Trichotomous Processes in Early Memory Development, Aging, and Cognitive Impairment: A Unified Theory

C. J. Brainerd
Cornell University
V. F. Reyna
Cornell University
M. L. Howe
Lancaster University

IN PRESS: *Psychological Review*

Abstract

One of the most extensively investigated topics in the adult memory literature, dual memory processes, has had virtually no impact on the study of early memory development. We remove the key obstacles to such research by formulating a trichotomous theory of recall that combines the traditional dual processes of recollection and familiarity with a reconstruction process. The theory is then embedded in a hidden Markov model that measures all three processes with low-burden tasks that are appropriate for even young children. These techniques are applied to a large corpus of developmental studies of recall, yielding stable findings about the emergence of dual memory processes between childhood and young adulthood and generating tests of many theoretical predictions. The techniques are extended to the study of healthy aging and to the memory sequelae of common forms of cognitive impairment, resulting in a theoretical framework that is unified over four major domains of memory research: early development, mainstream adult research, aging, and cognitive impairment. The techniques are also extended to recognition, creating a unified dual-process framework for recall and recognition.

Keywords: memory development, dual memory processes, aging, cognitive impairment, hidden Markov models

Trichotomous Processes in Early Memory Development, Aging, and Cognitive Impairment:
A Unified Theory

This paper has three objectives. The first is to resolve a fundamental problem in memory development research; the second is to apply that solution to a large corpus of developmental studies of recall; and the third is to extend the solution to memory changes that occur during healthy aging and in certain forms of cognitive impairment. The problem in question is the scant impact that dual-process distinctions have had on the study of early (child-to-young-adult) memory development. The result is a dramatic disparity in knowledge about dual memory processes in adults, which is vast, versus knowledge about how those processes first evolve, which is thin and inconsistent. Our solution, as will be seen, is (a) to propose a trichotomous theory of recall that subsumes traditional dual-process distinctions, (b) to implement that theory in a low-burden family of tasks that are appropriate for even very young children, and (c) to show that a new mathematical model of those tasks can be used to measure the development of trichotomous memory processes and to test theoretical predictions about them. Concerning the second objective, we show that an attractive feature of this solution is that a developmental data base exists that can be analyzed with the model, thereby closing the gap in knowledge about early memory development. With respect to the third objective, unlike early development, the study of memory changes during aging and in cognitive impairment have been strongly influenced by dual-process distinctions. We show that another attractive feature of our solution is that it is easily extended to both of these domains, yielding a theory that is unified over four fields of study: early memory development, mainstream adult memory research, aging, and cognitive impairment.

Returning to the problem that motivated this paper, the distinction between recollection- and familiarity-driven remembering has figured centrally in memory research with adults for some time (e.g., Atkinson & Juola, 1973; Mandler, 1980). Of late, this distinction has also come to play a prominent role in the study of aging (e.g., Anderson, Jennings, Cabeza, Ebert, Grady, & Graham, 2008; Dennis, Kim, & Cabeza, 2007; Parks, 2007; Skinner, & Fernandes, 2008; Toth & Parks, 2006) and cognitive impairment (e.g., Reyna & Mills, 2007; Schacter & Slotnick, 2004; Yonelinas, 2002). The picture in the memory development literature is quite different. On the one hand, some progress has been made in tracking the early ontogenetic course of dual memory processes, using methodologies that were originally devised to measure those processes in adults (e.g., Brainerd, Stein, & Reyna, 1998; Ghetti & Agelini, 2008; Ghetti & Castelli, 2006; Holliday & Hayes, 2000, 2002; Newcombe, & Lie, 1995). However, such research has been sparse and sporadic, and findings have been inconsistent. Consequently, dual-process distinctions have not penetrated mainstream developmental theory to any great degree. The sparseness problem is well illustrated by the published archive on Tulving's (1985) remember/know procedure. In the adult literature, hundreds of experiments have been reported in which this procedure has been used to separate recollection from familiarity (for a review of early work, see Donaldson, 1996; for reviews of subsequent work, see Dunn, 2008; Yonelinas, 2002), but as Ghetti and Angelini recently pointed out, there is only a single published study in which this procedure was used to compare the two forms of remembering in children of different ages (Billingsley, Smith, & McAndrews, 2002).[1] With respect to empirical inconsistencies, in the majority of studies that have used conventional adult methodologies, recollection has been found to develop between early childhood and young adulthood while familiarity has not. In some studies, however, familiarity has also been found to develop, and in certain studies, age changes have been more pronounced for familiarity than for recollection (Ghetti & Angelini, 2008). Such empirical inconsistencies are common features of literatures in which studies are few and methodological variability is high.

No doubt, the dearth of developmental dual-process research has many causes, but there is a pair of obvious impediments that will surely have to be removed if the situation is to change. The first is theoretical and is concerned with the types of memory performance in which dual processes figure. Beginning with Atkinson's and Mandler's early proposals (Atkinson & Juola, 1973; Mandler, 1980), dual-process distinctions have

been distinctions about mechanisms that underlie recognition. However, as Wixted (2007) concluded in a recent review, experimentation has failed to provide convincing evidence that recognition does, indeed, involve dual processes. Obviously, it would be questionable to focus developmental research on a form of memory performance that may be incapable of distinguishing dual processes in the first place. A further problem with recognition is that it has traditionally been of marginal interest to students of development, the reason being that age variability is limited during the child-to-young-adult age range. Although performance on standard recall tasks (cued, free, paired-associate, serial) improves dramatically across this age range (for reviews, see Bjorklund, 1987; Bjorklund & Muir, 1988), high levels of recognition for the same types of material are present during the preschool years (e.g., Merriman, Azmita, & Perlmutter, 1988; Morrison, Haith, & Kagan, 1980). [A parallel phenomenon occurs late in life, when recall tests are more sensitive to memory declines during healthy aging and in the emergence of cognitive impairment (e.g., Peterson et al., 1999, 2001; Spaan, Raaijmaker, & Jonker, 2004).] Consequently, null age effects have been ubiquitous in developmental studies of recognition (Ceci, Ross, & Toglia, 1987; Naus, Ornstein, & Kreshtool, 1977; Orstein & Corsale, 1979). Further, the developmental literature is replete with examples of important memory phenomena, such as forgetting, that display striking age changes when they are measured with recall but were once thought to develop minimally or not all because they had been measured with recognition (for a review, see Brainerd, Reyna, Howe, Kingma, & Guttentag, 1990).

The second impediment is the developmental inappropriateness of conventional dual-process methodologies. The adult paradigms that are most often used to effect separation of processes—such as remember/know (Tulving, 1985), process dissociation (Jacoby, 1991), and receiver operating characteristics (ROC; Lampinen, Odegard, Blackshear, & Toglia, 2005; Lampinen, Odegard, & Neuschatz, 2004; Yonelinas, 1994)—place high burdens on the capabilities of children. In those paradigms, the data that are diagnostic of dual processes are not old/new recognition but, rather, meta-cognitive judgments that supplement recognition. Such judgments require subjects to introspect on aspects of the mental experiences that recognition probes provoke, such as (a) vivid mental reinstatement of realistic details of prior presentations versus global feelings of familiarity (in the case of remember/know and process dissociation) and (b) feelings of confidence in old/new responses (in the case of ROC). The validity of these paradigms therefore turns on the assumption that subjects can introspect reliably on the mental states that probes stimulate and the assumption that subjects can comprehend and follow instructions as to how to perform introspections. Although each of these paradigms has been used in one or a few developmental studies, the extensive literature on meta-cognitive development (for reviews, see Bjorklund, 2004; Schneider & Bjorklund, 1998) argues that such assumptions are hazardous before adolescence. With the remember/know paradigm, for instance, Ghetti, Mirandola, Angelini, and Ciaramelli (2008; cited in Ghetti, 2008) found that children of age 7 and younger interpret remember/know instructions differently than older children and adults. Further, as Ghetti and Angelini (2008) have noted, using these paradigms with children also requires that they be simplified in various ways. That is problematical because child and adult measurements are then noncomparable, and any age differences that are detected may be artifacts of such noncomparability.

It seems, then, that a developmentally appropriate dual-process framework should have two features. First, it ought to be a theory of recall, both because recognition may not involve dual processes and because recognition displays limited variability during early memory development. Second, it ought to measure dual memory processes with low-burden tasks that elementary schoolers and preschoolers are capable of performing reliably. Thus, an ideal methodology would be one that extracts measurements of dual memory processes directly from performance on standard recall tasks. A developmentally appropriate framework should have a third feature, however: It ought to be easily extended to domains in which dual-process distinctions are already foci of research—notably, to adult memory, aging, and cognitive impairment. Otherwise, one problem (the dearth of developmental research) is being exchanged for another (noncomparable developmental research). In the remainder of this paper, we present and evaluate a theory that has all three of these properties, one that posits that recall is controlled by a pair of dissociated retrieval operations (direct access and reconstruction) and a slave judgment operation (familiarity) that is triggered whenever recall is based on reconstruction.

The presentation involves five steps. In the first section, we introduce a trichotomous theory of recall and summarize evidence from the adult literature that bears on its assumptions. In the second section, we implement the theory in a mathematical model of recall that separates and quantifies the three processes that are posited in the theory. In the third section, that model is exploited (a) to secure stable findings about age changes in trichotomous processes during child-to-adolescent development and adolescent-to-young-adult development and (b) to test core theoretical predictions about these processes. We rely on a corpus of developmental recall data sets that can be analyzed with the new model, which includes data from free, cued, and paired-associate recall tasks. In the fourth section, we show that this framework is readily extended to the study of memory changes that occur during healthy aging. Here, we rely on another corpus of recall studies of aging that can be analyzed with the new model. In the fifth section, we show that this framework is also readily extended to research on cognitive impairment and provide illustrative findings from studies of the memory sequelae of Alzheimer's dementia, mild cognitive impairment, depression, and schizophrenia.

Trichotomous Recall Processes

As we saw, there are three obstacles to using adult recognition methodologies to study the development of dual memory processes: Recognition may not involve dual processes, recognition displays minimal age variability, and adult methodologies place high demands on children. Those obstacles can be circumvented by using low-burden recall tasks. In that connection, dual-process distinctions have recently been proposed for recall in order to account for some otherwise puzzling findings. Following some historical remarks that place the present theory within the broader context of other theories of recall, we summarize those ideas in the second subsection below and then formalize a trichotomous theory of recall that will be used in developmental research. In the third subsection, we review experimental findings from the adult literature that motivate dual-process distinctions about recall.

Historical Background

Relative to other contemporary accounts of recall, the present theory has three distinguishing features: (a) It incorporates analogues of recognition processes, (b) it encompasses all of the standard recall paradigms (cued, free, paired-associate, serial), and (c) it is embedded in a mathematical model that separates the theory's processes and provides uncontaminated measurements of them. Concerning a, at one time recognition processes were central to theories of recall. That approach faded with the demise of generate/recognize theories, which posited that recall consists of whatever processes are involved in recognition, plus a generation process (e.g., Anderson & Bower, 1972; Kintsch & Morris, 1965). To recall a target, it was thought that subjects first, somehow, generate the item and then perform a subjective recognition test on it. The ground was cut from under such ideas by Tulving and Thomson's (1973) recognition-failure effect. If recall is just recognition plus an antecedent generation process, it will always be easier to recognize a target than to recall it (because the target is provided to subjects on recognition tests but they must first succeed in generating it on recall tests). Tulving and Thomson found that in certain types of paired-associate designs, it was easier to recall a target than to recognize it. A large body of experimentation on the recognition-failure effect then accumulated, and since, with the notable exception of some theories of stem completion (e.g., Bodner, Masson, & Caldwell, 2000; Jacoby, Toth, and Yonelinas, 1993), recognition processes have not figured centrally in recall theories. The present theory, unlike those in the generate/recognize vein, does not merely import recognition into its account of recall. Instead, as will be seen, it implements the *theoretical processes* that controversial recognition methodologies have sought to measure in recall tasks.

With respect to the second distinguishing feature, the modal contemporary theory of recall focuses on a single paradigm, such as free recall (e.g., Polyn, Norman, & Kahana, 2008), or on some particularly important recall effect, such as retrieval-induced forgetting (e.g., Norman, Newman, & Detre, 2007). Indeed, it is commonplace to restrict attention to particular effects that are produced by specific paradigms, with Kimball, Smith, and Kahana's (2007) fSAM theory of intrusions of semantic associates in free recall being a case in point. In contrast, the theory that is discussed in the present section can be applied to all standard recall paradigms (see the section

Development from Childhood to Young Adulthood, below), and it is assumed that different recall effects can be explained as parametric variations in the processes of direct access, reconstruction, and judgment.

The third distinguishing feature of the present theory is that it is embedded in a mathematical model whose parameters measure the processes of direct access, reconstruction, and judgment on a common ratio scale (see the section An Identifiable Model of Direct Access, Reconstruction, and Familiarity Judgment, below). When a theory is not embedded in a mathematical model, it poses some fundamental obstacles in the arena of experimental tests, two well-known examples being lack of quantitative fit and process impurity. Concerning fit, without a mathematical model it cannot be determined whether a theory is able give more than rough, qualitative accounts of empirical effects. Concerning process impurity, as Jacoby (1991) showed, separating the contributions of different processes to performance data with a model and then measuring those processes with the model's parameters are the only ways to ensure that the processes are not confounded with each other in experimental measurements. When process measurements are impure, manipulations that, theoretically, are predicted to affect a given process may do so or may fail to do for spurious reasons (i.e., owing to the contaminating influence of other processes). Despite these obstacles, some of the most influential theories of recall (e.g., Roediger et al., 2001; Koriat & Goldsmith, 1996) are not implemented in mathematical models.

Direct Access, Reconstruction, and Judgment

Barnhardt, Choi, Gerkens, and Smith (2006), Brainerd, Payne, Wright, and Reyna (2003), Brainerd, Wright, Reyna, and Payne (2002), and Reyna and Mills (2007) proposed that list items are recalled via a pair of dissociated retrieval operations, direct access and reconstruction, plus a slave judgment operation that evaluates the products of reconstruction. Here, we present a new theoretical account of direct access, reconstruction, and judgment that goes beyond such proposals, one that is sufficiently detailed to deliver several novel predictions about the behavior of parameters that measure these operations. Importantly, the necessity and sufficiency of these operations are tested later on, as part of the process of model validation. The tests that are performed will reject dual processes if they are not required by the data.

Direct Access

Direct access retrieves episodic traces of the prior presentations of *individual* items from a study list (verbatim traces). That is, this operation follows direct routes to traces of specific presentation events, and for that reason, it is assumed to be the faster and more accurate of the two retrieval methods. If this operation is the faster one, directly accessed items should predominate at the start of a free recall protocol (Barnhardt et al., 2006), and on paired-associate or cued recall tests, an item should be more likely to have been directly accessed if it is recalled quickly than if it is recalled slowly. Concerning accuracy, direct access supports errorless performance because an item's surface form is symbolically reinstated, so that the item can be recalled by merely reading out this surface information as it echoes in the mind's ear or flashes in the mind's eye—much as actors repeat lines that they hear from prompters or as readers pronounce words that they see on printed pages. Because episodic traces of surface forms are processed, direct access induces what is commonly termed recollective phenomenology; that is, vivid restoration of realistic details of items' prior presentations.

Despite these desirable properties, rememberers cannot rely solely on direct access, for two reasons: Experimental findings suggest that the types of traces that it accesses are quite sensitive to the output interference that accumulates during the course of recall and that those traces become rapidly unavailable as time passes (Reyna & Mills, 2007). Thus, direct access poses problems on the storage side of learning—the key problem being that for the operation to be successful, verbatim traces of item presentations must be available. Because such representations are interference-sensitive and labile, the major problem for learning is to store verbatim traces that are so

robust that, under the current experimental conditions, they are able to survive from one trial to the next until the overall performance criterion can be met, especially when the criterion is stringent (e.g., errorless recall).

Reconstruction

This retrieval operation regenerates targets from episodic traces of relational information about studied material, especially from gist traces of meaning content. Because such traces are not item-specific (e.g., "household pet" is not a specific animal, "Italian seasoning" is not a specific herb), an explicit mechanism that explains how rememberers get from such traces to individual candidate items for output is required. Here, we posit that reconstruction may be thought of as a classic delimited search operation (cf. Crowder, 1976), one that uses episodic traces of some of the targets' features (e.g., "household pet") to constrain the generation of candidate sets to ones that are restricted enough to be rapidly searched (e.g., *dog, cat, parakeet*). For any given list item, a *correct* search set is defined as one that contains that item.

Using episodic traces of information that does not uniquely identify specific targets (e.g., meaning features) to create correct search sets is the "construction" part of "reconstruction." With respect to this process, note that a good feature is one that delivers search sets that are simultaneously correct and small. Note, too, that the chief problem for reconstruction is not to identify target features that deliver correct search sets because any target is necessarily an exemplar of any of its features. Rather, the problem is to identify features that deliver small search sets. For instance, if *dog* and *tiger* are list words, "household pet" is an excellent dog feature, and "jungle cat" is an excellent tiger feature. However, "animal" is a poor dog feature and a poor tiger feature, notwithstanding that it is correct in both cases, because it over identifies a large search set.

Owing to the inherent speed differential between direct access and reconstruction, the probability that recall is due to the latter should be greater at the end of a free recall protocol than at the beginning (Brainerd et al., 2002), and this probability should increase with response latency on paired-associate and cued recall tests. Phenomenologically, reconstruction is experienced as subjective foraging for studied items, rather than as recollection of specific presentation events. Although reconstructive retrieval is focused on restricted sets of items, the fact that the information that delimits search is sketchy means that it is inevitable that some of the candidates that are identified for output will not have been part of the studied material. Interestingly, when this fact is combined with the notion that reconstruction is slower than direct access, the classic finding (e.g., Payne, 1986; Payne, Elie, Blackwell, & Neuschatz, 1996) that intrusions are usually concentrated at the ends of free recall protocols emerges as a straightforward prediction.

Reconstruction has two advantages, relative to direct access. First, the memory representations that it processes are less susceptible to output interference that accumulates during recall (Brainerd & Reyna, 1993; Reyna & Mills, 2007), and second, they are more stable over time (Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Reyna & Kiernan, 1994). Whereas the learning problem for direct access is mainly on the storage side, the learning problem for reconstruction is mainly on the retrieval side. The relational information that is processed to reconstruct items (e.g., "animal," "household pet," "jungle cat") does not need to be stored because it is already available in memory. Such information is simply activated and episodically tagged as targets are studied. However, subjects must learn how to use it to recover *specific* items. Because individual targets present multiple features from which contrasting correct search sets can be constructed, the problem for learning, as we said, is to identify features that deliver small correct search sets. Once such retrieval learning is complete for an item, reconstruction will succeed in finding it on a recall test. However, there is a remaining problem to be dealt with.

Judgment

That problem is that a search set that is small enough to be rapidly explored will typically contain non-target items as well as the target (e.g., *cat* and *parakeet* in addition to *dog*). To reduce the chances of outputting such plausible reconstructions, we have previously suggested that reconstruction is accompanied by a judgment operation that performs pre-output confidence checks (Brainerd et al., 2002; Reyna & Mills, 2007), but we have not proposed an explicit model of how such checks are performed. To remove that limitation, we assume that judgment is a signal detection process that consists of a familiarity signal and a bias parameter (cf. Snodgrass & Corwin, 1988). Explicitly, when reconstruction delivers a small set of candidate items, we assume that the items generate familiarity signals, much like probes on a

recognition test. Also on analogy to recognition probes, we assume that items' familiarity signals are processed by setting a decision criterion that executes confidence checks in the standard way: An item is output if the strength of its familiarity signal exceeds the decision criterion, but it is withheld otherwise. Thus, the judgment operation evaluates reconstructed items by passing them through a familiarity filter, with the probability of outputting such an item increasing as its familiarity signal becomes stronger and decreasing as the decision criterion becomes more stringent.

Although the judgment operation evaluates the products of reconstruction, it is a distinct process that can be affected by variables that either do not affect reconstruction or have opposite effects on reconstruction. For instance, instructions that encourage subjects to liberalize their decision criteria (e.g., Koriat & Goldsmith, 1996) should cause more reconstructed items to be released for output but not more items to be reconstructed. Further, Brainerd et al. (2002) proposed that increasing items' concreteness should make it more difficult to reconstruct them but should increase the subjective familiarity of any items that are reconstructed.

Basic Model

Consider the standard recall paradigms, in which trials consist of alternating study cycles and memory tests: free recall (subjects study lists of items and then recall as many as they can in any order); serial recall (subjects study lists of items and then recall as many as they can but in the order in which the items were presented); paired-associate recall (subjects study lists of item pairs and then attempt to recall the second member of each pair when the first is presented as a retrieval cue); and cued recall (subjects study lists of items that instantiate different semantic relations (e.g., taxonomic categories) and then recall as many exemplars of each relation as they can when that relation is presented as a retrieval cue). For such paradigms, the influence of direct access, reconstruction, and judgment can be formalized in the simple expression

$$P_i(Rc) = D_i + (1\text{-}D_i)R_i\, J_i, \tag{1}$$

where $P_i(Rc)$ is the probability of correctly recalling a target on the ith trial of an experiment, D_i is the probability of being able to directly access that target on the ith trial, R_i is the probability of being able to reconstruct that target on the ith trial, and J_i is the probability that a reconstructed target is familiar enough to pass the judgment check on the ith trial. (Throughout the remainder of this paper, an experimental "trial" will refer to one study cycle together with one or more recall tests.) In the present paper, as will be seen, the parameters D_i and R_i will be treated as *inter-state transition parameters*. That is, D_i will be the probability that an item has entered a state in which a verbatim trace can be directly accessed (which supports recall of that item with probability 1), and R_i will be the probability that an item has entered a state in which it can be reconstructed (which supports recall of that item with probability J_i). Once items have entered these states, they will not fall back to earlier states as long as the experimental conditions remain unchanged. Of course, values of D_i, R_i, and J_i will depend on those conditions, and as with any mathematical model, validity tests of process assumptions about these parameters are secured by introducing manipulations that embody those assumptions. Although Equation 1 expresses the probability of correct recall as a function of D_i, R_i, and J_i, these parameters are not identifiable; that is, they cannot be estimated from this equation because there is only one empirical degree of freedom, $P_i(Rc)$. Identifiability is a classic problem in model-driven theories of psychological processes (e.g., see Bamber & van Santen, 2000), and it refers to whether the parameters that measure such processes can in fact be estimated in the target data space (multi-trial recall experiments in this instance). We return to this question in a later section, where we introduce an identifiable model of direct access, reconstruction, and familiarity judgment.

Relations to Dual-Process Distinctions

Last, we explicate the relation between the present conceptualization of recall and traditional dual-process distinctions, that relation being to incorporate the customary distinction between recollection and familiarity, and to add a third process that is specific to recall. Direct access is the recall implementation of recollection, naturally. This process is accompanied by vivid reinstatement of targets' prior presentations, and it may be thought of as recollection that occurs in response to retrieval cues whose levels of specificity vary from items with which targets were paired on a study list (paired-associate recall) to items that appeared earlier in a study list (serial recall) to the names of semantic relations and categories that targets instantiate (cued recall) to generic list-identification cues (free recall). Turning to reconstruction, this operation has no analogue in traditional dual-process distinctions. Reconstruction regenerates items from

partial-identifying information about them, using that information to form delimited search sets that contain the items. Last, judgment, as conceptualized here, is the recall implementation of familiarity. It involves applying a decision criterion to the familiarity signals of items that may or may not have been studied. At a more explicit level, judgment can be viewed as a familiarity process that occurs in response to items that are cognitively regenerated (via reconstruction) rather than physically presented. A further important contrast between the trichotomous view of recall and traditional dual-process distinctions concerns the temporal sequencing of recollection and familiarity. In the standard view, familiarity is the faster of the two operations (e.g., Atkinson & Juola, 1973; Mandler, 1980), whereas in the present recall model, direct access occurs early and is backed up by familiarity if it is necessary to resort to reconstruction.

Summary

In the present conception, items are recalled either by directly accessing their verbatim traces or by reconstructing them from relational information and passing the reconstructions through a familiarity filter. A further consideration is that these two modes of retrieval seem to be antagonistic, from the perspective of learning. Direct access is a rote memorization process inasmuch as its aim is to store replicas of individual items that will survive long enough for subjects to meet performance criteria. Reconstruction, on the other hand, is a comprehension process, at least for the meaningful materials that are presented in developmental studies, because its aim is to understand how to use concepts, semantic features, and other relational information that studied items instantiate to create search sets that are small enough to regenerate the items rapidly on recall tests. These contrasting goals—rote memorization versus comprehension—suggest that measures of direct access and reconstruction will be dissociated in data, a possibility that is explored later in this paper.

Some Experimental Evidence

The foregoing distinctions are grounded in particular findings about recall. An early stimulus for the distinction between direct access and reconstruction was the cognitive triage effect, a puzzling U-shaped relation between the order in which items are output during free recall and their associated error rates—specifically, that items with lower error rates on previous trials tend to be recalled in middle positions, whereas items with higher error rates tend to be recalled in primacy and recency positions (Brainerd et al., 2002). Recently, however, the major impetus for this distinction is a series of dissociations between true and false recall in experiments in which subjects study lists of meaningfully-related items (e.g., Payne & Elie, 1997, 1998; Payne et al., 1996), with dissociations being particularly numerous for recall of Deese/Roediger/McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) lists and categorized lists. Such dissociations are predicted because one of the retrieval operations supports only true recall, whereas the other supports false as well as true recall. Thus, manipulations that increase the contribution of direct access to performance, ought to increase true recall and reduce false recall, whereas manipulations that increase the contribution of reconstruction to performance ought to increase false recall and may also reduce true recall or leave it unchanged (because reconstruction does not recover targets as reliably as direct access does; Brainerd et al., 2003). For instance, it is well known that intrusions that preserve the meaning of studied items tend to appear near the ends of free-recall protocols (e.g., Payne, 1987), which is congruent with the notion that intrusions are by-products of an error-prone operation that waxes during the later stages of recall. However, other recall patterns are more diagnostic of the direct access/reconstruction distinction, and we mention six examples from experiments in which subjects studied and recalled lists of meaningfully-related items.

First, when recall produces appreciable levels of intrusions, true and false recall probabilities are inversely related (e.g., Gallo & Roediger, 2003; Roediger, Watson McDermott, & Gallo, 2001). Negative correlations are expected because increased levels of false recall mean greater reliance on reconstruction, and greater reliance on reconstruction means less accurate true recall. Second, repeated testing has opposite effects on true and false recall: If subjects respond to a series of recall tests for a list, without further opportunities to study it, the false recall probability drifts upwards over tests while the true recall probability drifts downwards (e.g., Brainerd et al., 2003; Ceci & Bruck, 1995; Payne et al., 1996). This trend is expected because the sensitivity of direct access to the accumulation of output interference means that repeated testing will shift recall in the direction of reconstruction. Third, the length of the study list also has

opposite effects on true and false recall because as length increases, the false recall probability increases and the true recall probability decreases (for a review, see Brainerd, Reyna, & Ceci, 2008). This result is expected because longer lists generate more of the output interference that interferes with direct access, which shifts recall in the direction of (error-prone) reconstruction. Fourth, when subjects study lists of items that share meaning but recall is delayed for a few hours or days, the true recall probability declines steeply, the false recall probability remains relatively constant, and consequently, false recall increases substantially as a proportion of total recall (e. g., Brainerd et al., 2008; Gallo, 2006; Seamon et al., 2002a; Toglia, Neuschatz, & Goodwin, 1999). This pattern is expected on the ground that the representations that are processed by direct access versus reconstruction are forgotten at different rates, with those that support reconstruction being more likely to remain accessible as time passes. Fifth, when subjects of different ages study and recall lists of items that share meaning, the false recall probability increases dramatically between early childhood and young adulthood while the true recall probability increases more modestly, so that net recall accuracy declines with age (e.g., Howe, 2006; Metzger et al., 2008). This developmental trend is expected because the forms of semantic processing that extract the meaning relations that reconstruction operates on develop more slowly than the ability to store targets' surface features (e.g., Bjorklund, 1987, 2004). Sixth, encoding manipulations that make targets' surface features more distinctive while leaving their semantic content unchanged (e.g., presenting targets as pictures rather than as printed words, generating visual images of the orthographies of orally-presented words versus listening only) increase the true recall probability but suppress the false recall probability (for reviews, see Brainerd & Reyna, 2005; Gallo, 2006). This result is expected because making targets' surface forms more distinctive ought to shift recall in the direction of direct access.

Finally, a series of recent experiments by Barnhardt et al. (2006) tested two, rather precise, predictions of the direct access/reconstruction distinction. First, suppose that subjects study a list of words that share meaning, but within the list, a few unrelated words are also presented—e.g., a list of 25 words is studied that consists of 20 city names, plus 5 unrelated words inserted at random positions. The prediction is that unrelated words ought to be recalled relatively early in output because reconstructive retrieval waxes as recall proceeds, which favors targets that share salient meaning. Barnhardt et al. observed that pattern. Second, suppose that subjects study a list of words that share meaning. One of them, a word that is an especially good exemplar of the shared meaning content and is apt to be falsely recalled if it is not presented, appears on the study list for half the subjects but is omitted from the list for the other half. Theoretically, the former subjects can recall this word via either direct access or reconstruction, whereas the latter subjects can only recall it via reconstruction. The prediction is that when this single item is falsely recalled by the latter subjects, it will appear later in output, on average, than when it is correctly recalled by the former subjects. Barnhardt et al. also observed this pattern in their experiments.

In short, the direct access/reconstruction distinction about recall has proved to be quite productive in the study of false memory. Although this distinction has figured as a working hypothesis in several experiments, its application to the study of memory development demands a more formal treatment that embeds direct access, reconstruction, and familiarity judgment in a mathematical model that separates and

quantifies them, so that their relative contributions to age changes in performance can be determined. That is the matter to which we now turn.

An Identifiable Model of Direct Access, Reconstruction, and Familiarity Judgment

As we have seen, a developmentally appropriate framework for the study of dual memory processes should be focused on recall and should supply a low-burden methodology in which simple recall responses are used to measure those processes (rather than requiring children to perform supplementary meta-cognitive tasks). In this section, we show how the second criterion is met. To do that, we return to a problem that was mentioned in passing—namely, parameter identifiability. Equation 1, which expresses the probability of successful recall on trial *i* as a function direct access, reconstruction, and judgment, is not identifiable; there are more memory processes to estimate than there are empirical degrees of freedom.

Below, this limitation is removed by implementing Equation 1 as a two-stage absorbing Markov chain. In the first subsection, the basic features of such models and their history in memory research are briefly recounted. In the second subsection, the generic two-stage absorbing Markov chain is used to find an implementation of Equation 1 that will measure direct access, reconstruction, and familiarity judgment in low-burden recall designs. Although this initial implementation of Equation 1 greatly reduces the number of parameters that need to be estimated, we show that it, too, is not identifiable. In the third subsection, we present another Markov chain whose parameters are identifiable. The statistical machinery for estimating its parameters, for testing global fit, for generating predicted-observed comparisons of fine-grain performance statistics, and for testing hypotheses about parameters is also developed.

Markov Models of Memory and Cognition

It is commonplace to treat memory and reasoning processes as specifying distinct cognitive states, so that changes in those processes over experimental trials (learning) are conceptualized as transitions through a discrete state space. In such conceptualizations, finite Markov chains are the standard formalism for fitting data and extracting measurements of memory and reasoning processes (e.g., Busemeyer, Wang, & Townsend, 2006). The core assumptions of such models are just that (a) some type of performance (e.g., free recall, mental addition, probability judgment) consists of a small number of cognitive states C_1, C_2, …, C_k, each of which produces that performance with some average probability p_i, and that (b) the state that a subject occupies on Trial *i* of an experiment depends only on the state that was occupied on the immediately preceding trial. These assumptions are testable with the usual model-fitting procedures; that is, fits will be poor when either assumption is violated. Historically, data fits have usually been good (see illustrative recall fits in Figures 1 and 2). Consequently, finite Markov chains have long been popular devices for modeling memory processes (e.g., Bower & Theios, 1963; Greeno, 1968), a tradition that continues in contemporary research (e.g., Batchelder, Chosak-Reiter, Shankle, & Dick, 1997; Faglioni, Bertolani, Botti, Merelli, 2000a; Faglioni, Saetti, & Botti, 2000b; Katsikopoulos & Fisher, 2001), and they are an influential modeling technology in contemporary studies of judgment and decision making (see Busemeyer et al., 2006; Myung, Karabatsos, & Iverson, 2005).

A final consideration, one that is not widely appreciated, is that Markov models of memory and reasoning subsume many other commonly used modeling technologies. One especially popular technology, multinomial modeling, is a case in point. In a single-trial experiment, a multinomial model is the starting vector and response vector of a Markov chain, and in a multi-trial experiment, a multinomial model is the starting vector, response vector, and transition matrix of a Markov chain (see Riefer & Batchelder, 1988). Thus, such multinomial models as process dissociation (Jacoby, 1991), conjoint recognition (Brainerd, Reyna, & Mojardin, 1999), and source monitoring (Batchelder & Riefer, 1999) are all finite Markov chains.

The Generic Markov Chain for Recall

Returning to the identifiability problem with Equation 1, it is easy to see that this problem is a consequence of the assumption that there are distinct *D*, *R*, and *J* parameters for *each trial* of a recall

experiment. As there is only one empirical probability for each trial, *P*(*Rc*), nonidentifiability is inevitable. The solution, of course, is to reduce the number of memory parameters to a more manageable value—specifically, to a value that is below the number of empirical degrees of freedom—thereby securing an identifiable set of memory parameters *and* leaving some residual degrees of freedom for fit evaluation. This can be done by taking advantage of the history of modeling research on multi-trial recall, which has provided a modeling framework for recall data that has many fewer parameters than Equation 1 and is applicable throughout the lifespan.

That framework consists of a family of two-stage absorbing Markov chains. The fine-grain structure of adults' recall data—by which we mean the empirical distributions of various error and success statistics in standard recall paradigms—is known to conform closely to the predictions of such chains (for a review, see Brainerd, Howe, & Desrochers, 1982). Although Miller (1952) was the first to propose that finite Markov chains are applicable to memory paradigms, the earliest two-stage models of recall were Theios and Hakes' (1962) model of paired-associate recall and Waugh and Smith's (1962) model of free recall. Various investigators soon confirmed that two-stage absorbing Markov chains delivered excellent fits to paired-associate, cued, free, and serial recall data (e.g., Bower & Theios, 1963; Estes & DaPolito, 1967; Greeno, 1968; Halff, 1977; Kintsch, 1963; Kintsch & Morris, 1965; Pagel, 1973). Illustrations of the close correspondence between the predictions of two-stage models and the distributions of three statistics of recall data are shown in Figure 1.

Crucially for present purposes, these baseline results for adults were eventually extended to memory development, where it was found that, likewise, two-stage absorbing Markov chains delivered excellent fits to the recall data of younger children, older children, and adolescents (e.g., Brainerd & Reyna, 1991; Brainerd et al., 1990). Illustrations of such fits for child data are shown in Figure 2. It was also found that two-stage absorbing Markov chains provided excellent fits to the recall data of children with certain cognitive impairments (e.g., Howe, O'Sullivan, Brainerd, & Kingma, 1989; Kingma, 1987) and older adults (e.g., Batchelder et al., 1997; Howe & Hunter, 1985, 1986). As things stand, then, the accumulated literature on two-stage absorbing Markov chains shows that they fit the recall data of normal subject populations from the preschool years through late adulthood, and that they also fit the recall data of child populations with some forms of impairment.

This brings us back to the identifiability problem. In Equation 1, the number of memory parameters to be estimated increases linearly with the number of trials in a recall experiment. In contrast, two-stage absorbing Markov chains contain fewer free parameters, and *the number of free parameters does not increase as trials increase*. For instance, Greeno (1968) integrated various early examples of such chains into a generic Markov model of recall, using the canonical outcome space S_1T_1,S_2T_2, S_3T_3, …, where S_i is the *i*th study cycle, T_i is the *i*th recall test, and the ellipses mean that trials continue until recall is errorless. Intertrial changes in the probability of successfully recalling a target are controlled by transitions through a state space that consists of an initial state *U* (unlearned), in which the item cannot be recalled at all (success probability is 0), an intermediate state *P* (partially learned) in which successful recall occurs with some average probability $0 < p < 1$, and a terminal absorbing state *L* (learned), in which successful recall occurs with probability 1. For convenience, *P* is partitioned into a substate P_C, in which recall succeeds, and a substate P_E, in which recall fails. The number of identifiable parameters in this generic model, or in any Markov chain, can be determined with mathematical techniques that are in common use in the literature on hidden Markov models (or HMM; e.g., Bordes & Vandekerkhove, 2005; Chopin, 2007; Spezia, 2006; Welton & Ades, 2005).[2]

As this model fits recall data throughout the lifespan and contains a small, fixed number of parameters, it presents a tractable solution to the identifiability problem with Equation 1—namely, convert

the latter to the former by rewriting Equation 1 as the generic two-stage absorbing Markov chain, so that D_i, R_i, and J_i become parameters of that chain. We follow this tack in the next subsection. Along the way, however, it is shown that the identifiability problem remains because the resulting Markov model contains two more memory parameters than the number that can be estimated in the canonical outcome space. It is then shown that a fully identifiable model (i.e., all values of D_i, R_i, and J_i can be estimated) can be produced by slightly modifying the canonical outcome space.

Converting Equation 1 to the Generic Markov Chain for Recall

The generic Markov chain for recall is represented by a certain matrix expression (cf. Brainerd, Howe, & Kingma, 1982; Greeno, 1968):

$W_1 = [L(1), P_E(1), P_C(1), U(1)] = [a'b', a'(1-b')r, a'(1-b')(1-r), 1- a'];$

$$M_1 = \begin{array}{ccccc} & L(n+1) & P_E(n+1) & P_C(n+1) & U(n+1) \\ L(n) & 1 & 0 & 0 & 0 \\ P_E(n) & d & (1-d)(1-g) & (1-d)g & 0 \\ P_C(n) & c & (1-c)(1-h) & (1-c)h & 0 \\ U(n) & ab & a(1-b)e & a(1-b)(1-e) & 1-a \end{array}; \quad C_1 = \begin{array}{c} P(\text{correct}) \\ 1 \\ 0 \\ 1 \\ 0 \end{array}. \qquad (2)$$

The notation for Equation 2's parameter set {*a'*, *a*, *b'*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *r*} preserves the notation in the prior literature on this model. The use of this traditional notation means that the parameters are theory-neutral; that is, they are mathematical variables that do not yet involve any process interpretations. Later, when process interpretations from the trichotomous model are introduced, the notion will be altered to indicate that the parameters now have process definitions (see Equation 3). W_1 is a row vector (usually called a starting vector) that denotes the probabilities of being in each state on the first recall test (T_1), M_1 is a 4 × 4 matrix that denotes the probabilities of the possible interstate transitions on consecutive study-test trials of an experiment, and C_1 is a column vector that denotes the probability of successful recall in each state. As can be seen, there are four types of parameters: *U* escape probabilities (*a'* and *a*), *L* entry probabilities from state *U* (*b'* and *b*), *L* entry probabilities from state *P* (*c* and *d*), and recall probabilities while in state *P* (*e*, *g*, *h*, and *r*). Equation 2 describes the process of learning how to recall a target, in abstract terms, as consisting of two types of events: (a) escaping from the initial no-success state and (b) escaping from the intermediate partial-success state. That Equation 2 is an *absorbing* Markov chain is indicated by the top row of the M_1, which specifies that $p(L_{n+1}|L_n)$, the probability of being in the errorless recall state on test T_{n+1} if the process was in that state on test T_n, is unity, *as long as experimental conditions do not change*. Contrary to this assumption, it might be thought that *L* is not absorbing and that there is some probability that once items have reached *L*, they can fall back to *P* or *U*. This idea is testable. If it is incorrect, tests of global fit will fail because the top row of the M_1 must contain additional free parameters (rather than only the fixed parameters 0 and 1), and the observed asymptotes of curves like those in Figures 1 and 2 will be consistently higher than the asymptotes of the corresponding predicted curves. However, in recall experiments of the canonical form, it is well established that global fit tests produce satisfactory results and that, as can be seen in Figures 1 and 2, there is close correspondence between observed and predicted asymptotes. Therefore, empirically, the assumption that Equation 2 is absorbing has proved to be true to a statistically tolerable approximation.

Equation 1 can be converted to the generic Markov chain by mapping its parameter space with that of the generic model. This mapping is effected in three steps. First, it is assumed (a) that items that occupy state *L* of the generic model can be directly accessed (because direct access is *errorless* recall), (b) that items that occupy state *P* can be reconstructed but not directly accessed (because reconstruction is *imperfect* recall), (c) that items that occupy state *U* can be neither reconstructed nor directly accessed, and (d) that the level of recall in state *P* is an index of the familiarity level of reconstructed items and the stringency of decision criteria for outputting reconstructed items. Second, a subset of 10 parameters from Equation 1 are defined, which are exhibited in Table 1. Third, it is easy to see that the 10 parameters in Table 1 map with the 10 parameters of the generic Markov chain as follows: $D_1 = a'b'$, $D_2 = ab$, $D_{3E} = d$, $D_{3C} = c$, $R_1, = [a'(1-b')]/(1-D_1)$, $R_2 = [a(1-b)]/(1-D_2)$, $J_1 = 1-r$, $J_2 = 1-e$, $J_{3E} = g$, $J_{3C} = h$.

The mathematical relation between Equations 1 and 2 is that the parameter space of Equation 2 maps one-for-one with a subset of the parameter space of Equation 1. That is because for a recall experiment that consists of *k* trials, a series of *k* versions of Equation 1 can be written that contains a total of 3*k* parameters, whereas Equation 2 contains only 10 parameters. This mapping of Equation 2's parameters onto a subset of Equation 1's parameters does not in any way constrain empirical estimates that are ultimately obtained for Equation 1's parameters (or produce spurious relations between them) when Equation 2 is applied to recall data. The mapping is

merely a mathematical solution to the identifiability problem with Equation 1; neither model "knows" that its parameters have been mapped with those of the other model.

Worked Example of Learning to Recall via Reconstruction and Direct Access

We saw that once Equation 1 is converted to the generic Markov chain for recall, learning how to recall targets via direct access, reconstruction, and familiarity judgment can be described in two simple stages: escaping the no-success state and escaping the partial-success state. To make the memory processes that the parameters in Table 1 measure as concrete as possible, consider a simple experiment in which a group of subjects learns a list of 20 words to an errorless criterion under standard free recall conditions. That is, subjects study the list, then recall as many of the words as they can remember, then study the list again, then recall as many of the words as they can remember, and so on until all 20 words can be recalled.

Escaping the No-Success State

All words are assumed to begin in state U because subjects do not know the composition of the list before the first trial of the experiment. Each word can escape state U on the first study cycle or any subsequent study cycle if subjects learn how to directly access it or how to reconstruct it. On the first study cycle, escape from U is governed by the probabilities in the starting vector, W_1. A word can escape U on Trial 1 by becoming directly accessible, with probability D_1. If a word becomes directly accessible, it enters state L (errorless recall), and recall is successful with probability 1 on the first recall test and on all subsequent tests. If a word does not become directly accessible on the first study cycle, it can also escape U by becoming reconstructable, with probability $(1-D_1)R_1$. If a word becomes reconstructable but not directly accessible, it enters state P (imperfect recall) as it leaves U. If a word escapes U by becoming reconstructable, the judgment operation will output it with probability J_1 on the first recall test. If a word does not escape state U on the first study cycle, it can do so on any subsequent study cycle. Specifically, (a) a word can become directly accessible, with probability D_2, in which case it enters L as it leaves U and recall is successful with probability 1 on all subsequent tests, or (b) it can become reconstructable but not directly accessible with probability $(1-D_2)R_2$, in which case it enters P as it leaves U and the judgment operation outputs it with probability J_2.

Escaping the Partial-Success State

We know that if a word escapes U by becoming reconstructable but not directly accessible, it enters state P. Once a word has entered state P, it can be reconstructed but a familiarity judgment may not output it. Words that enter state P can escape from P to state L on some subsequent trial by becoming directly accessible. Escape from P to L is governed by the direct access parameters D_{3C} and D_{3E}, and the accuracy of recall while a word occupies state P is governed by the familiarity judgment parameters J_{3C} and J_{3E}. Whenever recall of a word is *un*successful in P, (i.e., familiarity judgment does not output reconstructions of the word), it either escapes to state L on the next study cycle, with probability D_{3E}, or it remains in P, with probability $1 - D_{3E}$. If a word escapes to L, recall is successful on the next recall test and on all subsequent tests with probability 1. If a word remains in P, the judgment operation either outputs the reconstruction of the word with probability J_{3E} on the next recall test or withholds it with probability $1 - J_{3E}$. On the other hand, whenever recall of a word is successful in P, it either escapes to state L on the next study cycle, with probability D_{3C}, or it remains in P, with probability $1 - D_{3C}$. If a word escapes to L, recall is successful on the next recall test and on all subsequent tests with probability 1. If a word remains in P, the judgment operation either outputs the reconstruction of the word with probability J_{3C} on the next recall test or withholds it with probability $1 - J_{3C}$.

Parameter Identifiability

Thus, for any recall experiment that follows the canonical design S_1T_1, S_2T_2, S_3T_3, …, Equation 1 can be converted to a two-stage absorbing Markov chain that contains only the 10 free parameters in Table 1—4 direct access parameters, 2 reconstruction parameters, and 4 judgment parameters. [The outcome space of such an experiment consists of sequences of responses to individual list items on T_1, T_2, T_3, …] This is a considerable reduction in Equation 1's parameter space and, hence, represents substantial progress in the direction of identifiability. Nevertheless, the parameters in Table 1 are not identifiable; none can yet be estimated from recall data. The reason is that when the generic Markov chain in Equation 1 is analyzed with mathematical techniques that are used to evaluate HMMs (Bordes & Vandekerkhove, 2005), it is found to have only 8 (rather than 10) identifiable parameters (see Appendix A, Equations A1, A2, and A3). As we show in Appendix A, these eight identifiable parameters, which are denoted by the set {*w*, *z*, *?*, *?*, *u*, *v, ?*, *?*}, are the only ones that can be independently estimated in recall experiments that follow the canonical design. This means that whenever a model that implements the

generic chain contains more than eight parameters, *those parameters will turn out to be expressible as functions of one or more of the parameters in the set* {*w*, *z*, *?*, *?*, *u*, *v*, *?*, *?*}. The exact functions that map the parameters in Table 1 onto the parameters of the identifiable set can be determined with algorithms that have been developed for this purpose (e.g., Bamber & van Santen, 2000; Rabiner, 1989).

We conducted such an analysis and derived the 8 functions that map the 10 parameters of Equation 2 onto the identifiable set of parameters, {*w*, *z*, *?*, *?*, *u*, *v*, *?*, *?*}. Those functions are displayed in Table 2, where the eight identifiable parameters appear on the left and the functions that map them with the 10 parameters that measure direct access, reconstruction, and familiarity judgment $\{D_1, D_2, D_{3E}, D_{3C}, R_1, R_2, J_1, J_2, J_{3E}, J_{3C}\}$ appear on the right. A glance at Table 2 reveals a deep conceptual difficulty. Whereas the second set contains *memory* parameters that have straightforward meanings that refer to direct access, reconstruction, and familiarity judgment, the identifiable parameters in the first set are merely mathematical variables that do not have process meanings. As can be seen in Table 2, each identifiable parameter is a function of *two or more* of the memory parameters, and except for *w* and *z*, those functions are rather complex. Further, four of the eight identifiable parameters (*?*, *?*, *?*, and *?*) are complex functions of the memory parameters for Trial 1 and for later trials.

In short, rendering the parameter space of Equation 1 more manageable by implementing it in the generic Markov chain for recall poses a dilemma. On the one hand, it is known that the generic chain is an appropriate HMM because it fits recall data to a very close approximation throughout the life span. Thus, direct access, reconstruction, and familiarity judgment should be measured by converting Equation 1 to a two-stage absorbing Markov chain. Once this conversion is effected, however, the result is a model (Table 1) whose parameters cannot be individually estimated because they are functions of a smaller set of identifiable parameters.

Converting Equation 1 to an Identifiable Markov Chain

Several developmental studies have been reported that used a slightly modified version of the canonical design over which the generic Markov chain for recall is defined (see Brainerd et al., 1990; Howe et al., 1989; Kingma, 1987; Reyna & Brainerd, 1995). This alternative design is $S_1T_1T_2$, S_2T_3, S_3T_4, …; that is, the only change is to insert an additional recall test (T_2) between the first recall test (T_1) and the second study cycle (S_2). It turns out that this solves the identifiability problem because it allows a slightly modified version of Equation 2 to be written, which contains 11 parameters rather than 10. When we analyzed this modified model, using the aforementioned HMM algorithms, its parameters proved to be fully identifiable. We introduce this identifiable model in the present subsection, and refer readers to Appendix A for the identifiability proof for its parameters and for the statistical machinery for estimating parameters, evaluating goodness of fit, and testing within- and between-condition hypotheses about the parameters.

Identifiable Markov Chain

Because we have already shown that the parameters in Table 1 map with the parameters of the nonidentifiable Markov chain (Equation 2), we use those parameters in developing the identifiable chain. For the alternative outcome space $S_1T_1T_2$, S_2T_3, S_3T_4, …, the modified version of Equation 2 is:

$$W_2 = [L(1)L(2), L(1)P_E(2), L(1)P_C(2), L(1)U(2), P_E(1)L(2), P_E(1)P_E(2), P_E(1)P_C(2), P_E(1)U(2), P_C(1)L(2), P_C(1)P_E(2), P_C(1)P_C(2), P_C(1)U(2), U(1)L(2), U(1)P_E(2), U(1)P_C(2), U(1)U(2)] =$$
$$[D_1, 0, 0, 0, 0, R_1(1-D_1)(1-J_1)(1-R_f)(1-J_{3E}), R_1(1-D_1)(1-J_1)(1-R_f)J_{3E}, R_1(1-D_1)(1-J_1)R_f, 0, R_1(1-D_1)J_1(1-R_f)(1-J_{3C}), R_1(1-D_1)J_1(1-R_f)J_{3C}, R_1(1-D_1)J_1R_f, 0, 0, 0, (1-D_1)(1-R_1)];$$

$$M_2 = \begin{array}{c} \\ L(n) \\ P_E(n) \\ P_C(n) \\ U(n) \end{array} \begin{array}{cccc} L(n+1) & P_E(n+1) & P_C(n+1) & U(n+1) \\ 1 & 0 & 0 & 0 \\ D_{3E} & (1-D_{3E})(1-J_{3E}) & (1-D_{3E})J_{3E} & 0 \\ D_{3C} & (1-D_{3C})(1-J_{3C}) & (1-D_{3C})J_{3C} & 0 \\ D_2 & R_2(1-D_2)(1-J_2) & R_2(1-D_2)J_2 & (1-D_2)(1-R_2) \end{array}; \quad C_2 = \begin{array}{c} P(\text{correct}) \\ 1 \\ 0 \\ 1 \\ 0 \end{array}. \qquad (3)$$

Note that Equation 3 is very similar, algebraically, to Equation 2. The key difference between them lies in their respective starting vectors: W_2 is a new starting vector that denotes the probabilities of being in each of the states on the first *two* recall tests (T_1 and T_2), whereas the previous starting vector, W_1, denotes the probabilities of being in each of the states on the first recall test. Thus, like the transition matrix, W_2 contains 16 cells, because it is defined over two consecutive tests on which an item can be in any of 4 memory states (L, P_E, P_C, and U) on each test. Note that all 10 parameters in Table 1 appear in Equation 3 *and measure exactly the same memory processes as before*. There is also an eleventh parameter, R_f. This is a forgetting parameter that allows for the possibility that if recall escapes from U on T_1 by becoming reconstructable, it may fall back to U on T_2 because there is no study cycle between them. Thus, our earlier description of how recall on Trial 1 is controlled by direct access, reconstruction, and familiarity is the same for Equation 3, save for the sole difference that items that have escaped state U by becoming reconstructable on the first study cycle are allowed to fall back from state P to U between T_1 and T_2, with some probability R_f. This single difference does not seem to be very important when it comes to actual recall performance. In the corpus of developmental recall studies that we explore in the next section of this paper, empirical estimates of R_f did not differ significantly from zero, except in rare instances. For that reason, this parameter is only considered in the later section on cognitive impairment. Importantly, however, note that the fact that the estimated value of this parameter is normally zero is consistent with the reconstructive interpretation of State P: If it is true that recall in this state involves processing highly stable meaning properties, forgetting in State P ought to be rare over such a short interval as a single recall test.

Identifiability Proof, Parameter Estimation, Goodness of Fit, and Hypothesis Testing

To prove that the parameters of any HMM, such as Equation 3, are identifiable in an outcome space, the standard mathematical technique (e.g., Bordes & Vandekerkhove, 2005) involves three steps. The first is to formulate an *observable*-states process (one in which the states are actual data events) that is implied by the HMM. Because such a process consists of data events, all of its parameters are necessarily identifiable, and by Bernoulli's theorem, a likelihood function can be written from which maximum likelihood estimates of those parameters can be computed for sample data. The second step is to analyze the (fully identifiable) parameter space of the observable-states process to derive a set of functions that maps its parameters onto the parameter space of the HMM (Equation 3 in this case); that is, a set of equations that expresses individual parameters of the observable-states process as functions of some of the parameters of the HMM. The third step is to solve this set of functions to determine whether they deliver *unique* estimators for each of the parameters of the Markov chain of interest; that is, an equation for each parameter of the HMM that expresses that *single parameter* as a function of some of the parameters of the observable-states process.

We conducted such an identifiability analysis of Equation 3 and found that all of its parameters were identifiable. As the identifiability proof is tedious, it has been relegated to Appendix A. Once identifiability is established, three other developments are necessary before Equation 3 can be applied to sample data: a method of estimating its parameters, a method of evaluating goodness of fit, and methods of testing within- and between-condition hypotheses about parameter values. These developments are also presented in Appendix A, following the identifiability proof.

Development from Childhood to Young Adulthood: Age Trends in Direct Access, Reconstruction, and Familiarity Judgment

We turn now to the second objective of this paper, which is to apply the foregoing theoretical distinctions and modeling techniques to a large corpus of developmental recall studies. This will begin to close the gap between the developmental and adult literatures on dual memory processes by securing stable evidence about age trends in direct access, reconstruction, and familiarity judgment during child-to-young-adult development. (Evidence about age trends in these processes during healthy aging is considered in a later section.) The principal resource for the production of these findings is a corpus of 207 developmental recall data sets from studies that used the $S_1T_1T_2$, S_2T_3, S_3T_4, … design that, as we have seen, is required by our identifiable model. These data sets are described more fully in the Appendix B. When our identifiable model was fit to these data sets, it gave statistically acceptable accounts of all of them. Actually, two types of fit tests were conducted for each data set: (a) a comparative fit test for the two-stage model versus a one-stage model followed by (b)

a fit test for the two-stage model. Concerning a, the logical alternative to the present model is one that assumes that learning to recall involves only a single process and, therefore, only one stage. A goodness-of-fit test for that alternative one-process model is described in Appendix A (Equation A22). That test was computed for the data sets in our corpus, and in all instances, the null hypothesis that learning to recall involves only a single stage was rejected at high levels of confidence. Concerning b, the fit test for the two-stage model, which is also described in Appendix A (Equation A17), evaluates the null hypothesis that learning to recall involves two stages, which are interpreted as learning how to directly access items versus learning how to reconstruct items. This test was computed for the data sets in our corpus, and the level of fit was satisfactory in each case (i.e., the value of the G^2 statistic in Equation A17 did not produce a null hypothesis rejection). As discussed in Appendix A, this pair of fit tests establishes that learning to recall did not involve less than two stages or more than two stages.

The corpus contains a subgroup of 152 data sets in which samples of children (mean age = 7-8) and young adolescents (mean age = 11-12) learned to recall the same lists under identical conditions. There is another subgroup of 44 data sets in which samples of young adolescents (mean age = 11-12) and young adults (mean age = 20-21) learned to recall the same lists under identical conditions. The former collection of data sets is the focus of discussion in the first subsection, below, and the latter collection is the focus of discussion in the second subsection.

Childhood to Adolescence

In this subsection, we examine the development of direct access, reconstruction, and familiarity judgment during the child-to-adolescent years, relying on the 152 data sets in which groups of children and young adolescents learned to recall lists of varying length and composition under paired-associate, free, or cued recall conditions (see Appendix B). The exploration of these data sets has two complementary aims. The first is to document that it is possible to make rapid progress on developmental questions about trichotomous memory processes by deriving estimates of the parameters of Equation 3 from the recall performance of subjects of different ages. The other, more fundamental purpose is to test theoretical hypotheses—explicitly, to produce findings that bear on the theoretical interpretations of the *D*, *R*, and *J* parameters. This is done by determining whether those parameters behave in ways that are consistent with the interpretations that we have discussed. In order to accomplish both objectives simultaneously, descriptive findings about age variability in the parameters are not the centerpiece of the presentation, although findings of that sort are considered. Rather, the focus is on particular predictions about the parameters that follow from theoretical conceptions of the corresponding memory processes and from recent analyses of how those processes might be expected to evolve during early memory development (e.g., Ghetti, 2008; Ghetti & Angelini, 2008; Holliday & Hayes, 2000, 2002; Lampinen, Leding, Reed, & Odegard, 2006). Hence, the evidence that we present is organized into four groups of parametric results, each of which bears on a particular group of theoretical hypotheses: (a) global developmental trends in direct access, reconstruction, and familiarity judgment; (b) negative relations between direct access and reconstruction; (c) direct access, with and without reconstruction; and (d) variations in familiarity judgment as a function of variations in items' familiarity.

Global Developmental Trends

We begin with overall changes in direct access, reconstruction, and familiarity judgment during childhood, testing three predictions about those changes that follow from theoretical distinctions and from extant developmental work on dual processes (using other methodologies). The first is an age-variance prediction about values that are observed for the *D* and *R* parameters. For two pairings of these parameters, D_1 vs. R_1 and D_2 vs. R_2, the parameters are estimated at the same time, as items escape state *U* on Trial 1 (D_1 vs. R_1) or on later trials (D_2 vs. R_2). Without a process theory of what these parameters measure, there is no basis for making predictions about whether it is easier to escape *U* by jumping to *L* or jumping to *P*, but the earlier theoretical distinctions provide a basis for directional predictions about both pairs of parameters. We saw that when items are in *U*, learning to directly access them is harder, in principle, than learning to reconstruct them because storing verbatim traces that can survive from trial to trial, in the face of accumulating interference, is a more difficult proposition than simply selecting meaning features of targets that will deliver small, correct search sets. Therefore, if D_1 and D_2 measure the first type of learning and R_1 and R_2 measure the second, the relations $D_1 < R_1$ and $D_2 < R_2$ should be observed.

The other two predictions are age-change predictions that are based on prior

developmental research and concern direct access vs. familiarity judgment. It was previously noted that although the number of extant developmental dual-process studies is small, recollection has usually been found to increase more than familiarity during childhood and familiarity has sometimes failed to increase at all (but cf. Ghetti & Angelini, 2008). Thus, the other two predictions are that recollection parameters will increase more than familiarity judgment parameters and that the latter may not increase at all.

The mean values of the four direct access parameters, the two reconstruction parameters, and the four familiarity judgment parameters for the complete collection of data sets are plotted by age level in the upper panel of Figure 3. Age differences in parameter values were tested for statistical significance with t tests. The average age improvement in direct access was reliable, $t\ (150) = 3.35$, $p < .001$, the average age improvement in reconstruction was reliable, $t\ (150) = 3.69$, $p < .0001$, but the average age improvement in familiarity judgment was not reliable, $t\ (150) = 1.52$. The first prediction, that R parameters ought to have larger values than D parameters, regardless of age level, was tested with a 2 (age) x 2 (parameters: the mean of R_1 and R_2 vs. the mean of D_1 and D_2) analysis of variance (ANOVA). There was a main effect in the predicted direction (reconstruction parameters being larger than direct access parameters), $F\ (1, 150) = 248.04$, $MSE = .02$, $p < .0001$, and the Age x Parameter interaction was not reliable, $F\ (1, 150) = 2.85$.[3] The absence of an interaction is instructive not only because it confirms the predicted parametric relation but also because it suggests that although direct access and reconstruction both become easier with age, memory development (during childhood at least) does not strongly favor one over the other. Concerning the other two predictions, both have already been confirmed by the t tests. Those tests showed that the mean value of the direct access parameters increased with age but the mean value of the familiarity judgment parameters did not, so that direct access increased more than familiarity judgment and, indeed, familiarity judgment did not exhibit reliable increases.

The global picture of developmental change during childhood, then, is that direct access and reconstruction both improve and by comparable amounts, whereas familiarity judgment does not. With respect to the latter result, age-invariance in the familiarity judgment parameters does not automatically mean that the familiarity of reconstructed items is age-invariant. Remember, here, that the probability that a reconstructed item is deemed to be familiar enough to output depends on two factors, the strength of the item's familiarity signal *and* the stringency of the decision criterion. These factors have opposite effects on the familiarity judgment parameters, with parameter values increasing as the familiarity signal becomes stronger and decreasing as the criterion becomes more stringent. However, interpretation is aided by the fact that there is a developmental literature on recognition in which signal detection estimates of criterion stringency have been computed for subjects who have ranged in age from young children to older adults (both healthy and impaired subjects). The life-span pattern runs as follows: (a) Decision criteria become much more stringent between childhood and adolescence (e.g., Holliday & Weekes, 2006); (b) decision criteria become slightly more stringent between adolescence and young adulthood (e.g., Brainerd & Mojardin, 1998); (c) decision criteria remain invariant between young and late adulthood (e.g., Budson, Todman, & Schacter, 2006b); (d) decision criteria are less stringent for older adults who are cognitively impaired than for healthy older adults (e.g., Budson et al., 2006b). Thus, in light of the first element of this pattern, the most likely interpretation of the age-invariance result for the familiarity judgment parameters is *not that familiarity is age-invariant* but, rather, that familiarity and criterion stringency both increase with age and these increases cancel each other out at the level of parameter values.

This interpretation can be tested by using the signal detection model in Figure 4 to separate familiarity from criterion stringency at both age levels. This model is just the standard signal detection representation (e.g., Snodgrass & Corwin, 1988) of the distributions of familiarity values of presented items versus unpresented items—in this case, the familiarity distributions of reconstructions of presented versus unpresented items. Thus, the signal detection parameters d' and C have the usual interpretations: d' is the measure of familiarity (specifically, of the distance between the two familiarity distributions), and C is the measure of criterion stringency. To estimate these parameters for recall, using signal detection equations, two empirical

quantities are required: the value of the *J* parameter and the intrusion probability (i.e., the probability of recalling unpresented items). The former quantity is available, but the intrusion probability is not because, in the original analyses of these data sets, recall protocols were not scored for intrusions. The raw protocols for most of 152 data sets were no longer available, and hence, they could not be rescored. However, raw protocols were still available for 28 matched data sets (14 for children and 14 for adolescents). When they were rescored, the mean intrusion probability was .051 for children and .016 for adolescents.[4] To compute values of *d'* and *C*, we used these two values, together with the mean value of the four *J* parameters for children (.61) and adolescents (.64). The estimated values of *d'* were 1.92 (children) and 2.51 (adolescents) and the estimated values of *C* were .68 (children) and 1.07 (adolescents). Thus, although estimates of the judgment parameters were age-invariant, the signal detection results suggest that familiarity signals were becoming stronger with age, while decision criteria were becoming more stringent.

Separability of Direct Access and Reconstruction

Next, we use these data sets to test some predictions about empirical dissociations between direct access and reconstruction that are analogous to predictions that have long been regarded as fundamental in adult dual-process research, though the latter predictions have focused on recollection and familiarity (rather than reconstruction). In the adult literature, since Mandler's (1980) seminal paper, a featured hypothesis has been that manipulations that stimulate the types of processing that underlie recollection ought to selectively affect measures of that memory process, whereas manipulations that stimulate the types of processing that underlie familiarity ought to selectively affect measures of that memory process (see Jacoby & Kelley, 1992). Mandler, for instance, discussed evidence of recollection/familiarity dissociations in patient populations in which one operation was thought to be deficient and the other was thought to be spared. In the subsequent literature, two basic forms of dissociation have been studied (for a review, see Yonelinas, 2002). By far the most commonly studied variety consists of single dissociations, circumstances in which theoretically-motivated manipulations affect measures of one process but not measures of the other (e.g., Donaldson, 1996; Jacoby, 1991; Gardiner & Java, 1991). The less commonly studied variety consists of double dissociations, circumstances in which theoretically-motivated manipulations drive measures of the two processes in opposite directions (e.g., Buchner & Wippich, 2000; Dunn & Kirsner, 1988; Howe, Rabinowitz, & Grant, 1993). Technically, as Dunn and Kirsner pointed out, there are two types of double dissociations, crossed and uncrossed, with circumstances in which a manipulation drives two measures in opposite directions being the crossed variety. Although, as Dunn and Kirsner also pointed out, it is possible to devise one-process models that will produce double dissociations, such interpretations can be ruled out with additional evidence that other manipulations drive the same two measures in the same direction (see below).

In the present conception of recall, the two methods of recovering items are distinct and separable inasmuch as they operate on dissimilar types of episodic traces, with the formation of one type of trace being akin to rote memorization and the formation of the other type being akin to conceptual understanding. For that reason, direct access and reconstruction pose different problems for learning, so that they ought to be dissociable within given age levels. We investigated this possibility in the broadest way that was open to us, by using our child versus young adolescent data sets to test for opposing effects of a manipulation that is present in all 152 data sets, repetition of study-test trials. We also evaluated the related prediction that *within each age level* estimates of *D* and *R* should be negatively correlated over the pool of data sets. We consider the two types of evidence separately.

Repetition Has Opposite Effects on Direct Access from Reconstruction

When trials are repeated, this means that both study and test cycles are repeated. Many findings show that these cycles have different effects on recall (e.g., Halff, 1977) and that some of the effects are negative (e.g., Brainerd & Reyna, 1993), even though sheer accuracy improves over trials. Test cycles, in particular, have been found to have negative as well as positive effects. On the negative side, recall tests generate output inference and because direct access is a rote memorization process that involves interference-sensitive verbatim traces, accumulation of interference impairs it (Brainerd et al., 2002). On the positive side, because learning how to reconstruct a target is a comprehension process, additional recall tests provide practice at using targets' features to quickly regenerate them (Brainerd et al., 2003).

Taken together, this means that, other things being equal, direct access will become more difficult but reconstruction will become easier as recall tests accumulate.

Now, consider the two parameters that measure the difficulty of learning how to directly access targets in the unlearned state, D_1 and D_2, and the two parameters that measure the difficulty of learning how to reconstruct targets in the same state, R_1 and R_2. Remember from Equation 3 that D_1 and R_1 measure the difficulty of direct access and reconstruction on the *first* recall test, while D_2 and R_2 measure the difficulty of direct access and reconstruction on the *third* recall test. The reconstruction practice provided by the first two recall tests should benefit R_2, relative to R_1, but the accumulated output interference from those tests should disadvantage D_2, relative to D_1. Thus, it should be harder for a target that is in state U to become reconstructable on the first trial (i.e., $R_1 < R_2$), but it should be easier for a target that is in state U to become directly accessible on the first trial (i.e., $D_1 > D_2$). The prediction, then, is that repetition should drive reconstruction up and direct access down. The second half of this prediction is counterintuitive, of course. Because the net accuracy of recall improves over trials, the baseline expectation is that all parameters that make performance more accurate should have larger values on later trials than on earlier trials. The theory says otherwise.

The relevant data are displayed in the upper panel of Figure 5. Pooling across the data sets for each age level, it can be seen that repetition of study-test trials has opposite effects on direct access and reconstruction at each age level. For both children and young adolescents, repetition *increased* the chances that targets would become reconstructable by roughly 60%, but it *decreased* the chances that they would become directly accessible by roughly 40%. To test this pattern for statistical reliability, we computed a 2 (age) x 2 (trial: first vs. later) x 2 (memory process: direct access vs. reconstruction) ANOVA, using estimates of the D and R parameters as dependent variables. The principal result, as Figure 5 implies, was a large Trial x Memory Process interaction, F (1, 150) = 207.36, MSE = .01, p < .0001. When the interaction was decomposed with post hoc tests (paired-samples t tests that controlled alpha at the .05 level), it was found that D_1 was larger than D_2 but that R_1 was smaller than R_2. There was also an instructive Age x Trial x Memory Process interaction, F (1, 150) = 11.57, MSE = .01, p < .001, the nature of which is apparent in Figure 5. A classic finding about memory development is that when children study and recall meaningful items, younger children are less likely than older children and adolescents to benefit from opportunities to extract meaning content (Bjorklund, 1987, 2004; Bjorklund & Muir, 1988). Another, less well-known, result is that output interference has more pronounced effects on older children's recall than on younger children's (e.g., Brainerd, Olney, & Reyna, 1993). Consistent with these results, post hoc analysis of the Age x Trial x Memory Process interaction revealed that (a) the age improvement in reconstruction was smaller on Trial 1 than on subsequent trials, whereas (b) the age improvement in direct access was larger on Trial 1 than on subsequent trials. In other words, repetition magnified developmental improvements in reconstruction, but it did the opposite to improvements in direct access.

Summing up, analysis of the parameter pairs (D_1, D_2) and (R_1, R_2) yielded evidence of the anticipated opposite effects of repetition of study-test trials. Although this pattern was present at both age levels, it was amplified by development. Specifically, the tendency of repetition to make reconstruction easier was more marked in older children, as was its tendency to make direct access more difficult. Further, this pattern cannot be interpreted as showing that the (D_1, D_2) and (R_1, R_2) pairs measure a *single* process and that they just happen to respond in opposite ways to the repetition variable. As Dunn and Kirsner (1988) showed, such an interpretation can be ruled out if there are other variables that produce *positive* associations between the (D_1, D_2) and (R_1, R_2) pairs. As can be seen in Figure 6A, age produces such a positive association; that is, all parameter values rise with age.

Parameter Correlations in Children and Adolescents

We test some further predictions about correlations within and between the parameter pairs (D_1, D_2) and (R_1, R_2). As mentioned, certain variables that are known to affect the difficulty of recall were manipulated across the list conditions in our corpus, such as the types of targets that were studied (e.g., abstract nouns, concrete nouns, pictures, category exemplars), the types of recall tests that were administered (e.g., paired-associate, cued, free), and the lengths of lists (see Appendix B). Such manipulations were included in the designs of individual experiments pursuant to the aims of those experiments and not with the objective of evaluating theoretical predictions about direct access, reconstruction, and familiarity judgment. Although those manipulations were not designed to test

theoretical predictions, the fact that they affected the accuracy of children's recall means that they must have affected some of the parameters of the present model. Because the nature of the manipulations varied widely, their parametric effects (i.e., which specific parameters were influenced) should correspondingly vary. This also allows one to ask how values of the reconstruction and direct access parameters should *covary* over these list conditions.

Here, two types of predictions bear on the earlier analysis of direct access and reconstruction. The first is about positive correlations. Ostensibly, D_1 and D_2 both measure the same process (the tendency of targets to become directly accessible when they escape state *U*), and likewise, R_1 and R_2 both measure the same process (the tendency of targets to become reconstructable when they escape state *U*). Thus, as parameter values vary over the list conditions in our data sets, D_1 and D_2 ought to covary positively and so should R_1 and R_2, and this should be true at both age levels. The other prediction is about negative correlations. Theoretically, direct access and reconstruction do not enable each other but, rather, rely on rather different forms of learning. Generally speaking, the list conditions that ought to help targets escape state *U* by becoming directly accessible are ones that make their surface forms more distinctive, whereas the list conditions that should help targets escape state *U* by becoming reconstructable are ones that make it easier to extract certain meanings. Such manipulations have been studied in the contemporary false-memory literature (for a review, see Brainerd & Reyna, 2005), and a common finding has been that making targets' surface forms more distinctive reduces semantic processing and making particular meanings more accessible reduces surface processing (e.g., Arndt & Gould, 2006; Hege & Dodson, 2004; Koutstaal, 2003; Reyna & Kiernan, 1995; Schacter, Israel & Racine, 1999; Seamon et al., 2002a, 2002b). The implication for our data sets is that conditions that enable direct access will be apt to interfere with reconstruction and vice versa, so that correlations between (D_1, D_2) and (R_1, R_2) should be negative.

To evaluate such predictions, we computed the 4 x 4 matrix of bivariate correlations for D_1, D_2, R_1, and R_2 for the pool of data sets for children, and we computed the same matrix of bivariate correlations for the corresponding pool of data sets for young adolescents. The results are reported in Table 3, where it can be seen that the bivariate correlations fell out as theory expects. Concerning the first prediction, there were positive correlations between D_1 and D_2 and between R_1 and R_2 at both age levels, which is congruent with the idea that the two *D* parameters are measuring the same type of memory process and the two *R* parameters are measuring the same type of memory process. With respect to the second prediction, there were negative correlations between D_1 and R_1, between D_2 and R_2, between D_1 and R_2, and between D_2 and R_1, at both age levels. Thus, the correlational evidence, like the findings for repetition, was consistent with the notion that the *D* parameters do not measure the same type of memory process as the *R* parameters and with the further conclusion that these processes can interfere with each other.

Direct Access, With and Without Reconstruction

Although R_1 and R_2 are the exclusive measures of reconstruction, there are two further measures of direct access, D_{3C} and D_{3E}. The latter parameters take account of the fact that (a) subjects can learn how to directly access a target when it is in state *U* or when it is in state *P* and the fact that (b) a target in state *P* can become directly accessible following a trial on which recall is successful (parameter D_{3C}) or a trial on which recall is unsuccessful (parameter D_{3E}). The present theoretical distinctions yield predictions about how difficult it is for a target to become directly accessible on trials when it is in state *U* versus state *P*, as well as further predictions about how difficult it is for a target to become directly accessible following successful versus unsuccessful recall in state *P.* With respect to state *U* versus state *P*, obviously it should be easier for targets to become directly accessible in state *P* than in state *U* because subjects receive additional covert study opportunities in state *P* by virtue of the fact that targets are reconstructable. To clarify this point, when targets are in state *U*, they are *not recovered* on recall tests, and thus, the only exposure that subjects receive to such targets is when they are physically presented on study cycles. For targets that enter state *P* when they leave U, the situation is different. Because subjects are able to reconstruct the targets on recall tests, they receive physical exposures on study cycles, plus additional covert exposures (via reconstructive retrieval) on recall tests. Therefore, across many data sets, such as those in our corpus, the relation between estimates of the parameters in the (D_1, D_2) set versus the (D_{3C}, D_{3E}) set should be (D_{3C}, D_{3E}) > (D_1, D_2). There is a further prediction about D_{3C} and D_{3E}. The difference between these parameters, it will be remembered, is that D_{3C} is the probability that a target becomes directly accessible following reconstruction and successful recall, whereas D_{3E} is the probability that a

target becomes directly accessible following reconstruction and unsuccessful recall. A well-established principle of rote memorization (e.g., Brainerd et al., 1993; Howe, 2004) is that active mnemonic processing (e.g., rehearsal, imagery, and elaboration) is helpful to such learning and that errors inform subjects that such effort is needed. According to this principle, when a target occupies state *P* and can therefore be reconstructed, unsuccessful recall (i.e., the judgment that a reconstructed item is not familiar enough to be output) is more likely to initiate active mnemonic processing of the item on the next study cycle. It follows that across many data sets, the relation $D_{3E} > D_{3C}$ should hold.

Results that bear on these predictions appear in Figure 7, where the mean values of the four direct access parameters are plotted for younger and older subjects. At each age level, the trends conform to the two predictions that were just mentioned. The mean values of D_{3E} and D_{3C} at each age level are larger than the corresponding means for D_1 and D_2, and the mean value of D_{3E} at each age level is larger than the corresponding mean for D_{3C}. To test these results for statistical reliability, we computed a 2 (age) x 4 (direct access parameters: D_1, D_2, D_{3E}, D_{3C}) ANOVA, using the parameter estimates in each data set as dependent variables. The key finding that confirmed the two predictions was a main effect for direct access, F (3, 150) = 280.75, MSE =.01, $p < .0001$, that when decomposed by post hoc tests, showed that the order of difficulty of learning to directly access targets was $D_2 < D_1 < D_{3C} < D_{3E}$ at both age levels. Thus, the data were consistent with the notion that $M_{D1/D2} < M_{D3C/D3E}$ because D_{3C} and D_{3E} involve additional covert study opportunities and with the notion that $D_{3E} > D_{3C}$ because unsuccessful recall initiates active mnemonic processing. There was no Age x Direct Access Parameter interaction, so that the beneficial effects on direct access of being able to reconstruct a target and of unsuccessful recall of reconstructed targets were comparable in younger and older children.

Familiarity Judgment Predictions

Up to this point, we have reported one important finding about familiarity judgment—namely, that its mean level did not increase with age in our data corpus. Beyond this null result, which is consistent with prior developmental findings on familiarity using conventional adult methodologies, positive predictions about familiarity judgment parameters can be made, which are concerned with inter-trial variations in this process. There are four parameters that measure familiarity judgment, J_1, J_2, J_{3E}, and J_{3C}, which is to say that each parameter measures the perceived familiarity of reconstructed targets and the stringency of the decision criterion, though at different stages of learning. The fact that the parameters apply to different stages of learning makes it possible to forecast an ordering, based on item selection. It is well established that item selection operates during the course of recall experiments: Items that are first recalled on earlier trials are in some sense easier than items that are first recalled on later trials, and the usual assumption is that easier items are more familiar (e.g., Greeno, James, & DaPolito, 1971). In that connection, J_1 measures familiarity judgment for items that become reconstructable on Trial 1, J_2 measures familiarity judgment for items that become reconstructable on Trial 2 or later, and J_{3E}, and J_{3C} measure familiarity judgment for items that have been waiting to escape state *U* for at least two trials. Assuming that the stringency of the decision criterion is constant, the theory says that the ordering of these parameters will reflect targets' familiarity, which should reflect item selection. Obviously, item selection has operated for items that occupy state *P*, relative to items that occupy state *U*, because the former are a more difficult subset of the latter. Thus, $M_{J1/J2}$ should be larger than $M_{J3C/J3E}$. For the same theoretical reasons, J_{3C} obviously should be larger than J_{3E}. The only difference between these two latter parameters is that J_{3C} measures familiarity judgment for reconstructed items that were deemed familiar enough to be output on the immediately preceding recall test, whereas J_{3E} measures familiarity judgment for reconstructed items that were *not* deemed familiar enough to be output on the immediately preceding test. As the latter items are, by definition, less familiar than the former, $J_{3E} < J_{3C}$ follows.

These predictions were evaluated by analyzing the estimated values of the familiarity judgment parameters at each age level in our data sets. The parameters' mean values are plotted by age level in Figure 8, where it can be seen that the pattern of inter-parameter variability was $J_{3E} < J_{3C} < J_2 < J_1$ at both age levels, which is consistent with both predictions ($M_{J1/J2} > M_{J3C/J3E}$ and $J_{3C} > J_{3E}$). To test the overall pattern for statistical significance, we computed a 2 (age) x 4 (judgment parameter: J_1, J_2, J_{3E}, J_{3C}) ANOVA, using the estimates of these parameters as dependent variables. The finding of principal interest was a main effect for judgment parameter, F (3, 150) = 169.92, MSE = .02, $p < .0001$, that when decomposed by post hoc tests, showed that the magnitude ordering of the judgment parameters was $J_{3E} < J_{3C} < J_2 < J_1$ at both age levels. In addition, there was an Age x Judgment Parameter interaction, F (3, 450) = 4.55, MSE = .02, $p < .005$. Post hoc tests revealed a simple pattern. On the one hand, consistent

with the developmental results that we reported earlier for the mean values of these parameters (Figure 3) three of the four parameters—specifically, J_1, J_2, and J_{3C}—were age-invariant. However, the fourth judgment parameter, J_{3E}, increased reliably with age (means = .39 and .47). Hence, there is some limited evidence of developmental improvement in familiarity judgment, though it is miniscule in comparison to the improvements in direct access and reconstruction. In this connection, it is important to remind ourselves that it does not follow that age increases in the underlying familiarity of reconstructed items are miniscule because, as we saw, increases in familiarity and increases in criterion stringency have opposite effects on the values of the familiarity judgment parameters. To illustrate, we used the signal detection model (Figure 4) and the intrusion probabilities that were mentioned earlier to compute *d'* and *C* values for the J_{3E} parameter. The estimated values of *d'* were 1.36 (children) and 2.07 (adolescents) and the estimated values of *C* were .96 (children) and 1.11 (adolescents). Thus, the results suggest that the developmental improvement in J_{3E} is due to a large increase in the strength of the familiarity signal that was not canceled by the smaller increase in criterion stringency.

A final important result is secured by comparing the plotted values of D_{3E} and D_{3C} in Figure 7 to the plotted values of J_{3E} and J_{3C} in Figure 8, which reveals a variable that has opposite effects on these parameter pairs. According to the present theory, the *D* parameters measure different processes than the *J* processes—specifically, the *D* parameters measure subjects' ability to store interference-resistant verbatim traces while the *J* parameters measure their willingness to output targets that have been semantically reconstructed. Consistent with the notion that the *D* and *J* parameters measure different processes, note that the two parameters react in opposite was to performance on the preceding recall test. While items are waiting in state *P*, a recall error makes direct access learning *easier* on the next trial than a success does, whereas a recall error makes familiarity judgment *harder* on the next trial.

Process Specificity of Model Parameters

We have considered several predictions about the *D*, *R*, and *J* parameters that follow from their process definitions. Although each is theoretically well specified, it might be argued that other results that bear narrowly on the parameters' process definitions would be desirable. More explicitly, it might be argued that the predictions that have been evaluated are general ones (e.g., that $M_{R1/R2}$ should be larger than $M_{D1/D2}$, that $M_{D3C/D3E}$ should be larger than $M_{D1/D2}$, that $M_{J1/J2}$ should be larger than $M_{J3C/J3E}$) that could be consistent with other (unspecified) process interpretations of the parameters. Thus, the argument continues, it would be desirable to determine how the parameters react to surgical manipulations that precisely embody their process definitions.

Some immediate evidence can be generated, owing to the presence of a highly surgical manipulation in several of the data sets in our corpus. This is a manipulation that should enhance the direct access process that was described earlier while simultaneously interfering with the reconstruction process--namely, category cuing. Suppose that subjects learn to recall a categorized list; that is, words belong to a few familiar taxonomic categories, such as animals, body parts, colors, and furniture. In an uncued condition, subjects study the list, and on test cycles, they recall it under standard free recall instructions. In a cued condition, the procedure is the same, except that on test cycles, each category label is presented in turn, and subjects are asked to output as many exemplars from that category as possible before moving on to the next category. Obviously, presenting such cues should facilitate the specific process that we described in connection with direct access, relative to the uncued condition. This procedure enriches the learning environment with further, salient contextual cues. Subjects can store these contextual cues as part of verbatim traces, which will make it easier to access those traces because the cues are presented as retrieval prompts during each recall cycle. Just as obviously, category cuing ought to interfere with the specific process that was described in connection with reconstruction. Here, remember from the earlier example of reconstructing *dog* and *tiger* using "animal" versus "household pet" and "jungle cat" that generic features such as "animal" are poor reconstruction features because they over identify large search sets. Category labels are broad features of this sort. Hence, they should interfere with the identification of more specific features that deliver small search sets. In sum, under the process definitions of what the

D and *R* parameters measure, category cuing should increase the former while decreasing the latter.

Our corpus contains several data sets in which groups of children and adolescents learned to recall the same categorized list with either cued or uncued recall. Specifically, 14 of the child data sets and 14 of the adolescent data sets were paired sets in which subjects learned either a two-category list under cued versus uncued recall conditions, or they learned a four-category list under those conditions. The results fell out as predicted: At both age levels, estimates of the *D* parameters were higher in cued conditions, whereas estimates of the *R* parameters were lower in cued conditions. Illustrative findings from an article by Howe et al. (1989), which reports four matched data sets for children in the normal ability range and four matched data sets for adolescents in the normal ability range, are shown in Figure 9. Concerning direct access, it can be seen in panel A that (a) category cuing nearly quadrupled the value of the D_1/D_2 pair in children and tripled it in adolescents and (b) category cuing increased the value of the D_{3C}/D_{3E} pair by roughly 50% in children and roughly 60% in adolescents. In contrast, it can be seen in panel B that category cuing decreased the value of the R_1/R_2 pair by roughly two-thirds in children and roughly one-quarter in adolescents.

Further evidence about process specificity comes from a new experiment that is not in our corpus. Brainerd et al. (2002) proposed that list length variations in free recall should affect direct access but not reconstruction. Their argument was simple. On the one hand, differences in list length ought to affect a specific variable to which direct access is sensitive, accumulating interference: The more words that subjects have to study and recall, the more interference will be generated from trial to trial. On the other hand, as long as lists of different lengths are matched on all other variables, there is no reason to expect that length will affect subjects' ability to construct search sets for individual items. Thus, *D* parameters should be larger for shorter than for longer lists, but *R* parameters should not be affected. We tested this prediction in an experiment in which subjects of two age levels (younger = 6- and 7-year-olds; older = 11- and 12-year-olds) learned to recall lists of familiar concrete nouns to an errorless criterion under standard free-recall conditions like those of the data sets in our corpus. At each age level, 25 subjects learned to recall a list of 12 words, and 25 subjects learned to recall a list of 16 words.

As with category cuing, the results for list length, which are plotted in Figure 10, fell out as predicted. At both age levels, estimates of the *D* parameters were higher for 12-item lists than for 16-item lists. Those data appear in panel A, where it can be seen that (a) decreasing list length more than doubled the value of the D_1/D_2 pair for children and increased it by roughly 50% for adolescents and (b) decreasing list length increased the value of the D_{3C}/D_{3E} pair by roughly 50% in children and by more than one-quarter in adolescents. In contrast, it can be seen in panel B that list length had no appreciable effect on the value of R_1/R_2 pair. Although the value of R_1/R_2 was larger for adolescents than for children, which is consistent with findings that we have already reported (Figure 3), it can be seen that this parameter pair was not affected by list length at either age level.

Additional evidence that bears on process specificity can be found in some adult experiments by Brainerd et al. (2002, 2003), which used a simplified version of the present model that is described in the fourth section of this paper (Cognitive Impairment). Brainerd et al. (2002) predicted (a) that presenting lists in distracting fonts should decrease *D* parameters (because it is harder to encode targets' surface forms) but should not affect *R* or *J* parameters (because neither targets' semantic features nor the familiarity of reconstructions is altered), (b) that recalling abstract nouns (e.g., *concept, mind*) rather than concrete nouns (e.g., *book*, *piano*) should simultaneously decrease *R* parameters (because concrete nouns divert processing away from semantic features by generating vivid visual images) and increase *J* parameters (because a reconstruction that is accompanied by vivid phenomenology will seem more familiar), and (c) that recalling longer lists rather than shorter ones should increase *D* parameters without affecting *R* or *J* parameters. All three patterns were obtained. Brainerd et al. (2003) predicted (d) that presenting three study cycles per recall test rather than the usual single study cycle should increase *D* parameters (because verbatim traces are labile) but should not affect *R* parameters (because semantic features are stable) and (e) that administering recall tests a few days after lists are studied rather than immediately after should decrease *D* parameters (again, because verbatim traces are labile) but should not affect *R* parameters (again, because semantic features are stable). Both patterns were obtained.

Taken together, these findings provide support for the process specificity of the parameters of our model. In the developmental data that were just reported (Figures 9 and 10), the parameters of the recall model reacted appropriately to surgical manipulations that precisely embody the process definitions of direct access and reconstruction. In prior adult experiments that Brainerd et al. (2002, 2003) reported, the parameters of a simplified model reacted appropriately to surgical manipulations that precisely embody the process definitions of direct access, reconstruction, and direct access.

Summary

We estimated the parameters of the trichotomous model of recall, using a corpus of 152 sets of recall data in which the subjects were children and young adolescents. This generated a picture of developmental changes in direct access, reconstruction, and familiarity judgment during childhood, and more important, it supplied tests of many theoretical predictions about these processes. The developmental picture ran as follows. The ease with which targets become directly accessible and reconstructable both increased between childhood and adolescence by comparable amounts. Although there was only a slight increase in familiarity judgment that was confined to one of the four J parameters, supplementary signal detection analyses revealed that the familiarity of reconstructed targets increased with age and that criterion stringency also increased. With these findings, we have progressed from the situation that was described at the start of this paper, in which our knowledge of the early development of dual memory processes was thin and inconsistent, to a situation in which child-to-adolescent trends in direct access, reconstruction, and familiarity judgment have been established for a large and varied data base.

Turning to theoretical predictions, the most fundamental one is the notion that the D and R parameters measure distinct processes—namely, the direct access and reconstruction operations of the present theory. Here, the data corpus produced two lines of evidence that were consistent with this hypothesis. First, repetition of study-test trials tests had opposite effects of the D and R parameters, decreasing the former (in line with idea that verbatim traces are sensitive to accumulating output interference) while increasing the latter (in line with the idea that recall tests provide reconstruction practice). Second, across list conditions, at both age levels, values of the direct access parameters correlated negatively with values of the reconstruction parameters. Such results cannot be interpreted as showing that direct access and reconstruction parameters simply measure opposite sides of a single process because both types of parameters are *positively* related to other variables, such as age.

Other results were consistent with (a) the hypothesis that the D and J parameters measure different processes and (b) the hypothesis that the R and J parameters measure different processes. Concerning a, successful recall had opposite effects on the D and J parameters for state P, decreasing the former while increasing the latter. Concerning b, repetition of study-test trials had opposite effects on the R and J parameters for items that are in state U: Repetition increased the reconstruction probability ($R_1 < R_2$) but *decreased* the probability that reconstructions would be output ($J_1 > J_2$).

Further predictions about the four D parameters and the four J parameters were evaluated that follow from our process conceptions of direct access and familiarity judgment. With respect to the D parameters, it was found, as predicted, that it is harder for items to become directly accessible when they are in state U than when they are in state P (because reconstruction produces covert target presentations in state P) and it was found, also as predicted, that items in state P are more likely to become directly accessible following unsuccessful recall than following successful recall (because errors are more informative than successes). With respect to familiarity judgment, it was found, as predicted, that J_1 and J_2 identified more items as being familiar enough to be recalled than either J_{3E} or J_{3C}, (because item selection has operated for items that occupy state P) and it was found, also as predicted, that J_{3C} identified more items as being familiar enough to output than J_{3E}. The specificity of the process definitions of the direct access and reconstruction parameters were investigated by studying the effects of a manipulation (category cuing) that ought to make direct access easier while making reconstruction harder. As predicted, D parameters had larger values but R parameters had smaller values in cued than in uncued conditions.

Early Adolescence to Young Adulthood

Next, we consider developmental changes in the same processes between early adolescence and

young adulthood. It is possible to do so because our corpus also includes 39 data sets in which young adults (college students) learned similar types of lists to criterion under paired-associate, free, or cued recall conditions (see Appendix B). In a subgroup of 22 of these 39 data sets, the experimental conditions for adults were identical to those in a subgroup of 22 of the 76 data sets for young adolescents. For these matched pairs of data sets, it is possible to chart the changes that occur between early adolescence and young adulthood by comparing estimates of the direct access, reconstruction, and familiarity judgment parameters that were obtained under identical experimental conditions.

The evidence that is presented in this section unfolds in two steps. First, young adults are a population of special interest because they provide the subject samples for most dual-process experiments in the mainstream memory literature. Thus, apart from developmental questions, it is important to know how the direct access, reconstruction, and familiarity judgment parameters behave in young adult samples, and, in particular, whether that behavior accords with theoretical prediction. Evidence of this sort is considered first, using the full complement of 39 young adult data sets. Second, we return to developmental questions by examining adolescent-to-young-adult trends in direct access, reconstruction, and familiarity judgment. Here, we rely on the 22 sets of young adult and 22 sets of adolescent data in which subjects learned to recall the same lists under identical conditions.

Behavior of Dual-Process Parameters in Young Adults

Estimates of the model's direct access, reconstruction, and familiarity judgment parameters for the 39 young adult data sets were used to examine the earlier predictions about relations among direct access, reconstruction, and familiarity parameters, about the ordering of direct access parameters, and about the ordering of familiarity judgment parameters.

Separability of direct access and reconstruction. The notion that direct access and reconstruction are distinct processes was evaluated as before. First, we consider the effects of repetition of study-test trials on the D and R parameters. As mentioned, repetition has the interesting property of simultaneously providing additional opportunities to process the meaning content of targets (which should enhance reconstruction) and generating additional output interference (which should impair direct access). As in the earlier results for children and adolescents, it was found that the mean value of R increased as a function of repetition (mean R_1 = .30, mean R_2 = .41), and the mean value of D decreased (mean D_1 = .22, mean D_2 = .19), though the latter difference was not reliable. Second, we computed bivariate correlations within and between the members of the (D_1, D_2) set and the members of the (R_1, R_2) set. The results appear at the bottom of Table 3, where it can be seen that the picture was similar to that for younger and older children. On the one hand, it appeared that in young adults, D_1 and D_2 measure the same process in adults and so do R_1 and R_2 because in each case, there was a strong positive correlation between the two parameters. On the other hand, it appeared that the D and R parameters do not measure the same process because there was a significant negative correlation between D_2 and R_2, while D_1 and R_1 were uncorrelated. In short, the same types of findings were present in the adult data sets as in the child data sets, with the only notable difference being that the evidence of interference between direct access and reconstruction was somewhat weaker in the adult data sets because D_1 and R_1 were not negatively correlated and the repetition-induced decline from D_1 and D_2 was slight in comparison to the declines that were previously reported for children and adolescents.

Ordering of direct access parameters. The theoretically expected ordering of the direct access parameters is still $D_2 < D_1 < D_{3C} < D_{3E}$. For the 39 adult data sets, the mean values of these parameters were D_2 = .19, D_1 = .22, D_{3C} = .41, D_{3E} = .56, which follows the expected ordering. A one-way ANOVA produced a significant main effect for these parameters, F (3, 114) = 63.15, MSE = .02, p < .0001, which shows that they differed reliably. Post hoc tests revealed that whereas the increases from D_1 to D_{3C} and from D_{3C} to D_{3E} were both reliable, the increase from D_2 to D_1 was not. Because the drop from D_1 to D_2 is a measure of the susceptibility of direct access to accumulating output interference, the indicated conclusion is that young adults are better able to resist the debilitating effects of such interference than children or adolescents.

Ordering of familiarity judgment parameters. The theoretically expected ordering of the familiarity judgment parameters is $M_{J1/J2} > M_{J3C/J3E}$ and $J_{3C} > J_{3E}$. A one-way ANOVA produced a significant main effect for these parameters, F (3, 114) = 12.27, MSE = .02, p < .0001, which shows that they differed

reliably. The relevant mean values were $M_{J1/J2}$ = .82, $M_{J3C/J3E}$ = .77, J_{3C} = .85, and J_{3E} = .68, which conform to prediction. Post hoc tests of the main effect revealed that J_{3E} was reliably smaller than each of the other three familiarity judgment parameters (.79, .85, and .85) but none of the other three parameters differed reliably. Importantly, note that the mean value of these parameters (.79) is much higher than the mean value that was reported earlier for adolescents (.64) and is very high in absolute terms. The indicated conclusion is that reconstructed items seem very familiar to young adults, so much so that the item-selection process produces much smaller differences between the judgment parameters. Finally, note that when the present results for familiarity judgment are combined with the above results for direct access, prior recall performance again had opposite effects on the two types of parameters: A prior recall error, relative to a prior success, increased the probability of direct access (from .41 to .56) but it decreased the probability of familiarity judgment (from .85 to .68).

Summary. The adult patterns for the direct access and reconstruction parameters were similar to those for children and adolescents. Once again, the two types of parameters reacted differently to repetition of study-test trials (although in this instance, reconstruction reacted but direct access did not), the two types of parameters were negatively correlated, and the values of the four direct access parameters followed the theoretically predicted ordering. Also as before, performance on prior recall tests (success versus error) had opposite effects on familiarity judgment and direct access parameters. Concerning familiarity judgment per se, the data were again consistent with the $M_{J1/J2} > M_{J3C/J3E}$ and $J_{3C} > J_{3E}$ predictions. Another important similarity between the adult, child, and adolescent patterns is that once an item had been reconstructed, adults were also very willing to judge it as being familiar enough to output, the overall mean of the four J parameters being .79.

Adolescent-to-Adult Changes

Next, we examine developmental trends in direct access, reconstruction, and familiarity judgment, between early adolescence and young adulthood. To do that, we focus on the 44 data sets—22 for young adults and 22 for young adolescents—in which the subjects at both age levels learned to recall the same items under identical conditions. As before, we begin with global developmental trends and then move to trends for specific parameters, emphasizing points of similarity and difference between the adolescent versus young adult changes and the changes that were reported earlier for childhood versus adolescence. A key prediction, based on extant research, concerns age changes in direct access versus reconstruction.

Global trends. We tested the same three predictions about global age trends as before: (a) that regardless of age level, D parameters should be smaller than R parameters, on average; (b) that D parameters should increase more with age than J parameters; and (c) that J parameters might not increase at all. The first prediction is based on theoretical differences between direct access and reconstruction that make the former inherently more difficult, and hence, this prediction should be confirmed at all age levels. The other two predictions, as we saw earlier, are grounded in prior developmental findings about the childhood years, using other paradigms (e.g., Ghetti & Angelini, 2008), and hence, different patterns could be obtained for other age ranges.

Mean values of the four direct access parameters, the two reconstruction parameters, and the four familiarity judgment parameters are plotted for young adults and adolescents in Figure 3B. Because the mean values for adolescents are based on a subset of 22 of the 76 data sets that were used to estimate the mean values of these parameters in Figure 3A, the estimates are not identical. They are quite similar, however: The average difference between the mean values in the two figures is only .05 for the four direct access parameters, .05 for the two reconstruction parameters, and .03 for the four familiarity judgment parameters. None of these small differences was reliable. We saw that during childhood (Figure 3A), mean direct access and mean reconstruction both increased with age, but mean familiarity judgment did not. In contrast, inspection of Figure 3B shows that between early adolescence and young adulthood, familiarity judgment increases more than either direct access or reconstruction and that reconstruction is age-invariant. To test the first prediction (that reconstruction parameters are larger than direct access parameters), we computed a 2 (age) x 2 (parameters: the mean of R_1 and R_2 versus the mean of D_1 and D_2) ANOVA, using parameter estimates as dependent variables. There was a main effect in the predicted direction for

parameters, F (1, 42) = 36.18, MSE = .03, p < .0001 (mean of R_1 and R_2 = .40 and mean of D_1 and D_2 = .17). There was also an age main effect, F (1, 42) = 4.10, MSE = .01, p < .05, and the Age x Parameter interaction was not reliable.

To test the other two predictions (direct access develops more than familiarity judgment and familiarity judgment is age-invariant), we computed a 2 (age) x 2 (parameters: the mean of the four D parameters versus the mean of the four J parameters), with the result of principal interest being an interaction that was expected between the two factors. Although there were large main effects for age, F (1, 42) = 26.94, MSE = .01, p < .0001, and parameter, F (1, 42) = 140.09, MSE = .02, p < .0001, the interaction failed to materialize. As can be see in Figure 3B, although the age increases in mean values of the D and J parameters did not differ reliably, the increase in mean J (from .61 to .79) was twice the size of the corresponding increase in mean D (from .29 to .37).

Thus, at the process level, the global picture of what controls improvements in recall between early adolescence and young adulthood was different than the picture during childhood. Direct access and reconstruction were responsible for improvements in recall during childhood: Both increased with age and by comparable amounts, while familiarity judgment was age-invariant. In contrast, familiarity judgment was responsible for age improvements in recall between adolescence and young adulthood. This difference in age trends for familiarity judgment between childhood and adolescence versus between adolescence and young adulthood is particularly instructive because it demonstrates that the perceived familiarity of reconstructed items increases substantially between early adolescence and young adulthood. Unlike the findings for child-to-adolescent development, there is no ambiguity in this interpretation arising from the fact that the J parameters are combined measures of familiarity signal strength and criterion stringency, and hence, supplementary analyses of the sort that were conducted earlier are not required to buttress this interpretation. That is, as noted earlier, there are many studies of the adolescent-to-young-adult age range in which the criterion parameter of signal detection theory has been estimated, a general finding of which has been modest increases in stringency during this age range. Because increases in criterion stringency would *decrease* estimates of the J parameters, the fact that those parameters *increase* between early adolescence and young adulthood means that the strength of the familiarity signal is increasing.

Direct access and reconstruction. The relations between age, repetition, direct access, and reconstruction are shown in Figure 5B. For items that are still in state U, we saw that during childhood and early adolescence (Figure 5A): (a) Repetition of study-test trials makes it easier to reconstruct targets ($R_1 < R_2$); (b) repetition makes it harder to directly access targets ($D_1 > D_2$); and (c) both effects interact with development, so that age increases in direct access and reconstruction are larger for D_1 and R_2 than they are for D_2 and R_1. In Figure 5B, it can be seen that all three effects were present at both age levels, but the nature of the interaction was different. With respect to the interaction, both the tendency of repetition to enhance reconstruction and to impede direct access *declined* with age, so that developmental increases in direct access are greater for D_2 than for D_1, and there was an Age x Trial cross-over such that R_1 increased with age but R_2 did not. These impressions were confirmed by a 2 (age) x 2 (trial: first vs. later) x 2 (memory process: direct access vs. reconstruction) ANOVA, using estimates of the D and R parameters as dependent variables. The main effects for age, F (1, 42) = 4.10, MSE = .003 p < .05, trial, F (1, 42) = 65.12, MSE = .003, p < .0001, and memory process, F (1, 42) = 36.18, MSE = .06, p < .0001, were all reliable. However, the key result was an Age x Trial x Memory Process interaction, F (1, 42) = 13.70, MSE = .02, p < .001. Post hoc analysis (Tukey HSD, .05 level of confidence) revealed that the mean values of D_2, and R_1 both increased with age, but age changes in the other two parameters were not reliable. The reasons for this pattern are evident in Figure 5B. Both of the previously noted effects of repetition (i.e., its tendency to make direct access harder and to make reconstruction easier) were more marked in adolescents than in young adults. These are quite sensible outcomes, considering that adults should be simultaneously less susceptible to the debilitating effects of output interference and more adept at identifying features that specify small, correct search sets.

As noted earlier, the opposite effects of study-test trial repetition on the (D_1, D_2) and (R_1, R_2) pairs that is produced by repetition cannot be interpreted as showing that the direct access and reconstruction parameters measure a *single* process, to which the two parameter pairs just happen to

respond in opposite ways. Once again, such an interpretation can be ruled out if there are other variables whose directional effects on the (D_1, D_2) and (R_1, R_2) are the same (Dunn & Kirsner, 1988). It can be seen in Figure 6B that age is such a variable.

Direct access, with and without reconstruction. The mean values of the four direct access parameters are plotted separately for the two age levels in Figure 7B. We already know that the mean values of these parameters follow the theoretically predicted ordering, $D_2 < D_1 < D_{3C} < D_{3E}$, for both adolescents and young adults. This ordering is apparent in Figure 7B, but so is another datum—namely, that none of the direct access parameters other than D_2 seem to develop during this age range. The ANOVA that was just reported established that D_2 increased with age, whereas D_1 did not. For the other two direct access parameters, we simply computed t tests to compare the parameters' mean values in adolescents versus young adults. We found that neither D_{3C}, t (42) = .79, nor D_{3E}, t (42) = .79, varied with age. Thus, in contrast to childhood, increases in the ability to directly access targets were quite limited between early adolescence and young adulthood. Another difference was concerned with the locus of age improvements in direct access. Those improvements were concentrated within later phases of learning during childhood (converting reconstructable items into directly accessible ones) but were concentrated within earlier phases of learning thereafter (converting items that are not reconstructable into directly accessible ones) during adolescence. Further, when the fact that only D_2 increased between early adolescence and young adulthood is combined with the fact (cf. Figure 7B) that the D_1-to- D_2 decline that is so marked in children and adolescents was very slight in adults, the improvements in direct access between early adolescence and young adulthood seem to be of a very specific sort: Subjects are becoming less sensitive to the effects of output interference.

Familiarity judgment. The mean values of the four familiarity judgment parameters are plotted separately for the two age levels in Figure 8B. We already know that these mean values follow the theoretical ordering $M_{J1/J2} > M_{J3C/J3E}$ and $J_{3C} > J_{3E}$ in children, adolescents, and young adults. A further point that becomes apparent by comparing Figures 8A and 8B is that developmental increases in familiarity judgment were far more pronounced between early adolescence and young adulthood than between childhood and early adolescence. We previously saw (Figure 3) that the mean value of the four familiarity judgment parameters did not increase reliably during childhood, but it increased considerably between early adolescence and young adulthood, t (42) = 5.53, p < .0001. For the early adolescent versus young adult data sets, we computed a 2 (age) x 4 (judgment parameters: J_1, J_2, J_{3C}, J_{3E}) ANOVA, using values of the judgment parameters as dependent variables. The main effect for age, F (1, 42) = 30.52, MSE = .05, p < .0001, and the main effect for parameter, F (3, 126) = 17.10, MSE = .02, p < .0001, were reliable, but the Age x Parameter interaction was not. Thus, J_1, J_2, J_{3C}, and J_{3E} all increased reliably between early adolescence and young adulthood, and they increased by equivalent amounts.

Summary. Developmental trends in memory processes that control age improvements in recall were different between adolescence and young adulthood than they were between childhood and adolescence. Between childhood and early adolescence, age improvements in recall consisted of improvements in direct access and reconstruction but not familiarity judgment. Between early adolescence and young adulthood, however, there are broad-based improvements in familiarity judgment without reliable overall increases in reconstruction and direct access. Considering that decision criteria are known to become somewhat more conservative during this age range, it appears that the superior recall of young adults, relative to adolescents, is chiefly the result of increased subjective familiarity of reconstructed targets.

Healthy Aging

In the next two sections, we take up the third objective of this paper, which is to extend the earlier theoretical distinctions and modeling techniques to the study of aging and cognitive impairment, thereby achieving a theoretical framework that is unified over the study of early development, mainstream adult research, and the study of aging and cognitive impairment. We consider changes that occur in direct access, reconstruction, and familiarity judgment during healthy aging in the present section and turn to cognitive impairment in the next section. With respect to aging, our corpus of recall data includes 16 data sets in which healthy older adults (who were screened for health problems and cognitive impairments; mean age = 70 - 80 years) learned word and picture lists to criterion under paired-associate, free, or cued recall conditions, and 16 parallel data sets

in which young adults learned to recall the same lists under identical conditions (see Appendix B). For these matched pairs of data sets, developmental changes between early and late adulthood can be explored by comparing estimates of the direct access, reconstruction, and familiarity judgment parameters for younger and older subjects.

In sharp contrast to the literature on early memory development, there is a substantial literature on developmental trends in dual memory processes between early and late adulthood, using remember/know judgments, process dissociation, and other conventional methodologies (Anderson et al., 2008; Parks, 2007; Skinner & Fernandes, 2008; Toth & Parks, 2006). A consistent finding has been that performance on recollection measures declines considerably during late adulthood, whereas performance on familiarity measures is either age-invariant or declines by small amounts. This well-established pattern provides yet another benchmark for the present model of recall because it predicts that the direct access parameters will decline far more than familiarity judgment parameters, if the process interpretations of those parameters are correct. Of course, nothing specific is known about aging trends in reconstructive retrieval because this operation is not measured in conventional dual-process methodologies. However, when the idea that reconstruction is semantically based is combined with the well-established finding (Budson et al., 2006b; Kensinger & Schacter, 1999; Koutstaal, 2003; Koutstaal & Schacter, 1997) that semantic aspects of memory are more likely than verbatim aspects to be spared during healthy aging, an obvious expectation is that aging declines in direct access ought to be more pronounced than aging declines in reconstruction.

Findings are presented in two steps. First, detailed information about how the direct access, reconstruction, and familiarity judgment parameters behave in older adults is presented, using the 16 older adult data sets. As before, the focus is on whether that behavior accords with theoretical prediction. Second, we consider questions about developmental change by examining early-to-late-adulthood trends in direct access, reconstruction, and familiarity judgment, using the 32 paired data sets.

Behavior of Parameters in Older Adults

Separability of Direct Access and Reconstruction

Evaluation of the hypothesis that direct access and reconstruction are distinct processes was confined to the earlier analyses of the effects of study-test trial repetition on (D_1, D_2) versus (R_1, R_2). (Unlike prior analyses, correlations within and between the two sets were not computed because the number of older adult data sets was too small for adequate power.) In the earlier results for children, adolescents, and young adults, repetition had opposite effects on the two types of parameters: The mean value of R increased with repetition, but the mean value of D decreased. In older adults, direct access was affected but reconstruction was not. A plot of the means for the two pairs of parameters appears in Figure 11B, where it can be seen that D_2 was markedly smaller than D_1, but R_1, and R_2 had comparable values. To evaluate these trends for statistical significance, we computed paired-samples t tests. It was found that like prior results, the decline in direct access as function of repetition was reliable, t (15) = 4.31, $p < .001$, but that unlike prior results, mean values of R_1 and R_2 did not differ reliably, t (15) = 1.24. The indicated theoretical conclusions are that when items escape state U, the chances that older subjects will learn to directly access them is sensitive to accumulating interference, which is also true of children and adolescents, but unlike other age levels, older adults do not exhibit intertrial improvements in learning how to reconstruct items.

With respect to the fundamental issue of whether the (D_1, D_2) and (R_1, R_2) pairs measure distinct, separable retrieval process, further evidence is provided by plotting *reversed associations* between these parameters. Dunn and Kirsner (1988) showed that a one-process interpretation of the fact that these parameter pairs react differently to selected experimental manipulations can be ruled out by reversed associations. A reversed association is a nonmonotonic relation between the two parameter pairs, which in the present case means that values of the (R_1, R_2) pair would sometimes increase when the (D_1, D_2) pair increases but, at other times, would *decrease* when the (D_1, D_2) pair increases. Consistent with the hypothesis that these parameter pairs measure distinct processes, a reversed association is detected when their mean values are plotted against each other, for the four age levels in our corpus. That relation is shown in Figure 12.

Ordering of Direct Access Parameters

The theoretically expected ordering of the direct access parameters is $D_2 < D_1 < D_{3C} < D_{3E}$, as it was for other age levels. For the 16 older adult data sets, the mean values of these parameters were D_2 = .15, D_1 = .25, D_{3C} = .42, D_{3E} = .27, which does not completely follow the expected ordering: Although the mean of D_1 and D_2 (.20) is smaller than the mean of D_{3C} and D_{3E} (.35), which conforms to prediction (because reconstruction provides additional covert presentations when items occupy state *P*), the mean values of D_{3C} and D_{3E} are the opposite of their predicted ordering (based on the principle that recall errors initiate active mnemonic processing). A one-way ANOVA produced a significant main effect for these parameters, F (3, 45) = 29.42, MSE = .02, p < .0001. Post hoc tests revealed that (a) the mean value of D_1 was smaller than the corresponding means for all of the other three parameters, (b) the mean value of D_2 was smaller than the mean value of D_{3C} and (c) the mean value of D_{3E} was smaller than the mean value of D_{3C}. Thus, on the one hand, the notion that it is more difficult to learn how to directly access items that are in state *U* than items that are in state *P* has again been confirmed (because the mean of D_1 and D_2 is smaller than the mean of D_{3C} and D_{3E}). On the other hand, the notion that once subjects have learned how to reconstruct items, it is easier to learn how to directly access them following an error than following a success is disconfirmed in older adult data sets (though it was confirmed at all other age levels). Evidently, errors no longer serve to initiate active mnemonic processing in late adulthood, and on the contrary, it is successful recall that may trigger such processing.

The shift in late adulthood from error-driven learning to success-driven learning is consistent with another recent finding about healthy aging, the positivity effect (e.g., Carstensen & Mikels, 2005; Mikels, Larkin, Reuter-Lorenz, & Carstensen, 2005). It has long been known (e.g., Storbeck & Clore, 2005) that the memories of young adults, adolescents, and children exhibit a negativity effect; that is, information with a negative emotional valence, such as errors and other forms of negative feedback, is preferentially processed and affects performance more than information with a positive or neutral valence. The positivity effect refers to the fact that, in contrast, across a broad range of tasks, older adults preferentially process information with a positive emotional valence, with successes and other positive feedback being examples of positively-valenced information.

Ordering of Familiarity Judgment Parameters

During late adulthood, the theoretically expected ordering of the familiarity judgment parameters is the same as during earlier segments of the life span—namely, $M_{J1/J2} > M_{J3C/J3E}$ and $J_{3C} > J_{3E}$. The mean parameter estimates for the present data sets are displayed in Figure 9D. The values conform to prediction because $M_{J1/J2}$ = .81, $M_{J3C/J3E}$ = .69, J_{3C} = .77, and J_{3E} = .60. A one-way ANOVA produced a significant main effect for these parameters, F (3, 45) = 11.55, MSE = .02, p < .005. Post hoc tests showed that J_{3E} was significantly smaller than any of the other parameters, J_2 was larger than J_{3C} or J_1, and that the latter two parameters did not differ. The finding that $J_2 > J_1$ was not predicted, and considering that this relation has not been observed in any of the prior analyses of the judgment parameters, it may be a statistical aberration. That interpretation is bolstered by the fact that this relation did not replicate in the longitudinal study of older adults that is reported below (see Table 6).

Summary

Parameter behavior in the older adult data sets conformed to theoretical expectations in most key respects. However, whereas reconstructive retrieval benefited from study-test trial repetition in young adults (and children and adolescents), this advantage was not present in older adults. Also, whereas it was easier for young adults (and children and adolescents) to learn to directly access reconstructable items that were *not* successfully recalled, it was easier for older adults to learn to directly access reconstructable items that *were* successfully recalled, a developmental change that is reminiscent of the shift from preferential processing of negative information to preferential processing of positive information in older adults.

Early-to-Late-Adulthood Changes

Next, we describe developmental differences in direct access, reconstruction, and familiarity judgment, between early and late adulthood, comparing parameter values for the matched pairs of data sets in which younger and older adults learned to recall identical lists under identical conditions. Once again, we begin with global developmental trends and then move to trends for specific processes. A key prediction, based on aging research with other dual-process methodologies, is that there will be marked

declines in direct access parameters, relative to reconstruction parameters, and that familiarity judgment parameters will be age-invariant.

Global Trends

Mean values of the four direct access parameters, the two reconstruction parameters, and the four familiarity judgment parameters are plotted for younger and older adults in Figure 11A. We have seen that (a) between childhood and adolescence (Figure 3A), direct access and reconstruction developed but familiarity judgment was age-invariant and that (b) between adolescence and young adulthood (Figure 3B), familiarity judgment displayed marked developmental change while reconstruction was age-invariant and direct access developed slightly. Inspection of Figure 11A suggests that a third pattern dominates the years between young and late adulthood: The age change in direct access (.19) is more marked than the corresponding change in either reconstruction (.11) or familiarity judgment (.04). To test this pattern for statistical significance, we computed a 2 (age) x 2 (parameter: D versus R versus J) ANOVA, using mean values of each of the three types of parameters as dependent variables. There were large main effects for age, F (1, 30) = 25.73, MSE = .01, p < .0001, and parameter, F (2, 60) = 184.77, MSE = .01, p < .0001, of course. However, the result of principal interest was an Age x Parameter interaction, F (2, 60) = 4.20, MSE = .01, p < .02. When this interaction was decomposed with post hoc tests (Tukey HSD, .05 level of confidence), it was found that the direct access and reconstruction parameters declined reliably with age, that the direct access parameters declined more than the reconstruction parameters, and that the familiarity judgment parameters were age-invariant.

Direct Access and Reconstruction

The relations between age, repetition, direct access, and reconstruction are shown in Figure 11B. We tested the earlier prediction that reconstruction parameters are larger than their corresponding direct access parameters. As before, we computed a 2 (age) x 2 (parameters: the mean of R_1 and R_2 vs. the mean of D_1 and D_2) ANOVA, using parameter estimates as dependent variables. There was a main effect in the predicted direction for these parameters, F (1, 30) = 17.05, MSE = .01, p < .0001 (mean of R_1 and R_2 = .42 and mean of D_1 and D_2 = .24). There was also an age main effect, F (1, 30) = 6.86, MSE = .03, p < .01, but the Age x Parameter interaction was not reliable. We just saw that the mean of the direct access parameters declines more with age than the mean of the reconstruction parameters, which is consistent with the standard finding that semantic aspects of memory are more likely than verbatim aspects to be spared during healthy aging. The absence of an Age x Parameter interaction shows that this pattern does not extend to the mean of D_1 and D_2. However, support for this pattern was present when age trends were analyzed at the level of individual parameters. Specifically, we computed a 2 (age) x 2 (parameter type: direct access versus reconstruction) x 2 (trial: 1 versus 2) ANOVA, using the estimates for D_1, D_2, R_1, and R_2 as dependent variables. There were main effects for age, F (1, 30) = 6.86, MSE = .05, p < .02, a main effect for parameter type, F (1, 30) = 17.05, MSE = .02, p < .0001, and an Age x Trial interaction, F (1, 30) = 8.42, MSE = .02, p < .0001.

The nature of the Age x Trial interaction can be seen in Figure 11B, where two features of importance can be noted. First, the repetition-induced decline from D_1 to D_2 was more pronounced in older adults, so much so that age declines in direct access were reliable for D_2 but not D_1. The other important feature is the relation between R_1 and R_2 at the two age levels. In all previous comparisons of these parameters, repetition increased the chances that items would escape state U by becoming reconstructable (i.e., $R_1 < R_2$). Although this effect was again present in young adults (R_1 = .37, R_2 = .45), it was absent in older adults. When these two features were combined, the result was that, overall, there were no reliable differences in the rates of age decline for either D_1 versus R_1 or D_2 versus R_2.

Direct Access, With and Without Reconstruction

The mean values of the four direct access parameters are plotted separately for the two age levels in Figure 11C. It can be seen that young adult data follow the theoretically predicted ordering, $D_2 < D_1 < D_{3C} < D_{3E}$, whereas we know that the older adult data follow the somewhat different ordering, $D_2 < D_1 < D_{3E} < D_{3C}$. To extract age trends from these data, we computed a 2 (age) x 4 (parameter: D_1, D_2, D_{3C}, D_{3E}) ANOVA, using parameter estimates as dependent variables. There were main effects for age, F (1, 30) = 14.93, MSE = .08, p < .001, and for parameter, F (3, 90) = 30.14, MSE = .02, p < .0001, and an Age x Trial interaction, F (3, 90) = 8.42, MSE = .02, p < .0001. Post hoc analysis (Tukey HSD, .05 level of confidence) of the interaction revealed that D_2, D_{3E}, and D_{3C} all declined reliably with age and by large amounts (roughly 50%, on average), but D_1 was age-invariant.

Familiarity Judgment

The mean values of the four familiarity judgment parameters are plotted separately for the two age

levels in Figure 11D. We already know that the mean values of these parameters conform to the theoretical ordering $M_{J1/J2} > M_{J3C/J3E}$ and $J_{3C} > J_{3E}$ in both younger and older adults. To measure age trends at the level of individual parameters, we computed a 2 (age) x 4 (parameter: J_1, J_2, J_{3C}, J_{3E}) ANOVA, using values of the familiarity judgment parameters as dependent variables. This ANOVA produced a parameter main effect, F (3, 90) = 21.94, MSE = .02, p < .0001, but no age main effect and no Age x Parameter interaction. Thus, during the course of healthy aging, familiarity judgment appears to be entirely spared. There was no evidence of decline at the level of the average value of the four parameters or at the level of the individual parameters. To interpret this finding, it is important to bear in mind, as already mentioned, that like early memory development, there is an aging literature in which signal detection estimates of criterion stringency have been computed for (e.g., Budson, Sullivan, Daffner, & Schacter, 2003; Budson et al., 2002, 2006a; Schacter et al., 1999). A common finding is that criterion stringency does not vary greatly between early and late adulthood. For example, in an experiment with DRM lists, Budson et al. (2006a) found that the C parameter did not differ reliably between the ages of 20 and 70. Thus, supplementary analyses of the sort that we reported earlier, in which the familiarity and criterion components of the judgment parameters are separated, are not required to establish the key conclusion: Age-invariance in these parameters cannot be explained on the ground that declines in the familiarity of reconstructed items are being canceled, at the level of the J parameters, by decreases in the stringency of decision criteria. Instead, with the common words and pictures that were used in the experiments in our data sets, the indicated conclusion is that the strengths of reconstructed items' familiarity signals are spared during health aging.

Summary

The age trends in parameter values suggest that memory declines during healthy aging are dominated by declines in direct access. Three of the four direct access parameters displayed declines that were large in absolute terms (around 50%). Although one of the reconstruction parameters declined with age, that result was due to the fact that in older adults, repetition of study-test cycles did not enhance the chances that subjects will learn how to reconstruct an item. Finally, familiarity judgment was completely spared from aging declines, the mean value of the four parameters being virtually identical in younger and older adults. Here, it is important to add that in addition to being spared from aging declines, familiarity judgment remains at a much higher level in older adults than in children or adolescents. The mean values of the four familiarity judgment parameters for these age levels were .76 (older adults), .79 (young adults), .64 (adolescents), and .61 (children).

Cognitive Impairment

So far, the earlier theoretical distinctions and modeling techniques have been applied to the recall of subjects in the normal ability range, which was possible because the model in which those distinctions are embedded delivered good fits to the data of all age levels. In the present section, our aim is more exploratory—specifically, to examine how these same theoretical distinctions and modeling techniques might be extended to the memory sequelae of cognitive impairment. At first blush, this might seem to be a rather speculative endeavor, in comparison to the work that figures in the preceding two sections. Actually, however, there is precedent for it. In prior research by Batchelder et al. (1997) and by Faglioni and associates (e.g., Faglioni et al. 2000a, 2000b), HMM's were successfully applied to the study of cognitive impairments that are associated with conditions such as Alzheimer's dementia (AD), Parkinson's disease, and multiple sclerosis.

A key methodological constraint in such research is that subjects may be unable to complete tasks that demand that they achieve errorless recall. Consequently, extending the present techniques to cognitive impairment requires that they be implemented in tasks that, unlike the experiments that figured in the preceding two sections, do not impose stringent acquisition criteria. We present two such implementations in this section and show how each can be used to isolate processes that are responsible for the memory sequelae of cognitive impairment. One of the tasks, which is considered in the first subsection below, involves a design of the form $S_1T_1T_2S_2T_3S_3T_4$. In other words, the design is the same as the one over which Equation 3 is defined,

except that the procedure stops after the third study-test trial. In the first subsection below, we show how direct access, reconstruction, and familiarity judgment can be measured with the data of two illustrative experiments in which this fixed-trials procedure was used. One experiment involved comparing the recall of healthy older adults to that of older adults who had been diagnosed with either mild AD or clinical depression. The other experiment involved comparing the recall of a single group of older adults over an 8- to 12-month interval, in order to detect the emergence of impairment. Such detection is an important objective because the rate of conversion to some form of cognitive impairment is substantial after age 70. In the second subsection below, we present an even simpler implementation of the dual-process model, which involves a single study cycle followed by three independent recall tests for the studied material (i.e., the design is $S_1T_1T_2T_3$). We show how this design can be used to measure direct access, reconstruction, and familiarity judgment with the data of an illustrative experiment in which the subjects were schizophrenic patients and age-matched control subjects.

Fixed-Trials Implementation

Comparisons of Healthy and Impaired Individuals

A familiar type of study in the neuropsychological literature involves administering conventional dual-process tasks (e.g., remember/know) to healthy older adults and older adults with some form of cognitive impairment. Common findings are that between-group differences are more pronounced for recollection measures than for familiarity measures and that between-group differences are often confined entirely to recollection measures (for a review, see Yonelinas, 2002). Equation 3 can be used in the same manner in such research, which can be illustrated by applying it to a fixed-trials recall experiment by Howe (1990). In Appendix B, as part of the identifiability proof for Equation 3's parameters, we show that it is possible to estimate all of its parameters and to test goodness of fit with the data of experiments that use the simplified design $S_1T_1T_2$, S_2T_3, S_3T_4. Howe's experiment employed this design.

Howe (1990) tested three groups of subjects: (a) six home-dwelling older adults who had been diagnosed with mild AD and were being treated through an outpatient clinic (mean age = 69), (b) six home-dwelling older adults who had been diagnosed with clinical depression and were being treated through the same outpatient clinic (mean age = 71), and (c) six healthy older adults who resided in the same community (mean age = 70). Each subject participated in two separate fixed-trials recall tasks, in which they studied and recalled a 16-item list. The data for the two lists were then pooled to measure overall performance. The accuracy of recall differed dramatically among the three groups. Healthy older adults performed best, with an accuracy level of slightly above 50% on T_1 and T_2, which rose to above 90% on T_4. AD patients' performance was worst, with an accuracy level of 5% on T_1 and T_2, which rose to roughly 10% on T_4. Depressed patients' performance was intermediate, with an accuracy level of slightly above 20% on T_1 and T_2, which rose to slightly below 50% on T_4.

The question of primary interest here is whether our model fits these data and, hence, whether estimates of its parameters can be used to isolate processes that are and are not responsible for the between-group differences in recall. The fit results are displayed in Table 4. These tests ask whether a model that assumes that learning to recall involves the two stages that are posited in Equation 3 gives statistically acceptable accounts of the data. These are likelihood-ratio tests of the form specified in Equation A17 that generate G^2 statistics with 4 degrees of freedom. According to the results in Table 4, Equation 3 fits these data well. This, in turn, provides grounds for optimism that when the simplified $S_1T_1T_2$, S_2T_3, S_3T_4 design is used to compare the recall of healthy older adults to that of cognitively impaired adults, the present model will fit the data and, therefore, its parameters can be used to pinpoint between-group differences in memory processes. Although that is the major conclusion, we also report estimates of the models' parameters in Table 5 in order to illustrate potential process-level differences between the healthy and impaired groups, differences that can be regarded as model-validity tests.

Consider, first, a validity result about familiarity judgment. It was just noted that a common finding about the memory sequelae of cognitive impairment is that familiarity is spared. Consistent with that datum, note that the mean value of the four J parameter

estimates in Table 5 was not appreciably higher for the healthy group (.69) than it was for the AD group (.67) or the depressed group (.79). Another validity result is concerned with direct access. Consistent with the common finding that recollection is not spared—that healthy older and impaired groups differ substantially on traditional measures of recollection—note that the mean of the four *D* parameters for the healthy group (.45) is much higher than the corresponding means for the AD group (.03) and the depressed group (.18).[5] A third validity result is concerned with group differences in reconstruction. Because this process has no counterpart in conventional methods of measuring dual memory processes, predictions about it cannot be derived from prior research. However, predictions follow from studies in which healthy and impaired subjects were compared on memory tasks that differed in semantic processing. Specifically, it is well established that recall tasks are reliable predictors of transitions to AD among older adults (for reviews, see Petersen et al., 1999; Spaan et al., 2004), and more recently, it has been found that the predictive power of recall tasks increases when they stress semantic processing (e.g., Benedict, Schretlen, Groninger, & Brandt, 1998). This has prompted the hypothesis that a hallmark of transition to AD is increasing difficulties with semantic processing (e.g., Budson et al., 2002, 2003; Pierce, Sullivan, Schacter, & Budson, 2005; Reyna & Mills, 2007). Under that hypothesis, reconstruction parameters ought to have lower values in AD patients than in healthy older adults because those parameters measure the processing of items' semantic content (Reyna & Mills, 2007). The data in Table 5 concur with this prediction: The mean of the two reconstruction parameters for the healthy group (.31) was much larger than the corresponding mean for the AD group (.05).

These parametric results must be cautiously interpreted, of course, because they are based on the data of only a single study of healthy older adults, AD patients, and depressed patients. With that proviso, it is significant that between-group parameter comparisons are consistent with what one would predict on the basis of the larger literature on cognitive impairment and our earlier assumptions about what the models' parameters measure.

Emergence of Cognitive Impairment

By recent estimates (e.g., Hebert, Scherr, Bienias, Bennett, & Evans, 2003), over 4.5 million people in the United States have AD, and more than twice that number have mild cognitive impairment (MCI). The combined risk of developing one or the other condition during the course of normal aging is above 20%. For instance, a study of 4000 adults between the ages of 70 and 90 by Petersen et al. (2001), using established clinical criteria, found that 12% to 15% of subjects were classified as MCI and another 8% were classified as AD. Although large-scale normative studies of the incidence of AD and MCI are currently in progress, extant data (e.g., Bennett et al., 2002; de Jager, Hogervorst, Combrinck, & Budge, 2003; Peterson et al., 2001) suggest that healthy adults who are 70 to 90 years old convert to MCI at a rate of 10% per year. The typical progression is for MCI to emerge first, followed at some later point by AD. For example, Petersen et al. compared a sample of older adults who had been diagnosed with MCI to a sample of healthy age-matched controls. Subjects in the MCI group converted to probable AD at a rate of 10%-15% per year, whereas the corresponding rate for controls was 1% to 2%.

The high baseline risk of MCI after age 70, coupled with its rate of conversion to AD, places a premium on early detection, which can lead to interventions that delay the progress of impairment to clinical MCI or AD. Because recall is the best single predictor of emerging cognitive impairment, it is desirable in clinical settings to use recall tests for that purpose (e.g., Benedict et al., 1998; Delis, Kramer, Kaplan, & Ober, 1987). Further, there are clinical incentives for restricting initial screening to recall tests because costs increase as the amount of testing increases, and other tests are less sensitive to impairment. Given these constraints, a practical clinical goal is to improve the detection power of recall tests (Benedict et al., 1998). That there is considerable room for such improvement follows from the idea that impairment is due to only a subset of the processes that control recall. Therefore, estimates of the present model's parameters should do a better job of predicting the emergence of impairment than raw recall performance because some of the processes that control raw performance are uncorrelated with impairment and therefore only contribute error variance to the detection

of impairment (Reyna & Mills, 2007). In particular, detection should be improved by relying on estimates of the reconstruction and direct access parameters because the familiarity judgment parameters do not seem to contribute to either the declines in recall that occur during healthy aging or to the declines that occur during transitions to impairment. Further, Reyna and Mills predicted that detection should be especially improved by estimates of reconstruction because declines in reconstruction are central to the emergence of impairment. According to Reyna and Mills, that is because diffuse cognitive impairment overcomes the neural redundancy that spares semantic processing (and therefore reconstruction) during healthy aging.

Although we are unaware of any data sets in which changes in parameter estimates over time can be estimated for individual older adults, the two-step detection procedure can be illustrated with *group* data via an experiment by Howe (1990). In this experiment, a sample of 25 healthy adults in the age range for emergence of MCI (mean age = 71) learned one of two word lists, using the $S_1T_1T_2$, S_2T_3, S_3T_4 design. Then, 8 to 12 months later, each subject learned another word list of the same general type, again using the $S_1T_1T_2$, S_2T_3, S_3T_4 design. Because this design was used on each occasion, direct access, reconstruction, and familiarity judgment parameters can be estimated for each test, with a view to determining whether any of these processes declined over the intervening months. On the hypothesis that these subjects were progressing towards impairment and that some had converted to MCI after 8-12 months, the expectation would be that the reconstruction parameters, in particular, ought to decline between Test 1 and Test 2 (Reyna & Mills, 2007). First, however, the question of whether the model provides statistically acceptable fits must be examined.

The fit results appear at the bottom of Table 4. As in the fixed-trial experiment that we considered above, the model fit the data of both list conditions and both testing sessions. For the $G^2(4)$ statistics in Table 4, a critical value of 9.49 is required to reject the null hypothesis that the model fits the data, and the mean value of that statistic (6.91) is well below the critical value. Obviously, these results provide further grounds for optimism that if the simplified $S_1T_1T_2$, S_2T_3, S_3T_4 design is used in research on aging and cognitive impairment, the present model will fit the data, so that its parameters can be used to identify between-group differences.

Turning to the parametric results, the relevant data appear in Table 6, where estimates of the direct access, reconstruction, and familiarity judgment parameters are reported for both initial and delayed sessions. The question of interest is whether any of these parameters declined between the two sessions. Examination of the mean values of the parameters for the two testing occasions yields findings that are consistent with the notion that emerging impairment is associated with declines in reconstructive retrieval. During the initial session, the mean estimates of the four direct access parameters, the two reconstruction parameters, and the four familiarity judgment parameters were .36, .45, and .78, respectively. After 8-12 months, the mean estimates of the three groups of parameters were .35, .26, and .79, respectively. Thus, just as current theoretical conceptions of cognitive impairment would expect, although the mean values of the direct access and familiarity judgment parameters did not decline over this interval, there was a substantial decline in the mean value of the reconstruction parameters, which suggests that changes in the estimates of these latter parameters may prove to be especially sensitive predictors of the emergence of impairment. These results are only suggestive because although subjects' levels of cognitive functioning were assessed at the start of the experiment (to establish that they were not impaired), subjects were not retested at the end (to establish conversion to impairment).

Repeated-Recall Implementation

Next, we consider an even simpler procedure that can be used with populations whose performance is seriously compromised. This procedure allows a reduced version of the trichotomous recall

model, which consists of a subset of 6 of the 11 parameters (D_1, R_1, R_f, J_1, J_{3C}, and J_{3E}). In this methodology (cf. Brainerd et al., 2002, 2003; Payne et al., 1996), a study list is presented only once, followed by three independent recall tests (three paired-associate tests or three cued recall tests or three free recall tests or three serial recall tests) for that list. If C denotes successful recall of an item and E denotes unsuccessful recall, each item on the study list must produce one of eight response sequences over the three recall tests—namely, $C_1C_2C_3$, $C_1C_2E_3$, $C_1E_2C_3$, $C_1E_2E_3$, $E_1C_2C_3$, $E_1C_2E_3$, $E_1E_2C_3$, and $E_1E_2E_3$. In Appendix B, we show how empirical probabilities of these response outcomes can be used to find estimates of the D_1, R_1, R_f, J_1, J_{3C}, and J_{3E} and to evaluate fit.

We now illustrate the application of the reduced model to the recall performance of schizophrenic patients. There is an extant literature in which conventional dual-process tasks have been applied to memory impairments in schizophrenic patients (e.g., Barch et al., 1996; Huron et al., 1995) and in subjects who exhibit high levels of certain schizophrenic symptoms but have not been diagnosed as schizophrenic (e.g., Brebion, Smith, Amador, Malaspina, & Gorman, 1997; Linscott, 1999). Examples of such symptoms are anhedonia (inability to experience enjoyment from pleasurable experiences, such as eating or sexual activity) and schizotypy (a collection of symptoms, including magical thinking, cognitive disorganization, and unstable moods). Two findings of general interest are that (a) such subjects consistently display reduced performance on conventional measures of recollection (Linscott & Knight, 2001), and (b) such subjects sometimes display elevated performance on traditional measures of familiarity (Linscott & Knight, 2004). We demonstrate that the repeated-recall procedure can be used to measure such differences by fitting the reduced model to the various conditions of an experiment by Korobanova (2008) and then estimating the parameters D_1, R_1, R_f, J_1, J_{3C}, and J_{3E} for schizophrenic and nonschizophrenic subjects.

Brainerd et al. (2002) reported some experiments in which young adult subjects studied either longer or shorter word lists and then responded to three independent recall tests under conditions of either paired-associate or free recall. Korobanova (2008) used a similar procedure, except that (a) her subjects were 20 schizophrenic patients and 87 age-matched controls and (b) each subject participated in all four cells of the List Length x Type of Recall design (during different sessions). The schizophrenic group consisted of patients who had been referred from mental health service providers, who had working diagnoses of schizophrenia, and who had no known neurological disorders. The working diagnosis was confirmed in separate clinical interviews in which participants met the DSM-IV (American Psychiatric Association, 1994) criteria for schizophrenia. All but two of the participants were taking antipsychotic medications. The control subjects were residents of the same community who were recruited through public advertisements. These subjects were screened for neurological disorders, traumatic brain injury, and personal and familial histories of psychosis. The mean ages of the schizophrenic and control groups did not differ reliably.

The results for the model fits and the estimates of the parameters of the reduced dual-process model are shown in Table 7. As with the fixed-trials implementation of the model, the fit results are the most important findings because they bear on whether the reduced model is apt to be a useful tool in future research with psychotic patients. The fit test (see Appendix B) is a $G^2(1)$ statistic, which evaluates the null hypothesis that the data conform to the reduced trichotomous model. The values of this statistic for the four schizophrenic conditions and four control conditions appear in the last column of Table 7. It can be seen that the reduced model's level of fit was excellent: None of the eight tests produced a null hypothesis rejection, and further, the mean of the eight tests (.88) was less that one-quarter of the critical value for null hypothesis rejection (3.84).

Turning to differences between schizophrenics and controls in underlying memory

processes, estimates of D_1, R_1, R_f, J_1, J_{3C}, and J_{3E} appear in the first 6 columns of Table 7, and the mean of the 3 familiarity judgment parameters appears in column 7. As with the earlier analyses of the fixed-trials implementation of the full model with AD patients, model validity is a key question: Do the results square with extant findings about schizophrenics, using conventional dual-process methodologies? They do. As mentioned, major findings from procedures such as remember/know and process dissociation are that schizophrenic subjects display impaired recollection but not impaired familiarity. Consistent with that pattern, the mean value of D_1 for schizophrenic patients across the four list conditions (.06) was less than one-third of the mean value for controls (.20). In addition, the overall mean value of the three familiarity judgment parameters for the schizophrenic patients (.70) was not smaller than the corresponding mean for the control subjects (.65), and the mean values of the reconstruction parameter for the two groups (.16 and .18) were nearly the same.

Thus, the reduced model produced findings that concur with prior results for schizophrenics using conventional dual-process methodologies. However, additional between-group findings are apparent when the effects of list length and type of recall test are considered. With these variables, it has been found with conventional dual-process methodologies that increasing list length decreases recollection without affecting familiarity, (e.g., Yonelinas, 1994). Also, because paired-associate recall provides subjects with less latitude to select features to be used in reconstruction than free recall provides, reconstruction has been found to be more difficult with paired-associate than with free recall in studies of normal adults with variants of the present model of recall (Brainerd et al., 2002). In line with those findings, for the control data in Table 7 note that (a) the mean value of D_1 for longer lists (.09) was less that one-third of the mean for shorter lists (.32), while the mean value of R_1 and of the three judgment parameters were not affected by length, and that (b) the mean value of R_1 for paired-associate recall (.07) was one-quarter of the mean for free recall (.28), while the mean value of D_1 and of the three familiarity judgment parameters were not affected by the type of recall test. The patterns for schizophrenic patients were different. Although reconstruction was influenced in the same manner by the type of recall test (mean R_1 = .08 for paired-associate versus .25 for free), direct access was not affected by list length (mean D_1 = .05 for longer lists versus .07 for shorter lists). Also, paired-associate recall produced noticeably higher levels of familiarity judgment (mean J = .71) than free recall (mean J = .58) in schizophrenic patients.

Summing up, application of the reduced model to Korobanova's (2008) repeated-recall study of schizophrenic patients yielded excellent fits, and it produced both validity results and novel findings. Concerning validity, prior reports of recollection deficits without familiarity deficits in schizophrenics leads one to expect that D_1 will be lower in the patient group, relative to the control group, but that the J parameters will not be suppressed in the patient group. That is how the results fell out. Concerning novel findings, in the patient group, unlike the control group, direct access was not affected by list length, and free recall suppressed familiarity judgment, relative to paired-associate recall.

Discussion and Conclusions

This paper began with three objectives. The first, which occupied us at the outset, was to resolve a fundamental problem in memory development research—namely, that dual-process distinctions have had virtually no impact on the study of early memory development. Key reasons for this situation are that conventional dual-process paradigms focus chiefly on recognition (which displays minimal improvement during this age span and may not involve dual processes), and those paradigms place high performance burdens on children (introspecting on mental states and making metacognitive judgments). Our solution was to devise a framework for studying dual memory processes with low-burden recall tasks. The framework consists of a theory that subsumes traditional distinctions, that incorporates a reconstruction operation that is specific to recall, and that implements these processes in a hidden Markov model that separates and quantifies them with the error/success data of standard recall tasks.

The second objective, which occupied us in the middle of this paper, was to close the knowledge gap on early memory development by putting the theoretical machinery to work with developmental data sets. We exploited a corpus of recall experiments in which the subjects were children, adolescents, and young adults. We secured reliable findings about the early development of direct access, reconstruction, and familiarity judgment, and equally important, we were able to test several theoretical predictions about these processes. The third objective, which occupied us in the last two sections, was to extend the framework to the study of healthy aging and cognitive impairment, thereby unifying it over: (a) the child-to-adolescent age range that is of primary interest in developmental research; (b) the young adult age level that dominates mainstream memory research; (c) the young-adult-to-older-adult age range that is of primary interest in aging research, and (d) some of the most commonly studied forms of cognitive impairment. With respect to healthy aging, we obtained reliable findings about lifespan trends in direct access, reconstruction, and familiarity judgment by analyzing a corpus of recall experiments in which the subjects were young and older adults. With respect to impairment, we demonstrated that the mathematical model that measures these processes can be implemented in two simplified tasks (fixed-trials recall and repeated recall) that are appropriate for subjects with clinically significant cognitive limitations. We also showed how estimates of the model's parameters can be used to pinpoint processes that are responsible for recall deficits that are sequelae of existing clinical conditions (e.g., AD, depression, and schizophrenia).

Thus, machinery is now in hand that can be used to track developmental and other between-group differences in direct access, reconstruction, and familiarity judgment and can also be used to test quantitative predictions about these processes. The former capability will be of greater interest to students of development, while the latter will be of greater interest to students of memory theory. Below, we briefly comment on what has been learned in connection with each of these capabilities. Then, the paper concludes by returning to the traditional dual-process conception of recognition to consider a surprising dividend of the trichotomous theory's modeling machinery. We show that the dual-process conception of recognition can be implemented in the same machinery, yielding a unified framework for the study of recall and recognition.

Development: The Big Picture

Research with conventional dual-process distinctions and methodologies has produced a portrait of developmental change that consists of scant evidence about child-to-young-adult development, coupled with more extensive evidence about young-adult-to-older-adult development. Within the former age span, where data are thin, the main question is, What do our techniques reveal about age trends in direct access, reconstruction, and familiarity judgment? Within the latter age span, where data are more extensive, the main question is, Do the age trends that are revealed by our techniques agree with the findings of conventional dual-process methodologies?

With respect to the first question, the principal outcomes were substantial improvements in all three of the processes during child-to-young-adult development, with improvements in specific processes being localized within different segments of this age range. During the child-to-adolescent segment, improvements in recall were dominated by improvements in direct access and reconstruction. Mean levels of both processes increased (and by roughly the same amount), whereas mean levels of familiarity judgment were age-invariant. If it is true, as our theory posits, that the chief difficulty factor for direct access is the interference sensitivity of verbatim traces and the chief difficulty factor for reconstruction is the ability to identify semantic features of targets that generate manageable search sets, the indicated conclusions are that (a) direct access is becoming less susceptible to the debilitating effects of interference and (b) the semantic

features that generate small search sets are becoming more readily accessible. Upon first impression, age-invariance in mean levels of familiarity judgment is consistent with two interpretations. On the one hand, the two processes that are assumed to control observed values of the familiarity judgment parameters (perceived familiarity of reconstructions and stringency of decision criteria) may both be invariant during child-to-adolescent development. On the other hand, because these processes have opposite effects on the familiarity judgment parameters, they may both be increasing, but the increases are canceling each other out at the level of parameter values. That is the more likely scenario, based on follow-up data that we reported (in which separate estimates of these processes were obtained) and on two further considerations. The first is that in recognition studies with subjects in this age range (e.g., Reyna & Kiernan, 1994, 1995), substantial increases in criterion stringency are ubiquitous findings. The other is that owing to general life experience, the subjective familiarity of list words ought to be increasing during this age range.

Turning to the adolescent-to-young-adult segment, age improvements in recall were dominated by improvements in familiarity judgment. All four familiarity judgment parameters increased reliably, and the mean increase in these parameters was approximately 30%. In contrast, overall mean values of the reconstruction and direct access parameters did not increase reliably (although there was some subsequent evidence that one of the four direct access parameters, the one that is most sensitive to accumulating output interference, may increase). To interpret the large increases in the familiarity judgment parameters, it is important to note that in developmental recognition studies, the marked increases in criterion stringency during child-to-adolescent development have been followed by small increases during adolescent-to-young-adult development (e.g., Brainerd & Mojardin, 1998). Thus, the perceived familiarity of reconstructed targets increases by substantial amounts during adolescent-to-young-adult development, and those improvements are not masked by parallel increases in criterion stringency. For early memory development as a whole, then, the overall conclusion is that the perceived familiarity of reconstructed targets increases throughout this age range, while increases in criterion stringency are centered within the child-to-adolescent part of the range. The former is a decidedly novel conclusion, as the thin literature on the early development of dual memory processes points to little improvement in familiarity.

Insofar as direct access and reconstruction are concerned, (a) the sensitivity of direct access to accumulating interference continued to decline, but by much smaller amounts, during adolescent-to-young adult development, and (b) the ability to use semantic and other relational features to construct small search sets was found to have completed its development by early adolescence. However, conclusion b is subject to the qualification that the lists in our corpus of recall studies were composed of words (or pictures) that are common in everyday discourse. Adolescent-to-young-adult improvements in reconstruction might be observed if lists were composed of less common items.

With respect to the second question, whether our techniques produce findings for the young-adult-to-older-adult development that concur with the findings of conventional dual-process methodologies, the answer is yes. The modal finding of conventional methodologies is that recollection declines during healthy aging but familiarity is spared. Analogously, parameter estimates for the aging studies in our corpus showed that the mean value of the direct access parameters declined dramatically (by more than 40%), whereas the mean value of the familiarity judgment parameters was virtually the same for 70-year-olds as for 20-year-olds. Because the mean value for older adults was high in absolute terms (.76), the implication is that older adults overwhelming regard their

reconstructions as being familiar enough to output. Concerning direct access, three of the four D parameters declined with age (by roughly 50%), with only the least interference-sensitive one (D_1) failing to decline reliably. Thus, it appears that increasing sensitivity of verbatim traces to accumulating interference is a major feature of healthy aging. Finally, what about reconstructive retrieval, which has no parallel in traditional dual-process research? Here, it was found that reconstruction declined with age but by much smaller amounts than direct access. Moreover, declines were limited to the tendency of reconstruction to benefit from repetition. In the present framework, reconstruction is treated as a comprehension processes, and as such, it ought to benefit from repeated opportunities to extract the meaning properties of targets. Consistent with this view, the studies in our corpus revealed that repetition of study-test trials increased the R measure (i.e., $R_2 > R_1$) among children, adolescent, and young adults. With older adults, however, repetition no longer had this effect, and hence, an important new finding about aging is that the ability of reconstruction to benefit from repeated opportunities to comprehend meaning declines during healthy aging.

Parametric Predictions

From the perspective of memory theory, the most important question about the present framework is, Do the numerical values of the recall model's parameters behave in ways that accord with prediction? This is a key question, of course, because parametric predictions are derived from theoretical conceptions of direct access, reconstruction, and familiarity judgment. Here, we used our corpus of developmental recall studies to test several predictions about the parameters that ostensibly measure these processes and to test those predictions at multiple age levels. The predictions of greatest theoretical significance are summarized in Table 8. Although various predictions are exhibited, there are three underlying themes: Predictions about the D parameters turn on the notion that direct access is sensitive to accumulating interference, predictions about the R parameters turn on the notion that reconstruction is sensitive to meaning comprehension, and predictions about the J parameters turn on the notion that familiarity judgment is sensitive to item selection when the decision criterion is held constant.

When predictions such as those in Table 8 were evaluated with many data sets, there was surprisingly little deviation between observed patterns of parameter behavior and predicted patterns. For instance, one dramatic point of agreement concerned the prediction that, under comparable conditions, learning to reconstruct targets ought to be easier than learning how to directly access them. That prediction was confirmed throughout the lifespan, with the grand average, from age 7 to age 70, of the R_1/R_2 pair (.36) being more than twice the grand average of the D_1/D_2 pair (.17). As another illustration, consider the prediction that the interference sensitivity of direct access means that repetition will suppress D_2 relative to D_1. Suppression effects of this sort were observed at all age levels, except the level at which, based on other research, subjects ought to be least susceptible to the effects of interference (young adults). As another example, consider the prediction that the comprehension sensitivity of reconstruction means that repetition of study-test trials will increase the value of R_2 relative to R_1. Enhancement effects of this sort were observed at all age levels, except late adulthood. Another instructive parametric result is that at all age levels where there were sufficient numbers of data sets to compute inter-parameter correlations, positive correlations were observed between matched pairs of D parameters and between matched pairs of R parameters, but negative correlations were observed between matched D/R pairs. Therefore, leaving aside the matter of exactly what these parameters measure, different D parameters seem to be measuring the same thing, different R parameters seem to be measuring the same thing, but D and R parameters seem to be measuring different things. Indeed, these data patterns suggest that the D and R parameters contrasting methods of retrieval. Last, the behavior of the J parameters likewise fell out in broad accordance with theoretical prediction. Two expected patterns that were confirmed throughout the lifespan concern (a) the relation between familiarity judgment when items first become reconstructable (the J_1/J_2 pair) versus while they are waiting in state P to become directly accessible (the J_{3C}/J_{3E} pair) and (b) the relation between familiarity judgments about reconstructions following prior successful recall (the J_{3C} parameter) versus following prior unsuccessful recall (the J_{3E} parameter). Owing to item selection, the mean of the J_1/J_2 pair should be larger than the mean of the J_{3C}/J_{3E} pair under prediction a, and J_{3C} should be larger than J_{3E} under prediction b. Between the ages of 7 and 70, the grand means of the J_1/J_2 pair and the J_{3C}/J_{3E} pair were .77 and .65, respectively, while the grand means of J_{3C} and J_{3E} were .72 and .54, respectively.

Findings such as these are crucially important validity results because the predictions arise from theoretical conceptions of direct access, reconstruction, and familiarity judgment. There is another group of validity results, however, that is empirical in origin and is concerned with the relation between the present framework and traditional dual-process frameworks. As we have noted, although dual-process research on early memory development is scarce, it is more extensive for aging and cognitive impairment. The standard dual-process finding about healthy aging is that recollection measures decline, but familiarity measures do not. With respect to cognitive impairment, AD has been the focus of multiple studies, with the modal pattern being further declines, relative to healthy older subjects, in recollection, coupled with continued sparing of familiarity. Also with respect to cognitive impairment, dual-process research with schizophrenic patients has produced a similar pattern of declines in recollection, relative to control groups, coupled with sparing of familiarity.

The available evidence suggests that the *D* and *J* parameters produce some of the same qualitative patterns as conventional measures do. The best evidence of such agreement is for healthy aging, where we analyzed 16 paired sets of recall data for younger versus older adults. Consistent with conventional methodologies, as mentioned above, there were marked declines in the *D* parameters, but the *J* parameters were age-invariant. With respect to cognitive impairment, the more limited amounts of data that were available to us also provided evidence of agreement between the present procedures and conventional methodologies: *D* parameters had smaller values for AD patients and for schizophrenic patients than for their respective control groups, but there were no between-group differences in *J* parameters.

Final Word: A Unified Framework for Recognition and Recall

We end where we began, with the traditional dual-process view of recognition. As we saw, a limiting feature of this theory is that conventional methodologies for estimating its recollection and familiarity components require that subjects make meta-cognitive judgments about introspections, which places excessive burdens on the capabilities of children and cognitively-impaired adults. That limitation can now be removed by implementing the dual-process approach to recognition in the machinery that has been developed for the trichotomous theory of recall. This yields a unified framework for recall and recognition, one that relies on a single modeling technology and uses low-burden tasks to study recognition as well as recall. A further desirable outcome is that the dual-process conception of recognition can now be represented as an HMM, whose fit can be rigorously evaluated and whose parameters measure dual memory processes on a common ratio scale.

To begin, remember that Equation 1 expresses the probability of correct recall as a function of direct access, reconstruction, and familiarity judgment for the canonical experiment $S_1T_1, S_2T_2, S_3T_3, \ldots$, where S_i is the *i*th study cycle and T_i is the *i*th recall test. Consider the corresponding canonical *recognition* experiment; that is, S_i is the *i*th study cycle but T_i is the *i*th recognition test. In the present framework, the theoretical expression for the hit probability, which is isomorphic to Mandler's (1980) original equation for recollection and familiarity, is

$$P_i(Rg) = D_i + (1-D_i)J_i. \tag{4}$$

$P_i(Rg)$ is the hit probability on the *i*th trial, D_i is the probability of being able to directly access a target's verbatim trace on the *i*th trial, and J_i is the probability that a target that cannot be directly accessed is familiar enough to pass a judgment check on the *i*th trial. Equation 4 is the same as Equation 1, except that R_i vanishes because items are physically presented on recognition tests, so that targets that cannot be directly accessed do not have to be reconstructed (hence, $R_i = 1$).

In order to identify the parameters of Equation 1, it was necessary to switch to the slightly modified design $S_1T_1T_2, S_2T_3, S_3T_4, \ldots$, which yielded Equation 3, whose parameter space is fully identifiable. Suppose that the same design is used for Equation 4 (except, of course, that the outcome space $T_1, T_2, \ldots$ consists of old/new recognition tests), and the constraint $R_i = 1$ is introduced in Equation 3. This yields the following dual-process equation for recognition:

$$W_3 = [L(1)L(2), L(1)P_E(2), L(1)P_C(2), P_E(1)L(2), P_E(1)P_E(2), P_E(1)P_C(2), P_C(1)L(2), P_C(1)P_E(2), P_C(1)P_C(2)] = [D'_1, 0, 0, 0, (1-D'_1)(1-J'_1)^2, (1-D'_1)(1-J'_1)J'_1, 0, (1-D'_1)(1-J'_1)J'_1, (1-D'_1)(J'_1)^2];$$

$$
M_3 = \begin{array}{c} \\ L(n) \\ P_E(n) \\ P_C(n) \end{array}
\begin{array}{ccc} L(n+1) & P_E(n+1) & P_C(n+1) \\ 1 & 0 & 0 \\ D'_{3E} & (1-D'_{3E})(1-J'_{3E}) & (1-D'_{3E})J'_{3E} \\ D'_{3C} & (1-D'_{3C})(1-J'_{3C}) & (1-D'_{3C})J'_{3C} \end{array}
\; ; \quad C_2 = \begin{array}{c} P(\text{correct}) \\ 1 \\ 0 \\ 1 \end{array} . \qquad (5)
$$

There is no longer a state U because hits are possible by guessing even before this list is studied for the first time. Equation 5 contains three parameters that measure direct access of verbatim traces (D'_1, D'_{3C}, and D'_{3E}) and three parameters that measure familiarity judgment about items that cannot be directly accessed (J'_1, J'_{3C}, and J'_{3E}). (Primes are appended to the parameters merely to indicate that memory processes are being measured for recognition rather than recall.) We know that all six parameters are identifiable because their identifiability has already been proved (see Appendix A, Equations A5-A8 and Equations A10 and A11).

In Equation 5, all targets are in state P before the first study cycle. A target can escape P on the first study cycle by becoming directly accessible, with probability D'_1. If it becomes directly accessible, it enters state L (errorless recognition) and recognition is successful on the first test and on all subsequent tests (i.e., the item absorbs in L). If a target does not become directly accessible on the first study trial, recognition on T_1 and T_2 is governed by the familiarity judgment parameter J'_1: On each test, the item is judged to be old (substate P_C) with probability J'_1 or new (substate P_E) with probability 1 - J'_1. On subsequent trials, escape from P to L is controlled by the direct access parameters D'_{3C}, and D'_{3E}. Specifically, on each study cycle, a target (a) becomes directly accessible (escapes from P to L) with probability D'_{3C}, if it was judged to be old on the immediately preceding test, or (b) it becomes directly accessible with probability D'_{3E}, if it was judged to be new on the immediately preceding test. If a target does not become directly accessible, recognition is governed by the familiarity judgment parameters J'_{3C} and J'_{3E}: The target is judged to be old with probability J'_{3C} or J'_{3E}, accordingly as it was judged to be old or new on the immediately preceding test.

Equation 5's fit to recognition data can be evaluated and its parameters can be estimated with the same machinery that we have used for recall. However, some further comments about fit and parameter estimation are in order. With respect to fit, note that Equation 5 makes the strong prediction that recognition data will be *one-stage* Markovian. This prediction is already known to be correct. Early attempts to model recognition data (e.g., Kintsch & Morris, 1965) showed that one-stage Markov chains gave excellent accounts of the sampling distributions fine-grained performance statistics (for a review of early studies, see Greeno, 1974). Interestingly, the fact that recognition data are one-stage Markovian may explain why, in the contemporary literature, some lines of evidence favor one-process models (e.g., Dunn, 2008) while others favor dual-process models (e.g., Yonelinas, 2002). According to Equation 5, recognition has both one- and two-process aspects: Learning to recognize an item involves a *single* interstate transition, but the overall probability of correct recognition is controlled by two memory processes (familiarity in state P and direct access of verbatim traces in state L) that contribute differentially to that probability during earlier versus later phases of learning.

Turning to parameter estimation, we noted earlier that the J parameters measure two processes, familiarity and response bias, and we showed how these processes can be separated with intrusion data. In recognition experiments, they can be separated with false-alarm data (e.g., see Snodgrass & Corwin, 1988). Specifically, each J parameter can be represented by the expression $J'_i = F'_i + (1- F'_i)?'_i$, where F'_i is the probability that a traget is familiar enough to be judged to be old and $?'_i$ is the probability that a target that is not familiar enough is nevertheless judged to be old on the basis of response bias. Obviously, the value of F'_i can be found for any of Equation 5's J parameters by simply substituting the relevant false-alarm rate for $?'_i$ and solving this expression for F'_i. In that connection, it is imporant to stress that the J parameters measure a weaker form of familiarity in recognition than in recall. In recall, as we saw, subjects only make familiaritiy judgments about items that are reconstructed; that is,

about items for which small correct search sets have been constructed from some of their features. In recognition, however, it is not necessary to reconstruct targets that cannot be directly accessed, and hence, subjects make familiarity judgments about items for which little or nothing has yet been learned. Thus, judged levels of familiarity ought to be lower in recognition than in recall, presumably leading to lower overall values of J parameters (if bias levels are comparable) and different relations between J parameters for earlier versus later phases of learning.

Together, Equations 3 and 5 supply a unified framework for studying dual-process conceptions with both recall and recognition data. This framework can be exploited to examine many fundamental issues, two of which we mention in closing. First, in keeping with the developmental objectives of this article, because low-burden tasks can now be used to estimate dual processes for recognition as well as recall, the two types of performance can be studied in tandem to track how these processes change during early memory development, healthy aging, and transitions to cognitive impairment. Further, because these low-burden recall and recognition tasks are so similar, the question of whether the same processes are being measured in recall and recognition can be explored. The D and J parameters of Equations 3 and 5 may or may not be measuring the same memory processes. If they are, patterns of change during early development, healthy aging, and transitions to impairment ought to be similar for comparable pairs of recall and recognition parameters, and within age levels or ability groupings, comparable pairs of parameters should react similarly to experimental manipulations. Such results would be expected for D_1, versus D'_1, D_{3C} versus D'_{3C}, D_{3E} versus D'_{3E}, J_1, versus J'_1, J_{3C}, versus J'_{3C}, and J_{3E} versus J'_{3E}. These are strong predictions, to which the only notable exception is that J'_1, J'_{3C}, and J'_{3E} will presumably be less sensitive than J_1, J_{3C}, and J_{3E} to manipulations (and subject characteristics) that affect familiarity because the former parameters measure a weaker form of familiarity.

Second, from the perspective of mainstream memory theory, surely the most attractive feature of the the unified framework is that traditional predictions of dissociation between dual memory processes can be studied simultanouesly for recognition and recall, using a single measurment technology and very similar experimental procedures. As is well known, dual-process ideas predict that in recognition, (a) certain manipulations (e.g., dividing attention at study or test) and certain subject characteristics (e.g., aging) will affect recollection but not familiarity, whereas (b) other manipulations (e.g., fluency, liberalty of response criteria) and subject characteristics (e.g., certain forms of brain damage) will affect familiarity but not recollection (for a review, see Yonelinas, 2002). Such predictions can now be tested with simple old/new recognition, by merely estimating the D' and J' parameters of Equation 5, and they can simultaneously be tested with simple recall data by merely estimating the D and J parameters of Equation 3.

When it comes to evaluating such predictions, a further attractive feature of the unified framework is that these parameters can be estimated for both earlier and later phases of learning. A limitation of the current dual-process recognition literature is that it is synonymous with early learning because multi-trial designs are rare. This is a key consideration because we saw with our corpus of recall data that D and J parameters react differently to experimental manipulations and to subject characteristics during earlier versus later learnng. A final attractive feature of the unified framework is that it can be used to determine why, at a process level, some manipulations have opposite effects on recall and recognition. Of course, the great bulk of manipulations that increase (or decrease) recall do likewise for recognition, though the magnitudes of the effects often differ. There are a few manipulations, however, that drive recall and recognition in opposite directions. Chief among them is word frequency, with lower-frequency words producing poorer recall but better recognition than higher-frequency words (e.g., Glanzer & Adams, 1985). A potential explanation that could be tested with the unified framework is that lower-frequency words increase recollection but seriously impair reconstruction. Thus, such words produce a net increase in recognition because reconstruction is not involved, but they produce a net decline in recall because they markedly deflate reconstruction. Under this hypotheis, it should be found that some of the D' parameters and some of the D parameters are larger for lower-frequency words, but that both of the R parameters are smaller for lower-frequency words. These are, of course, only illustrations of hypotheses that are motivated by dual-process distinctions that can be rigously explored within the unified framework.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97-123.

Anderson, N. D., Jennings, J. M., Cabeza, R., Ebert, P. L., Grady, C. L., & Graham, S. J. (2008). Recollection- and familiarity-based memory in healthy aging and amnestic mild cognitive Impairment. *Neuropsychology, 22*, 177-187.

Arndt, J., & Gould, C. (2006). An examination of two-process theories of false recognition. *Memory, 14,*814-833.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy in word recognition. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.

Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology, 44,* 20-40.

Barch, D. M., Cohen, J. D., Servan-Schreiber, D., Steingard, S., Steinhauer, S. S., & van Kammen, D. P. (1996). Semantic priming in schizophrenia: An examination of spreading activation using word pronunciation and multiple SOAs. *Journal of Abnormal Psychology, 105,* 592-601.

Barnhardt, T. M., Choi, H., Gerkens, D. R., & Smith, S. M. (2006). Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. *Journal of Memory and Language, 55*, 213–231.

Batchelder, W. H., Chosak-Reiter, J., Shankle, W. R., & Dick, M. B. (1997). A multinomial modeling analysis of memory deficits in Alzheimer's disease and vascular dementia. *The Journals of Gerontology, 52B,* 206-215.

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

Bartolucci, F., & Nigro, V. (2007). Maximum likelihood estimation of an extended latent Markov model for clustered binary panel data. *Computational Statistics and Data Analysis, 51,* 3470-3483.

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test – Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist, 12*, 43-55.

Bennett, D., Wilson, R., Schneider, J., Evans, D., Beckett, L., Aggarwal, N., Barnes, L., Fox, J. & Bach, J. (2002). Natural history of mild cognitive impairment in older persons. *Neurology, 59,* 198–205.

Billingsley, R. L., Smith, L. M., & McAndrews, P. T. (2002). Developmental patterns in priming and familiarity in explicit recollection. *Journal of Experimental Child Psychology, 82,* 251-277.

Bjorklund, D. F. (1987). How changes in knowledge base contribute to the development of children's memory. *Developmental Review, 7,* 93-130.

Bjorklund, D. F. (2004). *Children's thinking: Developmental and individual differences* (4th ed.). Pacific Grove, CA: Wadsworth.

Bjorklund, D. F., & Muir, J. E. (1988). Children's development of free recall memory: Remembering on their own. *Annals of Child Development, 5,* 79-123.

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

Bordes, L., & Vandekerkhove, P. (2005). Statistical inference for partially hidden Markov models. *Communications in Statistics: Theory and Methods, 34,* 1081-1104.

Bower, G. H., & Theios, J. (1963). A learning model for discrete performance levels. In R. C. Atkinson

(Ed.), *Studies in mathematical psychology* (pp. 1-31). Stanford, CA: Stanford University Press.

Brainerd, C. J. (1985). Model-based approaches to storage and retrieval development. In C.J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development*. New York: Springer-Verlag.

Brainerd, C. J., Desrochers, A., & Howe, M. L. (1981). Stages-of-learning analysis of picture-word effects in associative memory. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 1-14.

Brainerd, C. J., & Howe, M. L. (1982). Stages-of-learning analysis of developmental interactions in memory, with illustrations from developmental interactions in picture-word effects. *Developmental Review, 2*, 251-273.

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

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

Brainerd, C.J., Howe, M.L., Kingma, J., & Brainerd, S.H. (1984a). Explaining category interference effects in associative memory. *Canadian Journal of Psychology, 38*, 454-477.

Brainerd, C. J., Howe, M. L., Kingma, J. & Brainerd, S. H. (1984b). On the measurement of storage and retrieval contributions to memory development. *Journal of Experimental Child Psychology, 37*, 478-499.

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

Brainerd, C. J., Kingma, J., & Howe, M. L. (1986). Long-term memory development and learning disability: Storage and retrieval loci of disabled/nondisabled differences. In S. Ceci (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities*. Hillsdale, N.J.: Erlbaum.

Brainerd, C. J., & Mojardin, A. H. (1998). Children's spontaneous false memories for narrative statements: Long-term persistence and mere-testing effects. *Child Development, 69,* 1361-1377.

Brainerd, C. J., Olney, C. A., & Reyna, V. F. (1993). Optimization versus effortful processing in children's cognitive triage: Criticisms, reanalyses, and new data. *Journal of Experimental Child Psychology, 55,* 353-373.

Brainerd, C. J., Payne, D. G., Wright, R., & Reyna, V. F. (2003). Phantom recall. *Journal of Memory and Language, 48*, 445–467.

Brainerd, C. J., & Reyna, V. F. (1991). Forgetting in normal and learning disabled children: A disintegration/redintegration theory. In J. Obrzut & G. Hynd (Eds.), *Advances in the neuropsychology of learning disabilities* (pp. 147-178). New York: Academic Press.

Brainerd. C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100*, 42-67.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin, 134,* 343-382.

Brainerd, C. J., Reyna, V. F., Howe, M. L., Kingma, J., & Guttentag, R. E. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 33* (3-4), 1-109.

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

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

Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002). Dual-retrieval processes in free and associative recall. *Journal of Memory and Language, 46,* 120–152.

Brebion, G., Smith, M. J., Amador, X., Malaspina, D., & Gorman, J. M. (1997). Clinical correlates of memory in schizophrenia: Differential links between depression, positive and negative symptoms, and two types of memory impairment. *American Journal of Psychiatry, 154,* 1538-1543

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40,* 227-259.

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition, 51,* 251-261.

Budson, A. E., Sullivan, A. L., Mayer, E., Daffner, K. R., Black, P. M., Schacter, D. L. (2002). Suppression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. *Brain, 125,* 2750-2765, Pt. 12.

Budson, A. E., Todman, R. W., Chong, H.,Adams, E. H., Kensinger, E. A., Krangel, T. S., & Wright, C. I. (2006a). False recognition of emotional word lists in aging and Alzheimer Disease. *Cognitive and Behavioral Neurology, 19*, 71-78.

Budson, A. E., Todman, R. W., & Schacter, D. L. (2006b). Gist memory in Alzheimer's disease: Evidence from categorized pictures. *Neuropsychology, 20,* 113-122.

Busemeyer, J. R., Wang, Z., Townsend, J. T. (2006). Quantum dynamics of human decision-making. *Journal of Mathematical Psychology, 50,* 220-241.

Carstensen, L. L., & Mikels, J. A. (2005). At the intersection of emotion and cognition: Aging and the positivity effect. *Current Directions in Psychological Science, 14,* 117-121.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington, DC: American Psychological Association.

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility in children's memory: Psycholegal implications. *Journal of Experimental Psychology: General, 116*, 38-49.

Chopin, N. (2007). Inference and model choice for sequentially ordered hidden Markov models and model choice for sequentially ordered hidden Markov models. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 69,* 269-284.

Crowder, R. G. (1976). Principles of learning and memory. Hillsdale, NJ: Erlbaum.

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

de Jager, C. A., Hogervorst, E., Combrinck, M., & Budge, M. M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine, 33,* 1039-1050.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Research edition*. San Antonio, TX: Psychological Corporation.

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia, 45,* 3157-3166.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523-533.

Dunn, J. C. (2008). The dimensionality of the remember–know task: A state-trace analysis. *Psychological Review, 115,* 426–446.

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

Estes, W. K., & DaPolito, F. (1967). Independent variation of information storage and retrieval processes in paired-associate learning. *Journal of Experimental Psychology, 75,* 18-26.

Faglioni, P., Bertolani, L., Botti, C., & Merelli, E. (2000a). Verbal learning strategies in Parkinson's disease. *Cortex, 36,* 243-263.

Faglioni, P., Saetti, M. C., & Botti, C. (2000b). Verbal learning strategies in patients with

multiple sclerosis. *Neuropsychology, 14,* 456-470.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.

Gallo, D. A., & Roediger, H. L. (2003). Variability among word lists in eliciting memory illusions: evidence for associative activation and monitoring. *Journal of Memory and Language, 47,* 469-497.

Gardiner, J. M., & Java, R. I. (1991). Recollective experience in word and nonword recognition. *Memory& Cognition, 18,* 23-30.

Ghetti, S. (2008). Processes underlying developmental reversals in false-memory formation: Comment on Brainerd, Reyna, and Ceci (2008). *Psychological Bulletin, 134*, 764-767.

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: evidence from the dual-process signal detection model. *Child Development, 79,* 339-358.

Ghetti, S., & Castelli, P. (2006). Developmental differences in false-event rejection: Effects of memorability-based warnings. *Memory, 14,* 762-776.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory and Cognition*, *13*, 8–20.

Greeno, J. G. (1968). Identifiability and statistical properties of two-stage learning with no successes in the initial stage. *Psychometrika, 33,* 173-215.

Greeno, J. G. (1974). Representation of learning as discrete transition in a finite state space. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* San Francisco, Calif.: Freeman.

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

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

Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A., & Evans, D. A. (2003). Alzheimer disease in the U.S. population: Prevalence estimates using the 2000 census. *Archives of Neurology, 60,* 1119–1122.

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

Holliday, R. E., & Hayes, B. K. (2000). Dissociating automatic and intentional processes in children's eyewitness memory. *Journal of Experimental Child Psychology, 75,* 1-42.

Holliday, R. E., & Hayes, B. K. (2002). Automatic and intentional processes in children's recognition memory. *Applied Cognitive Psychology, 16,* 617-636.

Holliday, R. E., & Weekes, B. S. (2006). Dissociated developmental trajectories for semantic and phonological false memories. *Memory, 14,* 624–636.

Howe, M. L. (1988). Measuring memory development in adulthood: A model-based approach to disentangling storage-retrieval contributions. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood* (pp.39-64). New York: Springer-Verlag.

Howe, M. L. (1990). Development of a mathematical model of memory for clinical research applications in aging. In M. L. Howe, M. J. Stones, & C. J. Brainerd (Eds.), *Cognitive and behavioral performance factors in atypical aging* (pp. 3-68). New York: Springer-Verlag.

Howe, M. L. (2004). The role of conceptual recoding in reducing children's retroactive interference. *Developmental Psychology, 40,* 131-139.

Howe, M. L. (2006). Developmentally invariant dissociations in children's true and false memories: Not all relatedness is created equal. *Child Development, 77,* 1112-1123.

Howe, M. L., Brainerd, C. J., & Kingma, J. (1985a). Development of organization in recall: A stages-of-learning analysis. *Journal of Experimental Child Psychology, 39*, 230-251.

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

Howe, M. L., & Hunter, M. A. (1985). Adult age differences in storage-retrieval processes: A stages-of-learning analysis of developmental interactions in concreteness effects. *Canadian Journal of Psychology, 39,* 130-150.

Howe, M. L., & Hunter, M. A. (1986). Long-term memory in adulthood: An examination of the development of storage and retrieval processes at acquisition and retention. *Developmental Review, 6,* 334-364.

Howe, M. L., O'Sullivan, J. T., & Brainerd, C. J., & Kingma, J. (1989). Localizing the development of ability differences in organized memory. *Contemporary Educational Psychology, 14,* 336-356.

Howe, M. L., Rabinowitz, F. M., & Grant, M. J. (1993). On measuring the (in)dependence of cognitive processes. *Psychological Review*, *100*, 737-747.

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

Huron, C., Danion, J.-M., Giacomoni, F., Grange, D., Robert, P., & Rizzo, L. (1995). Impairment of recognition memory with, but not without, conscious recollection in schizophrenia. *American Journal of Psychiatry, 152,* 1737-1742.

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

Jacoby, L. L. & Kelley, C. M. (1992). Unconscious influences of memory: Dissociations and automaticity. In A. D. Milner & M. D. Rugg (Eds.),*The neuropsychology of consciousness*(pp. 201-233). San Diego, CA: Academic Press.

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

Katsikopoulos, K. V., & Fisher, D. L. (2001). Formal requirements of Markov state models for paired associate learning. *Journal of Mathematical Psychology, 45,* 324-333.

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of aging. *Cognitive Neuropsychology, 16,* 399-415.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review, 114,* 954–993.

Kingma, J. (1987). The new two-stage model of learning: A tool for analyzing the loci of memory differences in intellectually impaired children. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology* (pp. 38-85). New York: Springer-Verlag.

Kintsch, W. (1963). All-or-none learning and the role of repetition in paired-associate learning. *Science, 140,* 310-312.

Kintsch, W., & Morris, C. J. (1965). Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology,* 1965, *69,* 200-206.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133-159.

Koriat A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517.

Korobanova, D. (2008). *Automatic and controlled memory in schizophrenia*. Unpublished doctoral dissertation, the University of Otago, Dunedin, New Zealand.

Koutstaal, W. (2003). Older adults encode-but do not always use-perceptual details: Intentional versus unintentional effects of detail on memory judgments.

Psychological Science, 14, 189-193.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37,* 555-583.

Lampinen, J. M., Leding, J. K., Reed, K. B., & Odegard, T. N. (2006). Global gist extraction in children and adults. *Memory, 14,* 952-964.

Lampinen, J. M., Odegard, T. N., Blackshear, E., & Toglia, M. P. (2005). Phantom ROC. In F. Columbus (Ed.), *Progress in experimental psychology research.* Hauppauge, NY: Nova.

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cogntion, 30,* 332-342.

Linscott, R. J. (1999). *Recollection and automatic memory in schizophrenia and schizotypy: Evidence for hypokrisia in psychosis*. Unpublished doctoral dissertation, University of Otago, Dunedin, New Zealand.

Linscott, R. J., & Knight, R. G. (2001). Automatic hypermnesia and impaired recollection in schizophrenia. *Neuropsychology, 15,* 576-585.

Linscott, R. J., & Knight, R. G. (2004). Potentiated automatic memory in schizotyppy. *Personality and Individual Differences, 37,* 1503-1517.

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

Merriman, W. E., Azmita, M., & Perlmutter, M. (1988). Rate of forgetting in early childhood. International *Journal of Behavioral Development, 11,* 467-474.

Metzger, R. L., Warren, A. R., Price, J. D., Reed, A. W., Shelton, J., & Williams, D. (2008). Do children "D/R-M" like adults? False memory production in children. *Developmental Psychology. 44,* 169-181.

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging, 20,* 542-553.

Miller, G. A. (1952). Finite Markov processes in psychology. *Psychometrika, 17,* 149-167.

Morrison, F. J., Haith, M. M., & Kagan, J. (1980). Age trends in recognition memory for pictures: The effects of delay and testing procedures. *Bulletin of the Psychonomic Society, 16,* 480-483.

Myung, J. I., Karabatsos, G., Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology, 49,* 205-225.

Naus, M. J., Ornstein, P. A., & Kreshtool, K. (1977). Developmental differences in recall and recognition: The relationship between and memory as text expectation changes. *Journal of Experimental Child Psychology, 23,* 252-265.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida word association, rhyme, and word fragment norms*. Unpublished manuscript, University of South Florida, Tampa.

Newcombe, N. S., & Lie, E. (1995). Overt and covert recognition of faces in children and adults. *Psychological Science, 6,* 241-245.

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review, 114,* 887–953.

Ornstein, P. A., & Corsale, K. (1979). Organizational in children's memory. In C. R. Puff (Ed.), *Organization, structure, and memory.* New York: Academic Press.

Pagel, J. C. (1973). Markov analysis of transfer in paired-associate learning with high intralist similarity. *Journal of Verbal Learning and Verbal Behavior, 12,* 456-470.

Parks, C. M. (2007). The role of noncriterial recollection in estimating recollection and familiarity. *Journal of Memory and Language, 57,* 81-100.

Payne, D. G., (1986). Hypermnesia for pictures and words: Testing the recall level

hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition , 12,* 16–29.

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

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

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

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

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology, 56,* 303–308.

Petersen, R. C., Stevens, J., Ganguli, M., Tangalos, E., Cummings, J. & DeKosky, S. (2001). Practice parameter: early detection of dementia : mild cognitive impairment (an evidence-based review). *Neurology 56,* 1133–1142.

Pierce, B. H., Sullivan, A. L., Schacter, D. L., & Budson, A. E. (2005). Comparing source-base and gist-based false recognition in aging and Alzheimer's disease. *Neuropsychology, 19,* 411-419.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2008). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*, 129–156.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*, 257–286.

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

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

Reyna, V. F., Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity, 10,* 309-331.

Reyna, V. F., & Mills, B. A. (2007). Interference processes in fuzzy-trace theory: Aging, Alzheimer's disease, and development. In C. MacLeod & D. Gorfein (Eds.), *Inhibition in cognition* (pp. 185-210). Washington: APA Press.

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

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

Roediger H. L., Watson J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8,* 385-405.

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

Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron, 44,* 149–160.

Schneider, W., & Bjorklund, D. F. (1998). Memory. In R. S. Siegler & D. Kuhn (Eds.), *Handbook of child Psychology: Cognitive, language, and percetual development*. (Vol. 2, pp. 467-522). New York: Wiley.

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002a). Are false memories more difficult to forget than accurate

memories?: The effect of retention interval on recall and recognition. *Memory & Cognition, 30,* 1054-1064.

Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002b). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language, 46,* 323-340.

Skinner, E. I., & Fernandes, M. A. (2008). Interfering with remembering and knowing: Effects of divided attention at retrieval. *Acta Psychologia, 127,* 211-221.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Spaan, P. E. J., Raaijmaker, J. G. W., & Jonker, C. (2004). Early assessment of dementia: The contribution of different memory components. *Neuropsychology, 19,* 629-640.

Spezia, L. (2006). Bayesian analysis of non-homogeneous hidden Markov models. *Journal of Statistical Computation and Simulation, 76,* 713-725.

Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory. *Psychological Science, 16,* 785-791.

Theios, J., & Hakes, D. T. (1962, May). *Paired-associate response shifts in two-stage all-or-none learning*. Paper presented at Midwestern Psychological Association, Chicago, IL.

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

Toth, J. P., & Parks, C. M. (2006). Effects of age on estimated familiarity in the process dissociation procedure: The role of noncriterial recollection. *Memory & Cognition, 34,* 527-537.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1-12.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

Waugh, N. C., & Smith, J. A. (1962). A stochastic model for free recall. *Psychometrika, 27,* 141-145.

Welton, N. J., & Ades, A. E. (2005). Estimation of Markov chain transition probabilities and rates from fully and partially observed data: Uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making, 25,* 633-645.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152-176.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20,* 1341-1354.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-518.

Appendix A

Identifiability Proof for Equation 3

Equation 3 implies an observable-states process that consists of the following data events:

Q = the state on all recall tests after the last error for items with one or more precriterion errors and the state on all recall tests for items with no precriterion errors;

R = the state on all recall tests for an item, where the item is not recalled but the item has been recalled on at least one earlier recall test;

S = the state on all recall tests for an item, where the item is recalled but then it is not recalled on at least one later recall test;

E_i = the state on the *i*th recall test for an item, where the item is not recalled and it has not been recalled on any earlier recall test.

j is the maximum value of the index variable *i*. The starting vector, transition matrix, and response vector for this observable-states process are

$$W_3 = [Q(1)Q(2), Q(1)R(2), Q(1)S(2), Q(1)E_1(2), Q(1)E_2(2), \ldots, Q(1)E_j(2), R(1)Q(2), R(1)R(2), R(1)S(2), R(1)E_1(2), R(1)E_2(2), \ldots, R(1)E_j(2), S(1)Q(2), S(1)R(2), S(1)S(2), S(1)E_1(2), S(1)E_2(2), \ldots, S(1)E_j(2), E_1(1)Q(2), E_1(1)R(2), E_1(1)S(2), E_1(1)E_1(2), \ldots, E_j(1)E_j(2)] = [?_1, 0, 0, 0, 0, \ldots, 0, 0, 0, 0, 0, 0, \ldots, 0, 0, ?_2, ?_3, 0, 0, \ldots, 0, ?_4, 0, ?_5, 1 - ?_1 - ?_2 - ?_3 - ?_4 - ?_5, \ldots, 0];$$

$$M_3 = \begin{array}{c|cccccc} & Q(n+1) & R(n+1) & S(n+1) & E_2(n+1) & \ldots & E_{j+1}(n+1) \\ \hline Q(n) & 1 & 0 & 0 & 0 & \ldots & 0 \\ R(n) & u & (1-u)v & (1-u)(1-v) & 0 & \ldots & 0 \\ S(n) & 0 & z & 1 - z & 0 & \ldots & 0 \\ E_2(n) & ?_1 & 0 & ?_1 & 1-?_1-?_1 & \ldots & 0 \\ . & . & . & . & . & & . \\ . & . & . & . & . & & . \\ . & . & . & . & . & & . \\ E_j(n) & ?_{j-1} & 0 & ?_{j-1} = 1-?_{j-1} & 0 & \ldots & 0 \end{array} ; \quad C_3 = \begin{array}{c} P(\text{correct}) \\ \hline 1 \\ 0 \\ 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{array} . \quad \text{(A1)}$$

M_3 governs transitions from T_2 onward. For any recall experiment, *j* is defined as the length of the longest initial error run. Thus, if the latest occurrence of the first successful recall of any item is T_6, then j = 6, but if the latest occurrence of the first successful recall for any item is T_8, then j = 8. The starting vector W_3 gives the probabilities of the various pairings of the observable states on T_1 and T_2. Most of these pairings are impossible by the definitions of the observable states; only of six of the probabilities are nonzero. To illustrate, notice that all of the entries that begin with *R*(1) must be zero because *R* is defined as any error *after* an item has been recalled for the first time and T_1 is the first recall test. The transition matrix M_3 gives the probabilities of all inter- and intrastate transitions after T_2. M_3 also contains many zero entries that correspond to transitions that are impossible by the definitions of the states.

Because A1 involves only observable states (i.e., actual data events), each parameter in the starting vector and transition matrix is an observable quantity. There are 5 + 2*j* of these parameters, each of which is identifiable because, by Bernoulli's theorem, a unique maximum-likelihood estimator is available in the form of the proportion of events in any experiment that exhibits that data state (e.g., Bartolucci & Nigro, 2007). Such estimators are obtained from the likelihood function

$$L_{5+2j} = ?(?_i)^{N(i)} \times ?(?_{ij})^{N(I,j)}, \quad \text{(A2)}$$

where the $?_i$ range over the nonzero cells of W_3 and the $?_{ij}$ range over the nonzero cells of M_3. Note that as long as j ? 3, A1 has as many or more identifiable parameters than Equation 3.

As we know, Equation 3 contains 11 free parameters. If A1 is analyzed, using algorithms for analyzing HMMs (e.g., Chopin, 2007), in order to locate the 11 identifiable parameters of A1 that correspond to the parameters of Equation 3, that set of parameters is {$?_1$, $?_2$, $?_3$, $?_4$, $?_5$, u, v, z, w, ?, ?}. It should be stressed that this is a unique set of identifiable parameters—there is no other set containing all 11 identifiable parameters other than this one (for a proof, see Brainerd, Howe, & Kingma, 1982). As can be seen, the first eight parameters in this set appear in A1 because they are probabilities of simple data events. The last three parameters are not probabilities of simple data events, but they are involved, along with the parameters u and v in complex expressions for $?_i$ and $?_i$, which are probabilities of simple data events. Those expressions are:

$?_i$, = $(w^{i-1}[u - (1-?)?]) / ((w^{i-1} [? - (1-u)v]) / (w - (1-u)v)) + ((1 - (? - (1-u)v)) / (w - (1-u)v))((1-u)v)^{i-1})$; (A3)

$?_i$, = $(1-u)(1-v) - (w^{i-1} [? - (1-u)v][1 - (u - (1-?)?] / [? - (1-u)v]) / ((w^{i-1} [? - (1-u)v]) / (w - (1-u)v)) + ((1 - (? - (1-u)v)) / (w - (1-u)v))((1-u)v)^{i-1})$. (A4)

Because Equations 3 contains 11 free parameters and the set {$?_1$, $?_2$, $?_3$, $?_4$, $?_5$, u, v, z, w, ?, ?} also contains 11 identifiable parameters, to prove that Equation 3's parameters are identifiable it is only necessary to find a *unique* equation for each of its parameters that expresses it as a function of some of the members of {$?_1$, $?_2$, $?_3$, $?_4$, $?_5$, u, v, z, w, ?, ?}. Algebraic analysis of Equations 3 and A1, using algorithms for analyzing HMMs, yields such a series of 11 equations (A5-A15), each of which proves that the corresponding parameter of Equation 3 is identifiable. The complexity of the individual expressions can be greatly reduced by exhibiting the equations for D_{3C}, D_{3E}, J_{3C}, and J_{3E} first and then using these parameters as shorthand in the remaining seven equations. The identifiability equations for D_{3C}, D_{3E}, J_{3C}, and J_{3E} are:

$D_{3C} = (z?_4) / (?_4 + ?_5)$; (A5)

$D_{3E} = (u - [((1-v)\ ?_4) / (?_4(v-1) - ?_5)]) / (1 - [((1-v)\ ?_4) / (?_4(v-1) - ?_5)])$; (A6)

$J_{3C} = (1-z) / [1 - (z?_4) / (?_4 + ?_5)]$; (A7)

$J_{3E} = (1-v) / ([(v?_4) / (?_4 + ?_5)] - 1)$. (A8)

The forgetting parameter of Equation 3, R_f, can then be shown to be identifiable, as follows:

$R_f = ((?_4/?_3)(z - D_{3C})(J_{3E}/z) + J_{3C} - 1) / ((?_2/?_3)(z - D_{3C})(J_{3E}/z) - J_{3C})$. (A9)

Next, the identifiability expression for J_1 is:

$J_1 = 1 - (J_{3E} / (J_{3E} + (1- J_{3E})(?_4/?_3)))$; (A10)

The identifiability expressions for D_1 and R_1 are:

$D_1 = (?_1[R_f + (1- R_f)J_{3C}] - ?_2(1- R_f)J_{3C}[D_{3C}/(1-D_{3C})J_{3C}]) / (R_f - (1- R_f)J_{3C})$; (A11)

$R_1 = ?_2 / ((1-D_1)J_1 [R_f + (1- R_f)J_{3C}])$. (A12)

Last, the identifiability expressions for D_2, R_2, and J_2 are:

$D_2 = (1-w)((? -1)[?X - K - (1- D_1)(1- R_1)(1- D_2)(1- R_2)]) / (?Z - Y - R_2(1- R_1) + (? -1)[?X - K - (1- D_1)(1- R_1)(1- D_2)(1- R_2)])$; (A13)

$R_2 = 1 - (w / (1- D_2))$; (A14)

$J_2 = (?X - K - (1- R_1)(1- R_2)) / (D_2R_2(1- D_1)(1- R_1))$; (A15)

In A13 and A15, the variables K, X, Y, and Z are defined as $K = R_1(1- D_1)(1- J_1)(1 - D_{3E}(1-J_{3E})$, $X = (1- D_1)(1- R_1) + R_1(1- D_1)(1- J_1)$, $Y = (((R_1(1- D_1)(1- J_1))[D_{3C}J_{3C} + D_{3C}(1-J_{3C})]) / (1 - (1- D_{3C})J_{3C})$, and $Z = R_2(1- D_1)(1- R_1) + R_1(1- D_1)(1- J_1)[J_{3E} + (1- D_{3C})J_{3C}]$. This completes the identifiability proof for Equation 3.

Parameter Estimation, Goodness of Fit, and Hypothesis Testing

As Equation 3's parameter space is identifiable, a likelihood function can be written from which maximum likelihood estimates of its parameters are obtained from sample data. This function is also essential for evaluating goodness of fit and for testing within- and between-condition statistical hypotheses about the observed values of the model's parameters. Equation 3's likelihood function is:

$$L_{11} = [D_1R_1 (R_1(1- D_1)J_1(1- R_f)D_{3C}J_{3C}]) / (1 - (1- D_{3C})J_{3C})]^{N[Q(1)Q(2)]} \times$$
$$[(1- D_1)(1- R_1) + (R_1(1- D_1)(1-J_1)(R_f + (1- R_f)(1- J_{3E}))]^{N[E_1(1)E_1(2)]} \times$$
$$[(R_1(1- D_1)(1-J_1)(1- R_f)D_{3C}J_{3E}) / (1 - (1- D_{3C})J_{3C})]^{N[E_1(1)Q(2)]} \times$$
$$[(R_1(1- D_1)(1-J_1)(1- R_f)(1-D_{3C})J_{3E}(1- J_{3C})) / (1 - (1- D_{3C})J_{3C})]^{N[E_1(1)S(2)]} \times$$

$$[R_1(1- D_1)J_1(R_f + (1- R_f)(1-J_{3C}))]^{N[S(1)R(2)]} \times$$
$$[(R_1(1- D_1)J_1)(1- R_f)J_{3C}(1- D_{3C})(1- J_{3C})) / (1 - (1- D_{3C})J_{3C})]^{N[S(1)S(2)]} \times$$
$$[D_{3E} + ((1- D_{3E})D_{3C}J_{3E}) / (1 - (1- D_{3C})J_{3C})]^{N(R,Q)} \times$$
$$[1 - D_{3E} + ((1- D_{3E})D_{3C}J_{3E}) / (1 - (1- D_{3C})J_{3C})]^{N(R,R) + N(R,S)} \times$$
$$[(1-J_{3E})(1 - (1- D_{3C})J_{3C}) / ((1-J_{3E})(1 - (1- D_{3C})J_{3C}) + (1- D_{3C})(1-J_{3C})J_{3E})]^{N(R,R)} \times$$
$$[1 - (1-J_{3E})(1 - (1- D_{3C})J_{3C}) / ((1-J_{3E})(1 - (1- D_{3C})J_{3C}) + (1- D_{3C})(1-J_{3C})J_{3E})]^{N(R,S)} \times$$
$$[1 - (1- D_{3C})J_{3C}]^{N(S,R)} \times [(1- D_{3C})J_{3C}]^{N(S,S)} \times$$
$$\overset{j}{\underset{I=1}{?}} [(?_i)^{N[E_i(i)Q(i+1)]} \times (?_i)^{N[E_i(i)S(i+1)]} \times (1 - ?_i - ?_i)^{N[E_i(i)\, E_{i+1}(i+1)]}]. \quad \text{(A16)}$$

The exponent of each term in A16 is a data count—specifically, the raw frequency of one of the data events in A1. The exponents of the first six terms are just the total numbers of items that begin in the six possible starting combinations for T_1 and T_2, and the exponents of the remaining terms are just the total numbers of items that exhibit the indicated inter- or intra-state transitions on consecutive recall tests, from T_2 onward. As before, *j* is the maximum length of the initial error run, and the $?_i$ and $?_i$ are defined as in A3 and A4. Once counts from sample data are inserted for the exponents, A16 produces maximum likelihood estimates of the 11 memory parameters. This done is via computer search, using any standard search algorithm, such as EM or SIMPLEX.

A goodness-of-fit test for A3 is then obtained in the usual way by computing a likelihood ratio statistic (e.g., cf. Hu, 1998), which compares the *a posteriori* probability of sample data under Equation 3, which has 11 degrees of freedom, to the *a posteriori* probability of the same data under A1, which has 5 + 2*j* degrees of freedom. The exact statistic is twice the negative natural log of the ratio of A16 to A2:

$$G^2 = -2\ln[L_{11} / L_{5+2j}]. \quad \text{(A17)}$$

Because the difference in the degrees of freedom in the numerator and in the denominator is large whenever j ? 5, as it is in the data sets in our corpus (because subjects learned to recall to an errorless performance criterion), the model in Equation A16 is easy to reject if the data do not closely approximate a two-stage absorbing Markov process (Brainerd et al., 1990).

The test statistic G^2 has an asymptotic $?^2(2j\text{-}6)$ distribution, where $2j - 6$ is the difference between the number of degrees of freedom that are used to estimate L_{11} versus L_{5+2j}. Thus, as long as $j > 3$, the fit of Equation 3 to sample data can be tested with at least 2 degrees of freedom. This leads to an important simplification of the standard criterion-learning design for which Equation 3 is defined, a simplification that is exploited in the section of the present paper that deals with cognitive impairment: It is possible to estimate all of Equation 3's parameters and test goodness of fit using a fixed-trials design of the form $S_1T_1T_2$,S_2T_3,S_3T_4 because $j > 3$. Although most of the data sets that we analyze involve the criterion-learning design, we also analyze some data sets that involve the fixed-trials design, in which the subjects were patients with certain forms of cognitive impairment.

In research, principal interest does not attach to the technical issues of parameter identifiability, estimation, and fit that have been the focus of this appendix. Rather, interest centers on using the direct access, reconstruction, and familiarity judgment parameters to interpret recall performance. That, in turn, depends on being able to test within- and between-condition statistical hypotheses about the parameter values are estimated for sample data. As with any mathematical model, the aim of between-condition tests is to localize treatment effects within particular parameters (thereby pinpointing the memory processes that are responsible for the effects), whereas the purpose of within-condition tests is to decide whether one parameter is larger than another (say, whether $D_1 > D_2$ or $R_1 > D_1$). Between-condition tests are the more complicated of the two because they involve three steps. Consider an arbitrary experiment that contains *k* different conditions. The first step is to compute a conditionwise test that determines whether there is global statistical evidence that the parameters of Equation 3 differ between those conditions. The appropriate statistic is:

$G^2 = -2\ln[L'_{11} / (L(1)_{11} \times L(2)_{11} \times \ldots \times L(k)_{11})]$, (A18)
where the denominator contains the values of A17 that are computed for the data of each of the k conditions, and the numerator contains a single value of A17 that is computed for the pooled data of the k conditions. The G^2 statistic is asymptotically distributed as $?^2[11(k-1)]$, and it tests the null hypothesis that Equation 3's parameters do not vary between the k conditions. If this null hypothesis is rejected, the second step is to compute a conditionwise test for any pair of conditions that are of interest—say, conditions i and j. This test evaluates the null hypothesis that Equation 3's parameters do not differ between conditions i and j. That test statistic is:
$G^2 = -2\ln[L''_{11} / (L(i)_{11} \times L(j)_{11}]$, (A19)
where the denominator contains the values of A17 that are computed for conditions i and j and the numerator contains a value of A17 that is computed for the pooled data of the two conditions. This statistic is asymptotically distributed as $?^2(11)$. Last, if this statistic produces a null hypothesis rejection, parameterwise tests are computed to determine which specific pairs of parameters differ between conditions i and j. That test statistic is:
$G^2 = -2\ln[L(?_i)_{11} \times L(?_j)_{11}]L(?'_{11} / (L(i)_{11} \times L(j)_{j11}]$, (A20)
where $?$ is any one of Equation 3's parameters, the denominator is the same as the denominator of A19, and the numerator contains values of A17 for conditions i and j that are computed under the constraint that $?$ has the same value for both conditions. This statistic is asymptotically distributed as $?^2(1)$.

Turning to within-condition tests, such tests compare the values of different parameters within a single condition. Two tests of this sort are often of interest: exact numerical hypotheses and relational hypotheses. The former stipulate that some parameters must have predetermined values, such as 0 or 1, whereas the latter stipulate that some sort of numerical relationship (usually equality or inequality) must hold between pairs of parameters. (For instance, a relational hypothesis that figures prominently in our application of Equation 3 to developmental recall data is that learning to reconstruct items is easier than learning to directly access them, which generates the relational hypotheses $R_1 > D_1$ and $R_2 > D_2$.) For any condition i, all of these tests use the value of $L(i)_{11}$, which is computed in A18, when goodness of fit is evaluated. The test statistic for within-condition hypotheses is:
$G^2 = -2\ln[L(i)_{10} / (L(i)_{11}]$, (A21)
which is asymptotically distributed as $?^2(1)$. $L(i)_{10}$ is the likelihood of the data of condition i that is estimated with one less degree of free than $(L(i)_{11}$ because an exact or relational hypothesis about condition i imposes a single restriction on the freedom of Equation 3's parameters to vary.

The trichotomous theory assumes that recall involves two distinct retrieval operations and, therefore, that learning to recall is a two-stage process like that in Equation 3. On analogy to the current debate over one- versus two-process interpretations of recognition, it is possible to compare the fit of the two-stage model (Equation A17) to the fit of a model that assumes that learning to recall involves only one stage. That test is obtained as follows. First, we define three observable states of recall data: Q, R, and S. The definitions of these states are the same as in Equation A1. The comparative model fit test is
$G^2 = -2\ln[L_6/L_{11}]$, (A22)
which is asymptotically distributed as $?^2(5)$. L_{11} is the likelihood of the data under the two-stage model, as computed from Equation A16, and L_6 is the likelihood of the same data under the assumption that learning to recall involves only one stage. Therefore, the test statistic evaluates the null hypothesis that the one- and two-stage models fit the data equally well. The value of L_6 is computed from the expression

$$\begin{aligned} L_6 = \; & [m + (1 - m)(l - n)pc/(l - (1 - c)p)]^{N[Q(1)Q(2)]} \times [(1 - m)ns(c/(1 - (1 - c)p))]^{N[R(1)Q(2)]} \times \\ & [(1 - m)n(1 - s)]^{N[R(1)R(2)]} \times [(1 - m)ns(1 - c)(1 - p)/(1 - (1 - c)p)]^{N[R(1)S(2)]} \times [(l - m)(l - n)(l - p)]^{N[S(1)R(2)]} \times \\ & [(l - m)(l - n)p(l - c)(l - p)/(1 - (1 - c)p)]^{N[S(1)S(2)]} \times [d + (1 - d)sc/(1 - (1 - c)p)]^{N(R, Q]} \times \\ & [1 - d - (1 - d)sc/(1 - (1 - c)p)]^{N[R, R] + N[R, S]} \times \\ & [(1 - s)(1 - (1 - c)p)/((1 - s)(1 - (1 - c)p) + (1 - c)(1 - p)s)]^{N[R, R]} \times \end{aligned}$$

$$[1 - (1 - s)(1 - (1 - c)p)/((1 - s)(1 - (1 - c)p) + (1 - c)(l - p)s)]^{N[R, S]} \times$$
$$[1 - (1 - c)p]^{N[S, R]} \times [(l - c)p]^{N[S, S]}, \qquad (A23)$$

where $\{c, d, m, n, p, s\}$ is the parameter set of the generic one-stage Markov chain for recall; that is, the parameter space of the one-stage alternative to Equation 3. Thus, if the comparative fit test in Equation A22 produces a null hypothesis rejection, learning to recall involves more than one stage. If the comparative fit test in Equation A17 then fails to produce a null hypothesis rejection, learning to recall does not involve more than two stages.

Appendix B

Description of Data Corpus

The third and fourth sections of this paper focus on lifespan developmental trends in direct access, reconstruction, and familiarity judgment and on testing theoretical predictions about them. The reported findings are derived from a corpus of developmental recall studies. Although some of the data sets from these studies are unpublished, the great preponderance of them appeared in a series of articles that have been published over the past quarter-century. The published child and adolescent data sets appeared in articles by Brainerd (1985), Brainerd and Howe (1982), Brainerd, Howe, and Desrocher (1982), Brainerd, Howe, and Kingma (1982), Brainerd, Howe, Kingma, and Brainerd (1984b), Brainerd, Kingma, and Howe (1986), Brainerd et al. (1990), Howe, Brainerd, and Kingma (1985a, 1985b), and Howe et al. (1989). The published young adult data sets appeared in articles by Brainerd, Howe, and Kingma (1982), Brainerd, Howe, Kingma, and Brainerd (1984a), Brainerd, Kingma, and Howe (1985), Howe (1988), and Howe and Hunter (1985, 1986). The published data sets for older adults appeared in articles by Howe (1988) and Howe and Hunter (1985, 1986). As full methodological details may be found in those publications, we restrict attention here to general features of the procedures, subject samples, and fit analyses of the data sets in this corpus.

There are a total of 207 data sets in which the subjects learned to recall lists to a stringent acquisition criterion of one or two errorless tests. Of these, 183 are published and 24 are unpublished. In each data set, subjects from one of four age levels—children (7-8 years), adolescents (11-12 years), young adults (20-21 years), and healthy older adults (70-71 years)—learned to recall a list of items via the $S_1T_1T_2$, S_2T_3, S_3T_4, … procedure over which the identifiable model is defined. The sizes of the subject samples for the individual data sets range between 20 and 40. The corpus is subdivided into 71 data sets in which the subjects were children, 81 data sets in which the subjects were adolescents, 39 data sets in which the subjects were young adults, and 16 data sets in which the subjects were healthy older adults. The specific list that subjects learned to recall consisted of either unrelated pictures (43 data sets), unrelated words (96 data sets), or words that were exemplars of 1 to 4 familiar categories (68 data sets). List length varied from 10 to 24 targets. The specific learning procedure that was used was either paired-associate recall (68 data sets), cued recall (25 data sets), or free recall (122 data sets). In all data sets, the learning procedure incorporated standard short-term memory controls (e.g., buffering activities) between study cycles and recall tests and between the two recall tests on Trial 1. Thus, in all the data sets, the subjects first studied a list of targets (unrelated words, unrelated pictures, categorized words), then performed a brief buffer activity (e.g., 30 sec of letter shadowing) to empty short-term memory, then responded to the first recall test (free, cued, paired-associate), then performed another buffer activity, then responded to a second recall test, then studied the list for a second time, then performed another buffer activity, then responded to a third recall test, and so until a criterion of errorless performance on the recall test and been reached.

All 207 data sets passed rigorous fit evaluations. First, the comparative fit test in Equation A22 was computed. As we saw, this test evaluates a simpler one-process model, which assumes that learning to recall involves one stage rather than two. As we also saw, this test generates a G^2 statistic with an asymptotic ?^{2}(5) distribution. This test was computed for all 207 data sets, and in each instance, the hypothesis that recall performance could be accounted for by a one-process model was rejected at a high level of confidence. (The modal value of the $G^2(5)$ statistic was more than twice the critical value of 11.07, which is required to reject the one-process model at the .05 level of confidence.) Second, the comparative fit test in Equation A17 was computed, which evaluates the hypothesis that the data were generated by a two-process model that is isomorphic to Equation 3. This hypothesis could not be rejected for any of these data sets, and the modal value of the $G^2(2j\text{-}6)$ fit statistic was roughly half the value that would be required to reject this hypothesis at the .05 level of confidence. Third, when the one-process model is rejected (Equation A22) and the two-process model cannot be rejected (Equation A17), a further type of analysis provides instructive information about just how closely the data conform to the two-process model in Equation 3. This follow-up analysis involves deriving the sampling distributions of learning statistics such as those in Figures 1 and 2 from Equation 3 and then comparing the empirical distributions of these statistics (as they that are observed in individual data sets) to the corresponding distributions that are predicted by Equation 3. This allows one to detect whether, despite global fit results suggesting that learning to recall involves exactly two processes, there

are systematic discrepancies between predicted and observed features of more fined-grained aspects of the data. Predicted-observed comparisons of this sort have been conducted for more than half of the present data sets, and such systematic discrepancies have not been detected. Thus, the fit analyses of this corpus converge on the conclusion that the model in Equation 3 provides a very close approximation to the data of standard recall paradigms.

Repeated-Recall Version of Equation 3

Consider an experiment of the form $S_1T_1T_2T_3$; that is, the study list is presented once, followed by three recall tests for that list. For individual items, this design can produce eight distinct performance outcomes: $C_1C_2C_3$, $C_1C_2E_3$, $C_1E_2C_3$, $C_1E_2E_3$, $E_1C_2C_3$, $E_1C_2E_3$, $E_1E_2C_3$, and $E_1E_2E_3$, where C denotes that the item is recalled and E denotes that it is not recalled. It is possible to estimate six of the identifiable parameters of Equation 3 in this outcome space—namely, D_1, R_1, R_f, J_1, J_{3C}, and J_{3E}. This is because the probability of each of the data events can be expressed as a unique function of these parameters, as follows:

$$P(C_1C_2C_3) = D_1 + (1- D_1)R_1J_1(1- R_f)^2(J_{3C})^2; \quad \text{(B1)}$$
$$P(C_1C_2E_3) = (1- D_1)R_1J_1(1-R_f)J_{3C}[R_f + (1- R_f)(1- J_{3C})]; \quad \text{(B2)}$$
$$P(C_1E_2C_3) = (1- D_1)R_1J_1(1-R_f)^2(1-J_{3C})J_{3EC}; \quad \text{(B3)}$$
$$P(C_1E_2E_3) = (1- D_1)R_1J_1[R_f + (1-R_f)(1-J_{3C})R_f + (1-R_f)^2(1-J_{3C})(1-J_{3EC})]; \quad \text{(B4)}$$
$$P(E_1C_2C_3) = (1- D_1)R_1(1-J_1)(1-R_f)^2J_{3C}J_{3EC}; \quad \text{(B5)}$$
$$P(E_1C_2E_3) = (1- D_1)R_1(1-J_1)(1-R_f)J_{3C}[R_f + (1- R_f)(1- J_{3C})]; \quad \text{(B6)}$$
$$P(E_1E_2C_3) = (1- D_1)R_1(1-J_1)(1-R_f)^2(1- J_{3E})J_{3E}; \quad \text{(B7)}$$
$$P(E_1E_2E_3) = (1- D_1)(1-R_1) + (1- D_1)R_1(1-J_1) [R_f + (1-R_f)(1-J_{3E})R_f + (1-R_f)^2(1-J_{3EC})^2]. \quad \text{(B8)}$$

Estimates of D_1, R_1, R_f, J_1, J_{3C}, and J_{3E} are easily obtained for sample data by minimizing the following likelihood function, using a computer search program, such as GPT (Hu, 1998):

$$L_6 = [D_1 + (1- D_1)R_1J_1(1- R_f)^2(J_{3C})^2]^{N(C_1C_2C_3)} \times [(1- D_1)R_1J_1(1-R_f)J_{3C}[R_f + (1- R_f)(1- J_{3C})]]^{N(C_1C_2E_3)} \times$$
$$[(1- D_1)R_1J_1(1-R_f)^2(1-J_{3C})J_{3EC};]^{N(C_1E_2C_3)} \times [(1- D_1)R_1J_1[R_f + (1-R_f)(1-J_{3C})R_f +$$
$$(1-R_f)^2(1-J_{3C})(1-J_{3EC})]^{N(C_1E_2E_3)} \times P(E_1C_2C_3) = [(1- D_1)R_1(1-J_1)(1-R_f)^2J_{3C}J_{3EC};]^{N(E_1C_2C_3)} \times$$
$$[(1- D_1)R_1(1-J_1)(1-R_f)J_{3C}[R_f + (1- R_f)(1- J_{3C})]]^{N(E_1C_2E_3)} \times [(1- D_1)R_1(1-J_1)(1-R_f)^2(1- J_{3E})J_{3E}]^{N(E_1E_2C_3)} \times$$
$$[(1- D_1)(1-R_1) + (1- D_1)R_1(1-J_1) [R_f + (1-R_f)(1-J_{3E})R_f + (1-R_f)^2(1-J_{3EC})^2]]^{N(E_1E_2E_3)}. \quad \text{(B9)}$$

The exponents of the terms in B9 are data counts that correspond to the 8 possible data events. Specifically, the exponents are just the total numbers of times that each event is observed in sample data. Because six parameters are estimated, the likelihood value in B9 is computed with six degrees of freedom. A goodness-of-fit test for B9 is then obtained in the usual way by computing a likelihood ratio statistic that compares this value to the likelihood of the same data when all seven observable probabilities are free to vary. That test statistic, which is asymptotically distributed as ?^{2}(1), is

$$G^2 = -2\ln[L_6 / L_7], \quad \text{(B10)}$$

where L_7 is the likelihood of the data when all 7 observable parameters are free to vary. Last, within- and between-condition tests of hypotheses about the parameters can be tested using the procedures that were described earlier (Appendix A) for the full 11 parameter version of Equation 1.

Author Note

C. J. Brainerd and V. F. Reyna, Department of Human Development, Cornell University; M. L. Howe, Department of Psychology, Lancaster University.

Preparation of this article was supported in part by National Science Foundation Grant SBR973-0143 to the first and second authors and by Economic and Social Research Council of Great Britain Grant RES-062-23-0452 to the third author. The experiments were supported by grants from the Natural Sciences and Engineering Research Council to the first and third authors. We are indebted to A. F. Healy and J. C. Dunn for many incisive comments on earlier drafts of this article.

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

Footnotes

[1]Although remember/know judgments are the most common method of separating recollection and familiarity (see Yonelinas, 2002), there is continuing controversy about whether two memory processes, or only one, underlie such judgments. For instance, Donaldson (1996), in his early review of the literature, proposed a one-process signal detection model in which a more stringent decision criterion is used for remember judgments than for know judgments. He found that extant remember/know data were more consistent with this one-process interpretation than with a dual-process interpretation. In a recent meta-analysis, Dunn (2008) pointed out that the dual-process interpretation predicts that if criterion stringency is held constant, between-condition variability in remember/know judgments will exhibit a bi-dimensional structure. Contrary to this interpretation, the meta-analysis provided little or no evidence of bi-dimensionality.

[2]Recall data present a special case of a general research problem that is known as hidden Markov models (HMM; see Rabiner, 1989). In the biological and behavioral sciences, it often happens that the statistical structure of a target data space (sequences of responses to recall tests in the present case) exhibits Markovian properties. In this situation, some HMM is generating the data, but the exact model and whether its parameters are identifiable in the data space are unknown. The problem is to find the HMM and its identifiable parameters. This can be done by formulating an observable-states process that is implied by the HMM and then analyzing that process with algorithms that have been developed for this purpose (e.g., Bordes & Vandekerkhove, 2005). For instance, this procedure is used in Appendix A to locate the identifiable parameters of Equation 3.

[3]When attempting to identify reliable patterns across many sets of published data in which the same parameters of memory performance (e.g., signal detection parameters) have been reported, it is common practice to compute inferential statistics such as *t*, *F*, and *r* (e.g., see Donaldson, 1996). This practice is followed throughout the present section of the paper. In the ANOVAs of parameter values that we report, data sets are treated as subjects, age is treated as a between-subjects factor, and parameters are treated as repeated measures factors. These particular analyses are mixed-model ANOVAs that control for inter-experiment correlations between the values of different parameters.

[4]The intrusion rate was computed as follows. First, most intrusions are strong forward associates of individual words on study lists (e.g., Payne et al.., 1996). Hence, the first step in determining the intrusion rate for each list was to find the first forward associates of each list word, using the Nelson, McEvoy, and Schreiber (1999) norms of word association. We computed the intrusion probability for each of these unpresented words using a portion of the data for the corresponding list words. Specifically, for each list word, we used the data between the trial of first correct recall and the trial of last error. If the last error has not yet occurred, the word cannot yet be directly accessed, and if first success has already occurred, the word can be reconstructed. The number of words that meet this criterion on each trial somewhat underestimates the number words that are being reconstructed because, of course, some words are still being reconstructed after the last error has occurred. However, the value of the D_{3E} can be used to estimate the number words that are still being reconstructed on each trial following the trial of last error. On each of those trials for each list word, we counted the total number of times that the first associate was falsely recalled and divided by the number of trials. We then averaged over all list words to find the overall intrusion rate. Only trials between the first correct recall and the last error were used because those are the only trials on which, according to the theory, reconstruction must be occurring: (a) Before the first correct recall, reconstruction may not be occurring because the item may still be in state *U*, and (b) reconstruction may not be occurring after the trial of last error because the item may have entered state *L*.

[5]Throughout this section, the significance tests that are used to decide whether parameters differed reliably involved standard likelihood ratio comparisons of parameter estimates for individual conditions (see Appendix A and Appendix B), rather than ANOVAs of mean parameter values for several conditions (as in the prior two sections). Thus, in the present section, whenever parameter values are described as differing, this means that the relevant likelihood-ratio test produced a null hypothesis rejection at the .05 level of confidence, and whenever parameter values are described as not differing, this means that the relevant likelihood-ratio test failed to produce a null hypothesis rejection.

Table 1

Identifiable Parameters of the Trichotomous Model of Memory Development

Identifiable parameter	Definition
D_1	For an item that can neither be directly accessed nor reconstructed, D_1 is the probability that subjects learn how to directly access it on Trial 1.
D_2	For an item that can neither be directly accessed nor reconstructed, D_2 is the probability that subjects learn how to directly access it on any trial after Trial 1.
D_{3C}	If an item is reconstructed and is output on any trial, D_{3C} is the probability that it can be directly accessed on the next trial.
D_{3E}	If an item is reconstructed but not output on any trial, D_{3E} is the probability that it can be directly accessed on the next trial.
R_1	For an item that can be neither directly accessed nor reconstructed, R_1 is the probability that subjects learn how to reconstruct it on Trial 1.
R_2	For an item that can be neither directly accessed nor reconstructed, R_2 is the probability that subjects learn how to reconstruct it on any trial after Trial 1.
J_1	If subjects learn how to reconstruct an item on Trial 1 but not how to directly access it, J_1 is the probability that they are confident enough in the reconstruction to output it.
J_2	If subjects learn how to reconstruct an item but not how to directly access it on any trial after Trial 1, J_2 is the probability that they are confident enough in the reconstruction to output it.
J_{3C}	If subjects reconstruct an item on any pair of consecutive trials, J_{3C} is the probability that they will be confident enough to output it on the second trial if they output it on the first trial.
J_{3EC}	If subjects reconstruct an item on any pair of consecutive trials, J_{3E} is the probability that they will be confident enough to output it on the second trial if they did not output it on the first trial.

Table 2

Parameters of Equation 2 as Functions of the Identifiable Parameters of the Generic Markov Chain for Recall

Identifiable parameter	Function of parameters in Equation 2
w	$(1-D_1)(1-R_1)$
z	$1 - (1-D_{3C})J_{3C}$
?	$D_1R_1 + [((1-D_1)R_1J_1D_{3C}) / (1 - (1-D_{3C})J_{3C})]$
?	$((1-D_1)(1-R_1) + (1-D_1)R_1(1-J_1)) / [(1 - D_1R_1 - (1-D_1)R_1J_1D_{3C}) / (1 - (1-D_{3C})J_{3C})]$
u	$[D_{3E} + ((1- D_{3E})D_{3C}J_{3E})] / [1 - (1-D_{3C})J_{3C}]$
?	$[(1-J_{3E})(1 - (1-D_{3C})J_{3C})] / [(1-J_{3E})(1 - (1-D_{3C})J_{3C}) + J_{3E}(1-D_{3C})(1-J_{3C})]$
?	$[(1-D_1)(1-R_1)((1-D_2)(1-R_2) + (1-D_2)R_2(1-J_2) + (1-D_1)R_1(1-J_1)(1-D_{3E})(1-d)(1- J_{3E}))] / [(1-D_1)(1-R_1) + (1-D_1)R_1(1-J_1)]$
?	$[D_2R_2(1-D_1)(1-R_1)(D_2 + (1-D_2)J_2D_{3C}c) / ((1 - (1-D_{3C})J_{3C}) + (1-D_1)R_1(1-J_1)(D_{3C}J_{3C} + (1-J_{3C})D_{3E}))] / [R_2(1-D_1)(1-R_1)(1 – (1- D_2)(1- J_2)) + (1-D_1)R_1(1-J_1)(J_{3E} + (1- J_{3E})D_{3E}))]$

Table 3

Bivariate Correlations between Direct Access and Reconstruction in Children, Adolescents, and Young Adults

	D_1	D_2	R_1	R_2
		Children		
D_1	1	.75***	-.59***	-.55***
D_2	.75***	1	-.40**	-.49***
R_1	-.59***	-.40**	1	.79***
R_2	-.55***	-.49***	.79***	1
		Adolescents		
D_1	1	.44***	-.37**	-.26*
D_2	.44***	1	-.20	-.56***
R_1	-.37**	-.20	1	.47***
R_2	-.26*	-.58***	.47***	1
		Young adults		
D_1	1	.81***	.03	-.19
D_2	.81***	1	-.02	-.53***
R_1	.03	-.02	1	.63***
R_2	-.19	-.53***	.63***	1

***p < .0001; **p < .001; *p < .02

Table 4

Fit Evaluation of the Fixed-Trials Implementation of the Recall Model

Experiment/Group	Fit statistic
Experiment 1:	
Healthy	$G^2(4) = 3.55$
Alzheimer's	$G^2(4) = .02$
Depression	$G^2(4) = 3.73$
Experiment 2:	
Test 1: List 1	$G^2(4) = 6.21$
Test 1: List 2	$G^2(4) = 5.78$
Test 2: List 1	$G^2(4) = 7.22$
Test 2: List 2	$G^2(4) = 8.44$

*The critical value of the G^2 statistic for rejecting the null hypothesis that the model fits the data is 9.49.

Table 5

Parameter Estimates for Three Groups of Older Adults

Parameter	Group		
	Healthy	Alzheimer's	Depression
Direct access: state *U*			
D_1	.39	.02	.11
D_2	.38	.00	.01
Mean	.39	.01	.06
Direct access: state *P*			
D_{3C}	.56	.08	.05
D_{3E}	.47	.01	.53
Mean	.52	.05	.29
Reconstruction			
R_1	.22	.04	.31
R_2	.39	.05	.17
Mean	.31	.05	.24
Familiarity Judgment			
J_1	.62	.93	1
J_2	.83	1.0	.87
J_{3C}	.82	.71	.65
J_{3E}	.50	.03	.63
Mean	.69	.68	.79

Note. These results are based on a reanalysis of research that was reported by Howe (1990).

Table 6

Longitudinal Parameter Estimates for a Group of Older Adults

Parameter	Testing session			
	Initial		Delayed	
	List 1	List 2	List 1	List 2
Direct access: state *U*				
D_1	.15	.29	.16	.45
D_2	.26	.29	.02	.48
Mean	.21	.29	.09	.47
Direct access: state *P*				
D_{3C}	.50	.54	.56	.30
D_{3E}	.10	.72	.10	.73
Mean	.30	.63	.33	.52
Reconstruction				
R_1	.40	.41	.30	.17
R_2	.48	.51	.44	.13
Mean	.44	.46	.37	.15
Familiarity Judgment				
J_1	.85	.70	.94	.77
J_2	.85	.72	1.0	1.0
J_{3C}	.66	.56	.63	.45
J_{3E}	.98	.88	.75	.81
Mean	.84	.72	.83	.76

Note. These results are based on a reanalysis of research that was reported by Howe (1990).

Table 7
Dual-Process Parameter Estimates and Fit Tests for Schizophrenic and Control Subjects

Group/Condition	Parameter							
	D_1	R_1	R_f	J_1	J_{3C}	J_{3C}	M_J	$G^2(1)$
Schizophrenic								
Long								
Free	.10	.12	.14	.35	.57	.34	.63	0.07
Associative	.00	.07	.07	.72	.79	.73	.75	1.25
Short								
Free	.04	.38	.11	.81	.96	.44	.74	0.61
Associative	.10	.08	.12	.65	.82	.51	.66	0.01
M_{Long}	.05	.10	.11	.54	.68	.54	.59	
M_{Short}	.07	.23	.12	.73	.89	.48	.70	
M_{Free}	.07	.25	.13	.58	.77	.39	.58	
M_{Assoc}	.05	.08	.10	.69	.81	.62	.71	
Control								
Long								
Free	.08	.25	.03	.70	.87	.25	.61	3.03
Associative	.09	.06	.37	.52	.48	.83	.61	0.64
Short								
Free	.23	.31	.07	.77	.98	.52	.76	1.12
Associative	.40	.08	.12	.62	.85	.32	.60	0.28
M_{Long}	.09	.16	.20	.61	.68	.55	.61	
M_{Short}	.32	.20	.10	.70	.92	.42	.68	

Table 7—*continued*

M_{Free}	.16	.28	.05	.74	.93	.39	.69
M_{Assoc}	.25	.07	.25	.57	.67	.58	.61

Note. These results are based on a reanalysis of research that was reported by Korobanova (2008). The critical value of the $G^2(1)$ statistic to reject the null hypothesis that dual-recall model fits the data is 3.84.

Table 8

Summary of Numerical Predictions about Model Parameters

Predictions	Explanations
Dissociation	Members of matched pairs of direct access and reconstruction parameters (D_1/R_1 and D_2/R_2) should be singly and doubly dissociated by certain manipulations and values of pair members should be negatively correlated over list conditions.
Association	Members of matched pairs of direct access parameters (D_1/D_2) and matched pairs of reconstruction parameters (R_1/R_2) should be positively correlated over list conditions.
Ease of learning	Under comparable conditions, learning to reconstruct items ought to be easier than learning how to directly access them, so that the mean of the D_1/D_2 pair ought to be smaller than the mean of the R_1/R_2 pair.
Repetition dissociation	Because repetition simultaneously increases output interference and provides additional opportunities for meaning comprehension, it should doubly dissociate matched pairs of direct access parameters from matched pairs of reconstruction parameters ($D_1 > D_2$ but $R_1 < R_2$).
D variability	Because reconstructive retrieval generates covert presentations of targets, it should be easier to learn to directly access targets when those targets are reconstructable (as measured by D_{3C} and D_{3E}) than when they are not (as measured by D_1 and D_2).
J variability	Owing to item selection, items that are neither directly accessible nor reconstructable (waiting in state *U*) should be more familiar on average than items that are reconstructable but not direct accessible (waiting in state *P*). Hence, the average value of *J* should be greater when items first become reconstructable (as measured by J_1 and J_2) than for the subset of those items that are still waiting to become directly accessible (as measured by J_{3C} and J_{3E}).

Table 8—*continued*

For items that are waiting in state P, those that subjects recall on earlier trials in that state should be more familiar than those that they do not recall on earlier trials in that state ($J_{3C} > J_{3E}$).

Figure Captions

Figure 1. Predicted-observed comparisons of recall statistics for the conditions of an adult associative recall experiment reported by Brainerd, Desrochers, and Howe (1981). The fitted statistics are the probability of error runs of different lengths for a target after its first successful recall (Panel A), the probability of an error for a target on each trial of the experiment (Panel B), and the probability of error runs of different lengths for a target before its first successful recall (Panel C).

Figure 2. Predicted-observed comparisons of recall statistics for a free recall experiment reported by Brainerd, Howe, and Kingma (1982), in which the subjects were elementary school children. The fitted statistics are the probability of error runs of different lengths for a target after its first successful recall (Panel A), the probability of an error for a target on each trial of the experiment (Panel B), and the probability of error runs of different lengths for a target before its first successful recall (Panel C).

Figure 3. Global developmental trends in direct access, reconstruction, and familiarity judgment. Panel A = developmental trends for children versus adolescents. Panel B = developmental trends for adolescents versus young adults.

Figure 4. A signal detection model of familiarity and decision criteria for the judgment parameters of the recall model. For reconstructed items, there is a distribution of familiarity values for presented items and a distribution of familiarity values for unpresented items. The parameters *C* and *d'* have the usual interpretations: *C* is the decision criterion, which determines how high a reconstructed item's familiarity value must be before subjects are willing to recall it, and *d'* is the difference between the mean familiarity values of presented and unpresented reconstructions.

Figure 5. Opposite effects of study-test trial repetition on direct access and reconstruction. Panel A = children versus adolescents. Panel B = adolescents versus young adults.

Figure 6. Positive association of direct access and reconstruction by age. Panel A = children versus adolescents. Panel B = adolescents versus young adults.

Figure 7. Trends in direct access over trials of recall experiments. Panel A = children versus adolescents. Panel B = adolescents versus young adults.

Figure 8. Trends in familiarity judgment over trials of recall experiments. Panel A = children versus adolescents. Panel B = adolescents versus young adults.

Figure 9. Opposite effects of category cuing on direct access and reconstruction in children and adolescents. Panel A = effects of category cuing on the four direct access parameters. Panel B = effects of category cuing on the two reconstruction parameters.

Figure 10. Dissociative effects of list length on direct access and reconstruction. Panel A = effects of list length on the four direct access parameters. Panel B = effects of list length on the two reconstruction parameters.

Figure 11. Parametric results for the recall performance of younger adults and older adults. Panel A = global developmental trends in direct access, reconstruction, and judgment. Panel B = opposite effects of study-test trial repetition on direct access and reconstruction. Panel C = trends in direct access over trials of recall experiments. Panel D = trends in familiarity judgment over trials of recall experiments.

Figure 12. Reversed association between the mean of the two reconstruction parameters (R_1 and R_2) and the mean of the corresponding direct access parameters (D_1 and D_2), over the four age levels.