not recalled. The first two terms in A23 range over the starting vector, and the last two terms range over the transition matrix. The comparative model fit statistic for one versus two retrieval processes is

$$G^2 = -2\ln[L_2 / L_{15}], \tag{A24}$$

which is asymptotically distributed as $\chi^2(13)$. L_{15} is the likelihood in the numerator of A18 and A21, and L_2 is the likelihood of the same data under the assumption that learning to recall involves only a recollective process.

Note that if the fit statistic in A21 produces a null hypothesis rejection (showing that recall involves more than a single nonrecollective process), it is unnecessary to then compute the fit test in A24 (to determine whether recall involves a single recollective process). That is because the recollective one-process model (Equation A22) is obviously a submodel of the nonrecollective one-process model (Equation A19), which results by imposing the constraint $J'_1 = J'_{2C} = J'_{2E} = 0$ on A19 and setting $D'_2 = D'_{3E}$. Thus, for any set of sample data, A24 must produce a null hypothesis rejection if A21 produces a null hypothesis rejection.

Table 1

Parameters that Measure Recollective and Nonrecollective Retrieval in the Dual-Retrieval Model

Parameter	Definition
	Recollective retrieval
D_1	For an item that can neither be directly accessed nor reconstructed, D_1 is the probability that it becomes directly accessible on S_1.
D_2	For an item that can neither be directly accessed nor reconstructed, D_2 is the probability that it becomes directly accessible on any study cycle after S_1.
D_{3C}	If an item is reconstructed on any trial and familiarity judgment outputs it, D_{3C} is the probability it becomes directly accessible on the next study cycle.
D_{3E}	If an item is reconstructed on any trial and familiarity judgment does *not* output it, D_{3E} is the probability it becomes directly accessible on the next study cycle.
	Nonrecollective retrieval
Reconstruction:	
R_1	For an item that can be neither directly accessed nor reconstructed, R_1 is the probability that it becomes reconstructable on S_1.
R_2	For an item that can be neither directly accessed nor reconstructed, R_2 is the probability that it becomes reconstructable on any study cycle after S_1.
Familiarity:	
J_1	For an item that becomes reconstructable on S_1, J_1 is the probability that familiarity judgment passes it on for output.
J_2	For an item that becomes reconstructable on any study cycle after S_2, J_2 is the probability that familiarity judgment passes it on for output.
J_{3C}	For an item that is reconstructed on two consecutive recall tests T_i and T_{i+1}, J_{3C} is the probability that familiarity judgment outputs it on T_{i+1} if it was output on T_i.
J_{3EC}	For an item that is reconstructed on two consecutive recall tests T_i and T_{i+1}, J_{3E} is the probability that familiarity judgment outputs it on T_{i+1} if it was *not* output on T_i.

Note. The dual-retrieval model also contains an eleventh parameter, *f*, that measures forgetting between T_1 and T_2. As the value of this forgetting parameter is normally zero when the subjects are young adults (for a review, see Brainerd et al.,2009), we ignore it in the present research (see Appendix).

Table 2

Parameter Estimates for Experiment 1

Parameter	List type	
	Associative recall	**Free recall**
Recollective retrieval:		
D_1	.12	.11
D_2	.08	.04
$\boldsymbol{M_{D1/D2}}$	**.10**	**.08**
D_{3C}	.50	.16
D_{3E}	.53	.19
$\boldsymbol{M_{D3C/D3E}}$	**.52**	**.18**
Reconstructive retrieval:		
R_1	.24	.44
R_2	.28	.64
$\boldsymbol{M_{R1/R2}}$	**.26**	**.54**
Familiarity judgment:		
J_1	.60	.69
J_2	.43	.54
$\boldsymbol{M_{J1/J2}}$	**.52**	**.62**
J_{3C}	.71.	.73
J_{3E}	.44	.63
$\boldsymbol{M_{J3C/J3E}}$	**.58**	**.68**

Table 3

Observed and Predicted Distributions of Total Number of Errors and Trial Number of the Last Error

Condition/statistic	Probability				
	0	1	2	3	4
Experiment 1					
Associative recall:					
TE					
Observed	.189	.069	.168	.183	.391
Predicted	.189	.069	.168	,183	.391
TLE					
Observed	.189	.029	.174	.191	.417
Predicted	.189	.029	.177	.187	.418
Free recall:					
TE					
Observed	.234	.149	.236	.228	.153
Predicted	.234	.149	.236	.228	.153
TLE					
Observed	.234	.006	.259	.234	.267
Predicted	.234	.006	.260	.232	.268
Experiment 2					
Unrelated list:					
TE					
Observed	.248	.121	.265	.199	.167
Predicted	.248	.121	.265	.199	.167
TLE					
Observed	.248	.014	.284	.202	.252
Predicted	.248	.014	.284	.203	.251
Three category lists:					
TE					
Observed	.250	.155	.246	.239	.110
Predicted	.250	.155	.246	,239	.110

Condition/statistic	Probability				
	0	1	2	3	4
TLE					
Observed	.250	.022	.271	.218	.239
Predicted	.250	.042	.253	.216	.239
Six category lists:					
TE					
Observed	.269	.174	.311	.176	.070
Predicted	.269	.179	.305	.177	.070
TLE					
Observed	.269	.024	.327	.206	.174
Predicted	.269	.024	.333	.193	.181
Experiment 3					
UU:					
TE					
Observed	.327	.136	.315	.170	.052
Predicted	.327	.138	.313	.167	.055
TLE					
Observed	.327	.065	.298	.184	.126
Predicted	.327	.065	.317	.161	.130
UC:					
TE					
Observed	.263	.113	.277	.218	.129
Predicted	.263	.115	.275	.218	.129
TLE					
Observed	.263	.022	.238	.197	.280
Predicted	.263	.022	.239	.195	.281
CU:					
TE					
Observed	.196	.077	.266	.235	.226
Predicted	.196	.077	.266	.235	.226
TLE					

Condition/statistic	Probability				
	0	1	2	3	4
Observed	.196	.031	.257	.236	.280
Predicted	.196	.031	.259	.233	.281
CC:					
TE					
Observed	.108	.041	.170	.250	.431
Predicted	.108	.047	.159	.255	.431
TLE					
Observed	.108	.008	.132	.181	.571
Predicted	.108	.008	.132	.180	.572

Note. TE = the total number of errors, for each item (0–4 range). TLE = the trial number (0–4 range) on which the last error occurred, for each item.

Note. TE = total number errors per item. TLE = trial number of the last error per item.

Table 4

Parameter Estimates for Experiment 2

Parameter	List type		
	Unrelated	**3 Categories**	**6 Categories**
Recollective retrieval:			
D_1	.20	.15	.03
D_2	.24	.00	.09
$\boldsymbol{M_{D1/D2}}$	**.22**	**.08**	**.06**
D_{3C}	.17	.22	.23
D_{3E}	.61	.56	.40
$\boldsymbol{M_{D3C/D3E}}$	**.39**	**.39**	**.32**
Reconstructive retrieval:			
R_1	.31	.39	.51
R_2	.32	.65	.66
$\boldsymbol{M_{R1/R2}}$	**.32**	**.52**	**.59**
Familiarity judgment:			
J_1	.75	.81	.89
J_2	.81	.75	.86
$\boldsymbol{M_{J1/J2}}$	**.78**	**.78**	**.88**
J_{3C}	.54	.64	.77
J_{3E}	.50	.53	.64
$\boldsymbol{M_{J3C/J3E}}$	**.52**	**.59**	**.71**

Table 5

Parameter Estimates for Experiment 3

Parameter	List type			
	Unrelated-Unrelated	**Unrelated-Categorized**	**Categorized-Unrelated**	**Categorized-Categorized**
Recollective retrieval:				
D_1	.25	.22	.16	.08
D_2	.38	.21	.22	.02
$\boldsymbol{M_{D1/D2}}$	**.32**	**.22**	**.19**	**.05**
D_{3C}	.36	.11	.24	.13
D_{3E}	.34	.35	.50	.19
$\boldsymbol{M_{D3C/D3E}}$	**.35**	**.23**	**.37**	**.16**
Reconstructive retrieval:				
R_1	.49	.58	.23	.28
R_2	.98	.83	.39	.47
$\boldsymbol{M_{R1/R2}}$	**.74**	**.71**	**.31**	**.38**
Familiarity judgment:				
J_1	.50	.48	.49	.50
J_2	.36	.40	.43	.45
$\boldsymbol{M_{J1/J2}}$	**.43**	**.44**	**.46**	**.48**
J_{3C}	.62	.52	.62	.50
J_{3E}	.54	.26	.56	.16
$\boldsymbol{M_{J3c/J3E}}$	**.58**	**.39**	**.59**	**.33**