Published in final edited form as:

J Math Psychol. 2014 April 1; 59: 50–64. doi:10.1016/j.jmp.2013.07.003.

Markovian Interpretations of Dual Retrieval Processes

C. F. A. Gomes,

Department of Human Development, Cornell University, Ithaca, NY 14853, telephone: 607-793-6099, cf365@cornell.edu

C. J. Brainerd,

Department of Human Development, Cornell University, Ithaca, NY 14853, telephone: 607-255-2592, cb299@cornell.edu

K. Nakamura, and

Department of Human Development, Cornell University, Ithaca, NY 14853, telephone: 607-220-8324, kn252@cornell.edu

V. F. Reyna

Department of Human Development, Cornell University, Ithaca, NY 14853, telephone: 607-255-6778, vr53@cornell.edu

Abstract

A half-century ago, at the dawn of the all-or-none learning era, Estes showed that finite Markov chains supply a tractable, comprehensive framework for discrete-change data of the sort that he envisioned for shifts in conditioning states in stimulus sampling theory. Shortly thereafter, such data rapidly accumulated in many spheres of human learning and animal conditioning, and Estes' work stimulated vigorous development of Markov models to handle them. A key outcome was that the data of the workhorse paradigms of episodic memory, recognition and recall, proved to be one- and two-stage Markovian, respectively, to close approximations. Subsequently, Markov modeling of recognition and recall all but disappeared from the literature, but it is now reemerging in the wake of dual-process conceptions of episodic memory. In recall, in particular, Markov models are being used to measure two retrieval operations (direct access and reconstruction) and a slave familiarity operation. In the present paper, we develop this family of models and present the requisite machinery for fit evaluation and significance testing. Results are reviewed from selected experiments in which the recall models were used to understand dual memory processes.

Keywords

Markov chains; recollection; reconstruction; familiarity; fuzzy-trace theory

© 2013 Elsevier Inc. All rights reserved.

Correspondence to: C. J. Brainerd.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

In 1959, after spending a summer studying with J. Kemeny, a mathematician whose specialty was Markov chains (e.g., Kemeny & Snell, 1959), Estes published a comprehensive treatment of that topic for learning and memory researchers. The proximal reason for that essay was to expand the mathematical foundations of stimulus sampling theory to deal with what, at the time, was known as the small N problem. In stimulus sampling theory, a set of N stimulus elements is randomly sampled on each trial of a learning experiment, and following reinforcement or feedback or a study trial, the elements change their conditioning states in an all- or-none fashion. As long as N was assumed to be large, as it traditionally had been, the law of large numbers applied, and the ordinary mathematics of probability theory could be used to derive quantitative predictions from stimulus sampling theory. However, there were certain types of experiments, such as discrimination learning, for which it seemed plausible that N must be small—so that ordinary probability theory could no longer be used. In such circumstances, predictions could be derived by implementing stimulus sampling theory in finite Markov chains, and Estes developed the necessary mathematical machinery to accomplish that. During the ensuing years, those developments were widely used by other authors (e.g., Bower, 1961; Greeno & Steiner, 1964; Restle, 1962; Waugh & Smith, 1962), but not to derive predictions from stimulus sampling theory. Instead, finite Markov chains provided a natural framework for handling the all-or-none learning demonstrations that began to appear in the literature at about the same time as Estes' essay (e.g., Estes, 1960; Rock, 1958; Trabasso, 1963).

All-or-none learning refers to discrete changes in empirical response probabilities, rather than in theoretical entities, such as stimulus elements. Specifically, at the level of individual items, the probability of a correct response only takes on a small set of values, such as 0, some constant $0 < p < 1$, and 1. In the span of a few years, experiments in which subjects received multiple opportunities to study a focal list and to retrieve it produced data of that sort. Some tasks, such as associative recognition (Bower, 1961) and concept identification (Restle, 1962; Trabasso, 1963), were found to be one-stage Markovian; that is, variability in the probability of a correct response consisted of an initial unlearned state U, in which the probability was some constant $0 < p < 1$, and a terminal learned state L, in which the probability was 1. Other tasks, such as associative recall (Greeno, 1968) and free recall (Waugh & Smith, 1962), were found to be two-stage Markovian; that is, variability in the probability of a correct response consisted of an initial unlearned state U, in which the probability was 0, a partially learned state P, in which the probability was some constant $0 < p < 1$, and a terminal learned state L, in which the probability was 1. Using the developments in Estes (1959), it was a simple matter to derive the sampling distributions of a plethora of statistics for these models (e.g., trial number of the last error, trial number of first success, and auto-correlation of errors) and to use them to conduct rigorous quantitative evaluations of one- and two-stage Markov models for the indicated paradigms. Consequently, Markov chains dominated mathematical modeling work on basic memory paradigms for some years (for reviews, see Brainerd, Howe, & Desrochers, 1982; Greeno, 1974). Ultimately, they faded from view as more complex modeling efforts were mounted, with the appearance of global matching models of memory being a prominent case in point (for a review, see Clark & Gronlund, 1996).

Recently, however, there has been a resurgence of Markovian analyses of memory, for reasons not unlike those that motivated Estes' (1960) essay—namely, the need for a mathematical framework that will help us gain leverage on important theoretical problems. In this instance, the theoretical motivation is supplied by dual-process conceptions of episodic memory. According to such conceptions (e.g., Jacoby, 1991; Mandler, 1980), two forms of retrieval control memory for the events of our lives. The first is a recollective form, in which retrieval of episodic memories is accompanied by realistic phenomenology—vivid restatement of events' prior occurrences and details of the contexts in which events were experienced. The second is a nonrecollective form, which confers high confidence that the events were experienced but is not accompanied by realistic phenomenology. The distinction between recollective and nonrecollective forms of retrieval has been exploited, particularly over the last four decades, to explain memory deficits in patients with brain lesions (Schacter & Tulving, 1994) and neurodegenerative diseases (Bugajska, Morson, Moulin, & Souchay, 2011), brain activity during encoding and retrieval (Diana, Yonelinas, Ranganath, 2007; Ranganath, 2010), memory development (Gomes & Brainerd, 2013; Reyna & Mills 2007), and the effects of various experimental variables on memory, such as list length (Cary & Reder, 2003), word frequency (Gardiner & Java, 1990; Guttentag & Carroll, 1997), time of testing (Knowlton & Squire, 1995; Tulving, 1985), and subjects' level of attention during study (Gardiner & Parkin, 1990; Yonelinas, 2001), among others (for a review, see Yonelinas 2002).

Dual-process distinctions have been studied primarily with old/new item recognition. To do that, old/new recognition has been enriched with metacognitive tasks that supposedly disentangle its recollective and nonrecollective components. Remember/know judgments (Tulving, 1985) are far and away the most commonly used procedure, but confidence judgments, which allow recollective and nonrecollective components of the receiver operating characteristic (ROC) to be estimated (Yonelinas, 1994), and inclusion versus exclusion instructions, which allow the recollective and nonrecollective parameters of the process-dissociation model (Jacoby, 1991) to be estimated, have also been used. There are serious validity challenges to these recognition methodologies, however, and that has led to renewed interest in Markov models of *recall* as a response to the challenges—all of which forms the substance of the present paper.

The paper consists of four main major sections. The first sketches traditional recognition-based approaches to measuring dual memory processes, summarizes recent validity criticisms of those approaches, and discusses the need for alternative measurement procedures that respond to the criticisms. The second section presents a recall-based response that we have implemented in recent experimentation. This approach uses fuzzy-trace theory's (FTT) distinction between verbatim and gist traces of experience (e.g., Brainerd & Reyna, 2004; Reyna & Brainerd, 2011) to interpret the states of two-stage Markov chains in such a way that measures of dual memory processes can be extracted from raw recall data, without the need of additional metacognitive tasks. In the third section, the mathematical machinery for this family of models (matrix representations and fit tests) is developed. In the fourth section, two worked illustrations of how these techniques can be applied experimentally to answer theoretical questions about episodic memory are

presented. The first focuses on how factor analysis of parameter spaces over large numbers of data sets can be used to test the theoretical assumptions that lie behind this family of models. The second shows how the model can be extended to study dual memory processes in an important domain in which virtually no data on them are currently available, namely child-to-adult development.

Dual-Process Conceptions of Recognition and Recall

That episodic memory is controlled by two distinct forms of retrieval, recollective and nonrecollective, is, of course, one of the touchstones of contemporary memory research. Strong (1913) is usually credited with being the first to provide an experimental demonstration of this distinction, but current interest can be traced to a theoretical paper by Mandler (1980). Strong reported a surprisingly modern recognition experiment, in which his subjects studied a list of target words and then responded to an old/new recognition test. After that, subjects were asked to review the items that they had accepted as old, both hits and false alarms, to introspect on the phenomenology that each provoked, and to write a brief description of that phenomenology. Despite great variability in the specifics of what subjects wrote, Strong noticed that the responses were of two basic sorts. Sometimes, old responses were justified by describing realistic details that had been associated with their presentation (e.g., that a subject's stomach had growled or that a subject had remembered a dental appointment), whereas at other times, old responses were justified by assertions of confidence that were not backed up by realistic details (e.g., "just know it was on the list," "must have seen it"). Strong also noticed that the first type of report, now called recollection, was overwhelmingly stimulated by hits, whereas false alarms primarily stimulated the second type of report, now called familiarity. In his later paper, Mandler reviewed experimental demonstrations of the recollection-familiarity distinction and provided compelling anecdotal illustrations, such as the well-known butcher on the bus example. He also developed a simple multinomial model of how recollection and familiarity affect hit rates, in which their effects were jointly independent and additive.

The modern literature on recollection and familiarity preserves two key features of Strong's (1913) original demonstration: Old/new recognition is the focal memory paradigm, and the two forms of retrieval are measured by asking subjects to introspect on the phenomenological qualities of remembering (called metacognitive judgment nowadays). Tulving's (1985) remember/know procedure is the most popular form of metacognitive judgment. For instance, when we recently reviewed the published literature on this procedure, we found that it contained over 1,000 data sets. Two other procedures that have been extensively used are Jacoby's (1991) process-dissociation methodology and Yonelinas' (1994) dual-process ROC methodology. Together, these procedures have spawned vast behavioral and neuroscience literatures on recollection and familiarity. The former has been focused on identifying manipulations that dissociate the two forms of retrieval (e.g., word frequency, testing delays; see Yonelinas, 2002), while the latter has been focused on identifying different brain regions that are active during one versus the other (e.g., the perirhinal cortex versus the parahippocampal cortex; see Ranganath, 2010). The distinction between recollection and familiarity is also integral to contemporary explanations of

memory declines in healthy aging and neurocognitive impairment (e.g., Reyna & Mills 2007; Schacter & Slotnick, 2004).

Of late, however, these standard methodologies have been the focus of mounting validity criticisms. Beginning with an article by Donaldson (1996), these criticisms are extensive for the most widely used methodology, remember/know judgments (Dunn, 2004, 2008; Rotello, Macmillan, & Reeder, 2004). After reviewing several remember/know experiments, Donaldson concluded that the types of effects that were thought to support a dual-process view of retrieval could be handled by a one-process signal detection model. The model's key feature was to posit that old/new responses are governed by two decision criteria, a conservative one for remember judgments and a liberal one for know judgments. Dunn (2004) independently reviewed a large set of remember/know experiments, using them to evaluate five arguments for rejecting one-process signal detection models. He showed that each argument was flawed and concluded that the remember/know data in his sample of experiments provided no grounds for rejecting one-process models. Later, Dunn (2008) demonstrated that if two retrieval processes control old/new recognition, remember/know data should be bidimensional when decision criteria are held constant, but in fact, they are one dimensional.

Major critiques of the other two methodologies, process dissociation and ROC, have also appeared. Concerning process dissociation, for example, Ratcliff, Van Zandt, and McKoon (1995) noted that if this procedure actually separates two distinct retrieval operations, it should not be possible to recover estimates of its recollection and familiarity parameters from a one-process model. They conducted a simulation analysis of reported estimates of the procedure's recollection and familiarity parameters, using a well-known one-process model (Gillund & Shiffrin, 1984) and found that the prediction was wrong; estimates of the recollection and familiarity parameters were recovered. Turning to the dual-process ROC procedure, Wixted (2007) compared its fits to those of a signal detection model in which the noise and noise + signal theoretical distributions have unequal variances (Glanzer, Kim, Hilford, & Adams, 1999). Although the unequal variance signal detection model is compatible with a dual-process account in which the contributions of recollection and familiarity are combined into a unidimensional strength variable, it is also compatible with simpler, single-process accounts. However, on the grounds of parsimony, the latter (single-process) interpretation is preferred over the former (dual-process) as the reported ROC analyses did not differentiate between the two interpretations. Wixted found that whenever fit was different for the dual-process and unequal variance models, it was better for the unequal variance model, thus indicating that ROC analyses are compatible with single-process accounts.

A more recent criticism of all of these procedures centers on their inapplicability to certain subject populations that are of great interest in memory research—particularly, normal children and older adults with neurocognitive impairment (Brainerd et al., 2009). In contrast to the vast number of remember/know experiments in the main stream adult literature, only a single study with normal children has been published (Billingsley, Smith, & McAndrews, 2002) and only a handful of studies with impaired older adults (Bugaiska, Morson, Moulin, & Souchay, 2011). Likewise, virtually no research using the other two procedures has been

reported with such subjects. The dearth of research is due to the fact that remember/know, process dissociation, and ROC are high-burden methodologies that are beyond the capabilities of most subjects in these populations. With remember/know, for instance, subjects must read and comprehend complex instructions telling them how to introspect on their mental states, and they must implement those instructions accurately on recognition tests. Ghetti, Mirandola, Angelini, Cornoldi, and Ciaramelli (2008; cited in Ghetti, 2008) found that children could not follow even simplified versions of these instructions, and similar problems have been reported in studies of impaired older adults (Brainerd, Aydin, & Reyna, 2012).

In a seminal review, Malmberg (2008) surveyed the bulk of the dual process recognition literature. He concluded that the enterprise of separating dual-retrieval processes with old/new item recognition had misfired, that standard procedures for effecting such separation are demonstrably unreliable, and that contrary to Mandler's (1980) proposals, item recognition is controlled by a single familiarity process. Such conclusions leave us with a major methodological uncertainty: If the conventional tactic of enriching item recognition with metacognitive judgments fails to separate the recollective and nonrecollective forms of retrieval, how then can they be disentangled and quantified?

A solution that we have explored (Brainerd et al., 2009; Brainerd & Reyna, 2010; Brainerd et al., 2012) is to switch to recall-based measurement, which has two advantages relative to recognition. First and foremost, the uncertainty about whether and under what conditions single or dual processes control retrieval is avoided. Unlike recognition, it has long been known that dual-process models deliver better fits than one-process models to the data of associative, free, and cued recall experiments (Bower & Theios, 1964; Greeno, 1968; Greeno, James, & DaPolito, 1971; Kintsch & Morris, 1965). Second, the high-burden problem with metacognitive judgments is also avoided, allowing research on dual retrieval processes to be extended to populations with cognitive limitations. In that connection, as we shall see, the recall solution extracts measurements of recollective and nonrecollective retrieval *directly* from the error-success data of recall tests, without metacognitive intervention. Consequently, it can be applied to any population that performs recall tests reliably, which includes young children and older adults with neurocognitive impairments. With children, there is a long-standing developmental literature in which associative, free, and cued recall have been used to chart normal memory development, from the preschool years onward (for reviews, see Bjorklund & Muir, 1988; Schneider & Pressley, 1997). With older adults, recall tests are standard fare in the neuropsychological batteries that are used to diagnose cognitive impairment and dementia (e.g., Langa et al., 2005; Morris et al., 1989), and indeed, performance on such tests is the best single predictor of these diagnoses (Petersen et al., 1999).

Dual-Process Theory of Recall

The error-success data of recall paradigms are well fit by a two-stage Markov model whose general form was first developed by Greeno (1968). Over trials, the level of recall for an individual list item is described by three discrete performance states, with all-or-none escape from the first and second states—hence, “two-stage.” The probability of successful recall of

an item is effectively zero before the first study cycle. This is the initial no-recall state U. After the first study cycle or any subsequent study cycle, an item in state U can remain there, can escape to the second partial-recall state P, in which the probability of successful recall is some constant $0 < p < 1$, or it can escape to the third perfect-recall state L, where the probability of successful recall is 1. After the first study cycle or any subsequent cycle, an item in state P can remain there or can escape to state L.

As memory models go, a rather minimal assumption is all that is required to use two-stage Markov chains to measure dual retrieval processes—namely, that these processes can be mapped with occupancy of the two states in which the probability of successful recall is greater than 0. The specific assumption (Brainerd et al., 2009) is that recollective and nonrecollective retrieval can be mapped with occupancy of L and P, respectively. That interpretation was based on FTT's conception of the different types of memory traces that support recollective versus nonrecollective retrieval. In that conception, recollective retrieval is supported by verbatim traces whose content can be directly accessed and simply read out of consciousness on a recall test, whereas reconstructive retrieval processes traces of partial-identifying information about items (usually called gist traces). The more detailed theoretical interpretations are as follows.

In state L recollective retrieval is referred to as “direct access” because this retrieval operation accesses verbatim traces of a list items' prior presentation without searching through traces of other items. Consequently, it is the faster of the two operations. It is also the more accurate of the two because it supports errorless recall. Here, the idea is that when an item's verbatim trace is directly accessed, its surface form and contextual details of its presentation are consciously reinstated, so that the item can be recalled by merely reading it out of consciousness. The fact that an item's surface form is accompanied by vivid, realistic contextual details is what makes direct access a recollective operation, of course. Although direct access has speed and accuracy advantages, it has two limitations. The first is that during the initial phases of learning, verbatim traces are quite sensitive to sources of interference that are generated by recall tests, and the second is that because of such sensitivity, direct access rapidly degrades over the course of a recall test (Barnhardt, Choi, Gerkens, & Smith, 2006; Brainerd & Reyna, 1993). This means that recall cannot rely exclusively on direct access until learning is so far advanced that the verbatim traces that have been stored are robust enough to survive the interference that recall tests generate (Brainerd et al., 2009). Such reliance is not even an option in most recall experiments, which usually involve only a single study-test trial.

Turning to recall in state P, reconstruction is a nonrecollective operation and is accompanied by a slave judgment operation that performs familiarity checks on reconstructed items before outputting them. Reconstruction avoids the interference obstacle to direct access by regenerating list items from stable episodic traces of partial-identifying information, particularly semantic information (e.g., “clothing” for *shirt*). Such information is still available when verbatim traces are not, such as at the end of free recall protocols (Brainerd & Reyna, 1993), when output interference is maximal (Roediger & Schmidt, 1980), and after long intervals between study and test (Brainerd, Reyna, & Brandse, 1995; Toglia, Neuschatz, & Goodwin, 1999). That subjects are able to reconstruct items from partial-

identifying information is a fact, not a hypothesis, that is well established in the literatures on the tip-of-the-tongue phenomenon and feelings of knowing. In both literatures, subjects have been found to access various forms of partial information about list items and then process it in order to regenerate the items themselves (see Brown & McNeill, 1966; Hicks & Marsh, 2002; Koriat, 1993, 1995; Schacter & Worling, 1985). With respect to semantic features in particular, Koriat, Levy-Sadot, Edry, and Marcas (2003) found that subjects could access the semantic features in Osgood's (1952) model of meaning before recalling the corresponding items. In the present conception, reconstruction uses partial-identifying features to generate sets of matching candidate items (e.g., *shirt, coat, shoes, dress*) that are small enough to be processed rapidly on a recall test. Because such sets will normally include non-targets (*coat, shoes, dress*), a judgment operation performs pre-output familiarity checks on reconstructed items, in order to avoid high intrusion rates. Familiarity judgment is therefore a slave operation because it only operates on reconstructed items. Note that this is how the recall model incorporates the familiarity operation of dual-process theories of recognition.

Although nonrecollective retrieval avoids the interference problem that besets direct access, it poses a problem of its own. It is error-prone. On analogy to the familiarity operation of recognition theories, which permits false alarms, non-targets will sometimes survive the familiarity check and be passed on for recall. It follows that the advantage of recollection (errorless recall) ameliorates the problem posed by reconstruction, and the advantage of reconstruction (interference resistance) ameliorates the problem posed by recollection. Relying on both retrieval operations, then, increases net recall by compensating to some extent for each operation's limitations.

Markov Models of Dual Processes in Recall

Our dual-retrieval model posits that the probability of correctly recalling a target can be expressed as a function of recollective (direct access, D) and nonrecollective (reconstruction, R , and familiarity judgment, J) parameters. Specifically, in a one-trial study-test design, a target will occupy one of the following four states after study and immediately prior to recall: (a) the recollective state L , with probability D ; (b) a nonrecollective sub-state P_E in which it will *not* be recalled, with probability $(1 - D)R(1 - J)$; (c) a nonrecollective sub-state P_C in which it *will* be recalled, with probability $(1 - D)RJ$; and (d) the nonrecollectable and nonreconstructable state U , with probability $(1 - D)(1 - R)$. Therefore, the contribution of recollective and nonrecollective retrieval operations to correct recall, $P(Rc)$, is the probability of a target occupying either the first or third states, which is

$$P(Rc) = D + (1 - D)RJ. \quad (1)$$

As $P(Rc)$ is the only empirical degree of freedom, the model's parameters are not identifiable, but they can be rendered so by shifting from one-trial to multi-trial designs (Brainerd et al., 2009; Brainerd & Reyna, 2010), as follows.

Consider any standard recall paradigm (e.g., associative, free, cued, and serial) in which subjects are exposed to multiple successive study-test trials, and whose responses on recall

tests are either correct (1) or not (0). Over k successive trials, each target generates one out of 2^k possible sequences of observable recall patterns (e.g., 111, 110, ..., 000), and consequently, across all targets there will be a frequency distribution over recall patterns. Changes in recall over trials can be represented as transitions through a discrete state space, in which finite Markov chains provide a natural formalism for describing them by assuming that (a) learning over trials consists of making transitions through a finite set of memory states $\psi_1, \dots, \psi_s \in \Psi$ and (b) the state a target occupies on trial n (for $n = 1, \dots, k$) depends only on the state it occupied on trial $n - 1$. In the dual-retrieval model, there are four exhaustive and mutually exclusive memory states, denoted $L, P_E, P_C, U \in \Psi$. State U is the initial unlearned state in which neither gist traces that support successful reconstruction nor verbatim traces that support recollection can yet be retrieved, and thus a target occupying this state has probability 0 of being recalled. State P is the transient partially learned state, with sub-states P_E and P_C , in which gist traces that support successful reconstruction can be retrieved. Gist traces, however, support imperfect recall of targets, as familiarity judgment might fail to output list items that are successfully reconstructed, and it might authorize output of reconstructed items that are not actual list items. Therefore, a target occupying state P has average probability $0 < p < 1$ of being recalled. State L is the terminal absorbing learned state (i.e., once entered, an item cannot leave *as long as study-test trials continue*) in which the verbatim traces that support recollection can be retrieved, and a target is recalled with probability 1.

The transition diagram in panel A of Figure 1 is the model's depiction of learning in a recall paradigm. Learning to retrieve a target on recall tests consists of escaping from state U to the absorbing state L by either making a $U \rightarrow L$ transition or by first making a $U \rightarrow P$ transition and then a $P \rightarrow L$ transition. This class of two-stage absorbing Markov chain contrasts with the one-stage absorbing chains in panels B and C of Figure 1, which are used to assess the relative fit of simpler conceptions of recall in experimental applications (e.g., Brainerd et al., 2012). In panel B, learning to retrieve a target on recall tests consists of escaping from state U to the absorbing state L , while in panel C, learning to retrieve a target on recall tests consists of escaping from state P to the absorbing state L . It is the type of two-stage chain depicted in panel A, however, that provides close descriptions of the fine-grained structure of the recall data produced by children and adolescents (Brainerd et al., 1990, 2009; Reyna & Brainerd, 1991), younger adults (Brainerd, Howe, & Desrochers, 1982), and older adults (Batchelder et al., 1997; Howe & Hunter, 1985, 1986; Brainerd et al., 2012).

A Markov chain composed of the above states can be represented by a unit row vector

$\mathbf{w}^{(1)} = [w_j^{(1)}]_{1 \times 4}$, whose entries give the starting (unconditional) probabilities of each state, and a transition matrix $\mathbf{M} = [m_{ij}]_{4 \times 4}$, whose entries give the (conditional) probability of transitioning from state i on trial $n - 1$ to state j on trial n , as follows

$$\mathbf{w}^{(1)} = [P(L(1)), P(P_E(1)), P(P_C(1)), P(U(1))] = [w_1^{(1)}, w_2^{(1)}, w_3^{(1)}, w_4^{(1)}], \quad (2A)$$

$$\mathbf{M} = \begin{bmatrix} P(L(n)|L(n-1)) & P(P_E(n)|L(n-1)) & P(P_C(n)|L(n-1)) & P(U(n)|L(n-1)) & m_{11} & m_{12} & m_{13} & m_{14} \\ P(L(n)|P_E(n-1)) & P(P_E(n)|P_E(n-1)) & P(P_C(n)|P_E(n-1)) & P(U(n)|P_E(n-1)) & m_{21} & m_{22} & m_{23} & m_{24} \\ P(L(n)|P_C(n-1)) & P(P_E(n)|P_C(n-1)) & P(P_C(n)|P_C(n-1)) & P(U(n)|P_C(n-1)) & m_{31} & m_{32} & m_{33} & m_{34} \\ P(L(n)|U(n-1)) & P(P_E(n)|U(n-1)) & P(P_C(n)|U(n-1)) & P(U(n)|U(n-1)) & m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix}. \quad (2)$$

Because L is an absorbing state, Equation 2B can be simplified as

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \mathbf{a} & \mathbf{T} \end{bmatrix}, \quad (3)$$

in which $\mathbf{a}$ is the 3×1 column subvector of $\mathbf{M}$, whose entries give the probability of entering the absorbing state L on trial n from transient states on trial $n - 1$, and $\mathbf{T}$ is the 3×3 submatrix of $\mathbf{M}$, whose entries give the probabilities of making transitions through transient states between trials $n - 1$ and n when a transition to the absorbing state L does not occur. This indicates that, although many parameterizations of the entries of the transition matrix are possible (see Brainerd et al., 2010, 2012), direct access ought to be entries of the $\mathbf{a}$ subvector, while reconstruction and familiarity judgment ought to be entries of the $\mathbf{T}$ submatrix. Figure 1 illustrates how each cell of the transition matrix in Equation 3 relates to the transition diagrams posited by two- (panel A, Figure 1) and one-stage models (panels B and C, Figure 1).

Let $\mathbf{w}^{(n)} = [w_j^{(n)}]_{1 \times 4}$ be a unit row vector whose first entry gives the probability of a target occupying the recollective state L on the n^{th} trial, $P(L(n)) = w_1^{(n)}$, whose second entry gives the probability of a target occupying the state P_E on the n^{th} trial, $P(P_E(n)) = w_2^{(n)}$, whose third entry gives the probability of a target occupying the nonrecollective state P_C on the n^{th} trial, $P(P_C(n)) = w_3^{(n)}$, and whose fourth entry gives the probability of a target occupying the state U on the n^{th} trial, $P(U(n)) = w_4^{(n)}$. Then, for a canonical study-test design of form $S_1T_1 \dots S_kT_k$, in which S denotes a study cycle and T denotes are recall test, $\mathbf{w}^{(n)}$ can be expressed in terms of $\mathbf{w}^{(1)}$ and $\mathbf{M}$ as follows

$$\mathbf{w}^{(n)} = \mathbf{w}^{(1)} \times \mathbf{M}^{n-1}, \quad (4A)$$

whose entries are given by the equation

$$w_j^{(n)} = \begin{cases} w_j^{(1)}, & n=1 \\ \sum_{y=1}^4 w_y^{(n-1)} m_{yj}, & n>1 \end{cases}. \quad (4B)$$

Finally, the probability of correctly recalling a target on the n^{th} trial, $P_n(Rc)$, is given by the proportion of targets occupying either the recollection state L or the reconstruction state P_C on trial n and, therefore, it can be expressed as the inner product between $\mathbf{w}^{(n)}$ and the vector $\mathbf{c} = [1, 0, 1, 0]$ as follows

$$P_n(Rc) = \langle \mathbf{c}, \mathbf{w}^{(n)} \rangle = w_1^{(n)} + w_3^{(n)} = \begin{cases} w_1^{(1)} + w_3^{(n)}, & n=1 \\ \sum_{y=1}^4 w_y^{(n-1)} (m_{y1} + m_{y3}), & n>1 \end{cases} = \begin{cases} w_1^{(1)} + w_3^{(n)}, & n=1 \\ w_1^{(n-1)} + \sum_{y=2}^4 w_y^{(n-1)} (m_{y1} + m_{y3}), & n>1 \end{cases} \quad (5)$$

Its negation, $P_n(\neg Rc)$, is then given by the proportion of targets occupying either state P_E or state U on trial n , or the inner product between $\mathbf{w}^{(n)}$ and the vector $\mathbf{e} = [0, 1, 0, 1]$.

Both the one- and two-stage models depicted in Figure 1 are hidden Markov models (HMM) that describe a doubly embedded stochastic process; that is, a hidden stochastic process, called the theoretical-states model, which is only observed via a second set of stochastic process, called the observable-states model. It is the HMM that generates the sequences of recall patterns over trials (e.g., Brainerd et al., 1982; Greeno, Millward, & Merryman, 1971). In the two-stage model, for instance, more than one theoretical state is identified with correct recall—namely, L and P_C . However, a target that enters the theoretical state L on trial n can be identified with an observable state that *does not* produce an error thereafter, whereas a target that enters the theoretical state P_C on trial n can be identified with an observable state that *may* produce an error thereafter because state P is error-prone.

Next, we present six parameterizations of the starting vector and transition matrix, all of which have been used in recent research applications (Brainerd et al., 2009, 2010, 2012; Gomes, Brainerd, & Stein, 2013). Each delivers identifiable measures of direct access, reconstruction, and familiarity judgment. Such parameterizations depend on (a) the canonical and non-canonical forms of study-test designs and (b) assumptions about learning and forgetting. Following the six parameterizations, we present the statistical machinery for evaluating goodness of fit and computing maximum likelihood estimates of the parameters of each model.

Canonical Study-Test Designs

A canonical design has the form $S_1T_1 \dots S_kT_k$. A canonical design in which subjects receive a minimum of 3 study-test trials (e.g., $S_1T_1 S_2T_2 S_3T_3$) has been shown to identify 6 theoretical parameters of the dual-retrieval model (Brainerd et al., 2009, 2012; Gomes et al., 2013) by either assuming that (a) learning after the first trial only occurs on errors, (b) learning after the first trial only occurs on successes, or (c) learning after the first trial occurs both on errors and successes. All three model versions share the common starting vector

$$\mathbf{w}^{(1)} = [D_I, (1 - D_I)R_I(1 - J_I), (1 - D_I)R_IJ_I, (1 - D_I)(1 - R_I)], \quad (6A)$$

and therefore, all three posit that the probability of correctly recalling a target on the first trial (the special case $n = 1$) is $P_I(Rc) \langle \mathbf{c}, \mathbf{w}^{(1)} \rangle = D_I + (1 - D_I)R_IJ_I$, as in Equation 1. However, the three model versions differ with respect to the entries of the $\mathbf{a}$ subvector of the transition matrix $\mathbf{M}$. Specifically, let $\mathbf{M}_E$ be the transition matrix of the learning on errors model with $\mathbf{a}_E$ subvector, let $\mathbf{M}_S$ be the transition matrix of the learning on successes model with $\mathbf{a}_S$ subvector, and let $\mathbf{M}_B$ be the transition matrix of the learning on both errors and successes model with $\mathbf{a}_B$ subvector such that

$$\mathbf{M}_E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ D_2 & (1-D_2)(1-J_2) & (1-D_2)J_2 & 0 \\ 0 & (1-J_2) & J_2 & 0 \\ D_2 & (1-D_2)R_2(1-J_2) & (1-D_2)R_2J_2 & (1-D_2)(1-R_2) \end{bmatrix}, \quad (6B)$$

$$\mathbf{M}_S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-J_2) & J_2 & 0 \\ D_2 & (1-D_2)(1-J_2) & (1-D_2)J_2 & 0 \\ 0 & R_2(1-J_2) & R_2J_2 & (1-R_2) \end{bmatrix}, \quad (6C)$$

$$\mathbf{M}_B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ D_2 & (1-D_2)(1-J_2) & (1-D_2)J_2 & 0 \\ D_2 & (1-D_2)(1-J_2) & (1-D_2)J_2 & 0 \\ D_2 & (1-D_2)R_2(1-J_2) & (1-D_2)R_2J_2 & (1-D_2)(1-R_2) \end{bmatrix}, \quad (6D)$$

and thus, $\mathbf{a}_B = \mathbf{a}_E + \mathbf{a}_S$. Therefore, in the learning on errors version, a target can only make a transition to the recollective state L on trial n ($n > 1$) when it was not recalled on trial $n - 1$; that is, it occupies either state U or P_E on trial $n - 1$. Conversely, in the learning on successes version, a target can only make a transition to the recollective state L on trial n ($n > 1$) when it was recalled on trial $n - 1$; that is, when it occupies the nonrecollective state P_C on trial $n - 1$. This key difference between the two model versions can be shown via algebraic manipulation of (4B) with the entries of the absorbing state L, which in both models is the parameter D_2 . For the learning on errors version, D_2 can be described as a measure of the rate of change in the proportion of targets recalled recollectively between trials n and $n - 1$, denoted ${}_{(n,n-1)}w_I$, normalized by the proportion of targets that were not recalled on trial $n - 1$,

$$w_1^{(n)} = \sum_{y=1}^4 w_y^{(n-1)} m_{y1} = w_1^{(n-1)} + w_2^{(n-1)} D_2 + w_4^{(n-1)} D_2 \quad D_2 = \frac{w_1^{(n)} - w_1^{(n-1)}}{w_2^{(n-1)} + w_4^{(n-1)}} = \frac{\Delta_{(n,n-1)} w_1}{1 - P_{n-1}(Rc)}. \quad (7A)$$

For the learning on successes version, D_2 can be described as a measure of the rate of change in the proportion of targets recalled recollectively between trials n and $n - 1$, normalized by the proportion of targets that were recalled nonrecollectively on trial $n - 1$,

$$w_1^{(n)} = \sum_{y=1}^4 w_y^{(n-1)} m_{y1} = w_1^{(n-1)} + w_3^{(n-1)} D_2 \quad D_2 = \frac{\Delta_{(n,n-1)} W_1}{P_{n-1}(Rc) - w_1^{(n-1)}}. \quad (7B)$$

Brainerd et al. (2012) found that under the learning on errors version, changes in nonrecollective retrieval over trials can often be accounted for by inter-trial variability in familiarity judgment rather than reconstruction, which imposes two constraints on this model: (a) $R_I = R_2$, and (b) J depends on the trial number. More specifically, such alternative learning on errors model can be expressed by the row vector $\mathbf{w}_{E'}^{(1)}$ and transition matrix $\mathbf{M}_{E'}$ as follows

$$\mathbf{w}_{E'}^{(1)} = [D_1, (1 - D_1)R(1 - J_1), (1 - D_1)RJ_1, (1 - D_1)(1 - R)], \quad (8A)$$

$$\mathbf{M}_{E'} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ D_2 & (1 - D_2)(1 - J_n) & (1 - D_2)J_n & 0 \\ 0 & (1 - J_n) & J_n & 0 \\ D_2 & (1 - D_2)R_2(1 - J_n) & (1 - D_2)RJ_n & (1 - D_2)(1 - R) \end{bmatrix}. \quad (8B)$$

Non-Canonical Study-Test Designs

A non-canonical design is one in which subjects receive either more than one recall test between study cycles, or more than one study cycle between recall tests, or both. A particularly important non-canonical design is the one of form $S_1T_1T_2S_2T_3 \dots S_{k-1}T_k$ ($k \geq 4$). The importance of this design lies in the fact that it has been shown to identify 11 theoretical parameters of the dual-retrieval model (Brainerd et al., 2009, 2010) by adding an intermediate transition matrix $\mathbf{M}_f = [m_{fij}]_{4 \times 4}$ between the recall tests T_1 and T_2 that allows targets that escaped from state U to state P before T_1 to fall back to state U before T_2 . This matrix is not present on subsequent trials because the design reverts to canonical form after T_2 , and consequently, a single transition matrix $\mathbf{M}_C = [m_{cij}]_{4 \times 4}$ describes learning thereafter. The dual-retrieval model for the non-canonical design of form $S_1T_1T_2S_2T_3 \dots S_{k-1}T_k$ can be expressed by the row vector $\mathbf{w}^{(1)}$ and transition matrices $\mathbf{M}_f$ and $\mathbf{M}_C$ as follows

$$\mathbf{w}^{(1)} = [D_1, (1 - D_1)R_1(1 - J_1), (1 - D_1)R_1J_1, (1 - D_1)(1 - R_1)], \quad (9A)$$

$$\mathbf{M}_f = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & (1 - J_{3E})(1 - f) & J_{3E}(1 - f) & f \\ 0 & (1 - J_{3C})(1 - f) & J_{3C}(1 - f) & f \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (9B)$$

$$\mathbf{M}_C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ D_{3E} & (1 - D_{3E})(1 - J_{3E}) & (1 - D_{3E})J_{3E} & 0 \\ D_{3C} & (1 - D_{3C})(1 - J_{3C}) & (1 - D_{3C})J_{3C} & 0 \\ D_2 & (1 - D_2)R_2(1 - J_2) & (1 - D_2)R_2J_2 & (1 - D_2)(1 - R_2) \end{bmatrix}. \quad (9C)$$

For this model version, Equations 4 and 5 do not apply because by design, the Markov chain needs an additional transition matrix ($\mathbf{M}_f$) to describe transitions between the first two recall tests that are different than the transitions on subsequent canonical trials ($\mathbf{M}_C$). However, it is possible to revise Equations 4 and 5 to include the additional transition matrix, as follows

$$\mathbf{w}^{(n)} = \begin{cases} \mathbf{w}^{(1)}, & n=1 \\ \mathbf{w}^{(1)} \times \mathbf{M}_f \times \mathbf{M}_C^{n-2}, & n>1 \end{cases}, \quad (10A)$$

$$w_j^{(n)} = \begin{cases} w_j^{(1)}, & n=1 \\ \sum_{y=1}^4 w_y^{(1)} m_{f_{yj}}, & n=2 \\ \sum_{y=1}^4 w_y^{(n-1)} m_{c_{yj}}, & n>2 \end{cases}, \quad (10B)$$

$$P_n(Rc) = \begin{cases} w_1^{(1)} + w_3^{(1)}, & n=1 \\ w_1^{(1)} + \sum_{y=2}^3 w_y^{(1)} m_{f_{y3}}, & n=2 \\ w_1^{(n-1)} + \sum_{y=2}^4 w_y^{(n-1)} (m_{c_{y1}} + m_{c_{y3}}), & n>2 \end{cases}. \quad (11)$$

As before, Equation 1 is a special case of Equation 9 when $n = 1$. However, when $n = 2$, the model assumes that two situations can occur between T_1 and T_2 that cause a target that was recalled on T_1 to not be recalled on T_2 . First, a target that occupied state P_C on T_1 might fall back to state U on T_2 , with probability f . (In theoretical terms, the subject forgets how to reconstruct the item.) Second, a target that occupied state P_C on T_1 and does not fall back to state U on T_2 might occupy state P_E on T_2 , with probability $(1 - J_{3C})(1 - f)$. (In theoretical terms, the subject reconstructs the item on both trials, and it passes the familiarity check on T_1 but fails it on T_2 .) In addition, the model assumes that a situation can occur between T_1 and T_2 that causes a target that was not recalled on T_1 to be recalled on T_2 . A target that occupied state P_E on T_1 that does not fall back to state U on T_2 might occupy P_C on T_2 , with probability $J_{3E}(1 - f)$. (In theoretical terms, the subject reconstructs the item on both trials, and it fails the familiarity check on T_1 but passes it on T_2 .)

The non-canonical design $S_1T_1T_2 S_2T_3 \dots S_{k-1}T_k$ is theoretically richer than canonical designs because it allows the fitting of all five Markov chains to the same data space, whereas the non-canonical dual-retrieval model cannot be fit to canonical designs owing to parameter nonidentifiability (see Brainerd et al., 2009). More explicitly, Markov chains defined over canonical designs can be fit to the $S_1T_1T_2 S_2T_3 \dots S_{k-1}T_k$ design by separating it into two sequences of recall tests—namely $T_1T_3 \dots T_k$ and $T_2T_3 \dots T_k$ —thereby generating two sets of parameter estimates that will differ if subjects tend to forget targets between T_1 and T_2 . Later, we exploit this feature of non-canonical designs to investigate relations among the parameters of all five models.

Forgetting of Dual Retrieval Processes

Although forgetting is a classic topic in memory research, it is studied far less often than how we successfully retrieve past events. It turns out that the present model provides a rather seamless approach to theoretical interpretation of findings from both types of research. In the dual-retrieval models for canonical designs, forgetting between trials only occurs when a target that occupied state P_C on trial $n - 1$ makes a transition to state P_E on trial n . In other words, although the target is reconstructed on both trials, familiarity judgment authorizes its output on the previous trial but fails to authorize it on the current

trial. In the dual-retrieval model for the non-canonical design, an additional form of forgetting is incorporated between T_1 and T_2 —namely that a target that occupied state P_C on trial $n - 1$ might *fall back* to state U on trial n because the subject forgets how to reconstruct it. However, such forgetting has proved to be rare among healthy subjects, as the parameter f usually does not differ significantly from 0 in the $S_1T_1T_2S_2T_3 \dots S_{k-1}T_k$ design (Brainerd et al., 2009). Perhaps this is because there is only a short interval between T_1 and T_2 . Increasing the interval would be problematical because that might cause violations of the model's assumption about the absorbing feature of state L (i.e., targets that enter state L on T_1 might fall back if there is a long interval between T_1 and T_2 .) Therefore, the non-canonical $S_1T_1T_2S_2T_3 \dots S_{k-1}T_k$ design is not a suitable design to study forgetting of dual retrieval processes.

However, a design of form $S_1T_1 \dots S_{k-1}T_{k-1} I T_k$ ($k \geq 3$), in which I denotes a retention interval between two consecutive tests (e.g., an hour, a day, a week), has been shown to allow measurement of recollective forgetting and reconstructive forgetting, in addition to forgetting due to inter-trial transitions within state P , by dropping the absorbing features of the Markov chain between trials $k - 1$ and k (Brainerd et al., 2012). Recollective forgetting consists of a process in which targets that could be directly accessed in the past cannot be accessed any longer, whereas reconstructive forgetting consists of a process in which targets that could be reconstructed from partially-identifying information (e.g., semantic representations) in the past cannot be reconstructed any longer. In the aforementioned design, the former forgetting process occurs when a target occupying state L on trial $k - 1$ falls back to state U or state P on trial k (controlled by the parameter f_D), while the latter forgetting process occurs when a target occupying state P on trial $k - 1$ falls back to state U on trial k (controlled by the parameter f_R). Such a model can be expressed by the following starting vector $\mathbf{w}^{(1)}$ and transition matrices $\mathbf{M}_L$ and $\mathbf{M}_F$:

$$\mathbf{w}^{(1)} = [D, (1 - D)R(1 - J_1), (1 - D)RJ_1, (1 - D)(1 - R)], \quad (12A)$$

$$\mathbf{M}_L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ D & (1 - D)(1 - J_2) & (1 - D)J_2 & 0 \\ 0 & (1 - J_2) & J_2 & 0 \\ D & (1 - D)R(1 - J_2) & (1 - D)RJ_2 & (1 - D)(1 - R) \end{bmatrix}, \quad (12B)$$

$$\mathbf{M}_F = \begin{bmatrix} 1 - f_D & 0 & 0 & f_D \\ 0 & (1 - J_2)(1 - f_R) & J_2(1 - f_R) & f_R \\ 0 & (1 - J_2)(1 - f_R) & J_2(1 - f_R) & f_R \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (12C)$$

in which $\mathbf{M}_L$ controls transitions for trials $1 < n < k$, the learning trials, and $\mathbf{M}_F$ controls transitions for the trial $n = k$, the forgetting trial. Suppose, for example, that a subject receives a design of the form $S_1T_1S_2T_2IT_3$ ($k = 3$) and correctly recalls 59%, 65%, and 50% of the targets on the first, second, and third tests, respectively. Prior to application of the Markov chain in Equation 12, the process loci of learning on trials 1 and 2, and of forgetting between trials 2 and 3, are uncertain. However, a model in which the maximum

likelihood estimates of the parameters D , R , J_1 , J_2 , f_D , and f_R are .38, .64, .52, .28, .27, and .01, respectively, provides descriptions of the recall probabilities produced by the subject, and indicates that forgetting between trials 2 and 3, in particular, was chiefly due to recollective forgetting, as $f_D > f_R$. According to such estimates, recollective recall increased from 38% on the first trial to 53% on the second, and then decreased to 39% on the third trial, which indicates recollective forgetting between the second and third trials. Nonrecollective recall, however, decreased from 21% on the first trial to 12% on the second, and then remained at roughly the same level (11%) on the third trial, which indicates near absence of nonrecollective forgetting between the second and third trials.

Alternatively, consider the following experiment, which has two sessions. In the first session, subjects receive a study-test design with a minimum of 3 trials, such as the canonical design $S_1T_1 S_2T_2 S_3T_3$. For this session, any of the dual-retrieval models for canonical designs (Equations 6 and 8) can be fit to the data and will produce sets of parameter estimates that will describe changes in recall from T_1 to T_2 to T_3 . In the second session, which takes place after a retention interval I (e.g., $I = \text{a day}$), subjects receive a minimum of 3 consecutive recall tests *without further opportunities to study the target list*, such as the design of form $T_4 T_5 T_6$. As before, any of the dual-retrieval models for canonical designs can be fit to the data of this second session and will produce a different set of parameter estimates that will describe changes in recall from T_4 to T_5 to T_6 . Therefore, measuring the two forms of forgetting with such delayed repeated recall procedure consists of comparing the proportion of targets occupying state L or state P immediately *prior* to the retention interval with the proportion of targets occupying the respective states immediately *after* the retention interval. More specifically, for this example, recollective forgetting can be expressed as $w_1^{(3)} - w_1^{(4)}$, while reconstructive forgetting can be expressed as $(w_2^{(3)} + w_3^{(3)}) - (w_2^{(4)} + w_3^{(4)})$. If we were to fit the learning on errors model for canonical designs to the data generated in the second session, with parameters in the second session denoted by an additional f subscript, then such forgetting equations would reduce to $w_1^{(3)} - D_{1f}$ and $(w_2^{(3)} + w_3^{(3)}) - (1 - D_{1f})R_{1f}$, respectively. Consistent with prior research (e.g., Gardiner & Java, 1991; Sharot & Yonelinas, 2008), Brainerd et al. (2012) found that the memory traces that support the recollective form of retrieval are more susceptible to forgetting than the traces that support the nonrecollective form.

In addition to forgetting, notice that the delayed repeated recall procedure allows investigation of two forms of spontaneous, test-induced recovery of retrieval ability (reminiscence; Payne, 1987) by inspection of the parameters in the transition matrix of the model version that is fit to the data of the second session. This analysis taps the process loci of testing effects (e.g., Roediger & Karpicke, 2006). In the canonical design, learning can occur due to additional opportunities to study and to retrieve the focal list and, consequently, testing effects are confounded with study effects. In the non-canonical design, testing effects are separated from study effects between the first and second opportunities to retrieve the focal list, and confounded thereafter, as the design becomes canonical. In prior investigations (Brainerd et al., 2009, 2010, 2012; Gomes et al., 2013), however, recall has rarely differed between the first two tests in the non-canonical design, and when it does,

which usually only occurs in clinical samples, recall is lower on the second test than on the first test. Contrary to such designs, the delayed repeated recall procedure disentangles testing from study effects. For the learning on errors model, for instance, recollective reminiscence is measured with D_{2f} , while nonrecollective reminiscence is measured with R_{2f} and J_{2f} . Data on the relative magnitudes of these forms of reminiscence were reported by Brainerd et al. (2012) for associative recall tasks. Surprisingly, recollective retrieval, rather than nonrecollective retrieval, was subject to reminiscence on delayed associative recall tests, suggesting that repeated testing in delayed recall contributes to the restoration of items' verbatim traces rather than providing practice at reconstructing items from partial-identifying information.

Parameter Estimation and Goodness of Fit

Identifiable estimates of the parameters of all five dual-retrieval models can be obtained by maximizing the likelihood function

$$L_j = \prod_i P(i)^{N(i)}, \quad (13)$$

in which j is the number of free parameters in the model and $P(i)$ and $N(i)$ are, respectively, the probability posited by the model and the empirical data count of the i^{th} recall pattern over 2^k trials. A goodness of fit test is then obtained by computing a standard likelihood ratio statistic, G^2 , which compares Equation 13 to the posteriori probability of the sample data when the observable probabilities of all the recall patterns are free to vary, L_{2^k-1} , as follows

$$G^2 = -2 \ln \left(\frac{L_j}{L_{2^k-1}} \right). \quad (14)$$

Because G^2 is asymptotically distributed as χ^2 (Cressie & Read, 1984), with $df = 2^k - (1 + j)$, goodness of fit can be evaluated by comparing the observed to the critical $\chi^2(df)$ value. The statistic in Equation 14 evaluates the null hypothesis that a more complex model is not required to account for sample data (i.e., the dual-retrieval model is sufficient).

It is also necessary to evaluate the null hypothesis that a simpler model is adequate to account for sample data (i.e., the dual-retrieval model is not necessary). Here, one conducts fit tests for the simpler one-stage absorbing chains depicted in Figure 1, panels B and C. These one-stage models are derived from the dual-retrieval model by removing either the state P (panel B) or the state U (panel C). For instance, the one-stage version depicted in panel B of Figure 1 of the dual-retrieval models for canonical designs is derived by removing the state P as follows

$$\mathbf{w}^{(1)} = [P(L(1)), P(U(1))] = [D_1, (1 - D_1)], \quad (15A)$$

$$\mathbf{M}'' = \begin{bmatrix} P(L(n)|L(n-1)) & P(U(n)|L(n-1)) \\ P(L(n)|U(n-1)) & P(U(n)|U(n-1)) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ D_2 & (1 - D_2) \end{bmatrix}. \quad (15B)$$

Such a model has two free parameters and, consequently, the critical value to reject the null hypothesis of fit for is a $\chi^2(2^k - 3)$ value. Of course, one-stage models have fewer free parameters than the dual-retrieval model and are preferred on grounds of parsimony if they pass fit tests.

Two Applications: Factor Analyses and Developmental Analyses

Now, we turn to worked applications of the dual-retrieval model to recall data, in order to illustrate how it can be exploited to answer theoretical questions. This section consists of two subsections. In the first, we present findings from factor analyses of a large corpus of data sets from published and unpublished recall experiments that have accumulated since the early 1980s. In the second section, we sketch some applications of the model to the study of memory development between early childhood and adolescence, which illustrate how the present approach overcomes the high-burden problem with traditional metacognitive judgment methods of measuring dual processes.

Factor Structure of the Model's Parameter Space

Background—As noted, the non-canonical design $S_1T_1T_2 S_2T_3 \dots S_{k-1}T_k$ allows one to estimate parameters of the five versions (parameterizations) of the dual-retrieval model with the same data. Although there are conceptual distinctions among them (e.g., whether learning occurs on errors or successes), the two retrieval processes and the slave judgment process are the same. Previous studies in which such models were implemented (Brainerd et al., 2009, 2010, 2012; Gomes et al., 2013) have shown that D parameters are more sensitive than R or J parameters to manipulations that should affect recollective retrieval (e.g., list length, associative vs. free recall), while R and J parameters are more sensitive than D parameters to manipulations that ought to affect reconstruction or familiarity (e.g., unrelated vs. categorized lists, neutral vs. negative emotionally-valenced lists).

We expanded upon these traditional investigations by conducting a factor analysis of the parameter spaces of all five versions of the dual-retrieval model using a large corpus of multi-trial recall data. The corpus consists of 230 data sets from both published (206 data sets) and unpublished (24 data sets) developmental studies in which the $S_1T_1T_2 S_2T_3 \dots S_{k-1}T_k$ non-canonical design was used. The published data sets appeared in Brainerd (1985), Brainerd and Howe (1982), Brainerd, Howe, and Desrocher (1982), Brainerd, Howe, and Kingma (1982, 1984), Brainerd, Kingma, and Howe (1985, 1986), Brainerd, Reyna, Howe, and Kingma (1990), Howe (1988, 1990), Howe, Brainerd, and Kingma (1985a, 1985b), Howe and Hunter (1985, 1986), and Howe, O'Sullivan, Brainerd, and Kingma (1989). Depending upon the experiment, the subjects were children (5-, 7-, and 9-year-olds), adolescents (11- and 12-year-olds), younger adults (undergraduates), or older adults (65 years or older), but the majority of the data sets (79%) encompassed data from either children or adolescents. Subjects in those experiments performed free, associative, or cued recall tests on lists that ranged from 10 to 24 items. Study lists were either composed of familiar/unfamiliar words, abstract/concrete words, categorized/unrelated words, or unrelated pictures. In all experiments, each recall test was preceded by a brief buffer activity (e.g., 30 sec of letter shadowing) to minimize short-term memory effects, and subjects continued to receive study-test cycles until an errorless performance criterion had been

achieved (e.g., errorless recall of all targets across three consecutive tests). Maximum likelihood estimates of the parameters of each model version were obtained by maximizing Equation 13 with the EM algorithm (Hu & Batchelder, 1994). For the non-canonical model, the recall data consisted of the full sequence of $T_1 T_2 T_3 \dots T_k$ ($k = 4$) for each data set. For the canonical models, the recall data consisted of the two sequences, $T_1 T_3 T_4$ and $T_2 T_3 T_4$. However, in the non-canonical model, preliminary analysis showed that the parameter f , which measures forgetting between T_1 and T_2 , did not differ reliably from 0 in any data set, and consequently, it was set to 0 in all data sets. Similarly, in the canonical models, the parameter estimates for the two data sequences did not differ, and consequently, we computed a single set of parameter estimates for each model based on the pooled data from the two sequences.

Goodness of fit analyses indicated that the five models provided close descriptions of most data sets. For each data set, the critical value to reject the null hypothesis of fit is a $\chi^2(2^k - 11)$ value for the non-canonical model, and $\chi^2(1) = 3.84$ for the canonical models at the .05 level of confidence. The average of the observed $G^2(2^k - 11)$ statistic for the non-canonical model was roughly half the critical value to reject the null hypothesis of fit at the .05 level of confidence. Similarly, the canonical models fit most data sets well. The observed $G^2(1)$ values for one data set were 25 times the critical value, and it was removed from the subsequent analyses. For each canonical model, the distribution of observed $G^2(1)$ values across the remaining 229 data sets is shown in Figure 2. The mean $G^2(1)$ statistic was below the critical value for all models, namely the learning on errors model (panel A), $M = 3.15$, the alternative learning on errors model (panel B), $M = 1.00$, the learning on successes model (panel C), $M = 2.13$, and the learning on both errors and successes model (panel D), $M = 2.35$, and the respective proportions of the data sets in which the null hypothesis of fit could not be rejected were 80%, 94%, 86%, and 87%.

Factor analyses of these sets of parameter estimates supplied validity tests of a sort that is rarely reported for memory models. More particularly, the factor analyses evaluated FTT's mapping of different retrieval processes with the states of a two-stage Markov chain, across data sets that varied as to experimental conditions and subjects' memory abilities. Suppose that, indeed, recall involves two distinct retrieval operations plus a slave familiarity operation, that recollective retrieval controls recall in state L, and that reconstructive retrieval and familiarity judgment control recall in state P. If so, we expect that over data sets, a factor analysis of the parameter space of any of the five model versions should show that a three-factor solution accounts for most of the parameter variability. Another obvious validity prediction is that each factor ought to be readily interpretable as representing one of the three processes; that is, there should be a recollection factor (positive D loadings), a reconstruction factor (positive R loadings), and a familiarity judgment factor (positive J loadings).

Results—On the whole, the results were consistent with such predictions. We conducted a confirmatory factor analysis of the parameter space of each model version using the aforementioned data corpus. Principal components analysis was the extraction method and varimax was the criterion for rotation of the factor structure (see Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kaiser, 1958). If all five model versions measure three

processes (direct access, reconstruction, and familiarity judgment), a three-factor solution should account for the bulk of the variance in parameter estimates.¹ The results appear in Table 1, where the variance accounted for by each of the three factors and the parameters that loaded reliably on each are reported. Following the usual convention in factor analysis, only factor loadings $\geq .40$ were regarded as reliable.

Averaged across the model versions, a three-factor solution accounted for 81.0% of the variance in parameter estimates, as predicted. In comparison, a two-factor solution accounted for 18.0% less variance in parameter estimates, while a four-factor solution accounted only for 8.2% more variance in parameter estimates. Also as predicted, the overall pattern of factor loadings is congruent with the principle of separate recollection, reconstruction, and familiarity operations. Factor 1 is unambiguously a recollective retrieval factor. It is the only factor on which direct access parameters loaded positively, with the grand average loading across model versions being .83. In contrast, the R and J parameters never loaded reliably on Factor 1. Factor 2 can be interpreted as a reconstruction factor. In the four model versions with two R parameters, this is the only factor on which reconstruction on the first trial (R_1) ever loaded, and in the version with one R parameter, it is the only factor on which that parameter loaded. In the four model versions with two R parameters, reconstruction after the first trial (R_2) sometimes loaded positively on this factor and sometimes loaded positively on Factor 3. In addition, J parameters sometimes loaded positively on this factor, though they sometimes loaded negatively as well. Finally, Factor 3 can be interpreted as a familiarity judgment factor. Except for one of the five model versions, the J parameters always loaded positively on this factor. The D parameters never loaded positively on Factor 3, but the R_2 parameter sometimes loaded positively on this factor.

In general, parameters that control transitions from the no-recall state U to the correct-recall states P_C and L , denoted $U \rightarrow \{P_C, L\}$, tended to be most strongly related to their respective factors. In the non-canonical model, for example, $U \rightarrow \{P_C, L\}$ transitions are controlled by the parameters D_1 , D_2 , R_1 , R_2 , J_1 , and J_2 . Here, direct access on and after the first trial (D_1 and D_2) only loaded positively on Factor 1, reconstruction on and after the first trial (R_1 and R_2) only loaded positively on Factor 2, and familiarity judgment on and after the first trial (J_1 and J_2) only loaded positively on Factor 3. A similar pattern was observed for the other models, although the specific parameters that control $U \rightarrow \{P_C, L\}$ transitions are different. In the canonical model versions in which learning occurs either on errors or on successes or on both errors and successes, the parameters D_1 , R_1 , and J_1 are linked to $U \rightarrow \{P_C, L\}$ transitions because prior to the first study trial, all targets can be assumed to occupy state U . Similarly, in the learning on errors canonical model in Equation 8, $U \rightarrow \{P_C, L\}$ transitions are linked to parameters D_1 , R , and J_1 .

¹Such a prediction does not assume that all model versions are equally true. It assumes, instead, that the parameters in all model versions overlap in their conceptual definitions—that is, regardless of model version, the parameters should measure three distinct operations, namely direct access, reconstruction, and familiarity judgment.

Development of Recollective and Nonrecollective Recall

Next, we consider an application of the dual-retrieval model to memory development during childhood and adolescence. As mentioned, a major gap in the literature is that virtually no findings on the early development of dual processes have been reported, owing to the high-burden aspect of paradigms such as remember/know, process dissociation, and dual-process ROC. Also as mentioned, the dual-retrieval model avoids this limitation by extracting measurements of recollective and nonrecollective processes directly from the error-success data of recall tests. We illustrate how this advantage can be exploited by using the model to analyze data from some experiments reported by Brainerd and Howe (1982).

In these experiments, the subjects were 5-year-olds, 7-year-olds, and 11-year-olds, who participated in a canonical design of form $S_1T_1 \dots S_kT_k$. For individual subjects, study-test trials continued until a stringent learning criterion of three consecutive perfect recall tests was met. For purposes of the present application, however, we constrained the analyses to $k = 3$. On study cycles, subjects were presented with pairs of line drawings of familiar concrete objects (e.g., book – apple), and the names of both objects were pronounced to control for developmental differences in naming ability. After each study cycle, subjects responded to an associative recall test, on which one of the drawings of each studied pair was presented (e.g., book – ?), and the task was to recall the second object. Recall accuracy on the first, second, and third trials was, respectively, 40%, 53%, and 62% for the 5-year-olds, 39%, 53%, and 62% for 7-year-olds, and 47%, 71%, and 83% for 11-year-olds. Therefore, on average, 11-year olds recalled more items across trials (67%) than both 7- (51%) and 5-year olds (52%), while the latter two groups did not differ.

We then conducted model analyses to track variability at the level of retrieval processes, in order to explain this age trend. Although any one of the canonical models in (6) can be used with these data, we restrict attention to the version with one R parameter and three J parameters to simplify the presentation and avoid redundancy in results. First, goodness of fit analyses were performed to evaluate whether the dual-retrieval model is (a) necessary and (b) sufficient to account for the recall data produced by subjects of all three age levels. The necessity tests evaluated the null hypothesis that the *one-stage* versions of the alternative learning on errors model in Equation 8, which are illustrated in panels B and C of Figure 1, could account for the data. For each age group, the relevant tests produced a G^2 statistic with 5 degrees of freedom for the model in panel B and 2 degrees of freedom for the model in panel C. Therefore, for the experiment as a whole, the critical value to reject the null hypothesis of fit was 25.00 and 12.53 for the models in panels B and C, respectively. In both cases, the null hypothesis of fit was rejected. The observed G^2 value was 575.63 ($p < .001$) for the model in panel B and 23.56 ($p < .001$) for the model in panel C. The sufficiency test then evaluated the null hypothesis that the dual-retrieval model could account for the data. That test produced a G^2 statistic with 1 degree of freedom for each age group and, therefore, the critical value to reject the null hypothesis of fit was 7.81 for the experiment as a whole. The observed G^2 value, .62, was well below the critical value. Therefore, the fit analyses indicated that the dual-retrieval model was necessary and sufficient to account for the recall data produced by the subjects in this experiment.

Next, we computed maximum likelihood estimates of model parameters to pinpoint the process level differences among the age groups. Estimates of each parameter are reported by age level in Table 2. We first tested the omnibus null hypothesis that there was no reliable age variability in parameter values. This test is a G^2 statistic with 12 degrees of freedom, and thus, the critical value to reject the null hypothesis of age invariance in retrieval operations is 21.03. The observed G^2 statistic, 150.52 ($p < .001$), was roughly seven times the critical value, and hence, some of the parameters varied reliably with age. Visual inspection of Table 2 suggests that this variability was confined to 11-year-olds versus 5- and 7-year-olds. Such observation was confirmed when we tested the null hypothesis of no variability in parameter values between each pair of age groups. For each of the three possible age pairings, the relevant test is a G^2 statistic with 6 degrees of freedom and a critical value of 12.53. As anticipated, the null hypothesis was rejected for 11- versus 7-year-olds, $G^2(6) = 111.32$, and 11- versus 5-year-olds, $G^2(6) = 118.13$, but not for 7- versus 5-year-olds, $G^2(6) = 1.69$. Age comparisons for individual parameters were therefore restricted to 11-year-olds versus 5- and 7-year-olds.

Inspection of the parameter estimates in Table 2 suggests that there were developmental improvements in both recollection and reconstruction parameters, but statistical tests showed that only the former were reliable. The average value of the two direct access parameters increased from .29 among 5- and 7-year-olds to .43 among 11-year-olds. For each of these parameters, we tested the null hypothesis that the estimate was equal between age groups via a standard likelihood ratio test. The relevant test is a G^2 statistic with 1 degree of freedom for each pairwise age comparison, and thus, the critical value to reject the null hypothesis of invariance is 3.84. For the comparison between 11- and 5-year-olds, the null hypothesis was rejected for D_1 , $G^2(1) = 5.33$, but not for D_2 , $G^2(1) = 2.76$. Similarly, for the comparison between 11- and 7-year-olds, the null hypothesis was rejected for D_1 , $G^2(1) = 7.31$, but not for D_2 , $G^2(1) = 2.62$.

In contrast, the corresponding tests for the R and J parameters did not produce any null hypothesis rejections. Further analysis suggested that this may be due to floor effects in nonrecollective recall at all age levels, an analysis that was motivated by Brainerd and Reyna's (2010) finding that adults can exhibit rather low levels of nonrecollective recall with associative recall. When the parameter estimates in Table 2 are inserted in Equation 8, recall can be separated into its recollective and nonrecollective components via Equation 4 and to compute the probability of each type of recall on T_1 , T_2 , and T_3 . Averaged across age groups, the proportions of targets that were recalled recollectively on these trials were .41, .56, and .65, respectively, while the proportions of targets that were recalled nonrecollectively were .01, .04, and .04, respectively. Thus, levels of nonrecollective recall were near-floor, making it very difficult to detect developmental changes. Nevertheless, we did detect a reliable age trend in nonrecollective recall, as follows. Close inspection of Table 2 suggests that familiarity judgment improved with repetition and that the improvement was greatest for 11-year-olds. We then tested the null hypothesis that $J_1 = J_2 = J_3$ for the data of each age group, and found that this null hypothesis could only be rejected for the 11-year-olds, $G^2(2) = 10.34$ ($p < .006$).

Summing up, this worked example illustrates the analytical steps that are involved in identifying which retrieval processes are responsible for developmental trends in performance on recall tests. In this particular example, those analyses converged on two simple conclusions. First, all of the age improvement in retrieval processes occurred after age 7. Second, the retrieval locus of that improvement was quite narrow; it was confined to improvements in the recollective operation, though there was some evidence of an increasing tendency for reconstructed items to become more familiar over trials.

Concluding Comments

On the surface, modern dual process ideas about episodic memory—the verbatim and gist traces of FTT or the item and knowledge traces of Shiffrin’s retrieving effectively from memory theory (e.g., Shiffrin & Steyvers, 1997), for instance—are rather far afield from the ideas about conditioning states of stimulus elements that motivated Estes’ (1959) exegesis of Markov models. However, the underlying motivation for implementing such models is the same in both spheres, something that Estes would have been the first to recognize. That motivation is the desire for a simple, elegant mathematical framework in which important theoretical processes can be quantified and their ability to explain experimental effects can be studied systematically.

The present application of Markov models to the prototypical recall paradigms of the memory laboratory has both historical roots and contemporary features. On the historical side, it takes advantage of one of our field’s classic data archives. According to that archive, the error-success sequences of recall tasks are two-stage Markovian to close approximations because these models give quite good accounts of the fine-grain structure of such sequences. These models, then, supply an empirically grounded pathway to quantifying theoretical ideas about the processes that control recall. On contemporary side, the present application resolves three problems that are posed by the currently dominant approach to measuring dual retrieval processes, which consists of making certain types metacognitive judgments about old/new recognition decisions. The first two problems, which have been much discussed in the mainstream memory literature (e.g., Malmberg, 2008), are that (a) recognition may involve only a single familiarity operation and (b) even if it involves dual retrieval operations, metacognitive judgment tasks are not reliable methods of separating them. The third problem, which has cropped up in the developmental and neuropsychological literatures, is that metacognitive tasks are high-burden methodologies that are beyond the capabilities of young children and older adults with neurocognitive impairments. Nevertheless, the measurement of dual retrieval processes is of great interest with those populations because these processes have been invoked to explain their characteristic memory deficits.

To date, the application of two-stage Markov chains to the study of dual retrieval processes has proceeded along two lines. The first, which was illustrated in this article, is the analysis and interpretation of a corpus of previously reported recall experiments. Across that corpus, there is variability in important characteristics of subjects (e.g., chronological age, normal versus disabled), list content (e.g., categorized versus unrelated words, pictures versus words), and recall paradigm (e.g., associative versus free, immediate versus delayed) that

can advance theoretical understanding of what model parameters are measuring. The other line of research is the more traditional one of conducting new experiments that incorporate manipulations that are designed to test hypotheses about what the parameters are measuring, which leads to improved theoretical conceptions of dual retrieval processes.

Notwithstanding that these lines of research are recent ones, some instructive findings have already accumulated. This is not the place to undertake a comprehensive review of them, but we will mention a smattering of results that bear on three problems. The first is the possibility that, as with recognition, recall may involve only a one-stage operation. In that connection, we have studied a pair of one-stage alternatives to the dual-retrieval model recall (panels B and C of Figure 1). Fit tests have repeatedly rejected one-stage models at very high levels of confidence, whereas fit tests have failed to reject the dual-retrieval model. Beyond this, factor analyses of parameter estimates from large numbers of data sets have been in broad agreement with the notion that two distinct retrieval processes and a slave judgment operation figure in recall.

The second problem, that metacognitive judgment tasks may be unreliable methods of measuring dual retrieval processes, vanishes with the recall approach. That is because familiar recall paradigms are not altered in any way, and nothing other than the usual sequences of errors and successes supply inputs to the measurement process. The novel feature is analytical rather than methodological: One or more versions of the five Markov chains are used to split sequences into components in which reconstruction is operating or direct access is operating or neither is operating. Model parameters then index the mix of these components in particular data sets, which forms the basis for evaluating hypotheses about recollective and nonrecollective retrieval.

The third problem, that metacognitive judgment tasks are too demanding for certain populations that are of considerable theoretical interest, also vanishes with the recall approach. The populations in question, young children and impaired older adults, are quite capable of performing standard recall tasks. Indeed, ironically, the literature on memory development between the preschool years and young adulthood is dominated by recall experiments because recall exhibits marked age improvements while recognition does not (for reviews, see Schneider & Bjorklund, 1998; Schneider & Pressley, 1997). Likewise, recall experiments prevail in the literature on memory declines in elderly individuals who have transitioned to cognitive impairment or dementia, for similar reasons: Declines are far more apparent for recall than for recognition. Of course, the fact that recall tasks are within the capabilities of children and impaired older adults does not establish that the dual-retrieval model is as applicable to their recall data as to the data of subjects who do not have major cognitive limitations. It seems to be, however. This has been one of the foci of experimentation with the dual-retrieval model, and good model fits have been obtained in associative, free, and cued recall experiments with younger and older children, healthy older adults, older adults with mild cognitive impairment diagnoses, and older adults with dementia diagnoses.

In sum, the recall approach can be fairly said to still be on its shake-out cruise, but the results have been promising so far. The dual-retrieval model has delivered good fits to recall

data throughout the human life span, and the behavior of its parameters has been broadly consistent with theoretical conceptions of recollective and nonrecollective retrieval. The preliminary evidence, then, points to the conclusion that this model is a productive addition to existing techniques for disentangling recollective and nonrecollective retrieval and exploring their respective contributions to episodic memory.

Acknowledgments

Preparation of this article was supported by the CAPES Foundation (BEX 0328/12-0) to the first author, and by the National Institutes of Health Grant 1RC1AG036915-02 to the second and fourth authors.

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 & Language*. 2006; 55:213–231.
- Batchelder WH, Chosak-Reiter J, Shankle WR, Dick MB. A multinomial modeling analysis of memory deficits in Alzheimer's disease and vascular dementia. *Journal of Gerontology: Psychological Sciences*. 1997; 52:206–215.
- Billingsley RL, Smith LM, McAndrews PT. Developmental patterns in priming and familiarity in explicit recollection. *Journal of Experimental Child Psychology*. 2002; 82:251–277. [PubMed: 12093109]
- Bjorklund, DF.; Muir, JE. *Annals of child development*. Vol. Vol. 5. Greenwich, CT: JAI Press; 1988. Children's development of free recall memory: Remembering on their own; p. 79-123.
- Bower GH. Application of a model to paired-associate learning. *Psychometrika*. 1961; 26:255–280.
- 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. Three-state models of memory development - a review of advances in statistical methodology. *Journal of Experimental Child Psychology*. 1985; 40:375–394.
- Brainerd CJ, Aydin C, Reyna VF. Development of dual-retrieval processes in recall: Learning, forgetting, and reminiscence. *Journal of Memory and Language*. 2012; 66:763–788. [PubMed: 22778491]
- Brainerd CJ, Holliday RE, Reyna VF, Yang Y, Toglia MP. Developmental reversals in false memory: effects of emotional valence and arousal. *Journal of Experimental Child Psychology*. 2010; 107:137–154. [PubMed: 20547393]
- Brainerd CJ, Howe ML. Stages-of-learning analysis of developmental interactions in memory, with illustrations from developmental interactions in picture-word effects. *Developmental Review*. 1982; 2:251–273.
- Brainerd CJ, Howe ML, Desrochers A. The general theory of two-stage learning: A mathematical review with illustrations from memory development. *Psychological Bulletin*. 1982; 91:634–665.
- Brainerd CJ, Howe ML, Kingma J. An identifiable model of two-stage learning. *Journal of Mathematical Psychology*. 1982; 26:263–293.
- Brainerd CJ, Howe ML, Kingma J, Brainerd SH. On the measurement of storage and retrieval contributions to memory development. *Journal of Experimental Child Psychology*. 1984; 37:478–499.
- Brainerd CJ, Kingma J, Howe ML. On the development of forgetting. *Child Development*. 1985; 56:1103–1119.
- Brainerd CJ, Kingma J, Howe ML. Spread of encoding and the development of organization in memory. *Canadian Journal of Psychology-Revue Canadienne De Psychologie*. 1986; 40:203–223.
- Brainerd CJ, Reyna VF. Memory independence and memory interference in cognitive development. *Psychological Review*. 1993; 100:42–67. [PubMed: 8426881]
- Brainerd CJ, Reyna VF. Fuzzy-trace theory and memory development. *Developmental Review*. 2004; 24:396–439.

- Brainerd CJ, Reyna VF. Recollective and nonrecollective recall. *Journal of Memory and Language*. 2010; 63:425–445. [PubMed: 22279248]
- Brainerd CJ, Reyna VF, Brandse E. Are children's false memories more persistent than their true memories? *Psychological Science*. 1995; 6:359–364.
- Brainerd CJ, Reyna VF, Holliday RE, Nakamura K. Overdistribution in source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2012; 38:413–439.
- Brainerd CJ, Reyna VF, Howe ML. Trichotomous processes in early memory development, aging, and neurocognitive impairment: A unified theory. *Psychological Review*. 2009; 116:732–783.
- Brainerd CJ, Reyna VF, Howe ML, Kingma J, Guttentag RE. The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development*. 1990; 33:1–109. [PubMed: 2287345]
- Brown R, McNeill D. The “tip of the tongue” phenomenon. *Journal of Verbal Learning & Verbal Behavior*. 1966; 5:325–337.
- Bugajska A, Morson S, Moulin C-J-A, Souchay C. Metamemory, recollection and familiarity in Alzheimer's disease. *Revue Neurologique*. 2011; 167:3–13.
- Cary M, Reder LM. A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*. 2003; 49:231–248.
- Clark SE, Gronlund SD. Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review*. 1996; 3:37–60. [PubMed: 24214802]
- Cressie N, Read TRC. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B*. 1984; 46:440–464.
- Diana RA, Yonelinas AP, Ranganath C. Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*. 2007; 11:379–386. [PubMed: 17707683]
- Donaldson W. The role of decision processes in remembering and knowing. *Memory and Cognition*. 1996; 24:523–533. [PubMed: 8757500]
- 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. Component and pattern models with Markovian interpretations. In: Bush, RR.; Estes, WK., editors. *Studies in mathematical learning theory*. Palo Alto, CA: Stanford University Press; 1959. p. 3–52.
- Estes WK. Learning theory and the new mental chemistry. *Psychological Review*. 1960; 67:207–223. [PubMed: 13820869]
- Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 1999; 4:272–299.
- Gardiner JJ, Java RI. Recollective experience in word and nonword judgments. *Memory & Cognition*. 1990; 18:23–30. [PubMed: 2314224]
- Gardiner JJ, Java RI. Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*. 1991; 19:617–623. [PubMed: 1758306]
- Gardiner JM, Parkin AJ. Attention and recollective experience in recognition memory. *Memory & Cognition*. 1990; 18:579–583. [PubMed: 2266859]
- Ghetti S. Processes underlying developmental reversals in false-memory formation: Comment on Brainerd, Reyna, and Ceci. *Psychological Bulletin*. 2008; 134:764–767. [PubMed: 18729572]
- Gillund G, Shiffrin R. 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.
- Gomes, CFA.; Brainerd, CJ. Dual processes in the development of reasoning: The memory side of the theory. In: Barrouillet, P.; Gauffroy, C., editors. *The development of thinking and reasoning*. New York: Psychology Press; 2013. p. 221–242.

- Gomes CFA, Brainerd CJ, Stein LM. Effects of emotional valence on recollective and nonrecollective recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2013; 39:663–677.
- 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, Millward RB, Merryman CT. Matrix analysis of identifiability of some finite Markov models. *Psychometrika*. 1971; 33:173–215. [PubMed: 5242170]
- Greeno JG, Steiner TE. Markovian processes with identifiable states: General considerations and application to all-or-none learning. *Psychometrika*. 1964; 29:309–333.
- Guttentag R, Carroll D. Recollection-based recognition: Word frequency effects. *Journal of Memory and Language*. 1997; 37:502–516.
- Hicks J, Marsh R. On predicting the future states of awareness for recognition of unrecalable items. *Memory & Cognition*. 2002; 30:60–66. [PubMed: 11958355]
- Howe, ML. Measuring memory development in adulthood: A model-based approach to disentangling storage-retrieval contributions. In: Howe, ML.; Brainerd, CJ., editors. *Cognitive development in adulthood*. New York: Springer-Verlag; 1988. p. 39-64.
- Howe, ML. Development of a mathematical model of memory for clinical research applications in aging. In: Howe, ML.; Stones, MJ.; Brainerd, CJ., editors. *Cognitive and behavioral performance factors in atypical aging*. New York: Springer; 1990. p. 3-36.
- Howe ML, Brainerd CJ, Kingma J. Development of organization in recall - a stages-of-learning analysis. *Journal of Experimental Child Psychology*. 1985a; 39:230–251.
- Howe ML, Brainerd CJ, Kingma J. Storage-retrieval processes of normal and learning-disabled children: A stages-of-learning analysis of picture-word effects. *Child Development*. 1985b; 56:1120–1133. [PubMed: 4053734]
- Howe ML, O'Sullivan J, Brainerd CJ, Kingma J. Localizing the development of ability differences in organized memory. *Contemporary Educational Psychology*. 1989; 14:336–356.
- Howe ML, Hunter MA. Adult age differences in storage-retrieval processes: A stages-of-learning analysis of developmental interactions in concreteness effects. *Canadian Journal of Psychology*. 1985; 39:130–150.
- Howe ML, Hunter MA. Long-term memory in adult-hood: An examination of the development of storage and retrieval processes at acquisition and retention. *Developmental Review*. 1986; 6:334–364.
- Hu X, Batchelder WH. The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*. 1994; 59:21–47.
- Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*. 1991; 30:513–541.
- Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. 1958; 23:187–200.
- Kemeny, JG.; Snell, JL. *Finite Markov chains*. New York: Van Nostrand; 1959.
- Kintsch W, Morris CJ. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology*. 1965; 69:200–206. [PubMed: 14279778]
- Knowlton B, Squire L. Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology-Learning Memory and Cognition*. 1995; 21:699–710.
- Koriat A. How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review*. 1993; 100:609–639. [PubMed: 8255951]
- Koriat A. Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*. 1995; 124:311–333.

- Koriat A, Levy-Sardot R, Edry E, De Marcas G. What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2003; 29:1095–1105.
- Langa K, Plassman B, Wallace R, Herzog A, Heeringa S, Ofstedal M, Willis R. The aging, demographics, and memory study: Study design and methods. *Neuroepidemiology*. 2005; 25:181–191. [PubMed: 16103729]
- Malmberg KJ. Recognition memory: A review of the critical findings and an integrated theory of relating them. *Cognitive Psychology*. 2008; 57:335–384. [PubMed: 18485339]
- Mandler G. Recognizing: The judgment of previous occurrence. *Psychological Review*. 1980; 87:252–271.
- Morris JC, Heyman A, Mohs RC, Hughes JP, VanBelle G, Fillenbaum G, et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989; 39:1159–1165. [PubMed: 2771064]
- Osgood CE. The nature and measurement of meaning. *Psychological Bulletin*. 1952; 49:197–237. [PubMed: 14930159]
- Payne DG. Hypermnnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*. 1987; 101:5–27.
- Peterson RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*. 1999; 56:303–308. [PubMed: 10190820]
- Ranganath C. A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*. 2010; 20:1263–1290. [PubMed: 20928833]
- Ratcliff R, Van Zandt T, McKoon G. Process dissociation, single process theories, and recognition memory. *Journal of Experimental Psychology: General*. 1995; 124:352–374. [PubMed: 8530910]
- Restle F. The selection of strategies in cue learning. *Psychological Review*. 1962; 69:329–343. [PubMed: 14491579]
- Reyna VF, Brainerd CJ. Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*. 2011; 31:180–206. [PubMed: 22096268]
- Reyna VF, Brainerd CJ. Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*. 1991; 4:249–262.
- Reyna, VF.; Mills, BA. Interference processes in fuzzy-trace theory: Aging, Alzheimer's disease, and development. In: MacLeod, C.; Gorfain, D., editors. *Inhibition in cognition*. Washington: American Psychological Association; 2007. p. 185–210.
- Rock I. Repetition and learning. *Scientific American*. 1958; 199:68–76. [PubMed: 13580273]
- Roediger HL, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*. 2006; 17:249–255. [PubMed: 16507066]
- Roediger HL, Schmidt SR. Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning & Memory*. 1980; 6:91–105.
- 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]
- Schacter DL, Slotnick SD. The cognitive neuroscience of memory distortion. *Neuron*. 2004; 44:149–160. [PubMed: 15450167]
- Schacter, DL.; Tulving, E. *Memory systems*. Cambridge, MA: MIT Press; 1994.
- Schacter DL, Worling JR. Attribute information and the feeling of knowing. *Canadian Journal of Psychology*. 1985; 39:467–475.
- Schneider, W.; Bjorklund, DF. Memory. In: Damon, W.; Kuhn, D.; Siegler, RS., editors. *Handbook of child psychology: Vol. 2. Cognition, perception, and language*. 5th ed. New York: Wiley; 1998. p. 467–522.
- Schneider, W.; Pressley, M. *Memory development between two and twenty*. Englewood Cliffs, NJ: Earlbaum; 1997.

- Sharot T, Yonelinas AP. Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*. 2008; 106:538–547. [PubMed: 17451666]
- Shiffrin R, Steyvers M. A 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 M, Neuschatz J, Goodwin K. Recall accuracy and illusory memories: When more is less. *Memory*. 1999; 7:233–256. [PubMed: 10645381]
- Trabasso TR. Stimulus emphasis on all-or-none learning in concept identification. *Journal of Experimental Psychology*. 1963; 65:398–406. [PubMed: 13993915]
- Tulving E. Memory and consciousness. *Canadian Psychologist*. 1985; 26:1–12.
- 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 A. Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology-General*. 2001; 130:361–379. [PubMed: 11561915]
- Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*. 2002; 46:441–518.

Highlights

- We develop mathematical machinery for Markov models of dual retrieval processes.
- These processes are measured for learning, forgetting and reminiscence.
- The Markov models are applied to a corpus of 209 recall experiments.
- Use of model parameters to test theoretical hypotheses is illustrated.

Figure 1.

Transition diagrams of three absorbing Markov chains. Panel A shows the transition diagram of a Markov chain with one absorbing state L and two transient states U and P , called two-stage absorbing Markov model. Panels B and C show two types of one-stage absorbing Markov models. Panel B shows the transition diagram of a Markov chain with one absorbing state L and one transient state U , while panel C shows the transition diagram of a Markov chain with one absorbing state L and one transition state P . Transition probabilities correspond to the entries of the transition matrix shown in Equation 3.

Figure 2.

Distribution of the goodness of fit statistic across the 229 data sets of the corpus of multi-trial recall experiments for each canonical design model. Panel A shows the distribution of the learning on errors model. Panel B shows the distribution of the alternative learning on errors model. Panel C shows the distribution of the learning on successes model. Panel D shows the distribution of the learning on both errors and successes model. In each panel, the dashed line indicates the critical value (3.84) to reject the null hypothesis of fit.

Table 1

Factor Structure of the Parameter Space of the Five Versions of the Dual-Retrieval Model

Parameter	Factor		
	1 (direct access)	2 (reconstruction)	3 (familiarity judgment)
Non-canonical design model			
D_1	.90		
D_2	.82		
D_{3E}		.49	-.42
D_{3C}	.67		
R_1		.82	
R_2		.89	
J_1			.48
J_2			.73
J_{3E}			.79
J_{3C}		.57	.65
% Var	22.25	22.15	21.35
Canonical design, learning on errors model			
D_1	.94		
D_2	.95		
R_1		.97	
R_2			.87
J_1			.71
J_2			.81
% Var	31.84	23.86	32.65
Canonical design, learning on successes model			
D_1	.84		
D_2	.79		
R_1		.94	
R_2			.87
J_1		-.57	.71
J_2			.83
% Var	23.14	23.93	33.06
Canonical design, learning on both errors and successes model			
D_1	.93		
D_2	.91		
R_1		.95	

Parameter	Factor		
	1 (direct access)	2 (reconstruction)	3 (familiarity judgment)
R_2			.86
J_1		-.63	.70
J_2			.83
% Var	28.80	22.59	32.97
Canonical design, alternative learning on errors model			
D_1	.94		
D_2	.88		
R		.84	
J_1			.97
J_2		.86	
J_3		.89	
% Var	28.82	39.14	18.57

Note. % Var = % of variance in parameter estimates accounted by each factor. Only reliable factor loadings were included in the table.

Table 2

Maximum Likelihood Estimates of the Parameters of the Alternative Learning on Errors Dual-Retrieval Model

Parameter	5-year-olds	7-year-olds	11-year-olds
Direct access			
D_1	.39	.38	.45
D_2	.18	.20	.40
Reconstruction			
R	.09	.14	.27
Familiarity judgment			
J_1	.13	.12	.09
J_2	.34	.20	.29
J_3	.31	.18	.34