Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/13/$12.00 DOI: 10.1037/a0034057

Dual-Retrieval Models and Neurocognitive Impairment

C. J. Brainerd, V. F. Reyna, C. F. A. Gomes, A. E. Kenney, C. J. Gross, E. S. Taub, and R. N. Spreng
Cornell University

Alzheimer's Disease Neuroimaging Initiative

Advances in dual-retrieval models of recall make it possible to use clinical data to test theoretical hypotheses about mild cognitive impairment (MCI) and Alzheimer's dementia (AD), the most common forms of neurocognitive impairment. Hypotheses about the nature of the episodic memory declines in these diseases, about decline versus sparing of specific processes, and about which individuals will become impaired over time can all be rigorously tested. Basic theoretical principles, such as whether recollection and reconstruction are distinct retrieval processes, can also be evaluated. In 3 studies, measurements of recollective retrieval, reconstructive retrieval, and familiarity judgment were extracted from standard clinical instruments, for healthy subjects and for subjects with MCI and AD diagnoses. Differences in reconstructive retrieval consistently distinguished MCI and AD, in nationally representative subject samples as well as in highly educated samples, and recollective retrieval also distinguished them in highly educated samples. Dual-retrieval processes were accurate predictors of future conversion to MCI and AD over periods of 1.5–6 years and were better predictors than the best genetic marker of these conditions (the ε4 allele of the *APOE* genotype). The standard recollection-deficit account of memory declines in MCI and AD was not supported, but the data were consistent with an alternative account that stresses the increasing importance of reconstruction deficits as older adults convert to these diseases.

Keywords: dual-retrieval processes, memory deficits, Alzheimer's dementia, mild cognitive impairment, *APOE* genotype

Supplemental materials: http://dx.doi.org/10.1037/a0034057.supp

In this article, we illustrate how research on dual-process models of retrieval can be advanced by extending those models to the most common forms of neurocognitive impairment in older adults and to the types of clinical data that are used to diagnose these diseases. The work that we report had two aims, one theoretical and the other empirical. The theoretical objective was to connect dual-process models to neurocognitive impairment by using them to pinpoint the nature of memory declines in mild cognitive impair-

C. J. Brainerd, V. F. Reyna, C. F. A. Gomes, A. E. Kenney, C. J. Gross, E. S. Taub, and R. N. Spreng, Department of Human Development, Cornell University; Alzheimer's Disease Neuroimaging Initiative.

Preparation of this article was supported by National Institutes of Health (NIH) Grant 1RC1AG036915 to C. J. Brainerd and V. F. Reyna and by the CAPES Foundation (BEX 0328/12-0) to C. F. A. Gomes. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (Grant NIA U01AG009740) and is conducted by the University of Michigan. Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, by the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann–La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private-sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH Grants P30 AG010129 and K01 AG030514.

Some of the data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. Individual members of the Alzheimer's Disease Neuroimaging Initiative are listed in the supplemental materials.

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

ment (MCI) and Alzheimer's dementia (AD). Although there has been prior work on dual-process conceptions of such declines, the present research (a) implemented recall-based methods of measuring dual processes that avoid limitations of older recognition-based methods and (b) measured those processes with data from low-burden clinical instruments that are part of diagnostic batteries. The empirical objective was to determine whether measurements of dual-retrieval processes in well-characterized samples of healthy control (HC) individuals, individuals with MCI, and individuals with AD have predictive power in identifying older adults who are at risk of future HC → MCI and MCI → AD transitions. Here, we investigated how well those theory-driven measurements fare, relative to the best genetic predictor of such transitions, the ε4 allele of the apolipoprotein E (*APOE*) genotype.

In the first section below, we sketch recently developed techniques for measuring dual-retrieval processes with recall tasks. That section begins with a brief reprise of well-known criticisms of recognition techniques, continues with some problems that are posed when those techniques are used with impaired populations, and ends with a discussion of how recall-based measurement can be applied to the clinical memory instruments that figure in neurocognitive batteries. In the second section, we summarize current theoretical ideas about which processes are responsible for memory declines as individuals convert to MCI or AD. In the remaining sections, we report three studies in which recall-based measurement was used with samples of HC individuals and individuals with MCI or AD in order to tie particular retrieval processes to memory declines in particular conditions (Studies 1 and 3) and then to predict future conversion to MCI and AD over periods of 1.5 to 6 years (Studies 2 and 3).

Dual-Retrieval Processes in Recall

A key feature of the present research is that we used a recall-based technique for measuring dual-retrieval processes, which allows those processes to be measured with the clinical memory tests that are included in neuropsychological batteries. The most widely used of those instruments, such as the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; Morris et al., 1989) recall test and the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1941), involve a small number of study–test trials (e.g., 3–5) on supraspan lists (e.g., 10–15 words). Although recognition tasks have traditionally been used to measure dual-memory processes, alternative procedures that rely on recall (associative, free, serial) have recently been developed (Brainerd, Aydin, & Reyna, 2012; Brainerd & Reyna, 2010; Brainerd, Reyna, & Howe, 2009; Gomes, Brainerd, & Stein, 2013), for two reasons. The first was a series of critiques of conventional recognition methodologies (for a review, see Malmberg, 2008), and the second was that those methodologies are beyond the capabilities of subject groups with substantial cognitive impairment or dementia. We comment briefly on these two matters before sketching how dual processes are measured with recall tasks and discussing similarities and differences between the recall and recognition conceptions of dual processes.

Limitations of Recognition Techniques

The distinction between recollective and nonrecollective (usually called familiarity) retrieval has been studied almost exclusively with old/new item recognition (for a review, see Yonelinas, 2002). The dominant methodology, first used by Strong (1913), has been to enrich recognition with metacognitive judgments that supply information about whether the basis for old/new decisions is recollective or nonrecollective. Three procedures account for the bulk of the literature: (a) remember/know judgments (Tulving, 1985), which is by far the most extensively used method; (b) inclusion versus exclusion judgments, which allow the recollection and familiarity parameters of Jacoby's (1991) process-dissociation model to be estimated; and (c) confidence judgments, from which estimates of recollection and familiarity components of the receiver-operating characteristic (ROC) are extracted (Yonelinas, 1994).

Although these procedures have generated a vast literature, they have stimulated a series of validity challenges (e.g., Heathcote, Raymond, & Dunn, 2006). Such challenges, along with other sources of evidence (e.g., Ratcliff, Van Zandt, & McKoon, 1995), were examined in a seminal review by Malmberg (2008), who concluded that available evidence converges on the view that a single familiarity process is adequate to handle item recognition. Because remember/know, inclusion/exclusion, and confidence are defined over recognition, they become problematic as dual-process methodologies. In order to make progress on evaluating dual-process conceptions of memory declines in AD and MCI, then, alternative measurement procedures are desirable. However, they must be within the capabilities of subjects who have neurocognitive limitations. We discuss the second problem before summarizing the alternative procedure that we used.

Measuring Dual-Memory Processes in Subjects With Cognitive Limitations

Validity challenges aside, customary methods of separating recollective from nonrecollective retrieval share the practical limitation that they are high-burden methodologies. They require subjects to comprehend and remember instructions as to how to introspect on the phenomenological qualities of remembering and to perform those introspections reliably, which exceeds the capabilities of some populations—with young children (e.g., Ghetti & Angelini, 2008) and older adults with neurocognitive impairments (e.g., Brainerd et al., 2009) being prime examples. Nevertheless, those populations are important targets of dual-process research. For instance, dual-process hypotheses have been formulated for MCI and AD (Bugaiska, Morson, Moulin, & Souchay, 2011), with the standard proposal being that conversion to each condition is characterized by reductions in recollective retrieval coupled with sparing of familiarity (see below).

The high-burden problem can be illustrated by three features of remember/know, the dominant methodology. First, remember/know instructions are long and complex (see Rajaram, 1996), requiring high school levels of reading comprehension. Second, although instructions can be read aloud (e.g., Billingsley, Smith, & McAndrews, 2002), doing so ensures neither comprehension nor correct implementation when subjects have cognitive limitations. It also creates a new obstacle, relative to written instructions that can be consulted throughout a recognition test: An additional memory load is imposed, which is a key consideration with cognitively limited subjects. It might seem that both obstacles could be overcome by simplifying and compressing the instructions. How-

ever, the simplifications that are necessary to ensure comprehension by subjects with dementia are so drastic that it would be dubious to assume that the same processes are still being measured, and when such simplifications are effected, subjects may still fail to comprehend the instructions. The third obstacle is that remember/know judgments demand that subjects introspect on the contents of their mental states. Obviously, it is questionable to assume that subjects with the forms of brain atrophy that are associated with AD and MCI can introspect reliably.

It might also seem that these obstacles could be overcome by resorting to recognition tasks other than the conventional ones; that is, tasks that do not demand comprehension of complex instructions about how to introspect on the phenomenological qualities of remembering. Limited work of that sort has been reported in older adults with neurocognitive impairments (e.g., Gallo, Cramer, Wong, & Bennett, 2012), but such research raises the questions about process comparability that were just mentioned. The mainstream literature on dual processes in recognition is based squarely on remember/know, process dissociation, and dual-process ROC data, which means that these procedures are, in effect, the operational definitions of recollection and familiarity. Without validity studies linking these traditional metacognitive measures to simplified alternatives, there is a significant risk that very different processes are being measured (e.g., Ghetti & Angelini, 2008).

Recollective and Nonrecollective Processes in Recall

In contrast to metacognitive judgments, simple recall of lists is within the capabilities of impaired subjects. There is, of course, a large literature on associative, free, and cued recall in impaired as well as healthy older adults, and such tests are part of clinical neuropsychological batteries that are used to diagnose dementia (e.g., Langa et al., 2005). In the research that we report, a model that separates recollective and nonrecollective components of recall was fit to data from two of the most widely implemented clinical recall tests, the CERAD and the RAVLT, using large-scale studies of late-life impairment in which those tests were administered to well-characterized samples of HC individuals and individuals with MCI or AD. In addition to being low burden, the dual-retrieval model has the advantage that nothing is added to existing neuropsychological instruments. The method of data analysis is the only thing that changes.

The dual-retrieval model (Brainerd et al., 2009) posits that items are recalled via a recollective operation, called direct access, and a nonrecollective one, called reconstruction. Reconstruction is accompanied by a slave judgment operation that evaluates the familiarity of reconstructed items before outputting them. (It is a "slave" operation in the sense that it is not activated unless reconstruction is successful.) The recollective operation accesses verbatim traces of list items' prior presentations directly, without searching through traces of other items, and is therefore the faster of the two forms of retrieval. Direct access is also more accurate than reconstruction because it produces errorless recall: When an item's verbatim trace is directly accessed, its surface form is symbolically reinstated, so that the item can be recalled by simply reading it out of consciousness. Direct access is a recollective operation because it reinstates vivid, realistic details of prior presentations.

The nonrecollective operation, reconstruction, regenerates items from stable episodic traces of partial-identifying information, especially semantic information (e.g., "animal" and "farm" for *horse*). That subjects are able to reconstruct items in this manner is well documented in research on tip-of-the-tongue and feeling-of-knowing phenomena. In both cases, subjects have been found to access a range of partial-identifying information about list items before they can be recalled (e.g., Brown & McNeill, 1966; Hicks & Marsh, 2002; Koriat, 1993, 1995; Kurilla & Westerman, 2010; Schacter & Worling, 1985). For example, Koriat, Levy-Sadot, Edry, and de Marcas (2003) reported that the semantic features of Osgood's (1952) model of meaning can be accessed before items are recalled. Reconstruction searches for items that match partial-identifying features and generates sets of candidate items (e.g., *horse*, *cow, goat, sheep*) that are small enough to be processed within the time constraints of a recall test. As the features that generate such sets do not uniquely identify studied targets, the sets normally include nontargets (*cow, goat, sheep*). To avoid high intrusion rates, a judgment operation performs familiarity checks on reconstructed items before outputting them, which is how the dual-retrieval model implements the familiarity notion of dual-process conceptions of recognition. On analogy to signal detection models, familiarity signals from reconstructed items are processed by setting a decision criterion and outputting items whose familiarity exceeds that criterion.[1]

Recollective and nonrecollective retrieval are quantified by fitting two-stage Markov chains to the data of standard recall paradigms, including clinical instruments such as the CERAD and the RAVLT. Such models deliver tolerable fits to a wide range of recall data throughout the life span (for a review, see Brainerd et al., 2009). With respect to clinical instruments, Brainerd et al. (2012) and Gomes et al. (2013) showed that tasks in which subjects participate in only three study–test trials per list are adequate to conduct model fits and obtain identifiable estimates of its parameters. As the statistical machinery for fitting various permutations of the dual-retrieval model to data, estimating parameters, and conducting parameter significance tests has been presented in prior articles, it is not reprised in this paper. Here, the focus is on the retrieval processes that are measured by the model's parameters, which are defined in Table 1. It can be seen that there are separate parameters that measure recollective retrieval (D), reconstructive retrieval (R), and familiarity judgment (J). These parameters do not map with simple, observable features of recall performance; that is, there are no observable aspects of performance that one can point to and say that they are uniquely due to one of the parameters. Instead, parameters must be estimated with

[1] Reconstructive retrieval should not be confused with the notion of "partial recollection," which figures in recent dual-process recognition models (e.g., Rotello, Macmillan, & Reeder, 2004; Wixted & Mickes, 2010). As mentioned, reconstructive retrieval involves regenerating the studied item itself (e.g., *horse*), using stable stored information about it (e.g., some of its semantic features, such as "farm" and "animal"). Partial recollection shares neither property. The item itself is not regenerated because it is presented to subjects as a recognition probe. The stored information that subjects retrieve about that item is not stable stored information but, rather, consists of arbitrary contextual cues that were associated with its presentation on the study list. Thus, if *horse* were presented in blue script font inside a hexagonal box on the left side of the computer screen, retrieving some but not all of these contextual details (e.g., "blue" and "script" but not "hexagonal" or "left") is an example of partial recollection.

Table 1
Retrieval Processes That Are Measured With Repeated Recall Data and Delayed Recall Data

Process/parameter	Definition
	Learning to recall
Direct access (recollection):	
D	The probability that a verbatim trace of an item's presentation on a study list can be accessed on a recall test that follows a study cycle
Reconstruction:	
R	For any item whose verbatim trace cannot be accessed on a recall test following a study cycle, the probability that it can be reconstructed on that recall test
Familiarity judgment:	
J	For any item that is reconstructed on a recall test that follows a study cycle, the probability that the reconstruction is judged to be familiar enough to output
	Forgetting
Direct access (recollection):	
F_D	On a forgetting test, the probability that the direct access operation fails for items that could be directly accessed following the last study cycle
Reconstruction:	
F_R	On a forgetting test, the probability that the reconstruction operation fails for items that could be reconstructed following the last study cycle

a model that takes into account the fact that observed performance is not process pure.

Dual-Retrieval Processes in Recognition Versus Recall

Dual-process conceptions of recognition versus recall are similar in some respects and different in others. The major difference lies in nonrecollective retrieval, where the recall process is obviously more complex than the corresponding recognition process. In recognition, subjects need not recover target items because they are presented as test probes. When probes fail to provoke recollection, familiarity checks are executed to determine whether global memory strength is sufficient to warrant accepting them as old. In recall, targets are not presented as test probes, and, consequently, targets that cannot be recollected must be reconstructed from partial identifying information if familiarity checks are to occur. Thus, nonrecollective retrieval simply consists of familiarity checks in recognition, whereas it consists of item reconstruction plus familiarity checks in recall.

On the other hand, dual-process conceptions of recollective retrieval are similar in recognition and recall. In both instances, recollection involves becoming consciously aware, during the test phase, of what happened when particular target items were presented during the study phase. If recollection is indeed similar in recognition and recall, the straightforward prediction is that it should react similarly in the two domains to selected manipulations: Manipulations that increase conventional indexes of recollection in recognition ought to affect the *D* parameters of the recall model. We have reported several experiments that are consistent with this prediction. For example, consider six manipulations that have been consistently found to increase recollection in recognition (e.g., Yonelinas, 2002): (a) studying shorter versus longer lists; (b) studying lists of word pairs versus singletons; (c) studying lists of low- versus high-frequency words; (d) studying lists of emotionally valenced versus neutral words; (e) administering recognition tests to younger adults versus healthy older adults; and (f) administering immediate versus delayed recognition tests. In recall experiments with the dual-retrieval model, all of these manipulations were found to increase the *D* parameters (Brainerd et al., 2009, 2012; Brainerd & Reyna, 2010; Gomes et al., 2013).

In the same vein, some instructive findings on the relation between recognition and recall indexes of recollection were reported by Gomes (2013), who combined recall- and recognition-based measures in a single experiment and correlated them. His subjects learned to recall word lists, using a procedure that allowed the model in Table 1 to be fit to the data and its parameters to be estimated. After the last recall test, items that subjects had recalled were represented as test probes, and subjects made the traditional types of metacognitive judgments that separate recollection from familiarity in recognition. Some subjects made remember/know judgments, and others made source judgments (targets had been presented in distinctive contexts). For the first group, model parameters were used to compute the proportion of items that were recalled recollectively, and remember/know scores were used to estimate the proportion of items that were recognized recollectively. The recall and recognition indexes were strongly correlated. For the second group, model parameters were used to compute the proportion of items that were recalled recollectively, and accurate/inaccurate source judgments were used to estimate the proportion of items that were recognized recollectively. Again, the recall and recognition indexes were strongly correlated. Beyond the theoretical similarity between the recall and recognition conceptions of recollection, then, data on the effects of recollection-oriented manipulations and on correlations between recall and recognition indexes make a presumptive case for similarity.

Dual-Process Accounts of Memory Impairment in AD and MCI

AD is the most common variety of dementia, accounting for more than two thirds of dementia diagnoses after age 70 (e.g., Plassman et al., 2007) and affecting roughly 20% of individuals

after age 75 (Hebert, Scherr, Bienias, Bennett, & Evans, 2003). It is often preceded by MCI, a form of cognitive impairment no-dementia (CIND) that is more prevalent than AD. Both are quintessentially diseases of episodic memory, inasmuch as memory impairment is their diagnostic hallmark. With respect to AD, the diagnostic criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) require clinically significant impairment in episodic memory coupled with clinically significant impairment in at least one of four other domains (executive function, language, motor function, or object identification). In other words, individuals with AD must show memory impairment, which is why performance on clinical recall instruments such as the CERAD and the RAVLT is the best neuropsychological marker of this diagnosis (e.g., de Jager, Hogervorst, Combrinck, & Budge, 2003). Observed impairments must represent declines from earlier levels of functioning in order to exclude individuals who simply perform poorly on memory tests. Also, impairments cannot be due to other diseases that cause memory declines (e.g., alcoholism, cardiovascular disease, diabetes, stroke), but individuals may be assigned possible rather than probable AD diagnoses when such diseases are present but judged not to be causing impairments (e.g., McKhann et al., 2011).

Diagnostic criteria for MCI are weaker versions of AD criteria (see Petersen, 2004). Of those just mentioned, the last two (declines from earlier levels of functioning and exclusion of other diseases) are the same, but the first is different. For MCI, the first criterion requires impairment in episodic memory *or* one of the other domains, with the former designated as amnestic (a-MCI) and other forms as nonamnestic (n-MCI).[2] Thus, memory impairment is not necessary for MCI. Empirically, however, it accounts for roughly two thirds of such diagnoses (Petersen et al., 2010), which is why performance on clinical recall instruments is also the best neuropsychological marker of MCI diagnoses (e.g., Harel et al., 2011). Further, a-MCI but not n-MCI is prodromal to AD; there is a developmental progression from HC to a-MCI to AD (see Brainerd et al., 2013).

The memory declines in AD and MCI diagnostic criteria are purely descriptive, consisting simply of scores on memory instruments and self-reports of memory complaints. The obvious theoretical questions are: Which underlying processes are responsible for those declines? Are these processes different for MCI than for AD? Is there some sequence of process deterioration during the progression from HC to MCI to AD? Such questions are central to any theoretical characterization of these diseases. However, they are also of great clinical interest because theoretical characterization illuminates disease mechanisms, and disease mechanisms must be understood to produce successful treatments (Brainerd et al., 2009). To make progress on such questions, researchers have used dual-process conceptions to characterize memory impairments. That is a logical approach, because the brain regions that have been foci of dual-process research with healthy subjects (the perirhinal, parahippocampal, and entorhinal cortices and the hippocampus; e.g., Diana, Yonelinas, & Ranganath, 2007; Ranganath, 2010) are regions that exhibit pathology in postmortem studies of individuals with MCI or AD (e.g., Braak & Braak, 1995; Nelson, Braak, & Markesbery, 2009).

As discussed in a prior review (Brainerd et al., 2009), the predominant hypothesis about AD is a recollection-deficit notion: that conversion to AD is marked by declines in recollective retrieval, coupled with sparing of familiarity (Dalla Barba, 1997; Tse, Balota, Moynan, Duchek, & Jacoby, 2010; Westerberg et al., 2006; for a review, see Bugaiska et al., 2011). Recollective decline has also been used to explain collateral deficits that individuals with AD display in executive function, language, and object identification—the idea being that tests of these abilities require subjects to recollect specific details in order to perform the focal tasks (see Baudic et al., 2006; Creamer & Schmitter-Edgecombe, 2010). However, the dominant hypothesis about executive function is that there are declines in that ability that go beyond memory-induced deficits (e.g., Storandt, 2008). It has recently been proposed that conversion to MCI is also characterized by declines in recollection but not familiarity, with recollective deterioration simply being less severe than in AD (Anderson et al., 2008; Serra et al., 2010; Westerberg et al., 2006). There is, of course, a large literature, using traditional recognition methods of separating recollection and familiarity, that favors a recollective locus for age declines in episodic memory among healthy individuals (for reviews, see Light, Prull, La Voie, & Healy, 2000; Reyna & Mills, 2007; Yonelinas, 2002). Thus, this hypothesis about MCI and AD posits continuity in the recollection deficits that underlie episodic memory declines in these diseases and in healthy aging.

An alternative scenario that we have proposed (Brainerd et al., 2009; Reyna & Brainerd, 2011) involves qualitative shifts in the processes that are responsible for declines during healthy aging versus MCI and AD. That hypothesis is predicated on the following considerations. Before healthy adults reach age 70 and the MCI conversion rate accelerates, cumulative declines in traditional measures of recollection are already quite large. Further declines that would be commensurate with the levels of impairment in a-MCI (the diagnostic criterion is at least a 1.5 *SD* difference between individuals with a-MCI and HC individuals on episodic memory tests; Petersen, 2004) could drive recollection to near-floor levels. That might rule it out as a source of further memory decline in transitions to AD—leaving nonrecollective processes, such as familiarity or reconstruction, as remaining possibilities.

Comparisons of the two scenarios require measurement of recollective and nonrecollective remembering in samples of well-characterized HC individuals, individuals with MCI, and individuals with AD. Such samples are essential because diagnostic error is a persistent source of discrepant findings about these conditions (Brainerd, Reyna, Petersen, Smith, & Taub, 2011). Because MCI is a recent diagnostic category, there has been only limited time for data sets that meet that criterion to accumulate. There is a recent data set, however, that supplies large samples of well-characterized HC individuals, individuals with MCI, and individuals with AD: the Aging, Demographics, and Memory Study (ADAMS) of the National Institute on Aging's Health and Retirement Study (HRS; Health and Retirement Study, 2011). Another major advantage of the ADAMS is that it is the only nationally representative sample of HC individuals, individuals with MCI, and individuals with AD extant (Langa et al., 2005), so that what

[2] Clinically, the n-MCI diagnosis is further subdivided into individuals who exhibit impairment in a single nonmemory domain (sn-MCI) versus multiple nonmemory domains (md-MCI; see Panza et al., 2005). This other, less common, variety of MCI was not a focus of the present research, as it is not prodromal to AD.

is true for ADAMS diagnostic groups should be true in general for older adults in the United States. We used that data set in Studies 1 and 2. Another data set whose subject sample was not representative was used in Study 3 to examine some important follow-up questions that could not be studied with the ADAMS.

Overview of the Research

In the first two studies, the dual-retrieval model was fit to ADAMS subjects' recall on a standard clinical recall instrument (CERAD) in order to measure levels of recollection, reconstruction, and familiarity judgment in nationally representative HC, MCI, and AD groups. The aim in Study 1 was to pit different theoretical accounts against each other by pinpointing the process differences among the diagnostic groups during immediate and delayed recall. That was done, first, by showing that the dual-retrieval model provides acceptable fits for all three groups and, second, by comparing recollection, reconstruction, and familiarity judgment parameters (a) between the HC and MCI groups and (b) between the MCI and AD groups.

The next study focused on the problem of forecasting future disease. That problem, whose scope can be illustrated by studies of longitudinal progression to AD and CIND (e.g., Tabert et al., 2006), is that the neuropsychological tests that are used to diagnose MCI and AD do not perform especially well at predicting the future emergence of these diseases. At a minimum, neuropsychological batteries contain six types of tests (see, e.g., Langa et al. 2005): general cognitive ability tests, such as the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) and the Shipley Vocabulary Test (SVT; Zachary, 1986), and tests of the five specific abilities that are stipulated in the *DSM–IV* dementia criteria (episodic memory, executive function, language, motor function, and object identification). In a review of 73 studies of neuropsychological predictors of future AD, Twamley, Ropacki, and Bondi (2006) found that (a) predictive relations were weak, (b) only five of the six types of tests displayed significant predictive relations in some studies (general cognitive ability, executive function, language, memory, and motor function), and (c) none of the six types of tests displayed consistent predictive relations across all or even most studies. Concerning (c), the percentages of studies producing significant predictive relations were 38% for general cognitive ability, 50% for verbal memory, 28% for visual memory, 44% for executive function, 33% for language, and 17% for motor functioning. Given that predictive relations were weak and were only reliable about one third of the time, the need for reliable predictors, especially ones that are theoretically motivated, is clear.

That was our aim in Study 2, in which we focused on dual-process prediction of future AD in individuals with MCI diagnoses and future MCI in HC individuals. With regard to AD, we determined whether the measurements of recollection, reconstruction, and familiarity judgment in individuals with MCI that had been made in Study 1 would predict who would transition to AD after 16–18 months and after 4.5–6 years. Concerning MCI, we determined whether the measurements of recollection, reconstruction, and familiarity judgment in HC individuals that had been made in Study 1 would predict who would transition to MCI after 4.5–6 years. As a baseline, the ability of these processes to forecast future AD and MCI was compared to the best genetic predictor of these conditions, the ε4 allele of the *APOE* genotype.

Because Studies 1 and 2 revealed that dual-retrieval processes differentiated the ADAMS diagnostic groups and predicted future disease, it was important to investigate whether those processes could be successfully measured at the level of individual subjects. That was the aim in Study 3, which used another large-scale study of late-life impairment (Alzheimer's Disease Neuroimaging Initiative; ADNI) that provides recall data that are more extensive than those of the ADAMS. With these richer data, the model could be fit to the recall of individual HC, MCI, and AD subjects, and the questions that were examined in the first two studies could be reexamined with individualized measurements of dual-retrieval processes.

Study 1

The HRS involves a sample of over 30,000 subjects from over 70,000 households that encompass all geographical regions, racial groups, and ethnic groups in the United States. Subjects were sampled with a multistage clustered area probability model that ensured that the sample would accurately represent older adults from all regions, racial groups, and ethnic groups. ADAMS subjects are a nationally representative subsample of 856 individuals, age 70 and older, from the larger HRS pool (for details, see Plassman et al., 2007, 2008). HRS subjects participate in biennial interviews that gather information on a variety of demographic, employment, wealth, caregiving, family structure, and health measures (Juster & Suzman, 1995). The ADAMS subsample was constructed to include five levels of cognitive ability, ranging from low-functioning to high-normal functioning, based on subjects' performance during their most recent biennial interview (Langa et al., 2005).

ADAMS subjects received extensive neuropsychological testing and were diagnosed for the presence of neurocognitive impairment. Subjects participated in up to four separate waves (A, B, C, and D) of testing and diagnosis over 6 years. During Wave A, which is the focus of Study 1, 856 subjects (mean age = 81.6, range = 70–110) completed a 3- to 4-hr assessment. They (a) received a battery of neuropsychological tests, (b) received medical examinations to identify conditions that must be considered in diagnosing neurocognitive impairment (e.g., cardiovascular disease, diabetes), and (c) provided buccal tissue samples for genotyping. Concerning (a) and (b), these data were reviewed by a diagnostic team, who classified subjects according to three levels of functioning (HC, CIND, and demented [D]). Concerning (c), tissue samples were analyzed for the *APOE* genotype. The genetic data of 14 subjects were not usable, leaving a final sample of 842 subjects. The chronology of the ADAMS testing waves and the diagnostic composition of the subject samples of Study 1 and Study 2 are summarized in Table 2.

In Study 1, we used immediate and delayed recall data of HC subjects, CIND subjects with MCI diagnoses, and D subjects with AD diagnoses. Those data were generated by one of the clinical recall instruments mentioned earlier, the CERAD. To pinpoint process-level differences among the three groups, we fit the dual-retrieval model to the data, estimated its parameters (see Table 1) on an age-adjusted basis, and compared parameter values among

Table 2
Diagnostic Composition of the Subject Samples in Study 1 and Study 2 for the Four Waves of ADAMS Testing

	Study and testing wave		
	Study 1	Study 2	
Diagnostic group	A	B (16- to 18-month follow-up)	C and D (4.5- to 6-year follow-up)
HC	304	7	122
MCI	98	49	40
AD1	119	15	20
AD2	105	6	11

Note. These values are a combination of incident and prevalent diagnoses. ADAMS = Aging, Demographics, and Memory Study; HC = healthy control; MCI = mild cognitive impairment; AD1 = probable Alzheimer's dementia; AD2 = possible Alzheimer's dementia.

the groups.[3] As the HC group was large and covered a 24-year age range (70–94), we conducted follow-up analyses of how dual-retrieval processes varied throughout this age range by estimating the parameters separately for different age groups. Here, the objective was to determine whether, late in life, there are continuing age declines in recollection, reconstruction, or familiarity judgment when there is no evidence of disease.

Method

Subjects. The Wave A ADAMS sample with genetic data consists of 304 HC subjects (mean age = 78), 237 CIND subjects (mean age = 84.5), and 301 D subjects (mean age = 84.3). Within the latter two classifications, the diagnostic team also assigned subdiagnoses to CIND subjects and D subjects. There were 10 sub-CIND diagnoses, most of which contained very few subjects. There was a total of 17 sub-D diagnoses, most of which contained very few or no subjects.

In research with the ADAMS genetic data, Brainerd et al. (2011) found that the CIND group contained 98 MCI subjects (mean age = 83.2) and that the D group contained 224 AD subjects (mean age = 84.5). They also found that although the ADAMS MCI group is a mixture of a-MCI and n-MCI, the group's average performance on episodic memory tests meets the clinical criterion for a-MCI ($\geq$ 1.5 *SD*s below the performance of healthy agemates). Last, Brainerd et al. reported that the AD subjects could be further subdivided into 119 subjects with probable AD (AD1) diagnoses (mean age = 86.1) and 105 subjects with possible AD (AD2) diagnoses (mean age = 86.1).

Procedure. Summary data for the Wave A neuropsychological tests, genetic results, and psychiatric classifications are available from HRS. Our focus was on the immediate and detailed data of the clinical recall instrument, the CERAD. The immediate CERAD test consists of three trials of study plus free recall on a 10-word list, and the delayed CERAD test consists of one additional free recall test, which is administered after a 5-min filled retention interval.

The dual-retrieval model analyzes complete error–success sequences for individual list items (see Brainerd et al., 2009, Equation A1), and for the CERAD items, such sequences consist of three immediate tests plus one delayed test. The CERAD is a simple free recall task, in which subjects study a list of 10 familiar concrete nouns. (One list is *arm, butter, cabin, engine, grass, letter, pole, queen, shore, ticket.*) The list is studied three times, and each study cycle is followed by an oral free recall test. Approximately 5 minutes after the third test, a delayed recall test is administered, without further opportunities to study the list, as a forgetting measure. Other neuropsychological tests are interpolated during the delay. ADAMS subjects responded to constructional praxis tests, which measure the ability to draw two- and three-dimensional figures, during this interval. Responses sequences for recall Tests 1–3 were input to the dual-retrieval learning model in Brainerd et al. (2012; Equations A1–A9) and Gomes et al. (2013; Equations A1–A9), in order to fit it to the performance of the HC, MCI, AD1, and AD2 groups. As the fits were acceptable, the recollection, reconstruction, and familiarity judgment parameters were estimated for each group on an age-adjusted basis, and parameter significant tests were computed to determine which ones differentiated the groups. Next, the data sequences for recall Tests 2, 3, and 4 (delay) were input to the forgetting model in the Appendix, in order to estimate forgetting of recollection and reconstruction for each of the four groups on an age-adjusted basis. Parameter significance tests were computed to determine which ones differentiated the groups.

Finally, the 304 HC subjects were split into six chronological age groups, in order to identify any process-level declines that occur very late in life when there is no evidence of disease. Each group's Trial 1–3 data were input to the learning model to measure age variability in recollection, reconstruction, and familiarity judgment. Each group's Trial 2–4 data were input to the forgetting model to measure age declines in forgetting of recollection and reconstruction.

Results and Discussion

Before detailed results are presented, the qualitative patterns were as follows. At the most general level, the results provided no support for the continuity hypothesis that memory declines over HC $\rightarrow$ MCI $\rightarrow$ AD transitions are entirely recollection driven. In line with the hypothesis that loss of recollective ability among healthy older adults is too extensive for it to be the sole source of memory declines in neurocognitive impairment, absolute levels of recollective retrieval were low in this nationally representative sample of HC subjects. However, model fits showed that recollective as well as nonrecollective retrieval were needed to account for the data of all subject groups. During initial learning, there was only a slight decline in recollective retrieval for HC $\rightarrow$ MCI transitions. Instead, memory declines were dominated by deterioration in nonrecollective retrieval, especially the reconstruction component. Both reconstruction and familiarity components declined in HC $\rightarrow$ MCI transitions, and there were further and much larger declines in the reconstruction component in MCI $\rightarrow$ AD transitions. With the retention data, we found that reconstruction

[3] Individuals with AD are usually older than individuals with MCI or HC individuals, and individuals with MCI are usual older than HC individuals; hence, it is standard practice in neuropsychological research to use age-adjusted data in comparing these groups . Failure to do so has been a source of both false positive and false negative results (e.g., Brainerd et al., 2011). Although we report results for age-adjusted data, we also analyzed the data without age adjustment and found that the qualitative patterns were the same.

was again the key process that declined over diagnostic transitions. There was no reconstructive forgetting in HC subjects; preservation of that form of retrieval was perfect. However, such forgetting increased to nearly 50% in MCI subjects and to above 80% in AD subjects. There were also declines in recollective forgetting associated with AD but not MCI. Thus, in contrast to the continuity hypothesis, MCI and AD diagnoses were dominated by deterioration in nonrecollective retrieval. Further, these changes were unique to neurocognitive impairment: When HC subjects were stratified by age, levels of recollection, reconstruction, and familiarity remained constant from age 70 onward.

Differences in dual-retrieval processes: Learning. Mean probabilities of correct recall for the four recall tests appear by diagnostic group in Table 3. The percentage of total correct recall declined reliably from HC (60%) to MCI (38.5%) to AD (AD1 = 16% and AD2 = 20.2%), $F(3, 622) = 414.66$, $p < .0001$, partial $\eta^2 = .67$. Post hoc tests (Tukey honestly significant difference [HSD]) showed that MCI performance was worse than HC performance (mean difference = 6.40), that MCI performance was better than either AD1 performance (mean difference = 6.74) or AD2 performance (mean difference = 5.48), and that the two AD groups did not differ reliably (ps $< .0001$).

To measure differences in recollection, reconstruction, and familiarity judgment between diagnostic groups, we fit the dual-retrieval model to the age-adjusted Trial 1–3 CERAD data of each group. There are six slightly different versions of this model (see Brainerd et al., 2012; Gomes et al., 2013), which form a 2 × 3 matrix of submodels that differ along two dimensions. One dimension refers to the set of reconstruction and familiarity judgment parameters: A submodel may have either two of each (R_1, R_2, J_1, J_2), or it may have a single reconstruction parameter and three familiarity parameters (R, J_1, J_2, J_3). The second dimension refers to whether a submodel allows items that are reconstructable but not recollectable to become recollectable following a recall error, following a recall success, or following either. Tutorials and software for all of the modeling analyses that are reported in this article may be found at www.human.cornell.edu/hd/brainerd/research.cfm

Fits to the data of the HC, MCI, AD1, and AD2 groups showed that a single submodel fit the data of all groups. In this submodel, (a) recollective learning occurred following errors, and (b) there was a single reconstruction parameter for all three trials. Model fitting involved two steps, necessity tests followed by sufficiency tests. Necessity tests evaluate whether the dual-retrieval model is not parsimonious because recall data can be fit by a simpler one-process model; specifically, a model with a reconstructive retrieval operation and a familiarity judgment process. The fit test for this one-process model is provided in Brainerd et al. (2012; Equations A12–A15). When it is fit to the data of any group, it is fit with two degrees freedom. As there were four groups (HC, MCI, AD1, and AD2), there were eight degrees of freedom, so that the .05 critical value for the G^2 test of the null hypothesis that the model fits the data was 15.51. The observed value of $G^2(8)$ was 64.52, and, thus, the data were not generated by a single reconstruction process. An important implication is that although estimates of the recollection parameters were low in absolute terms, as expected, they were nevertheless reliable because the data could not be accommodated by reconstructive retrieval alone.

When necessity tests yield poor fits, sufficiency tests ask whether recall data are more complex than the dual-retrieval model posits and, therefore, cannot be fit by a model with two retrieval processes. The appropriate fit test is provided in Brainerd et al. (2012; Equation A11). For the study as a whole, the model was fit with four degrees of freedom (one per subject group), so that the .05 critical value of the G^2 test was 9.48. As the observed value (9.27) did not exceed the critical value, model fit was acceptable.

The retrieval processes that differentiate HC, MCI, and AD subjects can be seen in Table 4, where maximum likelihood estimates of the recollection, reconstruction, and familiarity judgment parameters are reported by diagnostic group. Although separate parameter estimates for the AD1 and AD2 groups are reported, a preliminary analysis revealed that they did not differ reliably between these diagnostic groups. Consequently, another set of estimates is reported for the pooled AD1 and AD2 data, and those values were used to test for differences in retrieval processes between MCI and AD subjects.

The procedure for identifying between-group differences in parameter values involves three steps (test statistics in Brainerd et al., 2012, Equations A11–A11c): a likelihood ratio test of the omnibus null hypothesis that none of the parameters differed reliably among the three groups (HC, MCI, AD), followed by likelihood ratio tests of the null hypothesis that none of the parameters differed reliably between specific pairs of groups (HC vs. MCI and MCI vs. AD), followed by likelihood ratio tests of the null hypothesis that specific parameters (say, R) did not differ reliably between specific pairs of groups (say, HC vs. MCI). The test of the omnibus null hypothesis was a G^2 statistic with 12 degrees of freedom and a .05 critical value of 21.03. This null hypothesis was rejected, $G^2(12) = 3{,}221.66$. The tests of the two-group null hypotheses were G^2 statistics with 6 degrees and a .05 critical value of 12.59. As the observed values for the HC–MCI test (342.35) and the MCI–AD test (953.68) were both more than

Table 3
Mean Correct Recall by Trial in Study 1 and Study 3

	Trial					
Diagnostic group	1	2	3	4	5	6
		Study 1				
HC	.43	.61	.73	.60		
MCI	.25	.42	.48	.25		
AD1	.10	.18	.20	.04		
AD2	.14	.23	.24	.07		
AD pooled	.12	.20	.22	.05		
		Study 3				
HC_{NC}	.35	.52	.62	.70	.74	.33
HC_C	.29	.43	.55	.60	.66	.34
a-MCI_{NC}	.30	.40	.44	.51	.55	.26
a-MCI_C	.25	.34	.39	.42	.44	.22
AD	.23	.29	.33	.33	.35	.19

Note. HC = healthy control; MCI = mild cognitive impairment; AD1 = probable Alzheimer's dementia; AD2 = possible Alzheimer's dementia; HC_{NC} = healthy control not converted to amnestic; HC_C = healthy control converted to amnestic; a-MCI_{NC} = amnestic mild cognitive impairment not converted to Alzheimer's dementia; a-MCI_C = amnestic mild cognitive impairment converted to Alzheimer's dementia; AD = Alzheimer's dementia.

Table 4
Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting for ADAMS Subjects During Wave A

Process/parameter	Diagnostic group				
	HC	MCI	AD1	AD2	AD pooled
Recollection					
D_1	.10	.04	.03	.03	.03
D_2	.04	.00	.02	.05	.03
M_D	.07	.02	.02	.04	.03
Reconstruction					
R	.69	.54	.18	.28	.22
Familiarity judgment					
J_1	.53	.40	.40	.43	.42
J_2	.65	.50	.44	.36	.39
J_3	.70	.51	.34	.24	.28
M_J	.63	.47	.40	.34	.36
Forgetting: Recollection					
F_D	.32	.59	.87	.82	.84
Forgetting: Reconstruction					
F_R	.00	.45	.75	.65	.70

Note. ADAMS = Aging, Demographics, and Memory Study; HC = healthy control; MCI = mild cognitive impairment; AD1 = probable Alzheimer's dementia; AD2 = possible Alzheimer's dementia.

25 times the critical value, these null hypotheses, too, were rejected.

At the level of retrieval processes, the HC–MCI and MCI–AD parameter comparisons identified specific processes that differed between diagnostic groups. The HC–MCI comparisons showed that two processes differed between these groups, reconstruction and familiarity judgment. The *R* parameter was significantly larger in HC subjects, as were all three *J* parameters (see Table 4), so that MCI memory decline was not due to loss of recollective ability. Rather, it was harder for MCI subjects to learn how to reconstruct items, and once they had done so, the reconstructions seemed less familiar to them. Although neither of the recollection parameters (D_1 and D_2) differed reliably for the HC–MCI comparison, follow-up analyses showed that there was a reliable difference in the proportion of items that could be recollected by Trial 3. Using the parameter estimates for these groups, one can calculate the total proportion of items that could be recollected on any learning trial. By Trial 3, those values were .14 for HC and .04 for MCI, which was a reliable difference ($p < .01$ by a proportions test). Thus, although the parameter comparisons failed to confirm that MCI memory decline is due to loss of recollective ability, there was some support for a recollective contribution at the level of cumulative performance.

The MCI–AD parameter comparisons failed to provide any support for the continuity hypothesis: Neither of the recollection parameters differed reliably between the MCI and AD groups, and follow-up analyses showed that the proportion of items that could be recollected by the last learning trial was the same. The process-level difference between these groups was simply that it was far harder for AD subjects to learn how to reconstruct items. The value of *R* in the MCI group was more than twice that in the AD group (.54 vs. .22). Also, reconstructed items were somewhat less familiar to AD subjects than MCI subjects. The mean value of the *J* parameters was smaller among AD subjects than among MCI subjects (.47 vs. .36), though the parameterwise tests showed that only J_3 was reliably smaller among AD subjects.

Differences in dual-retrieval processes: Forgetting. The fact that the CERAD includes a retention test 5 minutes after the third learning trial allows two forgetting processes, recollective forgetting (F_D) and reconstructive forgetting (F_R), to be measured with the forgetting version of the model (see Appendix). As background, the percentage of forgetting (decline in recall from the third learning trial to the retention test) increases as we move from HC (17%) to MCI (48%) to AD (AD1 = 74% and AD2 = 76%), $F(3, 622) = 262.53$, $p < .0001$, partial $\eta^2 = .56$. Post hoc tests (Tukey HSD) revealed that MCI forgetting exceeded HC forgetting (mean difference in recall probability = .31), that AD1 and AD2 forgetting both exceeded MCI forgetting (mean differences in recall probabilities = .26 and .28), but that AD1 and AD2 forgetting did not differ reliably ($ps < .0001$).

To measure differences in recollective and reconstructive forgetting, we analyzed the Trial 2–4 data, which provided estimates of both types of forgetting, F_D and F_R. Those estimates appear at the bottom of Table 4. First, however, we fit the model to the data of the HC, MCI, AD1, and AD2 groups. Because that analysis had eight degrees of freedom, the critical value of the G^2 statistic to reject the null hypothesis of fit at the .05 level was 15.51. The observed value, $G^2(8) = 14.22$, was below the critical value.

Before considering differences among diagnostic groups, we note a key validity result for the two forgetting parameters. Under their theoretical definitions, reconstruction is more resistant to forgetting than is recollection (Brainerd et al., 2012). If F_D and F_R are valid measures of those processes, the former should be larger than the latter, which has been found with young adults (Brainerd et al., 2012). Estimates of these parameters for the ADAMS reveal the same pattern. Across diagnostic groups, the mean value of F_D was .58 and the mean value of F_R was .38.

Inspection of the parameter estimates reinforces the previous finding that declines in reconstructive retrieval are hallmarks of neurocognitive impairment, because, on the retention test, group differences were dominated by reconstructive forgetting. The most dramatic change occurred in HC → MCI transitions. Among HC subjects, there was no reconstructive forgetting, which is consistent with the notion that the memory representations that it processes are very stable in healthy individuals (e.g., semantic features). In sharp contrast, the forgetting rate for reconstruction was nearly 50% in MCI subjects, and a parameterwise significance test (see above) confirmed that this increase was highly reliable, $G^2(4) = 16.13$, $p < .003$. With respect to recollective retrieval, it can be seen in Table 4 that forgetting was substantial among HC subjects ($F_D = .32$) and was higher still among MCI subjects ($F_D = .59$). However, the difference was not reliable. The percentage of items that MCI subjects learned to retrieve recollectively (4% by Trial 3) was so low that there was little power to detect reliable HC–MCI differences in F_D.

For AD subjects, because forgetting rates did not differ between the AD1 and AD2 groups, estimates of F_D and F_R are reported for the pooled data as well as for the individual groups. First, it can be seen that the forgetting rate for reconstructive retrieval rose to 70% in AD subjects. Parameterwise tests showed that the MCI → AD increase in F_R was highly reliable, $G^2(1) = 52.26$, $p < .0001$. Second, it can be seen that the forgetting rate for recollective

retrieval, which was .59 in MCI subjects, rose to .84 among AD subjects. Again, however, the percentage of items that MCI and AD subjects learned to retrieve recollectively was so small that there was little statistical power with which to detect group differences in F_D, and this difference was not reliable.

Developmental changes in healthy older adults. The findings reported so far show that dual-retrieval processes—especially reconstruction—are sensitive to neurocognitive impairment. What about the other dimension of diagnostic separation, specificity? The findings might not be specific, and, instead, the same declines that were detected in HC–MCI and MCI–AD comparisons may also be occurring as healthy individuals continue to age. The specificity question can be answered with the data of the HC sample, which covers a broad age range (70 to 94).

We split the HC sample into six adjacent age groups that contained reasonably equal numbers of subjects. Their mean ages were 71.5 years, 73.5 years, 75.5 years, 79 years, 81 years, and 89 years. We fit the learning model to each group's Trial 1–3 data to estimate levels of recollection, reconstruction, and familiarity judgment, and we fit the forgetting model to each group's Trial 2–4 data to estimate levels of forgetting for recollection and reconstruction. Third, we computed an omnibus likelihood ratio test of the null hypothesis that none of the learning model's parameters differed reliably among the six age groups, and we computed the same test for the forgetting model. Neither test produced a null hypothesis rejection. Thus, the earlier findings are specific to neurocognitive impairment because these age group comparisons showed that dual-retrieval processes did not decline after age 70 as long as subjects remained healthy.

Estimated levels of recollection, reconstruction, familiarity judgment, recollective forgetting, and reconstructive forgetting appear by HC age group in Table 5. Consistent with the omnibus tests, visual inspection reveals that these processes, as measured by the CERAD at least, were remarkably constant from age 70 to the early 90s. The caveat should be added that these results are cross-sectional, and, hence, they may be contaminated by selective survival effects. It may be that subjects who survive to older and older ages are progressively healthier, on average, than age-mates who do not survive and that this is masking age declines that would be detected with longitudinal comparisons.

Table 5
Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting for ADAMS Healthy Control Subjects During Wave A

Process/parameter	Age group (years)					
	71.5	73.5	75.5	79	81	89
Recollection						
D_1	.08	.07	.22	.11	.14	.09
D_2	.00	.00	.35	.04	.00	.10
M_D	.04	.04	.29	.07	.07	.04
Reconstruction:						
R	.69	.69	.91	.66	.83	.63
Familiarity judgment						
J_1	.60	.57	.32	.50	.40	.54
J_2	.67	.69	.40	.63	.60	.63
J_3	.71	.74	.44	.71	.67	.67
M_J	.66	.67	.39	.61	.54	.61
Forgetting: Recollection						
F_D	.30	.21	.38	.33	.32	.36
Forgetting: Reconstruction						
F_R	.00	.00	.00	.00	.00	.00

Note. ADAMS = Aging, Demographics, and Memory Study.

Summary of process-level results. Application of the dual-retrieval model to the data of HC, MCI, AD1, and AD2 subjects produced three general patterns. First, as subjects were learning to recall, declines in reconstruction (R) were the most salient markers of HC → MCI and MCI → AD transitions, with the J parameters also declining by smaller amounts. Second and similarly, over the retention interval, increases in reconstructive forgetting (F_R) were hallmarks of HC → MCI and MCI → AD transitions over a retention interval. Third, changes in reconstruction and, to a lesser extent, familiarity judgment were specific as well as sensitive to neurocognitive impairment. As long as older adults remained healthy, aging itself did not produce reliable changes in the model's parameters.

Study 2

Now that process-level differences among HC, MCI, and AD diagnoses have been identified, we turn to whether such differences will forecast future disease. Concerns about possible cognitive impairment or dementia are among the most frequent reasons that older adults present for medical treatment (Petersen, 2004). Hence, current data that predict future impairment are of great clinical interest and are a key test of the clinical applicability of basic research on memory declines (Brainerd et al., 2009). Considering that AD is the most prevalent form of dementia and that MCI is its prodrome, data that predict future MCI in HC individuals and future AD in MCI individuals are of special interest (Summers & Saunders, 2012).

We shift attention to such questions in Study 2 by investigating whether measurements of dual-retrieval processes will predict longitudinal transitions to MCI and AD, over periods of 16–18 months and 4.5–6 years. This was done by combining dual-retrieval measurements that were made during Wave A of the ADAMS with Wave B, C, and D diagnostic data. We also compared the predictive power of dual-retrieval processes to that of the best genetic marker of MCI and AD, the ε4 allele of the *APOE* genotype.

Recall that during Wave A, older adults had been diagnosed as HC, CIND, or D on the basis of neuropsychological tests and medical examinations. Wave B focused specifically on transitions from CIND to D over a 16–18 month interval following Wave A, the aim being to retest and rediagnose all subjects who received CIND classifications. Ultimately, three quarters of them participated, with death being the modal reason for nonparticipation. Our interest lies with the subset of CIND subjects who were diagnosed with MCI because they account for roughly two thirds of new AD diagnoses (Brainerd et al., 2013). The objective was to determine whether the dual-retrieval measurements that were made during Wave A would predict who would be more likely to progress to AD 16–18 months later. As genetic data were available for these subjects, the predictive power of dual-retrieval processes was compared to that of the ε4 allele.

Waves C and D focused on transitions from CIND to D and from HC to CIND or D by retesting and rediagnosing subjects who

had been previously classified as CIND or HC (see Plassman et al., 2011). Slightly more than half of the Wave A HC subjects participated in Wave C, with death being the modal reason for nonparticipation, and slightly less than half of the Wave A CIND subjects participated in Wave C, with the modal reasons for nonparticipation being death or a dementia diagnosis during Wave B. (ADAMS subjects who received dementia diagnoses in a given wave did not participate in subsequent waves.) We were concerned with the following questions about HC subjects: (a) Can dual-retrieval processes predict which of them are at risk of converting to MCI 4.5–6 years later? (b) Can dual-retrieval processes predict which of them are at risk of converting to AD 4.5–6 years later? (c) How does the predictive ability of dual-retrieval processes compare to that of the ε4 allele? With respect to MCI subjects, although dual-process and genetic prediction had already been examined for the Wave B data, we reexamined these questions over the entire 6-year interval with the Wave C and D data.

Method

Subjects. With respect to the Wave B data, 77 of the 98 Wave A MCI subjects were rediagnosed during Wave B. We focused on those who either received a second cognitive impairment diagnosis (MCI or some other form of CIND) or were rediagnosed as AD (AD1 or AD2). There were 49 subjects who received a second CIND diagnosis, 35 of whom received a second MCI diagnosis and 14 of whom were diagnosed with some other form of CIND. The latter subjects were ones for whom some medical condition that can cause impairment was present that had not been detected earlier. There were 21 subjects who received a Wave B AD diagnosis, 15 of whom were diagnosed as AD1 and 6 of whom were diagnosed as AD2. The mean age of subjects who progressed did not differ significantly from the mean age of subjects who did not progress. In addition, some subjects ($N = 7$) were rediagnosed as HC during Wave B. Although that number is too small for direct statistical comparisons, indirect statistical comparisons are possible, as we show.

With respect to the Wave C and D data, the subjects of interest were those who had either (a) a Wave A MCI diagnosis or (b) a Wave A HC diagnosis. Concerning the first group, of the 77 MCI subjects who were rediagnosed during Wave B, 36 were rediagnosed during Wave C and/or Wave D. Of these 36 subjects, 15 received a final diagnosis of MCI in Wave C (if they were tested only in that wave) or Wave D (if they were tested in both waves), 14 received a final diagnosis of AD1 or AD2, and 7 received a final HC diagnosis in Wave C or D.

Concerning the second group, of the 304 subjects with a Wave A HC diagnosis, 223 were retested and rediagnosed during Wave C and/or Wave D. The data of 164 subjects were of interest. Of these subjects, 122 received a final diagnosis of HC in Wave C (if they were tested only in that wave) or Wave D (if they were tested in both waves), 25 received a final diagnosis of MCI, and 17 received a final diagnosis of AD1 or AD2.

Procedure. During Waves B, C, and D of ADAMS testing, the procedures (neuropsychological testing, medical examination, and diagnosis by a consensus panel) were the same as during Wave A. The diagnostic team that reviewed the data of each wave and assigned diagnoses was blind with respect to subjects' earlier diagnoses and data. We used the Wave A dual-retrieval measurements to predict Wave B, C, and D diagnoses of HC and MCI subjects, and we did likewise for the Wave A genetic data.

Results and Discussion

There were three important qualitative patterns for Wave B. First, mean recall accuracy did not differ reliably among the three transition groups (MCI → MCI/CIND, MCI → AD, and MCI → HC) on either the learning or the forgetting parts of the CERAD. However, second, the three groups did differ reliably at the level of retrieval processes, and as in Study 1, reconstructive retrieval differentiated them. In other words, measurements of reconstructive retrieval that had been made during Wave A were able to predict which MCI subjects would be more to likely to convert to AD over the next 16–18 months. Third, in contrast, the best genetic marker of AD failed to predict conversion to AD among the same subjects.

There were four important qualitative patterns for Waves C and D. First, as with the Wave B data, mean accuracy on the learning and forgetting portions of the CERAD did not differ among the groups of MCI subjects (MCI → MCI/CIND, MCI → HC, and MCI → AD), and it also did not differ reliably among the groups of HC subjects (HC → HC, HC → MCI, and HC → AD). Second, as with the Wave B data, measures of reconstructive retrieval during Wave A predicted which MCI subjects would convert to AD over the next 6 years, and, now, recollective forgetting also predicted such transitions. Third, reconstructive retrieval did not predict future transitions to impairment among HC subjects, but those transitions were predicted by another nonrecollective process, familiarity judgment, and also by recollective forgetting. Hence, measurements of dual-retrieval processes were able to predict which HC subjects would be more likely to transition to cognitive impairment or dementia 4.5–6 years after those measurements were taken. Fourth, the best genetic marker of neurocognitive impairment did not perform as well. The frequency of the ε4 allele failed to predict either HC → AD or MCI → AD transitions, although it did predict HC → MCI transitions.

Predicting future disease: Learning. For the Trial 1–3 data, the percentages of total correct recall for the MCI → MCI/CIND and MCI → AD transition groups were virtually the same (39% vs. 37% for Wave B subjects; 39% and 43% for Wave C/D subjects) and did not differ reliably (by *t* tests). With respect to HC subjects, the percentages of total correct recall for the three HC transition groups were 64% (HC → HC), 57% (HC → MCI), and 56% (HC → AD). An *F* test showed that these values did not differ reliably. In short, recall data by themselves, before retrieval processes were measured, did not predict future transitions to MCI or AD. It should be noted that the ADAMS contains two other tests of episodic memory, a word list recognition test and the Wechsler Story Memory Test (Wechsler, 1997), and two tests of general cognitive ability that are administered in most neuropsychological batteries, the MMSE and the SVT. Performance on those tests also failed to predict future transitions to MCI or AD. However, retrieval processes were reliable predictors.

The relevant data appear in Tables 6 (Wave A MCI → Wave B groups), 7 (Wave A MCI → Wave C/D groups), and 8 (Wave A HC → Wave C/D groups). First, however, we fit the learning model to the data of the three Wave B groups (MCI → HC, MCI → MCI/CIND, and MCI → AD) and the five Wave B/C groups

Table 6
Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting for ADAMS Wave B Diagnoses of Wave A MCI Subjects

Process/parameter	Wave B diagnosis of Wave A MCI subjects			
	MCI/other CIND	not AD	AD1/AD2	HC
Recollection				
D_1	.04	.04	.06	.08
D_2	.00	.00	.00	.18
M_D	.02	.02	.03	.13
Reconstruction				
R	.58	.60	.42	.98
Familiarity judgment				
J_1	.40	.39	.43	.21
J_2	.47	.48	.52	.42
J_3	.50	.50	.54	.41
M_J	.46	.46	.49	.33
Forgetting: Recollection				
F_D	.66	.61	.69	.44
Forgetting: Reconstruction				
F_R	.00	.00	.00	.01

Note. ADAMS = Aging, Demographics, and Memory Study; HC = healthy control; MCI = mild cognitive impairment; AD1 = probable Alzheimer's dementia; AD2 = possible Alzheimer's dementia; CIND = cognitive impairment no-dementia; not AD = pooled MCI/CIND and HC groups.

(MCI → MCI/CIND, MCI → AD, HC→HC, HC→MCI, and HC→AD). Thus, for the study as a whole, the model was fit with eight degrees of freedom, so that the goodness of fit test was $G^2(8)$, with a critical value of 15.51 to reject the null hypothesis of fit at the .05 level. As before, fit was acceptable, the observed value of the $G^2(8)$ statistic, 9.85, being well below the critical value.

MCI transitions. That the retrieval processes that MCI subjects used 16–18 months to 6 years earlier predicted who would be rediagnosed as MCI/CIND versus AD can be seen in Tables 6 and 7, where maximum likelihood estimates of the recollection, reconstruction, and familiarity judgment parameters are reported for the Wave B transition groups and the Wave C/D transition groups. Also, parameter estimates are reported for MCI → not-AD, which is an amalgamation of the MCI → MCI/CIND and MCI → HC groups. The MCI → not-AD group is actually of greater clinical interest than the MCI → MCI/CIND or MCI → HC group, because once an MCI diagnosis is assigned, the crucial question is the likelihood of a later dementia diagnosis relative to all other possibilities that do not involve dementia.

As before, the procedure for identifying between-group differences in parameters consisted of (a) a likelihood ratio test of the null hypothesis that none of the parameters differed reliably among the groups, and (b) likelihood ratio comparisons of parameter estimates for the MCI → MCI/CIND and MCI → AD groups only (the Wave B MCI → HC group was too small for direct comparisons). Concerning (a), this test produced a null hypothesis rejection for both Wave B, $G^2(12) = 48.97$, $p < .0001$, and Wave C/D, $G^2(6) = 16.86$, $p < .01$. Hence, some of the Wave A parameters could predict future transitions in both sets of data. With regard to (b), a glance at the values in Tables 6 and 7 shows that recollective retrieval and familiarity judgment did not predict which MCI subjects would transition to AD. Indeed, none of the G^2 tests for between-group differences in the D or J parameters was reliable. In contrast, it appears that reconstructive retrieval predicted future dementia, as well as transitions back to normal functioning in Wave B, in the same manner that it differentiated HC, MCI, and AD groups during Wave A: The ordering of the values of the R parameter was (MCI → HC) > (MCI → MCI/CIND) > (MCI → AD) in Wave B and (MCI → MCI/CIND) > (MCI → AD) in Wave C/D. Only the MCI → MCI/CIND and MCI → AD transition groups contained sufficient subjects for direct statistical comparison, and likelihood ratio tests showed that the estimate of R was larger in the MCI → MCI/CIND group in both Wave B and Wave C/D, $G^2(1) = 6.51$, $p < .01$, and $G^2(1) = 6.03$, $p < .02$.

Although the small number of subjects in the Wave B MCI → HC group precludes direct statistical comparison to the MCI → MCI/CIND group, an indirect comparison is possible. In particular, a parameter-invariance test can be computed for the Wave B MCI → MCI/CIND group in which the value of R is fixed at its estimated value for the MCI → HC group, rather than letting R be a free parameter (whose value is .58 for MCI → MCI/CIND subjects). The fit of the model is recalculated for the MCI → MCI/CIND group under this constraint, which yields a G^2 statistic with two degrees of freedom that is then compared to the G^2 statistic obtained when the model was originally fit with one degree of freedom (i.e., when R was free to vary). The difference between the two statistics is a $G^2(1)$ test, with a critical of value of 3.84 to reject the null hypothesis. $G^2(1) = 4.49$, and, thus, the R value that was observed for MCI → MCI/CIND subjects was reliably smaller than the value that was observed for MCI → HC subjects. Thus, there was statistical support for the conclusion that although total correct recall did not differ among the groups, the ability to retrieve items reconstructively declines steadily as one moves from MCI subjects who later transition back to HC to subjects who remain impaired to subjects who progress to AD.

Table 7
Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting for ADAMS Wave C or D Diagnoses of Wave A MCI Subjects

Process/parameter	Type of transition		
	MCI → MCI/CIND	MCI → not AD	MCI → AD
Recollection			
D_1	.05	.07	.08
D_2	.02	.06	.00
M_D	.04	.07	.04
Reconstruction			
R	.50	.65	.48
Familiarity judgment			
J_1	.45	.32	.40
J_2	.54	.45	.51
J_3	.48	.47	.53
M_J	.52	.41	.48
Forgetting: Recollection			
F_D	.58	.52	.71
Forgetting: Reconstruction			
F_R	.00	.00	.00

Note. ADAMS = Aging, Demographics, and Memory Study; MCI = mild cognitive impairment; CIND = cognitive impairment no-dementia; not AD = pooled MCI/CIND and HC groups; AD = Alzheimer's dementia.

Finally, with respect to the clinically important MCI → not-AD group, we repeated the sequence of likelihood ratio tests for this group versus the MCI → AD group, in order to determine whether there was support for a conclusion that seems apparent from inspecting the groups' parameter estimates: Reconstructive retrieval differentiates these transition groups, but recollective retrieval and familiarity judgment do not. There was support for that conclusion. For both data sets, the MCI → AD versus MCI → not-AD parameterwise G^2 test for R was reliable, but the corresponding tests for the D and J parameters were not. With respect to the critical question of which MCI subjects will not progress to dementia over the next 6 years, then, the answer was those who were better at reconstructive retrieval.

HC transitions. With the Wave C/D data, we investigated whether dual-retrieval processes also forecast future transitions to impairment and dementia among subjects who are currently healthy. The retrieval processes that, 4.5–6 years earlier, predicted which HC subjects would convert to MCI or AD can be seen in Table 8, where parameter estimates are reported for each of the three transition groups and also for the pooled data of the HC → MCI and HC → AD groups. The procedure for identifying reliable parameter differences among HC transition groups was the same as before; that is, an omnibus likelihood ratio test of the null hypothesis that none of the parameters differed reliably among the three groups, followed by omnibus likelihood ratio tests of the null hypothesis that none of the parameters differed reliably between the HC → HC and HC → MCI groups and of the null hypothesis that none of the parameters differed reliably between the HC → HC and HC → AD groups, followed by likelihood ratio tests of the null hypothesis that specific parameters (say, R) did not differ reliably between specific pairs of groups. The first test produced a null hypothesis rejection, $G^2(12) = 43.02$, $p < .0001$, establishing that at least some of the parameters differed reliably among the groups. The second pair of tests produced null hypothesis rejections for HC → HC versus HC → MCI and for HC → HC versus HC → AD, both $G^2(6)$ tests > 12.59, establishing that these transitions could be predicted by at least some of the parameters. Glancing at Table 8, it appears that (a) the recollection and reconstruction parameters do not differentiate any of the groups, but (b) the three familiarity judgment parameters differentiate the HC → HC group from the HC → AD group but not from the HC → MCI group. Parameterwise likelihood tests confirmed this. When the J parameters were compared between the HC → HC and HC → AD groups, J_1, J_2, and J_3 were larger in the group that remained healthy, the values of the $G^2(1)$ tests being 7.87, 13.62, and 10.32, respectively (all ps $< .01$). The values of $G^2(1)$ for the parameter tests that did not produce null hypothesis rejections were all < 2.50, which is well below the critical value of 3.84.

Table 8
Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting for ADAMS Wave C or D Diagnoses of Wave A Healthy Control Subjects

Process/parameter	Type of transition: HC → HC	HC → MCI/AD	HC → MCI	HC → AD
Recollection				
D_1	.16	.10	.09	.12
D_2	.17	.14	0	.29
M_D	.17	.12	.05	.21
Reconstruction				
R	.63	.70	.71	.71
Familiarity judgment				
J_1	.58	.46	.51	.41
J_2	.63	.52	.61	.39
J_3	.70	.59	.68	.45
M_J	.64	.52	.60	.42
Forgetting: Recollection				
F_D	.24	.43	.48	.47
Forgetting: Reconstruction				
F_R	0	0	0	0

Note. ADAMS = Aging, Demographics, and Memory Study; HC = healthy control; MCI = mild cognitive impairment; AD = Alzheimer's dementia.

Predicting future disease: Forgetting. As in Study 1, we used the Trial 2–4 data to estimate recollective forgetting (F_D) and reconstructive forgetting (F_R) for each group. The percentage of forgetting was virtually the same for the MCI → MCI/CIND and MCI → AD groups (13% vs. 10% for Wave B subjects; 28% and 28% for Wave C/D subjects). The same was true for the for HC → HC, HC → MCI, and HC → AD groups in Waves C/D (11%, 19%, and 16%). t and F tests showed that none of the between-group differences was reliable. Thus, raw forgetting, before retrieval processes were measured, did not predict future transitions to AD. Estimates of the two forgetting parameters are reported by transition group at the bottoms of Tables 6, 7, and 8.

For Wave B, the values in Table 6 reveal little between-group variability in recollective or reconstructive forgetting parameters. Consistent with that suggestion, a likelihood ratio test of the null hypothesis that neither forgetting parameter differed reliably among the three Wave B groups, which was a $G^2(6)$ statistic with a critical value of 12.59, did not produce a null hypothesis rejection.

The picture is different for Wave C/D, with recollective forgetting predicting future disease in both MCI and HC subjects. An omnibus likelihood ratio test indicated that at least some of the parameters differed reliably between the MCI → MCI and MCI → AD groups, $G^2(5) = 36.63$, $p < .0001$. For HC subjects, some parameters differed reliably among the three diagnostic groups, $G^2(10) = 20.02$, $p < .03$, some differed reliably between the HC → HC and HC → MCI groups, $G^2(5) = 15.95$, $p < .007$, and some differed reliably between the HC → HC and HC → AD groups, $G^2(5) = 15.55$, $p < .009$. Glancing at the values of the recollective forgetting parameter in Table 7, the rate of recollective forgetting was higher for MCI → AD subjects than for MCI → MCI/CIND subjects, and a parameterwise test showed that the difference was reliable, $G^2(1) = 5.61$, $p < .02$. Glancing at the values of the recollective forgetting parameter in Table 8, it also seems to predict future transitions to MCI and AD among HC subjects: F_D was larger in the HC → MCI group than in the HC → HC group, and it was larger in the HC → AD group than in the HC → HC group. Parameterwise tests confirmed that F_D was larger in the HC → AD group than in the HC → HC group, $G^2(1) = 19.03$, $p < .0001$, and that it was larger in the HC → MCI group than in the HC → HC group, $G^2(1) = 19.20$, $p < .0001$. Qualitatively, HC subjects who will progress to MCI or AD a few years later forget about half the items that they learn how to retrieve recollectively,

whereas HC subjects who will remain HC forget about one quarter of those items.

Predicting future disease with APOE. If measurements of dual-retrieval processes predict MCI → AD, MCI → HC, HC → MCI, and HC → AD transitions, what about the ε4 allele? For Wave B, we computed the frequencies of carriers of this allele in the MCI → MCI/CIND, MCI → not-AD, and MCI → AD groups. The frequencies were .31, .37, and .38, respectively. The MCI → not-AD versus MCI → AD frequencies and the MCI → MCI/CIND versus MCI → AD frequencies did not differ reliably, with $\chi^2(1) < 1$ in each instance. For Wave C/D, we computed frequencies of ε4 carriers in the MCI → MCI, MCI → not-AD, and MCI → AD groups, which were 43%, 44%, and 38%, respectively. Neither the MCI → MCI versus MCI → AD difference nor the MCI → not-AD versus MCI → AD difference was reliable, $\chi^2(1) < 1$ in each instance. Also for Wave C/D, we computed frequencies of ε4 carriers in the HC → HC, HC → MCI, and HC → AD groups, which were 20%, 50%, and 29%, respectively. The difference in ε4 frequencies was reliable for HC → MCI conversion, $\chi^2(1) = 12.32$, $p < .005$, but not for HC → AD conversion, $\chi^2(1) = 1.46$.

Thus, the predictive story for dual-retrieval processes versus ε4 was simple. The former were successful predictors, but the latter mostly were not. Whereas the memory processes forecast all transitions, including MCI → HC reversion, ε4 predicted only HC → MCI conversion.

Study 3

In Study 2, measurements of dual-retrieval processes identified subgroups of HC subjects who were at increased risk of conversion to AD and MCI and subgroups of MCI subjects who were at increased risk of conversion to AD. Measurements of those processes proved to be better predictors of future AD and future MCI than the best genetic marker of these conditions. Such findings suggest that dual-retrieval processes can be useful tools when it comes to identifying individuals who are at increased risk of MCI or AD, but predictions of that sort require parameter estimates for individuals. Although the dual-retrieval model is applicable to individuals as well as groups (Brainerd et al., 2009), the CERAD does not generate sufficient data to fit the model to the performance of individuals; hence, only group-level analyses are possible with the ADAMS. While that suffices to test theoretical hypotheses about AD and MCI and to study predictive power, it is unsatisfactory because predicting impairment is a question about individuals.

To make progress on that question, we report a final study in which another standard clinical instrument, the RAVLT, was used to (a) fit the model to the performance of individuals and obtain individualized estimates of *D, R,* and *J*, (b) test theoretical hypotheses about AD and MCI with individualized rather than group parameter estimates, and (c) test the ability of individualized estimates to predict future transitions to MCI and AD. The RAVLT generates considerably more data for individuals than does the CERAD, for two reasons: The study list consists of 15 words rather than 10, and there are five study–test cycles rather than three. As noted previously, three cycles are sufficient to produce identifiable estimates of all six parameters. With more than three, as we have shown elsewhere (Brainerd et al., 2012), reliable individualized parameter estimates can be obtained, even with fairly short lists, using a sliding window bootstrap with resampling procedure. The resulting estimates of *D, R,* and *J* are averages of the separate parameter values for each of these sequences.

With respect to the first of the three objectives, we sought clinical data in which the RAVLT had been administered to large samples of HC individuals, individuals with MCI, and individuals with AD. Although establishing individualized model fits for healthy older adults would be significant progress by itself, the ultimate question is whether the model fits the performance of individuals with impairment or dementia as well as healthy individuals, which would allow direct comparisons of retrieval processes among diagnostic groups. A data set that meets this specification is the ADNI, in which the RAVLT, along with some of the same ADAMS neuropsychological tests, was administered to older adults. With respect to the second objective, a consistent finding of the ADAMS HC–MCI and MCI–AD comparisons was that recollective retrieval did not decline with increasing impairment. Although that is a valid normative pattern, because the subject sample on which it is based is nationally representative, it might not hold for some important subgroups of older adults. In particular, it might not hold for more highly educated subgroups, who exhibit higher levels of recollection than representative samples (Brainerd et al., 2009). Mean educational level is considerably higher for ADNI subjects than for ADAMS subjects (see below).

With respect to the third objective, predicting future MCI and AD, the ADNI, like the ADAMS, has a longitudinal component. Subjects who were initially diagnosed as HC or MCI were followed for 2 years and were rediagnosed at intervals of 6, 12, 18, and 24 months. At the end of 2 years, 44% of subjects who had initially been diagnosed as MCI had converted to AD, and 8% of subjects who had initially been diagnosed as HC had converted to MCI. Consequently, it was possible to determine whether individualized measurements of dual-retrieval processes, which were obtained at the start, were able to predict future HC → MCI and MCI → AD conversion. ADNI subjects, like ADAMS subjects, were genotyped, so it was also possible to compare the predictive power of individualized measurements of dual-retrieval processes to that of the ε4 allele. Because measurements of dual-retrieval processes were individualized, they could be combined with the genetic data in logistic regressions to compute the levels of sensitivity and specificity with which future disease was predicted.

Method

ADNI design and subjects. Data used in the preparation of this article were obtained from the ADNI database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and nonprofit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treat-

ments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California–San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 subjects, but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1,500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org

The ADNI resembles the ADAMS in that older adults were administered a battery of neuropsychological tests, which consisted of the RAVLT, a word list recognition test, the Wechsler Story Memory Test, the MMSE, forward and backward digit span, the Boston Naming Test, the Clocks Test, the Geriatric Depression Scale, and the Category Fluency Test. Except for the RAVLT, all of these instruments were administered in the ADAMS, the main difference between the two batteries being that the ADAMS included other instruments that were not in the ADNI. Similar to the ADAMS, subjects received a diagnosis at the start of the ADNI, but the subject sample was then restricted to individuals who received one of just three diagnoses: HC ($N = 207$), a-MCI ($N = 368$), and probable AD ($N = 173$).

The focal recall task, the RAVLT, differs from the CERAD in three ways. First, as mentioned, the list consists of 15 words rather than 10, and, second, learning consists of 5 study–test cycles rather than three. Third, the forgetting test occurs 30 minutes after the last learning trial. As with the CERAD, the RAVLT lists consist of familiar concrete nouns, free recall tests are oral, the second forgetting test occurs without further opportunities to study the original list, and the forgetting interval is filled with interpolated neuropsychological tests. The interpolated tests included the Wechsler logical memory test, forward and backward digit span, category fluency, digit-symbol substitution, Boston naming, and rating scales for depression and dementia.

There were four notable differences between the ADNI and the ADAMS subjects. First, the ADAMS sample was nationally representative, but the ADNI sample was not. ADAMS subjects were obtained via representative sampling from a national pool of over 60,000 older adults, whereas ADNI subjects were individuals who responded to recruitment efforts at various clinical sites. Second, ADNI subjects were more educated than ADAMS subjects. The average ADAMS subject had not completed high school (mean education = 11.2 years), but the average ADNI subject had completed nearly 4 years of college (mean education = 15.6 years). Third, although the mean age of both samples was above 70, ADAMS subjects were roughly five years older than ADNI subjects (mean ages = 80 years 11 months vs. 75 years 5 months). Fourth, the ADNI MCI sample was restricted to the AD prodrome, a-MCI, and the AD sample was restricted to probable AD. The Wechsler logical memory test was used to distinguish a-MCI subjects from other MCI subjects during initial testing: Subjects were required to have scores in the 2–8 range on the delayed part of this test (the maximum score is 25) to be included in the MCI sample. As a result, although ADAMS MCIs were a mixture of a-MCI and n-MCI, ADNI MCIs were exclusively a-MCI. This is important when it comes to testing theoretical hypotheses about the retrieval processes that differentiate HC, MCI, and AD groups and that predict future HC → MCI and MCI → AD transitions. Those processes might be different when the MCI group is restricted to subjects with significant memory impairment and when the AD group is restricted to probable AD.

As mentioned, ADNI subjects were rediagnosed at regular intervals over the next 2 years. Nearly half the a-MCI subjects converted to AD, and the rest did not (designated as the a-MCI_C and a-MCI_{NC} subgroups below). A small proportion (8%) of the HC subjects converted to a-MCI, and the rest did not (designated as the HC_C and HC_{NC} subgroups below).

Results and Discussion

Our principal concern, of course, was whether the model fit the recall data of individual subjects, and, crucially, whether fit was acceptable for impaired and demented subjects as well as for healthy ones. Assuming that fit could be established, we intended to reexamine the question of which retrieval processes differentiated those groups and the question of whether any of the retrieval processes were able to predict future disease; this time with individual rather than group parameter estimates.

Individualized model fits and parameter estimates. Individualized data are inherently noisier than group data, and models that work well with the latter often fail with the former. When the dual-retrieval model is applied to the recall of an individual subject, the test has one degree of freedom, just as with group data. As there were RAVLT data for 748 subjects, the total degrees of freedom for fitting the model to the complete data space were 748, and the critical value of the G^2 statistic to reject the null hypothesis of model fit at the .05 level was 812.76. The observed value was 700.88, so that the model delivered acceptable fit to individual recall protocols, for the study as whole. Equally important, fit was acceptable for each diagnostic group, considered separately, allowing direct parameter comparisons among them. There were 173 AD subjects, 368 a-MCI subjects, and 207 HC subjects, so that the critical .05 values of the G^2 statistic were 204.69, 413.73, and 241.57, respectively. The observed values were 182.92, 354.67, and 183.31, respectively.

Distributions of observed values of the G^2 statistic for individual subjects in each diagnostic group are shown in Figure 1. As nearly half of the a-MCI group ultimately progressed to AD, the distributions for the a-MCI_C and a-MCI_{NC} subgroups are shown separately. For individual subjects, the critical value of G^2 to reject the null hypothesis of model fit is 3.84, which appears as a dashed line in each panel. It can be seen that the mean G^2 values of these distributions (AD = 1.06, a-MCI_C = 0.99, a-MCI_{NC} = 0.95, and HC = 0.78) are all far below this cutoff, with virtually all of the distribution for each diagnostic group falling below it. It can also be seen that each distribution has a pronounced positive skew.

Diagnostic differences in dual-retrieval processes: Learning and forgetting. Mean probabilities of correct recall for the six recall tests appear by diagnostic group in Table 3. Naturally, average recall differed among the diagnostic groups. During the

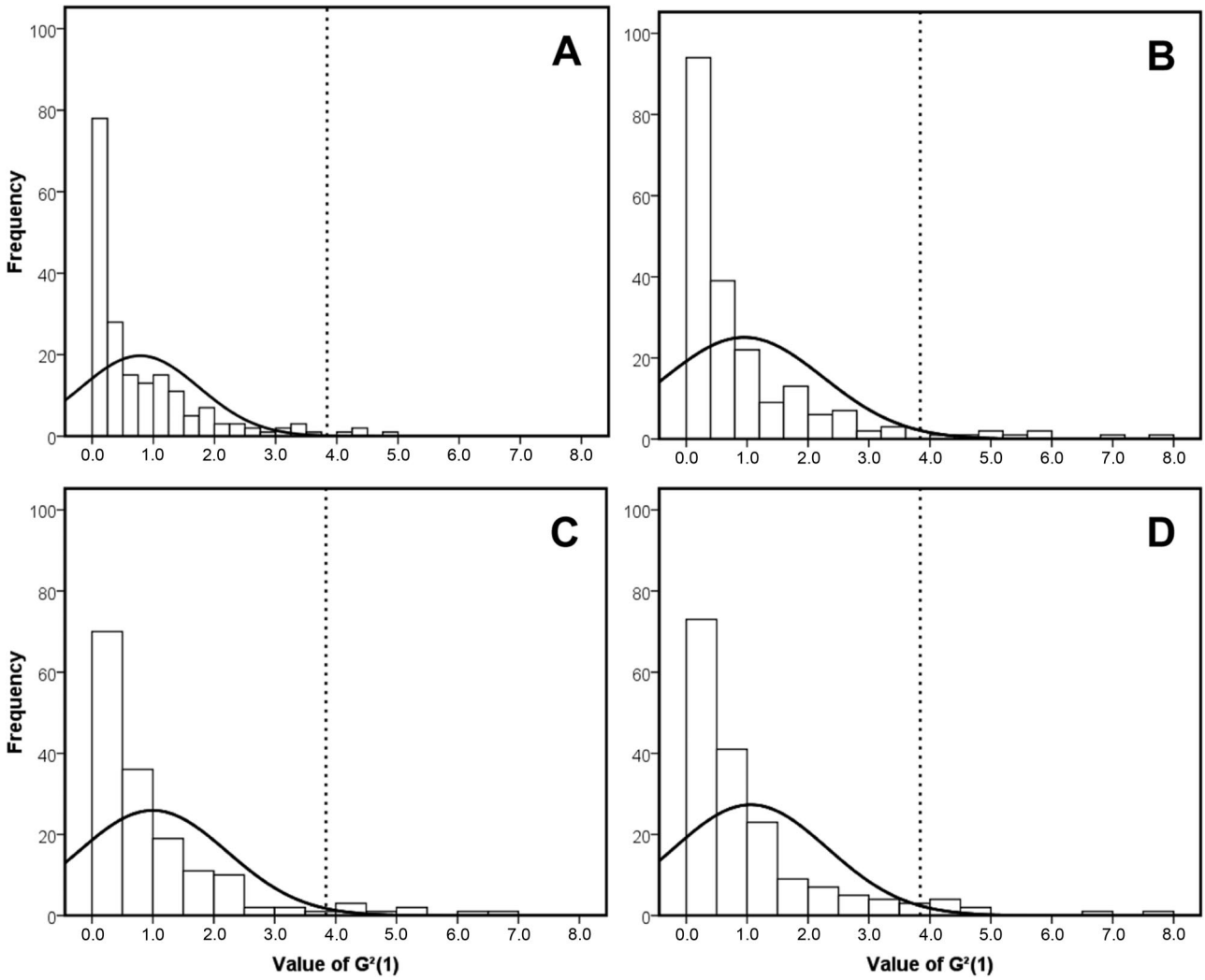

Figure 1. Distributions of the individualized goodness-of-fit statistics for healthy control subjects (A), a-MCI subjects who remain a-MCI (B), a-MCI subjects who would later convert to probable AD (C), and probable AD subjects (D). The dotted line in each panel is the critical value for rejection of the null hypothesis that dual-retrieval model fits the data. a-MCI = amnestic mild cognitive impairment; AD = Alzheimer's dementia.

learning phase, mean recall (out of 15) was 8.71 (HC), 6.20 (a-MCI), and 4.54 (AD), $F(2, 724) = 219.89$, $MSE = 3.16$, $p < .0001$. During the forgetting phase, it was 6.58 (HC), 3.77 (a-MCI), and 2.30 (AD), $F(2, 724) = 209.57$, $MSE = 3.61$, $p < .0001$. Mean values of individualized parameter estimates for the AD, a-MCI_C, a-MCI_{NC}, and HC groups are reported in Table 9, and separate estimates are reported for the HC_C and HC_{NC} subgroups. Visual comparisons reveal similarities and differences with the earlier comparisons for the ADAMS diagnostic groups (see Table 4), which were confirmed with significance tests. With respect to similarities, the most obvious one is that reconstructive retrieval is again the retrieval process that consistently discriminates all diagnostic groups from each other: As we move from HC to a-MCI to AD, its value declines from .48 to .38 to .28. Another similarity is that recollective forgetting was considerably lower for HC than for a-MCI or AD. Additional similarities with earlier findings are that (a) reconstructive retrieval was easier than recollective retrieval, (b) rates of forgetting were much higher for recollection than for reconstruction, (c) familiarity judgment did not differentiate any of the diagnostic groups, and (d) estimated values of the familiarity judgment parameter for HC and a-MCI subjects were close to their values in the ADAMS data.

With respect to differences between the estimates in Table 9 versus Table 4, there are three notable ones. First, concerning recollective retrieval, we thought that the ADNI data would increase the values of the recollection parameters and that this process might now differentiate diagnostic groups. It did. The mean values of D in Table 9 were larger than the corresponding values in Table 4, and it can be seen that M_D declined as subjects transitioned from HC to a-MCI and from a-MCI to AD, though the latter decline was small relative to the former. Second, although reconstructive retrieval was again easier than recollective retrieval, the values of the reconstruction parameter were smaller for all diagnostic groups in this study—especially for the HC and a-MCI

Table 9
Individualized Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting Parameters for ADNI Subjects

Process/parameter	Diagnosis						
	HC_{NC}	HC_C	HC	a-MCI_{NC}	a-MCI_C	a-MCI	AD
Recollection							
D_1	.27	.20	.26	.19	.15	.17	.14
D_2	.20	.16	.20	.11	.07	.09	.06
M_D	.24	.18	.23	.15	.11	.13	.10
Reconstruction							
R	.48	.34	.48	.40	.36	.38	.28
Familiarity judgment							
J_1	.72	.85	.73	.74	.74	.74	.81
J_2	.51	.55	.51	.48	.46	.47	.45
J_3	.43	.44	.43	.39	.30	.38	.30
M_J	.56	.61	.56	.54	.50	.53	.52
Forgetting: Recollection							
F_D	.16	.33	.17	.30	.32	.31	.39
Forgetting: Reconstruction							
F_R	.01	.00	.01	.02	.03	.02	.04

Note. ADNI = Alzheimer's Disease Neuroimaging Initiative; HC = healthy controls; HC_{NC} = healthy controls who did not convert to a-MCI; HC_C = healthy controls who converted to a-MCI; a-MCI = amnestic mild cognitive impairment; a-MCI_{NC} = a-MCI subjects who did not convert to AD; a-MCI_C = a-MCI subjects who converted to AD; AD = Alzheimer's dementia.

groups. Item selection is a likely explanation. The RAVLT has two more learning trials than the CERAD, so that more items can be recollected by the end of learning. This means that the average difficulty of the remaining items, which must be reconstructed to be recalled, will be higher for the RAVLT than the CERAD. Third, forgetting levels for both recollection and reconstruction were lower than before, and reconstructive forgetting was near floor in all diagnostic groups. Again, a likely explanation lies with the additional learning trials of the RAVLT, which should make both retrieval operations more resistant to forgetting because learning is more thorough.

Turning to significance tests of these similarities and differences, because parameters were estimated for individuals, standard analysis of variance (ANOVA) procedures can be used rather than likelihood ratio comparisons. To simplify the ANOVAs, we first computed a mean value of the two *D* parameters and a mean value of the three *J* parameters for each subject, which meant that there would be four parameters to submit to ANOVA, three learning parameters (M_D, R, and M_J) and a forgetting parameter (F_D). (ANOVAs for the reconstructive forgetting parameter were not computed because values were near floor.) Using each parameter as a dependent variable, we computed a 3 (diagnostic group: HC, a-MCI, AD) × 2 (ε4: noncarrier vs. carrier) ANOVA to determine if it differed reliably as a function of diagnosis. The *APOE* factor was included because we planned to investigate the retrieval processes' ability to predict future disease (see below), and to interpret those data, it is essential to know whether the processes interact with genotype. (In Study 2, retrieval processes and genotypes were independent predictors.)

The ANOVA for reconstructive retrieval produced a main effect for diagnostic group, $F(2, 742) = 32.63$, $MSE = .05$, $p < .0001$, no main effect for ε4, and no interaction. Post hoc comparisons (Tukey HSD) for the diagnostic group effect revealed (all ps $< .0001$) that R was larger for HC than for a-MCI and larger for a-MCI than for AD, as in Study 1. The ANOVA for recollective retrieval produced a main effect for diagnostic group, $F(2, 742) = 84.05$, $MSE = .01$, $p < .0001$, no main effect for ε4, and no interaction. Post hoc comparisons revealed (all ps $< .001$) that M_D was larger for HC than for a-MCI and larger for a-MCI than for AD. The ANOVA for familiarity judgment produced no main effects and no interaction. The ANOVA for recollective forgetting produced a main effect for diagnostic group, $F(2, 742) = 30.25$, $MSE = .07$, $p < .0001$, no main effect for ε4, and no interaction. Post hoc comparisons for the diagnostic group factor revealed (all ps $< .001$) that F_D was larger for HC than for a-MCI and larger for a-MCI than for AD.

In summary, the process-level differences among the diagnostic groups resembled those in Study 1 in three major respects: Reconstructive retrieval separated all three diagnostic groups from each other, recollective forgetting separated HC from a-MCI and AD, and familiarity judgment did not separate the diagnostic groups. There were two notable differences as well. First, recollective retrieval now separated all of the diagnostic groups, and, second, recollective forgetting now separated a-MCI from AD. In connection with the latter finding, remember that there were large differences between MCI and AD F_D estimates for the ADAMS, but they were not reliable owing to low levels of initial recollective learning among AD subjects. Thus, the only result that is substantially different than the ADAMS data is that recollective retrieval also separated the diagnostic groups.

Because retrieval processes were estimated for individuals, formal statistical measures of the accuracy with which they separate the HC group from the a-MCI group and the a-MCI group from the AD group can be computed. This is done via logistic regressions, in which diagnostic groups are the dependent variables and individualized estimates of the parameters that differentiate those groups are the predictor variables. This delivers measures of sensitivity and specificity, from which an overall accuracy estimate,

the mean of the two measures, can be calculated. We computed two such analyses, one for HC–a-MCI discrimination and one for a-MCI–AD discrimination. Estimates of the model's learning and forgetting parameters were the predictor variables in both analyses.

To conduct the logistic regression for HC-a-MCI, we first deleted the data of the small group of subjects ($N = 25$) whose fit test equaled or exceed the critical value, leaving a total of 550 subjects. The logistic regression for those subjects produced a reliable fit statistic, $\chi^2(8) = 173.06$, $p < .0001$. The overall accuracy with which dual-retrieval processes discriminated the two groups was good (73.5%), with good sensitivity (88.8%; chance = 63.8%) and specificity (58.3%; chance = 32.6%). The frequency of ε4 differed reliably among the three diagnostic groups (HC = 25.1%, a-MCI = 54.9%, AD = 64.7%), and we saw in the earlier ANOVAs that the effects of ε4 carrier status were independent of retrieval processes (i.e., there was never an *APOE* × Retrieval Process interaction). Therefore, it was possible that sensitivity and/or specificity might improve if ε4 carrier status were added as a predictor variable. They did not improve. When ε4 carrier status was added, the new sensitivity and specificity values, 86.3% and 59.8%, were virtually the same as without this predictor. Thus, retrieval processes alone did a good job of differentiating the two diagnostic groups, and the genetic data could be dispensed with because they did not improve differentiation.

To conduct the logistic regression for a-MCI–AD, we first deleted the data of the small group of subjects ($N = 26$) whose fit test equaled or exceed the critical value, leaving a total of 515 subjects. The logistic regression for those subjects produced a reliable fit statistic, $\chi^2(8) = 98.94$, $p < .0001$. The overall accuracy of group differentiation was moderately good (65%). However, although the specificity of group differentiation was good (89.7%; chance = 68%), sensitivity was not (40.2%; chance = 31.8%). As before, it was possible that sensitivity and/or specificity might improve if ε4 carrier status were added as a predictor variable. However, when it was added, sensitivity and specificity values, 89.5% and 39.6%, were virtually unchanged. In differentiating the a-MCI group from the AD group, then, retrieval processes did a good job of avoiding false positives but an unsatisfactory job of avoiding false negatives. As with HC–a-MCI differentiation, the genetic data could be dispensed with because they did not improve differentiation.

Developmental changes in HC subjects. In Study 1, we found that although dual-retrieval processes differed among HC, MCI, and AD groups, they did not differ among HC subjects of different ages. To determine whether this was also true for ADNI subjects, we split the HC group into four adjacent age levels with reasonably equal numbers of subjects per group: mean ages = 70.4 years (range = 62–72), 73.7 years (range = 73–75), 76.9 years (range = 76–78), and 82.6 years (range = 79–89). We fit the model separately to the recall protocols of individual subjects in each age group. The critical value for rejection of the null hypothesis of fit was 3.84 for each subject, and the mean values of the fit statistic for the age four groups were .90, .78, .62, and .82, respectively.

The means of the individualized estimates of the D, R, J, F_D, and F_R parameters for each age group appear in Table 10. Visual inspection indicates that they do not vary perceptibly. That was confirmed with a 4 (age group) × 8 (parameter) ANOVA, which produced neither an age main effect nor an Age × Parameter interaction. Once again, therefore, there was no support for the hypothesis that dual-retrieval processes decline with age during late adulthood, as long as subjects remain healthy. Hence, the declines that were observed for the HC versus a-MCI and a-MCI versus AD are specific to disease.

Table 10
Mean Estimates of Recollective Retrieval, Reconstructive Retrieval, Familiarity Judgment, and Forgetting for ADNI Healthy Control Subjects

	Age group (years)			
Process/parameter	70.4	73.7	76.9	82.6
Recollection				
D_1	.28	.28	.25	.24
D_2	.23	.19	.20	.16
M_D	.26	.24	.23	.20
Reconstruction				
R	.48	.51	.47	.47
Familiarity judgment				
J_1	.72	.69	.72	.73
J_2	.51	.50	.51	.52
J_3	.43	.44	.43	.44
M_J	.55	.54	.55	.56
Forgetting: Recollection				
F_D	.16	.13	.16	.18
Forgetting: Reconstruction				
F_R	0	.01	.02	.01

Note. ADNI = Alzheimer's Disease Neuroimaging Initiative.

Predicting HC → a-MCI and a-MCI → AD conversion with dual-retrieval processes. Recall that nearly half of the a-MCI subjects converted to AD over the 2-year interval, while a small proportion of HC subjects converted to a-MCI. Thus, the question of whether dual-retrieval processes predict future disease reduces to whether any of them are able to distinguish a-MCI_C subjects from a-MCI_{NC} subjects or to distinguish HC_C subjects from HC_{NC} subjects. To answer these questions, we computed a 5 (diagnostic group: HC_C, HC_{NC}, a-MCI_C, a-MCI_{NC}, AD) × 2 (ε4: noncarrier vs. carrier) × 4 (parameter: M_D, R, M_J, F_D) ANOVA, which produced a main effect for diagnostic group and a Diagnostic Group × Parameter interaction. We then analyzed the interaction to determine whether any of the retrieval processes distinguished MCI_C from a-MCI_{NC} subjects or HC_C from HC_{NC} subjects. They did. Recollective retrieval (M_D) predicted future transitions from a-MCI to AD, whereas reconstructive retrieval (R) and recollective forgetting (F_D) predicted future transitions from HC to a-MCI.

As with the ADAMS data, then, measurements of dual-retrieval processes that were taken at the start were able to predict future transitions to a-MCI and to AD over a period of 2 years. Recollective forgetting was more common among subjects who ultimately converted from HC to a-MCI, similar to the ADAMS, but there was a further predictor, reconstructive retrieval. Note that reconstructive retrieval was not reliable predictor in Study 2. With respect to AD conversion, in contrast to the ADAMS data for a-MCI → AD, where learning how to retrieve reconstructively was a reliable predictor, it was not so for ADNI subjects. However, recollective retrieval was: a-MCI subjects who converted to AD were less able to retrieve recollectively at the start of the 2-year interval.

Because dual-retrieval processes were estimated for individuals, it is possible to compute formal statistical measures of the sensi-

tivity and specificity with which they predict future disease. As before, this was done with logistic regressions, in which diagnostic groups supplied the dependent variables, and the predictor variables were individualized estimates of the model parameters. Also as before, the small groups of subjects whose fit statistics equaled or exceeded the critical value did not figure in the logistic regressions.

For HC → a-MCI transitions, computation of a binary logistic regression in which the dependent variable was HC_C versus HC_{NC} and the predictor variables were estimated values of the learning and forgetting parameters produced a reliable fit test, $\chi^2(8) = 21.04$, $p < .008$. However, that analysis is not meaningful. Because the number of HC_C subjects (16) was so small, relative to the number of HC_{NC} subjects (191), the level of chance prediction of a future HC_{NC} diagnosis is 92.3%. Hence, it is impossible to achieve a statistically reliable level of specificity with HC subjects, which means that the formal predictive power of dual-retrieval processes cannot be evaluated with these data when it comes to future conversion to a-MCI.

In contrast, a formal evaluation is possible with respect to prediction of future AD, because the sizes of the a-MCI_C ($N = 158$) and a-MCI_{NC} ($N = 93$) groups are not as extremely unbalanced as those of the HC_C and HC_{NC} groups. When we computed a binary logistic regression with a-MCI_C versus a-MCI_{NC} as the dependent variable and the model's learning and forgetting parameters as predictor variables, the fit test was reliable, $\chi^2(8) = 47.81$, $p < .0001$. The overall accuracy with which dual-retrieval processes predicted future AD was moderately good (66%), with sensitivity being reliably above chance (62%; chance = 45%) and specificity also being reliably above chance (69.9%; chance = 55%). As in prior logistic regressions, we added ε4 carrier status as a predictor variable to determine if it contributed further predictive variance. It did not. Sensitivity (60.1%) and specificity (68.9%) were virtually the same as before, and, hence, the genetic data could once again be dispensed with.

Summary. Overall, application of the dual-retrieval model to the ADNI data set produced four patterns that are worthy of note. By far the most important one from the perspective of theory-driven research on neurocognitive impairment is that individualized model fits were good. The model fit individual RAVLT protocols as well for individuals with a-MCI and individuals with AD as it did for HC individuals (see Figure 1). Second, as in Study 1, dual-retrieval processes differentiated the diagnostic groups. Reconstructive retrieval again declined steadily as one moved from HC to a-MCI to AD. In these more educated subjects, who exhibited higher levels of recollective retrieval than those in the ADAMS sample, recollection declined and recollective forgetting increased, from HC to a-MCI to AD. Sensitivity–specificity analyses showed that by themselves, dual-retrieval processes produced good HC–a-MCI differentiation, though sensitivity was not reliable for a-MCI–AD differentiation. Third, dual-retrieval processes again predicted future disease, with reconstruction and recollective forgetting predicting conversion to a-MCI and recollection predicting conversion to AD. Dual-retrieval processes' levels of sensitivity and specificity in predicting future AD were both reliable. The genetic data contributed no predictive variance beyond what was contributed by dual-retrieval processes alone.

General Discussion

We began with two broad objectives. One was to extend current work on dual-process models of recall to theoretical hypotheses about neurocognitive impairment, a domain in which recall decline is the most reliable neuropsychological marker of MCI and AD. There, our focus was on comparing different dual-process conceptions of memory declines in MCI and AD. The other objective was to use dual-process measurements to predict future transitions to MCI and AD and to compare their predictive power to that of the ε4 allele. Thus, our research may be thought of as a proof of concept—the concept being that mathematical models of recall are productive tools for pinpointing the retrieval processes that define "memory decline" in MCI and AD and for predicting future emergence of these diseases. To conclude, we discuss what was accomplished in each of these spheres.

Dual-Process Conceptions of MCI and AD

Dual-retrieval models of recall were introduced in response to two uncertainties that are posed by recognition-based measurement. One, which is explicated in Malmberg's (2008) review, is that the standard methods of separating dual processes in recognition data (remember/know, ROC, process dissociation) may not do so. Many investigators (e.g., Heathcote et al., 2006) have shown, for one or another of these methods, that the resulting data can be handled by models that posit only a single familiarity process. The other uncertainty (Brainerd et al., 2009; Ghetti & Angelini, 2008) is that these methods require subjects to make complex metacognitive judgments, and, hence, they are too demanding for some populations—notably, older adults with CIND or dementia. Dual-retrieval models of recall avoid both problems: Recall indexes two distinct forms of retrieval, and the simple tasks over which these models are defined are within the capabilities of subjects with impairment or dementia.

These models have been applied in recent developmental studies (Brainerd et al., 2009, 2012) and experiments with young adult samples (Brainerd & Reyna, 2010; Gomes et al., 2013). Levels of fit have been good, and model parameters have behaved in accordance with theoretical expectations (e.g., *D* estimates are smaller than *R* or *J* estimates, forgetting rates are steeper for *D* than for *R* or *J*, *D* responds to verbatim cuing manipulations while *R* responds to semantic cuing manipulations). The present studies yielded a similar picture for healthy and impaired older adults. At both the group and individual levels, fit statistics did not exceed critical values for model rejection and were usually well below critical values. Also, parameter behavior was theoretically coherent inasmuch as the patterns among *D, R, J,* F_D, and F_R estimates that have been observed in young adults were observed in MCI and AD subjects, as well as in healthy older adults. Study 3 provided data on these points that went well beyond prior research, because prior research has not involved individualized fits or individualized parameter estimates.

Turning to theoretical hypotheses about memory declines, across the three studies, several parametric patterns that bear on those hypotheses were reported, and for convenience of reference, all of them have been summarized in qualitative language in Table 11. Hypotheses about memory declines in neurocognitive impairment are motivated by the fact that the brain regions of main interest in dual-process research with healthy subjects are also

Table 11
Diagnostic Differentiation and Prediction of Future Disease by Dual-Retrieval Processes

Process	Diagnostic differentiation		Prediction of future disease				
	HC vs. a-MCI	a-MCI vs. AD	HC → a-MCI	HC → AD	a-MCI → HC[c]	a-MCI → not AD[c]	a-MCI → AD
			Studies 1 and 2 (ADAMS)				
Retrieval							
Direct access	Yes[a]	No	No	No	No	No	No
Reconstruction	Yes	Yes	No	No	Yes	Yes	Yes
Familiarity judgment	Yes	Yes[b]	No	Yes	No	No	No
Forgetting							
Recollective	No	No	Yes[d]	Yes[d]	No	No	Yes[d]
Reconstructive	Yes	Yes	No	No	No	No	No
			Study 3 (ADNI)				
Retrieval							
Direct access	Yes	Yes	No	NA	NA	NA	Yes
Reconstruction	Yes	Yes	Yes	NA	NA	NA	No
Familiarity judgment	No	No	No	NA	NA	NA	No
Forgetting							
Recollective	Yes	Yes	Yes	NA	NA	NA	No
Reconstructive	No	No	No	NA	NA	NA	No

Note. HC = healthy controls; a-MCI = amnestic mild cognitive impairment; AD = Alzheimer's dementia; NA = not analyzed; ADAMS = Aging, Demographics, and Memory Study; ADNI = Alzheimer's Disease Neuroimaging Initiative.
[a] The only reliable difference was in the proportion of items recalled recollectively by the last trial (Trial 3). [b] Based on reliable differences in the J_3 parameter only. [c] Based on indirect statistical comparisons necessitated by small sample sizes. [d] Predicts only 4.5- to 6-year transitions.

regions that exhibit pathology in postmortem studies of individuals with MCI and individuals with AD (e.g., Nelson et al. 2009). We saw that the dominant continuity view extends a well-established trend in healthy aging, positing that conversion to MCI and AD are characterized, respectively, by further declines in and almost complete loss of recollective capability, coupled with sparing of nonrecollective processes (e.g., Bugaiska et al., 2011; Serra et al., 2010). The alternative discontinuity view posits that the shift from healthy aging to neurocognitive impairment is accompanied by a shift from recollective to nonrecollective declines (e.g., Brainerd et al., 2009). Here, a key consideration is that recollective declines in healthy subjects are so substantial by age 70 that recollection may not be a realistic source of the severe declines in episodic memory that are specified in diagnostic criteria for MCI and AD. Another consideration is that the brain atrophy that is associated with these conditions, particularly with AD, has spread beyond the regions that are associated with recollection. Nevertheless, the hypothesis that nonrecollective declines become prominent in conversion to MCI and AD is surprising from the perspective of the traditional continuity hypothesis.

When the subjects were nationally representative samples of HC individuals, individuals with MCI, and individuals with AD (Studies 1 and 2), none of the dual-process comparisons among the groups could be said to favor a continuity view. We measured both learning and forgetting of recollective and nonrecollective retrieval. On the learning side, there was little support for the notion that recollective retrieval is connected to neurocognitive impairment. (The only evidence of that was a small difference in the total amount of recollective retrieval by the last learning trial for HC versus MCI but not for MCI versus AD.) Instead, both the HC–MCI and MCI–AD comparisons were dominated by declines in nonrecollective retrieval, especially reconstruction. Both the reconstruction and familiarity judgment components declined in HC versus MCI. The reconstruction component declined by a much larger amount in MCI versus AD, and the familiarity judgment component declined by a much smaller amount. On the forgetting side, reconstruction was again the dominant discriminator of the three diagnostic groups. First, there was a qualitative difference between healthy and impaired subjects in the sense that there was no reconstructive forgetting among HC subjects. There was marked reconstructive forgetting in impaired subjects, however, with the level being higher for AD than MCI (80% vs. 50%). All of these results are more consistent with the discontinuity view of memory declines than the continuity view. Another noteworthy datum is that the process-level changes that were hallmarks of MCI and AD were specific to disease. When we divided the HC sample into adjacent age groups, recollection, reconstruction, and familiarity judgment were invariant between age 70 and the early 90s.

Because the subject sample was nationally representative, these results provide a normative picture of process differences among diagnostic groups for the average older adult. In Study 3, we continued to examine process differences with the data of the ADNI, which provides another large sample of HC individuals, individuals with MCI, and individuals with AD. This sample is not nationally representative, and the MCI and AD groups are restricted to a-MCI and probable AD individuals. Unlike the ADAMS, then, the ADNI does not provide a valid normative picture. However, it has a key advantage over the ADAMS—namely, that with the ADNI data, one can address the question of whether the model performs well with individual recall protocols. The ADNI used a recall instrument (the RAVLT) that generates sufficient data for individualized fits and individualized parameter estimates. We found that the individualized fits were good and that they were as good for subjects with impairment or dementia as they were for healthy ones (see Figure 1). Further, the individualized parameter estimates differed reliably among diagnostic groups. Another useful feature of the ADNI is that one can evaluate the extent to which the

components of the ADAMS normative picture of process differences among diagnostic groups hold for different subpopulations of older adults. Explicitly, although declines in recollective retrieval are not markers of MCI and AD in nationally representative samples, they might be in more highly educated samples, in which baseline levels of recollection are higher.

In that connection, we found that most process differences were the same as in the ADAMS, but some were different. With respect to similarities, the process difference that dominated group comparisons in the ADAMS was also present in the ADNI: Reconstructive retrieval declined steadily from HC to MCI to AD. Also, familiarity judgment, which displayed small diagnostic group differences in the ADAMS, was completely spared in the ADNI. Thus, the ADAMS findings for reconstructive retrieval are candidates for universal aspects of memory decline in MCI and AD. As expected, the main difference between the ADNI and ADAMS data lay in the findings for recollection. During learning, there was a substantial HC–MCI difference in the mean value of the *D* parameters (.23 vs. .13) and a small but reliable MCI–AD difference. During forgetting, the differences among diagnostic groups were entirely due to differences in recollective forgetting, whereas reconstructive forgetting had contributed to differences among diagnostic groups in the ADAMS.

With ADNI data, the availability of individualized parameter estimates allowed us to conduct formal analyses of the ability of dual-retrieval processes to differentiate individuals with different diagnoses by computing sensitivity and specificity statistics. Statistically speaking, the model did a good job of differentiating HC individuals from a-MCI individuals, because sensitivity and specificity were both well above chance, but not of differentiating a-MCI individuals from individuals with AD, because sensitivity was not above chance (although specificity was excellent). Adding genetic data as a further predictor variable failed to increase diagnostic differentiation, relative to dual-retrieval processes alone.

Dual-Process Prediction of Longitudinal Transitions to MCI and AD

There is a need for reliable, theoretically motivated predictors of future conversion to impairment among healthy older adults and of future conversion to dementia among older adults with CIND diagnoses. With respect to future dementia, some neuropsychological tests have been found to be reliable predictors, but predictive relations are weak and variable (for a review, see Tabert et al., 2006). On the latter point, tests of verbal memory are the only ones that have proved to be reliable predictors in even half of extant studies. With respect to genetic predictors of conversion to dementia, early-onset AD is associated with defects on chromosome 21, particularly mutation of the amyloid precursor protein gene, and on chromosomes 1 and 14, which are due to mutations in the presenilin-2 and presenilin-1 genes (Kawas & Katzman, 1999). Combined, however, all three mutations account for only 2% of early-onset AD diagnoses (Pericak-Vance et al., 2000), and in any event, the great preponderance of AD diagnoses are late-onset (after age 65). There, the ε4 allele fares better, with roughly 50% of individuals with probable AD and roughly 30% of individuals with possible AD being carriers as compared to roughly 20% of healthy age-mates, in nationally representative samples (Brainerd et al., 2011). However, recent research suggests that this allele may not predict conversion to AD among individuals with CIND diagnoses (Brainerd et al., 2013).

With respect to conversion to MCI, the predictive picture is weaker still and is complicated by the fact that following MCI diagnosis, more than 20% of individuals usually transition back to HC (Fisher et al., 2011). On the neuropsychological side, the fact that MCI is a relatively new diagnostic category means that a substantial literature on its predictors has not yet had time to accumulate (Summers & Saunders, 2012). On the genetic side, there is no established marker of MCI, and there is an ongoing controversy as to whether ε4 is elevated in MCI as well as AD (Brainerd et al., 2011; Small, Rosnick, Fratiglioni, & Backman, 2004).

Against this backdrop, we examined the ability of dual-retrieval processes to forecast future MCI and future AD in ADAMS subjects in Study 2 and to forecast future a-MCI and probable AD in ADNI subjects in Study 3. With the ADAMS, a finding of general significance is that recollective retrieval, which failed to differentiate the diagnostic groups during Wave A, also failed to predict future AD among MCI subjects and failed to predict either future AD or future MCI among HC subjects. We studied prediction of MCI → AD conversion over a 16–18 month interval and prediction of MCI → AD, HC → MCI, and HC → AD conversion over a 4.5–6 year interval. With respect to MCI → AD, one of the processes that differentiated the MCI group from the HC and AD groups during Wave A was a reliable predictor of conversion after 16–18 months and after 4.5–6 years; namely, reconstructive retrieval. Values of the *R* parameter were smaller for MCI subjects who would convert to AD than for those who would not. Further, the predictive power of reconstructive retrieval was better than the best genetic marker of AD, because the statistical association between MCI subjects' ε4 carrier status and their tendency to progress to AD was not reliable.

With respect to conversions over a 4.5- to 6-year interval, reconstructive retrieval was again a reliable predictor of which MCI subjects would convert to dementia over that interval. Recollective forgetting was also a reliable predictor. With respect to HC subjects, the familiarity judgment parameters predicted which subjects would convert to dementia over the 4.5- to 6-year interval, and the recollective forgetting parameter predicted which subjects would convert to MCI or dementia. Once again, the dual-retrieval model did a better job of forecasting longitudinal transitions than ε4 did. Whereas the model's parameters predicted all of the transition types, ε4 carrier status did not predict HC → AD transitions or MCI → AD transitions. It did predict HC → MCI transitions, however. For this transition group, additional analyses showed that ε4 carrier status was unrelated to model parameters, which meant that the two were independent predictors of conversion to MCI.

Turning to Study 3, once again dual-retrieval processes were reliable predictors of future conversion, both to a-MCI and to probable AD, and prediction was not improved by the addition of genetic data. Although dual-retrieval processes were reliable predictors of both HC → a-MCI and a-MCI → probable AD, the proportion of subjects who made the former transition over a 2-year period was far too small to conduct a formal sensitivity–specificity analysis. However, the proportion of a-MCI subjects who converted to probable AD was adequate for such an analysis. Both sensitivity and specificity were above chance, and overall, the dual-retrieval model was a moderately good predictor of con-

version from a-MCI to probable AD. As in the sensitivity–specificity analysis of diagnostic group differentiation, the genetic data added nothing to dual-retrieval processes' ability to predict future AD: When ε4 carrier status carrier status was added as a further predictor variable, there was no improvement in either sensitivity or specificity.

Summing up the predictive picture for MCI and a-MCI, the process that did the best job of differentiating the MCI group from the HC group during Wave A of the ADAMS, reconstructive retrieval, also predicted future a-MCI → HC and a-MCI → not AD transitions during Waves B, C, and D, and in the ADNI, reconstructive retrieval even predicted the small ($N = 16$; 8%) group of HC → a-MCI transitions. Summarizing the predictive picture for AD and probable AD, reconstructive retrieval also predicted MCI → AD transitions during Waves B, C, and D of the ADAMS, but recollective retrieval predicted a-MCI → probable AD transitions in the ADNI.

Concluding Comments

A large number of findings about dual-retrieval processes in MCI and AD have been reported, and to avoid losing the forest for the trees, we close by highlighting five items of general significance. The most important one, surely, is that mathematical models of memory can extract reliable measurements of underlying retrieval processes from simple instruments that are administered every day in clinics worldwide. It is difficult to overestimate the importance of the fact that massive clinical databases can now be analyzed to pinpoint process loci of diseases. Indeed, such analyses could easily be incorporated as routine features of existing neuropsychological batteries.

The second item is the theoretical importance of the overall picture that emerged for reconstructive retrieval. Those results provide broad support for the hypothesis that the shift from healthy aging to neurocognitive impairment is accompanied by a shift to declines in nonrecollective processes, rather than a continuation of the sparing that characterizes healthy aging. Declines in reconstructive retrieval differentiated all diagnostic groups in both the ADAMS and the ADNI, whereas declines in recollective retrieval usually did not. Declines in reconstructive retrieval also predicted future MCI and future AD in the ADAMS and future a-MCI in the ADNI. Indeed, declines in reconstructive retrieval, always on the learning side and also on the forgetting in the ADAMS, could be fairly said to be the prime memory marker of neurocognitive impairment in these studies.

The third item of general significance, considering the mixed and inconsistent literature on neuropsychological predictors of MCI and AD, is that dual-retrieval processes so consistently predicted future emergence of these diseases. Moreover, when the data allowed a formal sensitivity–specificity analysis to be conducted, sensitivity and specificity values were both well above chance. The fourth item of general significance, considering the ε4 allele's stature as a biomarker of AD and the expense of performing genetic tests on older adults, is that the genetic data added nothing to dual-retrieval processes when it came to predicting future disease. In the ADAMS, ε4 was simply not associated with conversion to AD, whereas dual-retrieval processes were. In the ADNI, ε4 was associated with conversion to AD, but it did not improve the sensitivity or specificity of prediction, relative to dual-retrieval processes alone.

The final item is concerned with healthy rather than impaired older adults, though it illustrates the role of impairment in late-life cognitive declines. In the nationally representative ADAMS sample, when the model was fit to the memory performance of adjacent HC age groups, dual-retrieval processes remained stable between age 70 and the early 90s. Likewise, in the more highly educated ADNI sample, fitting the model to the memory performance of adjacent HC age groups showed that dual-retrieval processes did not vary with age. Thus, although there are substantial declines in episodic memory among healthy individuals prior to age 70, subsequent declines seem to be associated with disease rather than aging.

References

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

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 doi:10.1037/0894-4105.22.2.177

Baudic, S., Dalla Barba, G., Thibaudet, M. C., Smagghe, A., Remy, P., & Traykov, L. (2006). Executive function deficits in early Alzheimer's disease and their relations with episodic memory. *Archives of Clinical Neuropsychology, 21,* 15–21. doi:10.1016/j.acn.2005.07.002

Billingsley, R. L., Smith, M. L., & McAndrews, M. P. (2002). Developmental patterns in priming and familiarity in explicit recollection. *Journal of Experimental Child Psychology, 82,* 251–277. doi:10.1016/S0022-0965(02)00007-3

Braak, H., & Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging, 16,* 271–284. doi:10.1016/0197-4580(95)00021-6

Brainerd, C. J., Aydin, C., & Reyna, V. F. (2012). Development of dual-retrieval processes: Learning, forgetting, and reminiscence. *Journal of Memory and Language, 66,* 763–788. doi:10.1016/j.jml.2011.12.002

Brainerd, C. J., & Reyna, V. F. (2010). Recollective and nonrecollective recall. *Journal of Memory and Language, 63,* 425–445. doi:10.1016/j.jml.2010.05.002

Brainerd, C. J., Reyna, V. F., & Howe, M. L. (2009). Trichotomous processes in early memory development, aging, and neurocognitive impairment: A unified theory. *Psychological Review, 116,* 783–832. doi:10.1037/a0016963

Brainerd, C. J., Reyna, V. F., Petersen, R. C., Smith, G. E., Kenney, A. E., Gross, C. J., . . . Fisher, G. G. (2013). The apolipoprotein E genotype predicts longitudinal transitions to mild cognitive impairment but not to Alzheimer's dementia: Findings from a nationally representative study. *Neuropsychology, 27,* 86–94. doi:10.1037/a0030855

Brainerd, C. J., Reyna, V. F., Petersen, R. C., Smith, G. E., & Taub, E. S. (2011). Is the apolipoprotein E genotype a biomarker for mild cognitive impairment? Findings from a nationally representative study. *Neuropsychology, 25,* 679–689. doi:10.1037/a0024483

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337. doi:10.1016/S0022-5371(66)80040-3

Bugaiska, A., Morson, S., Moulin, C. J. A., & Souchay, C. (2011). Metamemory, recollection and familiarity in Alzheimer's disease. *Revue Neurologique, 167,* 3–13. doi:10.1016/j.neurol.2010.03.001

Creamer, S., & Schmitter-Edgecombe, M. (2010). Narrative comprehension in Alzheimer's disease: Assessing inferences and memory opera-

tions with a think-aloud procedure. *Neuropsychology, 24,* 279–290. doi:10.1037/a0018107

Dalla Barba, G. (1997). Recognition memory and recollective experience in Alzheimer's disease. *Memory, 5,* 657–672. doi:10.1080/741941546

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. doi:10.1017/S0033291703008031

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11,* 379–386. doi:10.1016/j.tics.2007.08.001

Fisher, G. G., Franks, M. M., Plassman, B. L., Brown, S. L., Potter, G. G., Llewellyn, D., . . . Langa, K. M. (2011). Caring for individuals with dementia and cognitive impairment, not dementia: Findings from the Aging, Demographics, and Memory Study. *Journal of the American Geriatrics Society, 59,* 488–494. doi:10.1111/j.1532-5415.2010.03304.x

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method of grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. doi:10.1016/0022-3956(75)90026-6

Gallo, D. A., Cramer, S. J., Wong, J. T., & Bennett, D. A. (2012). Alzheimer's disease can spare local metacognition despite global anosognosia: Revisiting the confidence–accuracy relationship in episodic memory. *Neuropsychologia, 50,* 2356–2364. doi:10.1016/j.neuropsychologia.2012.06.005

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. doi:10.1111/j.1467-8624.2007.01129.x

Gomes, C. F. A. (2013). *On the relation between memory and metamemory in free recall: The effects of list length and word frequency on dual processes* (Unpublished master's thesis). Cornell University.

Gomes, C. F. A., Brainerd, C. J., & Stein, L. M. (2013). Effects of emotional valence and arousal on recollective and nonrecollective recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 663–677. doi:10.1037/a0028578

Harel, B. T., Darby, D., Pietrzak, R. H., Ellis, K. A., Snyder, P. J., & Maruff, P. (2011). Examining the nature of impairment in visual paired associate learning in amnestic mild cognitive impairment. *Neuropsychology, 25,* 752–762. doi:10.1037/a0024237

Health and Retirement Study. (2011). Unpublished dataset, University of Michigan.

Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language, 55,* 495–514. doi:10.1016/j.jml.2006.07.001

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. doi:10.1001/archneur.60.8.1119

Hicks, J. L., & Marsh, R. L. (2002). On predicting the future states of awareness for recognition of unrecallable items. *Memory & Cognition, 30,* 60–66. doi:10.3758/BF03195265

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541. doi:10.1016/0749-596X(91)90025-F

Juster, F. T., & Suzman, R. (1995). An overview of the Health and Retirement Study. *Journal of Human Resources, 30,* S7–S56. doi:10.2307/146277

Kawas, C., & Katzman, R. (1999). Epidemiology of dementia and Alzheimer disease. In R. D. Terry, R. Katzman, K. L. Bick, & S. S. Sisodia (Eds.), *Alzheimer disease* (pp. 95–116). New York, NY: Raven Press.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639. doi:10.1037/0033-295X.100.4.609

Koriat, A. (1995). Dissociating knowing and feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124,* 311–333. doi:10.1037/0096-3445.124.3.311

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1095–1105. doi:10.1037/0278-7393.29.6.1095

Kurilla, B. P., & Westerman, D. L. (2010). Source memory for unidentified stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 398–410. doi:10.1037/a0018279

Langa, K. M., Plassman, B. L., Wallace, R. B., Herzog, A. R., Heeringa, S. G., Ofstedal, M. B., . . . Willis, R. J. (2005). The Aging, Demographics and Memory Study: Study design and methods. *Neuroepidemiology, 25,* 181–191. doi:10.1159/000087448

Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). Oxford, England: Oxford University Press.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57,* 335–384. doi:10.1016/j.cogpsych.2008.02.004

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7,* 263–269. doi:10.1016/j.jalz.2011.03.005

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., VanBelle, G., Fillenbaum, G., . . . Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD): Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39,* 1159–1165. doi:10.1212/WNL.39.9.1159

Nelson, P. T., Braak, H., & Markesbery, W. R. (2009). Cognitive impairment in Alzheimer disease: A complex but coherent relationship. *Journal of Neuropathology and Experimental Neurology, 68,* 1–14. doi:10.1097/NEN.0b013e3181919a48

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin, 49,* 197–237. doi:10.1037/h0055737

Panza, F., D'Introno, A., Colacicco, A. M., Capurso, C., Del Parigi, A., Caselli, R. J., . . . Solfrizzi, V. (2005). Current epidemiology of mild cognitive impairment and other predementia syndromes. *American Journal of Geriatric Psychiatry, 13,* 633–644.

Pericak-Vance, M. A., Grubber, J., Bailey, L. R., Hedges, D., West, S., Santoro, L., . . . Haines, J. L. (2000). Identification of novel genes in late-onset Alzheimer's disease. *Experimental Gerontology, 35,* 1343–1352. doi:10.1016/S0531-5565(00)00196-0

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256,* 183–194. doi:10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., . . . Rocca, W. A. (2010). Prevalence of mild cognitive impairment is higher in men: The Mayo Clinic study of aging. *Neurology, 75,* 889–897. doi:10.1212/WNL.0b013e3181f11d85

Plassman, B. L., Langa, K. M., Fisher, G. G., Heeringa, S. G., Weir, D. R., Ofstedal, M. B., . . . Wallace, R. B. (2007). Prevalence of dementia in the United States: The Aging, Demographics, and Memory Study. *Neuroepidemiology, 29,* 125–132. doi:10.1159/000109998

Plassman, B. L., Langa, K. M., Fisher, G. G., Heeringa, S. G., Weir, D. R., Ofstedal, M. B., . . . Wallace, R. B. (2008). Prevalence of cognitive

impairment without dementia in the United States. *Annals of Internal Medicine, 148,* 427–434. doi:10.7326/0003-4819-148-6-200803180-00005

Plassman, B. L., Langa, K. M., McCammon, R. J., Fisher, G. G., Potter, G. G., Burke, J. R., . . . Wallace, R. B. (2011). Incidence of dementia and cognitive impairment, not dementia in the United States. *Annals of Neurology, 70,* 418–426. doi:10.1002/ana.22362

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 365–377. doi:10.1037/0278-7393.22.2.365

Ranganath, C. (2010). Binding items and contexts: The cognitive neuroscience of episodic memory. *Current Directions in Psychological Science, 19,* 131–137. doi:10.1177/0963721410368805

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single process theories, and recognition memory. *Journal of Experimental Psychology: General, 124,* 352–374. doi:10.1037/0096-3445.124.4.352

Rey, A. (1941). Psychological examination of traumatic encephalopathy. *Archives de Psychologie, 28,* 286–340.

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review, 31,* 180–206.

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, DC: American Psychological Association. doi:10.1037/11587-010

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review, 111,* 588–616. doi:10.1037/0033-295X.111.3.588

Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling-of-knowing. *Canadian Journal of Experimental Psychology, 39,* 467–475. doi:10.1037/h0080074

Serra, L., Bozzali, M., Cercignani, M., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2010). Recollection and familiarity in amnesic mild cognitive impairment. *Neuropsychology, 24,* 316–326. doi:10.1037/a0017654

Small, B. J., Rosnick, C. B., Fratiglioni, L., & Backman, L. (2004). Apolipoprotein E and cognitive performance: A meta-analysis. *Psychology and Aging, 19,* 592–600. doi:10.1037/0882-7974.19.4.592

Storandt, M. (2008). Cognitive deficits in the early stages of Alzheimer's disease. *Current Directions in Psychological Science, 17,* 198–202. doi:10.1111/j.1467-8721.2008.00574.x

Strong, E. K. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20,* 339–372. doi:10.1037/h0072087

Summers, M. J., & Saunders, N. L. J. (2012). Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. *Neuropsychology, 26,* 498–508. doi:10.1037/a0028576

Tabert, M. H., Manley, J. J., Liu, X. H., Pelton, G. H., Rosenblum, S., Jacobs, M., . . . Devanand, D. P. (2006). Neuropsychological prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Archives of General Psychiatry, 63,* 916–924. doi:10.1001/archpsyc.63.8.916

Tse, C. S., Balota, D. A., Moynan, S. C., Duchek, J. M., & Jacoby, L. L. (2010). The utility of placing recollection in opposition to familiarity in early discrimination of healthy aging and very mild dementia of the Alzheimer's type. *Neuropsychology, 24,* 49–67. doi:10.1037/a0014887

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1–12. doi:10.1037/h0080017

Twamley, E. W., Ropacki, S. A. L., & Bondi, M. W. (2006). Neuropsychological and neuroimaging changes in preclinical Alzheimer's disease. *Journal of the International Neuropsychological Society, 12,* 707–735. doi:10.1017/S1355617706060863

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.

Westerberg, C. E., Paller, K. A., Weintraub, S., Mesularn, M. M., Holdstock, J. S., Mayes, A. R., & Reber, P. J. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer's disease. *Neuropsychology, 20,* 193–205. doi:10.1037/0894-4105.20.2.193

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117,* 1025–1054. doi:10.1037/a0020874

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. doi:10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517. doi:10.1006/jmla.2002.2864

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised manual*. Los Angeles, CA: Western Psychological Services.

(*Appendix follows*)

Appendix

Forgetting Model

The data space that was analyzed for the CERAD delayed test was the event sequence $S_2T_2S_3T_3T_4$. (A similar data space, $S_4T_4S_5T_5T_6$, was analyzed for the RAVLT delayed test, and the developments in this section apply to both the CERAD and RAVLT data spaces.) The $S_2T_2S_3T_3$ part is the last two learning trials, and the T_4 part is the delayed recall test. The dual-retrieval model for this event sequence has five memory parameters: *D, R, J,* F_D, and F_R. The first three parameters have the same definitions as in Table 1. F_D is a forgetting parameter that measures loss of the ability to retrieve an item's verbatim trace between T_3 and the delayed test, and F_R is a forgetting parameter that measures loss of the ability to reconstruct an item from traces of partial identifying information between T_3 and the delayed test. Over the three recall tests, there are eight error–success patterns for any item: $C_2C_3C_4$, $C_2C_3E_4$, . . . , $E_2E_3E_4$. The probabilities of these patterns can be expressed as functions of the memory parameters:

$$P(C_2C_3C_4) = D(1 - F_D) + (1 - D)RJ^3(1 - F_R); \quad \text{(A1)}$$

$$P(C_2C_3E_4) = DF_D + (1 - D)RJ^2[F_R + (1 - F_R)(1 - J)]; \quad \text{(A2)}$$

$$P(C_2E_3C_4) = (1 - D)RJ^2(1 - J)(1 - F_R); \quad \text{(A3)}$$

$$P(C_2E_3E_4) = (1 - D)RJ(1 - J)[F_R + (1 - F_R)(1 - J)]; \quad \text{(A4)}$$

$$P(E_2C_3C_4) = (1 - D)(1 - R)RJ^2(1 - F_R) + (1 - D)R(1 - J) D(1 - F_D) + (1 - D)R(1 - J)(1 - D)J^2(1 - F_R) + (1 - D)(1 - R)D(1 - F_R); \quad \text{(A5)}$$

$$P(E_2C_3E_4) = (1 - D)^2(1 - R)RJ[F_R + (1 - F_R)(1 - J)] + (1 - D)(1 - R)DF_D + (1 - D)R(1 - J)DF_D + (1 - D)R(1 - J)(1 - D)J[F_R + (1 - F_R)(1 - J)]; \quad \text{(A6)}$$

$$P(E_2E_3C_4) = (1 - D)^2(1 - R)R(1 - J)(1 - F_R)J + (1 - D)R(1 - J)^2(1 - D)(1 - F_R)J; \quad \text{(A7)}$$

$$P(E_2E_3E_4) = (1 - D)^2(1 - R)^2 + (1 - D)^2(1 - R)R(1 - J)[F_R + (1 - F_R)(1 - J)] + (1 - D)^2R(1 - J)^2[F_R + (1 - F_R)(1 - J)]. \quad \text{(A8)}$$

The likelihood of any sample of data and estimates of the five parameters are obtained by maximizing the following likelihood function:

$$L_5 = P(p_i)^{N(i)}. \quad \text{(A9)}$$

The p_i are the eight expressions on the right sides of Equations A1–A8. Because 5 memory parameters are estimated, the likelihood in A9 is computed with 2 degrees of freedom. A goodness-of-fit test of this model is obtained by computing a likelihood ratio statistic that compares the likelihood in A9 to the likelihood of the same data when all 7 free empirical probabilities are free to vary. That test statistic, which is asymptotically distributed as $\chi^2(2)$, is

$$G^2 = -2\ln[L_5/L_7], \quad \text{(A10)}$$

where L_7 is the likelihood of the data when all empirical probabilities are free to vary.

This test statistic is used to evaluate within- and between-condition hypotheses about differences in parameter values. For between-condition tests, consider an experiment that contains k conditions. As a test of hypotheses about whether a parameter (say, F_D) differs between a pair of conditions, (a) an experimentwise test is computed to determine whether there is global statistical evidence that the parameter differs among the k conditions, and (b) if that test yields a null hypothesis rejection, conditionwise tests are computed to determine whether the parameter differs between specific pairs of conditions. The first test statistic is

$$G^2 = -2\ln\{L'_{i5}/[L_5(1) \times L_5(2), \times \ldots, L_5(k)]\} \quad \text{(A11)}$$

where the denominator contains the values of the numerator of A10 that are computed for the data of each of the k conditions and the numerator contains a single value of the numerator of A10 that is computed for the pooled data of the k conditions under the constraint that the value of the target parameter is the same in all conditions. The G^2 statistic is asymptotically distributed as $\chi^2(2k)$. The second test statistic, for two conditions i and j, which is asymptotically distributed as $\chi^2(2)$, is

$$G^2 = -2\ln[L'_{ij5}/(L_{i5} \times L_{j5})]. \quad \text{(A12)}$$

The numerator is a value of the numerator of A10 that is computed for the pooled data of the two conditions under the constraint that the value of the parameter is the same in those conditions, and the denominator is the two values from the denominator of A11 for these conditions.

For within-condition tests, these tests compare the values of different parameter pairs (say, F_D vs. F_R) within a condition. Such a test stipulates that a relation of equality or inequality holds between the members of the pair. The test statistic is just

$$G^2 = -2\ln[L_4/L_5], \quad \text{(A13)}$$

which is asymptotically distributed as $\chi^2(1)$ because the likelihood in the numerator is estimated with one less degree of freedom than when all five parameters are free to vary.

Received March 19, 2013
Revision received June 13, 2013
Accepted June 17, 2013 ■