Universität Potsdam

Reinhold Kliegl, Ulman Lindenberger

Modeling intrusions and correct recall in episodic memory : adult age differences in encoding of list context

Modeling Intrusions and Correct Recall in Episodic Memory: Adult Age Differences in Encoding of List Context

Reinhold Kliegl and Ulman Lindenberger

A model for correct recall and intrusions in cued recall of word lists is introduced. Intrusions are false responses that were correct in an earlier list. The model assumes 3 exclusive states for memory traces after encoding: with a list tag (i.e., with information about list origin), without list tags, and missing. Across lists, a trace can lose its list tag or its content. For retrieval, an optimal strategy of response selection was assumed. Younger and older laboratory-trained mnemonists participated in 2 experiments in which recall of permutations of a single word list across a single set of cues was held constant with individually adjusted presentation times. With correct recall equated to younger adults, older adults were more susceptible to intrusions. Age differences were restricted to model parameters estimating the probability of generation of list tags. Alternative accounts of age differences in context memory are discussed.

When several lists of words must be memorized successively, previously learned words often interfere with the most recently learned ones. For example, sometimes a false answer is given that was correct in an earlier list. Such errors are called *intrusions* in the present article. We aim to account for intrusions and correct recall by means of a mathematical model. The model differs from previous accounts of list recall (a) in its explicit assumptions about the fate and influence of memory traces from earlier lists and (b) in the simultaneous use of correct recall and intrusion errors for parameter estimation. Theoretically, intrusions are assumed to be the consequence either of failing to encode or of losing information about the context in which a memory trace was generated. In aging research, evidence is accumulating that older adults are doing less well than young ones in retrieving such information (see Light, 1991, for a review). Therefore, we expected an age difference in the susceptibility to intrusion errors. Moreover, we hoped to be able to trace the age difference to theoretically informative parameters of the model.

Memory for Context Information

There are two functions of episodic memory. One of them is to acquire new information and to keep it available for some time. There is, however, also information that is required only for a short span of time or that is highly variable. In this regard, a good episodic memory is characterized by being up to date (Bellezza, 1982; Bjork, 1978; Bjork & Landauer, 1978). For example, we only have to remember our current phone number or the current parking spot of our car. Context information (e.g., temporal and spatial cues) related to a target object is critical for keeping memory up to date. There are numerous studies in cognitive aging which show that older adults have problems in remembering context information related to, for example, presentation modality (Lehman & Mellinger, 1984), sex of voice (Kausler & Puckett, 1981), color of target information (Park & Puglisi, 1985), and whether facts were learned in an experiment or were already known prior to it (McIntyre & Craik, 1987). Older adults' tendency to generate more text-extraneous elaborations than younger adults in recall of prose may also be related to their poorer memory for context (Hultsch & Dixon, 1983; Tun, 1989). Similarly, older adults have problems with temporal and spatial markers. For example, they do less well than young adults in picking which of two items was presented last (McCormack, 1982) or in identifying the list of origin of an item (McCormack, 1984).

A comprehensive review of this research was provided by Light (1991). Light pointed out that memory for target information and memory for context information are usually conjointly observed to be worse in older than in younger adults. Therefore, it remains unclear whether age-related impairment of memory for context is simply part of a general age-related decline of memory or whether memory for context is disproportionately affected. One strategy to overcome this problem is to equate age groups in memory for target information (or even to induce better memory for targets in older adults) before comparing their memory for context. In two recent studies these conditions happened to be met. Older adults still scored significantly lower in memory for the source of facts (Schacter, Kaszniak, Kihlstrom, & Valdiserri, 1991) or contextual detail (Mäntylä & Bäckman, 1992). (For reasons not quite clear this was true only for one

of several comparisons in Schacter et al. and primarily for unexpected items in Mäntylä & Bäckman.) Thus there is some support for the specificity of the age deficit. An additional, related issue that has not been addressed so far is how differences in encoding, storing, or retrieving context information affect memory for target information. In the experiments to be reported, we experimentally controlled age differences in recall of target information by adjusting presentation times at the individual level. In addition, we present a model that traces the consequences either of failing to encode or of losing contextual information for the recall of targets.

In earlier memory training experiments with the method-of-loci mnemonic, informal observations suggested that participants, especially older adults, frequently recalled words that they had learned in earlier lists, even if these words were not part of the current list or had appeared in combination with a different locus of the mental map. We call these errors *intrusions*. Because permutations of a single word list across a single list of cues had to be memorized in the following experiments, this concept has a more restricted meaning than usual: Intrusions are false responses that originated at the same location cue but were in a previous list. Thus, intrusions refer to word–location combinations. Table 1 lists a verbal protocol of how intrusions might arise. The participant attempted to recall the word associated with one particular landmark (i.e., church) in the third list of an experimental session. She started with the remark that she recalled an elephant stomping through the church but had no recollection whether the word *elephant* had been presented in the current or in an earlier list. Consequently she searched for other words that had been presented at this landmark. She retrieved the word *lemon* along with the list in which it occurred and the word *mother* without a list tag. List tags are defined as contextual information about list membership of a given trace, that is, information about the list in which the trace was generated during an experimental session. Depending on whether the word was presented in the current list, responding with *mother* or *elephant* could lead either to a correct response or to an intrusion.

If the participant did not recall *mother* (i.e., Illustration 3b in Table 1), it would be best to pick *elephant* as an answer. As she still remembers the image of the first list, the probability that *elephant* is correct is still .50. Obviously, the probability of recall success in this experiment depended on the quality of both earlier and current memory traces. Aside from the content of the trace, the availability of information about the temporal context of the encoding situation (i.e., the list of origin) increases correct recall.

Hypotheses Compatible With Age-Differential Susceptibility to Intrusions

Relatively little attention has been paid to intrusion errors in past research on list learning or on list recall. One reason was that intrusion errors appeared to be uncorrelated with level of recall (Keppel & Rauch, 1966). Second, there were often too few intrusions for substantive analyses (Keppel, 1968). As far as we know, there was only one study that reported an age-related intrusion effect in list learning (i.e., Hartley & Walsh, 1980). In this study, however, age differences in level of recall complicated the interpretation of the intrusion effect.

The relative lack of age-differential evidence for intrusion errors is somewhat puzzling because greater susceptibility to intrusion errors is compatible with, if not predicted by, various accounts of cognitive aging. Two hypotheses that lead to this prediction come to mind. First, older adults may have greater difficulty in encoding or integrating contextual information than younger adults (e.g., Burke & Light, 1981; Hashtroudi, Johnson, & Chrosniak, 1989; McIntyre & Craik, 1987; Simon, 1979). List tags are one example of contextual information. Second, older adults may forget contextual information faster than younger adults. Although in general there is little evidence for age-related differences in forgetting rates (cf. Kausler, 1982; Wickelgren, 1975), the case may be different for such contextual features.

Table 1
Illustration of Intrusions in Cued-Recognition Task

Illustration	Comment
1. "I see an *elephant* stomping through the *church* but I have no idea whether this occurred in this or the last list."	List tag was lost for *elephant at church*.
2. "Ah, now I also remember the priest in the *church* eating a *lemon*. This happened in the first list of today."	List tag was available for *lemon at church*.
3a. "Hm, I think my *mother* was baptized in the *church*. I am unsure though when I formed this image. I guess it must be either an *elephant* or my *mother*."	List tag was lost for *mother at church*.
3b. "I have no idea what the other word was that I had at the *church* today."	A trace (e.g., *mother at church*) was lost completely.

Formal Development of a Mathematical Model

In this section, we develop a formal representation of the processes described in Table 1. This model distinguishes between generation, storage–forgetting, and retrieval–selection of memory traces.

Generation of Traces

Encoding leads to memory traces with a list tag with probability a, to traces without a list tag with probability b, or to no trace at all with probability $1-a-b$. Illustration 2 in Table 1 represents a trace with a list tag; Illustrations 1 and 3a are examples of traces without tags. Of course, we do not know whether they were never generated or whether they were lost. For example, construction of mental images may fail because no suitable relation could be found between word and landmark within the available presentation time.

Trace Transitions: A Two-Stage Markov Process

Traces can change in the course of an experimental session. A trace may lose its list tag and change to a trace without a tag; a trace may also be lost completely or no longer be accessible. These changes can be represented as a two-stage Markov process. Table 2 summarizes the transition probabilities. Initial states are listed in rows and destination states in columns. Thus, the tag of a trace stays intact with probability p; the content of a trace stays intact with probability q. The entries of the transition matrix follow from the assumptions (a) that lost traces and lost tags can not be recovered, (b) that transition probabilities are invariant across lists, and (c) that losing trace content and losing a trace tag are statistically independent events.

Under the assumption that the transition matrix is applied recursively to available traces each time a new list of words is encoded, Table 3 represents the microgenesis of a field of traces across four lists in an experimental session. With each new list, new traces are introduced into the field of traces ("Current list" column). Thus, time in the session flows from top to bottom. Traces of earlier lists change in accordance with the transition probabilities of Table 2. Traces that lose their tags are added to the pool of traces without tags. As more lists are added, the set of available responses changes. These changes are reflected in the expansion across columns as a function of list number.

Table 2

Transition Probabilities for States of a Memory Trace Between Lists

State of trace in list n	State of trace in list $n + 1$		
	2	1	0
Trace with tag (2)	p	$q(1-p)$	$(1-p)(1-q)$
Trace without tag (1)	0	q	$1-q$
No trace (0)	0	0	1

Note. p = probability that the tag of a trace stays intact. q = probability that the content of a trace stays intact.

The entries associated with lagged lists (i.e., with the past of the system) are generated by the following two recursive functions for traces with tags and traces without tags, respectively:

$$a_{ml} = a_{(m-1)(l-1)}p \tag{1}$$

$$b_{ml} = qb_{(m-1)(l-1)} + (1-p)a_{(m-1)(l-1)}, \tag{2}$$

for $m > 1$ and $l > 0$; m refers to list number and l indicates how many lists back the trace originated, that is, the lag of the list of intrusion origin. We make the assumption—at least initially—that the probability with which traces of a certain type are generated does not depend on the list in the experimental session. For the probabilities of Table 3, this assumption implies $a_{10} = a_{20} = a_{30} = a_{40} (= a)$ and $b_{10} = b_{20} = b_{30} = b_{40} (= b)$. Thus, the probabilities of Table 3 can be generated with four free parameters (i.e., a, b, p, q).

Cognitive Control Structure for Response Selection

The third model component formalizes the process of a response selection that maximizes probability of correct recall. Recall is constrained by the state of the memory system prior to response selection. Theoretically possible constellations of states of memory traces in the second list of an experimental session and the probability of different types of responses that make optimal use of the available information are given in Table 4 for one landmark of the mental map. On the basis of the assumption that traces are available in one of three mutually exclusive states, the number of combinatorially possible constellations of memory states for any single landmark cue can be expressed as 3^n, where n is the number of lists memorized. The left part of Table 4 contains the nine constellations of possible states of traces. To each of these constellations there corresponds a probability with which a certain type of response will be given; they are listed in the right part of Table 4. After two lists, we distinguish between correct responses, intrusions from the first list, and other responses; the probabilities associated with a constellation add up to 1.0. Theoretically possible constellations and associated probabilities for recall of the third list are shown in Table 5.

As mentioned previously, the assignment of probabilities to trace constellations assumes a response strategy that makes use of all the information available. For example, if there are two traces without tags (see Constellation 5 in Table 4), it is best to respond with one of them. In the long run, probabilities are .50 for a correct response and .50 for an intrusion response. Given the available information, 50% is the best one can do. If only one trace without a tag is available (i.e., in Constellation 6 or in Constellation 8), then this trace should determine the response. For traces generated in the current list (i.e., Constellation 6), it always leads to a correct response; for traces generated in the first list (i.e., in Constellation 8), it always leads to an

Table 3
Microgenesis of a Field of Memory Traces in an Experimental Session With Four Lists

List	State of trace	Current list lag-0	List of lag-1	List of lag-2	List of lag-3
1	2	a_{10}			
	1	b_{10}			
	0	$1 - a_{10} - b_{10}$			
2	2	a_{20}	$a_{21} = a_{10}p$		
	1	b_{20}	$b_{21} = b_{10}q + a_{10}q(1-p)$		
	0	$1 - a_{20} - b_{20}$	$1 - a_{21} - b_{21}$		
3	2	a_{30}	$a_{31} = a_{20}p$	$a_{32} = a_{21}p$	
	1	b_{30}	$b_{31} = b_{20}q + a_{20}q(1-p)$	$b_{32} = b_{21}q + a_{21}q(1-p)$	
	0	$1 - a_{30} - b_{30}$	$1 - a_{31} - b_{31}$	$1 - a_{32} - b_{32}$	
4	2	a_{40}	$a_{41} = a_{30}p$	$a_{42} = a_{31}p$	$a_{43} = a_{32}p$
	1	b_{40}	$b_{41} = b_{30}q + a_{30}q(1-p)$	$b_{42} = b_{31}q + a_{31}q(1-p)$	$b_{43} = b_{32}q + a_{32}q(1-p)$
	0	$1 - a_{40} - b_{40}$	$1 - a_{41} - b_{41}$	$1 - a_{42} - b_{42}$	$1 - a_{43} - b_{43}$

Note. Rows (each set of 3 lines) show the expansion of the field of traces due to the processing of word lists. States: 2 = trace with tag; 1 = trace without tag; 0 = no trace.

intrusion. Again, in the long run the number of correct responses is maximized. The optimal response strategy can be summarized in three rules:

Rule 1: If the constellation contains a trace with a tag (2) from the current list, then take this trace as the response.
Rule 2: If Rule 1 does not apply and if there are traces without list tags, then take one of them at random.
Rule 3: If neither Rule 1 nor Rule 2 applies, then respond at random with a nontagged word.

The probability of correct responses or of various intrusions that are due to Rule 3 was set to zero for the following computations; it was assumed to lead to responses classified as "other." In reality the probability of a correct response or of a specific type of intrusion that is due to Rule 3 depends on the number of response alternatives; with a list length of 30 it is 1/30 = .033. In computations of response probabilities for correct recall and intrusions, however, the corresponding expressions would enter as factors in products of probabilities (see Table 6), and, consequently, would yield values very close to zero.

Theoretical Probabilities for Constellations of States of Memory Traces

Tables 4 and 5 (and tables constructed accordingly for other lists) contain response probabilities for each theoretically possible constellation of states of memory traces. The probability of each of these constellations in an experiment can be computed from the entries of Table 3. The expected probability of a specific constellation, then, is simply the product of the theoretical probabilities (taken from Table 3) corresponding to the states of memory traces in the constellation.

For example, the probability of Constellation 5 in Table 4 ("1 1") is the product of probabilities of traces without tags of the current (second) list (b_{20}) and of traces without tags of the lag-1 list (b_{21}), that is:

$$\text{prob ("1 1")} = b_{20}b_{21}. \tag{3}$$

The probability for Constellation 16 ("1 0 2") in Table 5 is

$$\text{prob ("1 0 2")} = b_{30}(1 - a_{31} - b_{31})a_{32}. \tag{4}$$

Table 4
Combinatorial Constellations of Trace States for a Cue and Associated Probabilities for Response Categories After Two Lists

	Constellations of trace states		Probability of response category		
No.	State in current list	State in list of lag-1	Correct	Lag-1 intrusion	Other
1	2	2	1	0	0
2	2	1	1	0	0
3	2	0	1	0	0
4	1	2	1	0	0
5	1	1	.50	.50	0
6	1	0	1	0	0
7	0	2	0	0	1
8	0	1	0	1	0
9	0	0	0	0	1

Note. States: 2 = with tag; 1 = without tag; 0 = no trace.

Table 5

Combinatorial Constellations of Trace States for a Cue and Associated Probabilities for Response Categories After Three Lists

	Constellations of trace states			Probability of response category			
No.	Lag-0	Lag-1	Lag-2	Correct	Lag-1 correct intrusion	Lag-2 intrusion	Other
1	2	2	2	1	0	0	0
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
9	2	0	0	1	0	0	0
10	1	2	2	1	0	0	0
11	1	2	1	.50	0	.50	0
12	1	2	0	1	0	0	0
13	1	1	2	.50	.50	0	0
14	1	1	1	.33	.33	.33	0
15	1	1	0	.50	.50	0	0
16	1	0	2	1	0	0	0
17	1	0	1	.50	0	.50	0
18	1	0	0	1	0	0	0
19	0	2	2	0	0	0	1
20	0	2	1	0	0	1	0
21	0	2	0	0	0	0	1
22	0	1	2	0	1	0	0
23	0	1	1	0	.50	.50	0
24	0	1	0	0	1	0	0
25	0	0	2	0	0	0	1
26	0	0	1	0	0	1	0
27	0	0	0	0	0	0	1

Notes. States: 2 = tag; 1 = without tag; 0 = no trace.

Accordingly, probabilities can be computed for all possible constellations of states of memory traces in an experimental session as a function of the four theoretical parameters.

Theoretical Probabilities for Response Categories

In the next step, the expected probabilities for the various types of responses are derived. They depend on the expected probabilities of constellations of memory states and on the probabilities implied by the optimal response strategy. Specifically, the probability of a certain type of response in the context of a specific constellation is determined as the product of the probability for the constellation and the corresponding probability of the response. The theoretical probability for a certain type of response, finally, is simply the sum of those probabilities across constellations of memory states.

For example, with two lists the joint probability of an intrusion and Constellation 5 is

$$\text{prob ("1 1" \& intrusion)} = .50 \, b_{20}b_{21}. \tag{5}$$

Intrusions also occur in Constellation 8:

$$\text{prob ("0 1" \& intrusion)} = 1.0 \, (1 - a_{20} - b_{20})b_{21} \tag{6}$$

For the two-list case, the probability of an intrusion is zero for all other constellations. Therefore, the probability of a lag-1 intrusion in the second list of an experimental session is computed as the sum of Equations 5 and 6:

$$\text{prob (intrusion in List 2)} = b_{21} - a_{20}b_{21} - (b_{20}b_{21})/2 \tag{7}$$

Analogously, probabilities were computed for each response category. Specifically, with four lists, there were four scores of correct recall, three scores of lag-1 intrusions, two scores of lag-2 intrusions, and one score of a lag-3 intrusion; probabilities for "other" responses are not free to vary because within each list probabilities sum to one. In general, the number of nondetermined dependent variables is $n(n + 1)/2$, where n is the number of lists in the experimental session. Computational formulas for the 14 response categories as a function of model parameters are listed in Table 6.

Experiment 1

Modeling of intrusions requires an experimental paradigm that maximizes the frequency of their occurrence. In addition, for the investigation of age differences, younger and older adults should not differ in their mean level of correct recall to minimize chances that a higher number of intrusions among older adults is just a side effect of less accurate performance. Therefore, in the first experiment the following four design characteristics were implemented: (a) maximization of proactive interference, (b) generation of stable memory traces, (c) criterion-referenced testing, and (d) instruction in an optimal guessing strategy.

Maximizing Proactive Interference

Proactive interference refers to the negative effects of earlier on later learning (Müller & Pilzecker, 1990). Memory

Table 6
Theoretical Probabilities for 14 Response Categories as a Function of Model Parameters

Response Category	Computational formula
	List 1
1	$p(\text{correct}) = a_{10} + b_{10}$;
2	$p(\text{other}) = 1 - a_{10} - b_{10}$;
	List 2
3	$p(\text{correct}) = a_{20} + b_{20} - (b_{20}b_{21})/2$;
4	$p(\text{lag-1 intrusion}) = b_{21} - a_{20}b_{21} - (b_{20}b_{21})/2$;
5	$p(\text{other}) = 1 - a_{20} - b_{20} - b_{21} + a_{20}b_{21} + b_{20}b_{21}$;
	List 3
6	$p(\text{correct}) = a_{30} + b_{30} - (b_{30}b_{31})/2 - (b_{30}b_{32})/2 + (b_{30}b_{31}b_{32})/3$;
7	$p(\text{lag-1 intrusion}) = b_{31} - a_{30}b_{31} - (b_{30}b_{31})/2 - (b_{31}b_{32})/2 + (a_{30}b_{31}b_{32})/2 + (b_{30}b_{31}b_{32})/3$;
8	$p(\text{lag-2 intrusion}) = b_{32} - a_{30}b_{32} - (b_{30}b_{32})/2 - (b_{31}b_{32})/2 + (a_{30}b_{31}b_{32})/2 + (b_{30}b_{31}b_{32})/3$;
9	$p(\text{other}) = 1 - a_{30} - b_{30} - b_{31} + a_{30}b_{31} + b_{30}b_{31} - b_{32} + a_{30}b_{32} + b_{30}b_{32} + b_{31}b_{32}$ $- a_{30}b_{31}b_{32} - b_{30}b_{31}b_{32}$;
	List 4
10	$p(\text{correct}) = a_{40} + b_{40} - (b_{40}b_{41})/2 - (b_{40}b_{42})/2 + (b_{40}b_{41}b_{42})/3 - (b_{40}b_{43})/2 + (b_{40}b_{41}b_{43})/3$ $+ (b_{40}b_{42}b_{43})/3 - (b_{40}b_{41}b_{42}b_{43})/4$;
11	$p(\text{lag-1 intrusion}) = b_{41} - a_{40}b_{41} - (b_{40}b_{41})/2 - (b_{41}b_{42})/2 + (a_{40}b_{41}b_{42})/2 + (b_{40}b_{41}b_{42})/3$ $- (b_{41}b_{43})/2 + (a_{40}b_{41}b_{43}/2) + (b_{41}b_{42}b_{43})/3 - (a_{40}b_{41}b_{42}b_{43})/3$ $- (b_{40}b_{41}b_{42}b_{43})/4$;
12	$p(\text{lag-2 intrusion}) = b_{42} - a_{40}b_{42} - (b_{40}b_{42})/2 - (b_{41}b_{42})/2 + (a_{40}b_{41}b_{42})/2 + (b_{40}b_{41}b_{42})/3$ $- (b_{42}b_{43})/2 + (a_{40}b_{42}b_{43})/2 + (b_{40}b_{42}b_{43})/3 + (b_{41}b_{42}b_{43})/3 - (a_{40}b_{41}b_{42}b_{43})/3$ $- (b_{40}b_{41}b_{42}b_{43})/4$;
13	$p(\text{lag-3 intrusion}) = b_{43} - a_{40}b_{43} - (b_{40}b_{43})/2 - (b_{41}b_{43})/2 + (a_{40}b_{41}b_{43})/2 + (b_{40}b_{41}b_{43})/3$ $- (b_{42}b_{43})/2 + (a_{40}b_{42}b_{43})/2 + (b_{40}b_{42}b_{43})/3 + (b_{41}b_{42}b_{43})/3 - (a_{40}b_{41}b_{42}b_{43})/3$ $- (b_{40}b_{41}b_{42}b_{43})/4$;
14	$p(\text{other}) = 1 - a_{40} - b_{40} - b_{41} + a_{40}b_{41} + b_{40}b_{41} - b_{42} + a_{40}b_{42} + b_{40}b_{42} + b_{41}b_{42} - a_{40}b_{41}b_{42}$ $- b_{40}b_{41}b_{42} - b_{43} + a_{40}b_{43} + b_{40}b_{43} + b_{41}b_{43} - a_{40}b_{41}b_{43} - b_{40}b_{41}b_{43}$ $+ b_{42}b_{43} - a_{40}b_{42}b_{43} - b_{40}b_{42}b_{43} - b_{41}b_{42}b_{43} + a_{40}b_{41}b_{42}b_{43} + b_{40}b_{41}b_{42}b_{43}$;

Note. For definition of coefficients, see Table 3. Formulas were computed with *Mathematica* (Wolfram, 1988).

traces without list tags from earlier lists are one potential source of proactive interference. Intrusions should be more frequent when the confusability of word lists is high. In the (A–B, A–Br) paradigm of verbal learning, the same set of responses is paired with a single set of stimuli with different assignments of responses to stimuli across lists. Because the experimental material is the same in all lists, only some form of time tagging or list tagging can help to discriminate between memory traces that were generated in successive lists. In our study, 30 Berlin landmarks served as stimuli (cues), and different orders of a single set of 30 concrete nouns served as the to-be-remembered material. As the words to be recalled were known to participants, the experiment simulated an updating situation; the task was to know with which location a specific word was paired in the current list.

Stable Memory Traces

The theoretical specification implies that one necessary condition for intrusions is the availability of memory traces without list tags from earlier lists. Thus, the content of memory traces, but not the list tags, needs to have stability across lists. Mnemonic techniques represent one way to generate stable memory traces with imaginable material. In the current study concrete nouns were used as material to be remem-

bered. Participants were experienced laboratory "mnemonists"; prior to this study they had participated in a training study that was based on the method of loci (Kliegl, Smith, & Baltes, 1989).

Criterion-Referenced Testing

A second necessary condition for intrusions is a less-than-perfect availability of traces from the current list. Therefore, the generation of powerful mental images must be constrained to an intermediate level of recall success. If recall is too low, memory traces may be too weak to interfere in a later list. If recall is too high, then memory traces may not lose list tags fast or often enough, and the pool of potential intrusions may be too small. To achieve a medium level of correct recall, we dynamically adjusted presentation time contingent on the individual participant's level of recall in the previous list. Items were presented faster or slower depending on whether 50% of the words in a given list had been recalled. This testing strategy also served to equate younger and older adults in overall recall. Age differences in basic memory ability should be reflected in the presentation time needed to achieve 50%. Most important, given equal recall, it is possible to investigate age-differential susceptibility to different types of er-

rors. We expected older adults to commit more intrusion errors than younger adults.

Response Format and Instruction in Optimal Guessing Strategy

The response selection component of the model introduced earlier presumes that participants are rational about choosing their answers. Moreover, younger and older adults should not differ in this aspect of task performance. One method to guard against response biases is to use a forced-choice format; participants were asked to select one of the 30 words, even if they were uncertain. A second method for guarding against response biases is to instruct participants in a guessing strategy that would maximize their probability of correct recall within the confines of this experiment.

In summary, we expected large age differences in the amount of presentation time needed to achieve 50% recall in this (A–B, A–Br1, A–Br2, ...) paradigm. Of central concern was the demonstration of a higher degree of intrusions in older than younger adults for comparable levels of correct recall. In a second step, the data of this experiment were used to estimate parameters of the formal model introduced earlier.

Method

Participants

In the Kliegl et al. (1989) study, 18 younger and 19 older adults had participated. Of this original sample, 3 older adults did not adapt to a 50% level of recall within three sessions; 1 older adult was not available for this experiment. Therefore, data from only 15 older adults were used for the analyses. Younger adults (7 women, mean age = 23.9 years, range = 19–29 years) were university students specializing in various fields of study. They all had completed 13 years of school education plus 1 to 6 years of university studies. Older adults (10 women, mean age = 71.9 years, range = 65–80 years) had on average completed 12.3 years of education (SD = 3.0). All participants were volunteers and physically able to come to the laboratory. Reported subjective health was rated as above average on a 5-point self-report scale relative to that of age peers and did not differ between groups (younger: mean rating = 4.2, SD = 0.7; older: M = 4.1, SD = 0.7).

The typical age-related pattern of psychometric intelligence was obtained on the Hamburg-Wechsler Intelligence Test (HAWIE): Age groups were equal in verbal scores (younger: mean score = 63.3, SD = 3.8; older: M = 64.5, SD = 4.9) but older adults scored significantly lower than younger adults on performance scores (younger: M = 62.2, SD = 6.4; older: M = 54.9, SD = 7.6), $t(31) = 3.0, p < .01$.

Apparatus and Material

Apple IIe computers were used for display of lists and response collection. Words and landmark cues were presented in standard 40-column Apple font in the center of the computer monitor. The landmark cues selected for the present study consisted of a set of 30 well-known landmarks in Berlin (e.g., Memorial Church, Television Tower). These landmarks had been learned according to the order in which they would be visited on a sightseeing trip of the city. Photographs of different aspects of each landmark had been used

during initial instruction. Thirty to-be-remembered nouns were selected from a pool of nouns for which the German equivalent of Paivio, Yuille, and Madigan's (1968) nouns for concreteness and imagery were available (Baschek, Bredenkamp, Oehrle, & Wippich, 1977). The selected nouns represented the most concrete and imaginable items. Five sets of 30 lists each comprising the set of 30 nouns were constructed. Order of nouns within lists was random with the constraint that, in each set, each of the 30 nouns appeared once with each landmark (i.e., at each serial position) across the 30 lists. Assignment of sets to participants was counterbalanced across age groups. Lists 1–12 (3 sessions × 4 lists) were used in this experiment; Lists 13–24 were used in Experiment 2. An index card with these nouns and associated two-digit codes was placed below the monitor throughout the experiment.

Instruction in the Method of Loci

Participants had been instructed in the method of loci in a previous experiment. Instructions followed the recommendations by Bower (1970), which emphasize the generation of interactive, dynamic images or thoughts that connect landmarks and to-be-remembered nouns (see Kliegl et al., 1989; Kliegl, Smith, & Baltes, 1990, for detailed information).

Instruction in Response Selection

At the beginning of this experiment, participants received instruction in optimal response selection for the current task. Specifically, they were told that a particular noun would occur only once in combination with a landmark cue. Therefore, if they knew that a word had appeared with a landmark in a previous list, then they would increase their chances of being correct by choosing an alternative item. Also, they were told that each word appeared only once in a given list. Therefore, if they knew that a word had occurred at a different landmark, then an alternative choice would increase their chances of being correct.

Design and Procedure

Three sessions were scheduled. At the beginning of each session, participants practiced entering two-digit codes on the keypad in response to nouns that were displayed on the monitor. Errors and response latencies were recorded. Then, four lists comprising different permutations of the set of 30 nouns were presented for cued recall.

In the first session, performance was calibrated to a 50% recall level for each individual by dynamic adjustment of presentation time. All older adults started with 5 s per word; all younger adults started with 3 s per word. Criterion for decrease or increase of presentation time on the next list was whether the 50% mark was passed on the present list. For rates below 3 s, increase and decrease occurred in steps of 0.5 s; for times slower than 3 s, a 1-s interval was used. The dynamic adjustment was also used in the second and third sessions.

Participants initiated word-by-word presentation of a list of 30 nouns with a button press on the computer keyboard. Above each stimulus word, a verbal label denoting the Berlin landmark of corresponding serial position was displayed simultaneously. The same order of landmarks was used on all lists.

After the presentation, landmark cues were presented in a randomized order to discourage chaining of words during encoding. In response to the landmark prompt, participants entered the two-digit

code corresponding to the noun recalled. Responses had to be entered within 40 s. Response latencies to the first keypress were collected.

Estimation of Model Parameters

Parameters a, b, p, and q were determined with respect to the following maximum-likelihood statistic for young and old adults:

$$G^2 = -2 * \sum_{i=1}^{14} \ln(x_i/(N_o * N_s)) * \mathrm{prob}_i(a, b, p, q) \qquad (8)$$

where $N_o = 60$ (i.e., words per list summed across two sessions), N_s is the number of subjects in the respective age groups, x_i is the observed frequency of response i, and prob_i (a, b, p, q) is the expected probability of response i as a function of model parameters (with $i = 1$–14, see Table 6). There were six degrees of freedom for each age group because four parameters were used to reconstruct 10 nonredundant observations. The model can be cast as a List (4) × Type of Response (5) contingency table. Response categories are "correct," "lag-1," "lag-2," "lag-3," and "other." There are six cells with structural zeros in this table (e.g., "lag-1," "lag-2," and "lag-3" for List 1, "lag-2," and "lag-3" for List 2, and "lag-3" for List 3). The four cells of "other" responses are not free to vary because there was a fixed marginal for each list (i.e., 60 words per list and person). The theoretical frequencies of the remaining 14 cells are specified by Equations 1–14 in Table 6. The 240 words (i.e., 30 words × 4 lists × 2 sessions), then, were a random sample from a pool of 900 word–landmark combinations stratified according to type of word and landmark cue; data from the first session were used for calibration. The total number of observations was 7,920 (i.e., 240 words × 33 persons).

The function was determined with the nonlinear regression module CNLR in SPSS-X (SPSS, Inc., 1988). To estimate parameters as probabilities (i.e., for a range between 0 and 1), we used the following mathematically equivalent parameterization for construction of memory traces: a, $b(1-a)$, and $(1-a)(1-b)$ for traces with list tag, without list tag, and failed construction, respectively.

Results

The experiment allowed a description of task performance and of related age differences in terms of presentation time, correct recall, intrusions, and other false responses. For the following analyses, data were averaged across Session 2 and Session 3 (i.e., after calibration to 50% recall). For all analyses, the alpha level was set at 5%. All significant effects were also obtained when sessions were analyzed separately. After that, the results pertaining to the estimation of parameters of the formal model are reported.

Analyses of Performance Indicators

Presentation time. Presentation time for a list of words varied across lists (and across participants) as a function of percentage correct in the previous list. Thus, in one sense it was an experimental condition, but in another sense, it could also be treated as a dependent variable. On average, younger adults needed 1.8 s and older adults required 5.3 s per word to maintain a 50% criterion of correct recall. This difference between age groups was significant, $F(1, 31) = 65.3$, $MS_e = 6.05$. Despite the quasi-random, criterion-referenced se-

lection of presentation times after each list, there was also a significant main effect of list, $F(3, 93) = 5.2$, $MS_e = 0.14$. As shown in Figure 1 (left panel), presentation times for the first and last list of an experimental session were slower than presentation times for the two middle lists. The interaction between list and age group was not significant, $F(3, 93) = 1.3$.

Correct recall. With dynamically adjusted presentation times there were no significant age differences in number of words recalled (younger: 15.0; older: 14.4), $F(1, 31) = 0.9$. As shown in Figure 1 (middle panel), recall differed by list of experimental session, $F(3, 93) = 9.3$, $MS_e = 16.5$. The first list was recalled significantly better than the remaining lists. There was no interaction of this effect with age group, $F(3, 93) = 1.4$. Thus, the individual adjustment of presentation time was successful in establishing comparable levels of recall in younger and older adults. There remained, however, significant differences between lists, which, at least in part, were probably due to longer presentation times in the first than the second list.

Intrusions per list. The number of intrusions from lists of the current session is shown as a function of list and age group in the right panel of Figure 1. There was a strong age difference in all four lists of the experimental session, $F(1, 31) = 26.3$, $MS_e = 4.7$. Also, the increase in intrusions across lists was significant, $F(2, 62) = 21.2$, $MS_e = 1.9$, and did not interact with age group, $F(2, 62) = 0.6$. The experiment demonstrated a clear age-differential susceptibility to intrusions with equivalent levels of recall.

Intrusions per list of origin. The increase in intrusions across lists reflects the increase in the number of intrusion opportunities. The number of lists constituting the pool for potential intrusions was one less than the list number (i.e., one for List 2, two for List 3, three for List 4). Computing the average number of intrusions per list of intrusion origin controls for the differences in intrusion opportunities across lists. These percentages were used to estimate model parameters in the next section. The observed and estimated

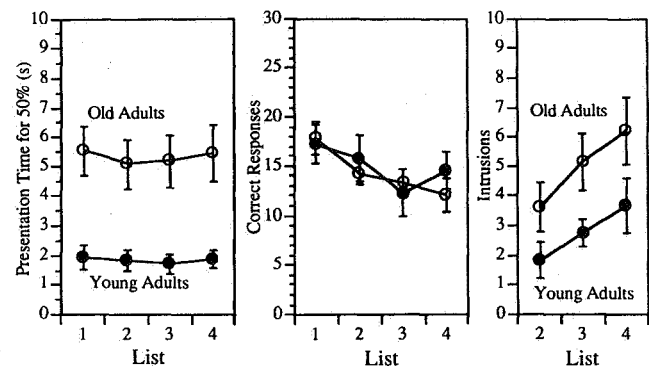*Figure 1.* Criterion-referenced presentation time (left panel), correct recall (middle panel), and intrusions (right panel) as a function of list in experimental session and age group in Experiment 1. (Continuous adjustment of presentation times eliminated age differences in recall and revealed age-differential susceptibility to intrusions. Data were averaged across two experimental sessions.)

scores that were based on this criterion are displayed in Figure 2 (right panel). The estimated scores (i.e., the open symbols in Figure 2) indicate the predictions of the model for the current set of data. In the following analysis, a measure of between-sessions intrusions (i.e., intrusions from Session 1 into 2 and from Session 2 into 3) was included as well. A list of origin (7) by age group (2) repeated measures analysis of variance (ANOVA) resulted in a significant effect of age group, $F(1, 31) = 27.2$, $MS_e = 2.0$, a significant effect of list of intrusion origin, $F(6, 186) = 14.2$, $MS_e = 1.0$, and a significant interaction between the two factors, $F(6, 186) = 2.5$.

Intrusions by lags and lists: contrasts. Six contrasts were specified a priori on the list factor: The first three tested differences between successive lists in the session (i.e., averaging across lists of intrusion origin within List 3 and within List 4). The decline in number of intrusions per list of intrusion origin was significant from List 2 to List 3, List 3 to List 4, and between List 4 and between-sessions intrusions, $F(1, 31) = 27.1$, 15.3, and 48.9 with corresponding $MS_e = 1.3$, 0.9, and 0.4, respectively. This decline is consistent with the model. Of the interactions of these contrasts with age group, the first and third one were significant, $F(1, 31) = 5.3$, $MS_e = 1.3$, and $F(1, 31) = 7.7$, $MS_e = 0.4$, respectively. The first interaction showed that the age difference in intrusions per list of origin was larger in List 2 than in List 3. The second interaction was due to a significant age difference for intrusions per list of origin in List 4 (younger: 1.2; older: 2.1), $t(31) = -3.5$, but there was no age difference for intrusions from the last session (younger: 0.7; older: 0.8) $t(31) = -1.4$. Thus, the intrusion effect was a disturbance confined to the current experimental session. The model also predicted an interaction with age for the contrast between the third and fourth list but this was not supported by the data.

The second three contrasts tested the expected decline in frequency of intrusions with increasing lag of the list of or-igin. One contrast tested differences between lags of list of intrusion origin within List 3, and the remaining two tested differences between lags of list of intrusion origin within List 4. The only significant effect obtained was for the comparison within List 4 between lag-1 intrusions and the average of lag-2 and lag-3 intrusions as origin, $F(1, 31) = 8.2$, $MS_e = 1.6$. In List 4, lag-1 intrusions were more frequent than intrusions of higher lag (see Figure 2). Neither the difference between lag-2 and lag-3 intrusions within List 4 nor any of the interactions with age was significant. These differences were predicted by the model but, as is apparent from Figure 2, the expected effects were very small. Therefore, the lack of significance in the ANOVA could be due to low statistical power. In general (perhaps with the exception of the interaction between age and lag within List 3), when effects predicted by the model were large, they were significant in the ANOVA of intrusion errors.

Other false responses. Given equal recall and a significant difference in intrusions between age groups, younger adults had to commit more other false responses (i.e., incorrect responses that were not pairs in previous lists of the current session) than older adults. On average, there were 10.5 $(SD = 2.2)$ and 12.2 $(SD = 2.5)$ false responses for older and younger adults, respectively. This difference was significant, $F(1, 31) = 5.2$, $MS_e = 22.1$. False responses also increased significantly across lists, $F(3, 93) = 3.0$, $MS_e = 11.8$, but there was no interaction with age. Because of the 40-s time limit for responses there were also a few cases of omission errors. On average, older adults omitted 0.6 $(SD = 0.6)$ and younger adults 0.1 $(SD = 0.2)$ responses per list. This difference was significant, $F(1, 31) = 9.7$, $MS_e = 0.8$. Omissions decreased across lists, $F(3, 93) = 3.4$, $MS_e = 0.2$. The interaction between list and age group was not significant. Thus, less than one response per list was omitted in either age group.

Response format. The response format in this experiment required the subjects to enter two-digit codes extracted from an index card. Data from the control task without memory demand (performed at the beginning of each session) yielded no age difference in error rate, which was always below an average of 1 out of 30, but a significant difference in response time (younger: 3.8 s; older: 4.9 s per item), $t(31) = 2.7$. Age differences for correct and intrusion responses remained significant when this variable was included as a covariate in the ANOVAs just reported. Thus, age differences were not related to this component of the response format.

Estimation of Model Parameters

To our knowledge, these results document for the first time an age-differential susceptibility to intrusion errors in list recall when overall level of recall was controlled. The results were consistent with an age-differential encoding of list context (i.e., list tags), an age-differential loss of such list tags, or both of these. In both cases, older adults' memory would be flooded by traces without tags, thereby leading to a higher propensity for intrusions. We distinguished between these explanations by using frequencies associated with various types of responses, which were fit to the mathematical model.
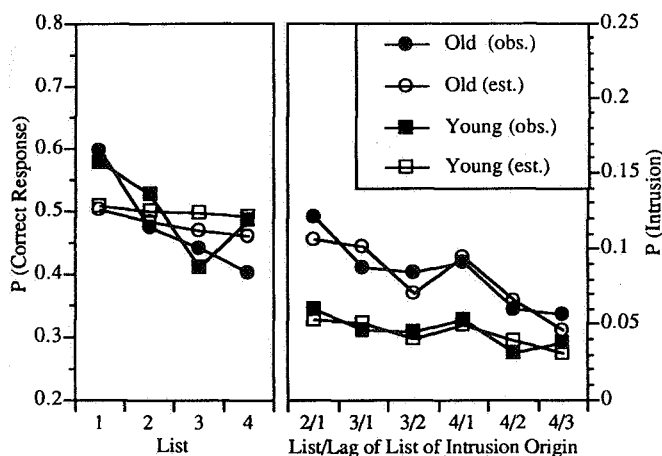

Figure 2. Observed (filled symbols) and estimated (open symbols) probabilities of correct and intrusion responses for younger and older adults in Experiment 1. (Intrusions are given per lag of intrusion origin [e.g., 4/1 means intrusion that was observed in the fourth list, and that the response was correct in the preceding List 3]).

Results are organized in three sections. First, we report results pertaining to the fit of the four-parameter model (a, b, p, q) to the 10 nonredundant response categories. Second, we test the necessity of each parameter for the model. Finally, we examine model parameters for significant age differences.

Observed and expected probabilities for response categories. Figure 2 contains observed and expected probabilities associated with the 10 response categories for the two age groups; correct responses are given in the left panel and intrusions in the right panel. The model recovered the age-differential susceptibility to intrusions; compare, for example, the effect associated with lag of list, that is the higher probability of intrusions from the immediately preceding (i.e., lag-1) list compared with earlier lists. The model fit appears to be adequate for intrusions, but there are systematic deviations for correct responses. Most notable is the much higher probability of younger and older adults' observed correct recall in the first list compared with model expectations. This issue is taken up in the *Discussion* section. The dip of correct recall in younger adults' third list was due to 3 participants who recalled at most three words in both sessions in this list; when they were left out of the analyses, the anomaly disappeared, but the parameter estimates reported in the following sections did not change by more than .01.

Estimates of four-parameter model (Model 0). In the top block of Table 7 the G^2 statistic for the overall goodness of fit and parameter estimates are displayed for the four-parameter baseline model (i.e., Model 0), with parameters estimated separately for younger and older adults. The G^2 statistic was highly significant; the fit was not very good. The deviations associated with correct responses were the pri-

mary reason. We evaluated the relevance of the parameters and associated age differences by specifying the submodels of Model 0 and by evaluating the change in goodness of fit.

Submodels of Model 0. A test of Parameter p (Model 1, $p = 1$) showed that results of this experiment could be reproduced equally well under the assumption that memory traces with list tags do not alter their states across subsequent lists; the change in the G^2 statistic was only 0.3 ($p > .50$). For each of the other three parameters, however, there were significant losses in goodness of fit when they were fixed at extreme values (Model 2, $q = 1$; Model 3, $a = 0$, $p =$ undefined; Model 4, $b = 0$; statistics are listed in Table 7). Results indicate that the distinction between traces with and without list tags appears to be justified with respect to the construction process.

Age differences in parameter estimates. Results of model tests applied to both age groups. This establishes the validity of the model for a broad spectrum of interindividual differences in memory ability. For the following tests of age differences in parameter estimates, Model 1 was used as a baseline model to avoid the problem with boundary estimates. Results are summarized in Table 8. Forcing Parameter a (Model 5, $a_{\text{young}} = a_{\text{old}}$) or Parameter b (Model 6, $b_{\text{young}} = b_{\text{old}}$) to be equal in both age groups resulted in a significant decline in goodness of fit. Thus, the probability of constructing traces with tags was significantly lower, and the probability of constructing traces without tags was significantly higher, in older compared with younger adults. In contrast, the probability that traces without list tags will be maintained between lists was not statistically different for the age groups (i.e., Model 7, $q_{\text{young}} = q_{\text{old}}$). This result is in agreement with other studies showing similar forgetting functions for

Table 7

Goodness of Fit and Estimates of Four-Parameter Model (i.e., Model 0) and Four Submodels in Experiment 1

Model	G^2	df	Probability	a	b	p	q
Model 0	124.1	12	< .01				
Older Adults				.26	.24	.86	.60
Younger adults				.38	.12	.91	.59
Model 1 ($p = 1$)	125.8	14					
Older adults				.27	.23	*1.00*	.72
Younger adults				.38	.12	*1.00*	.81
Model 1 − Model 0	1.7	2	> .25				
Model 2($q = 1$)	160.6	14					
Older adults				.34	.15	1.00	*1.00*
Younger adults				.42	.09	1.00	*1.00*
Model 2 − Model 0	36.5	2	< 0.1				
Model 3 ($a = 0$, $p = $ nd)	391.7	16					
Older adults				*.00*	.51	*nd*	.40
Younger adults				*.00*	.51	*nd*	.25
Model 3 − Model 0	267.6	4	< .01				
Model 4 ($b = 0$)	168.0	14					
Older adults				.48	*.00*	.29	.55
Younger adults				.50	*.00*	.52	.41
Model 4 − Model 0	43.9	2	< .01				

Note. Values set in italic were fixed for model estimation. nd = not defined.

Table 8

Age Differences in Parameters a, b, and q in Experiment 1

Model	G^2	df	Probability	a	b	p	q
Model 1 ($p = 1$)	125.8	14					
Older adults				.27	.23	*1.00*	.72
Younger adults				.38	.12	*1.00*	.81
Model 5	147.5	15					
($a_{\text{old}} = a_{\text{young}}$)							
Older adults				.33	.18	*1.00*	.85
Younger adults				*.33*	.17	*1.00*	.64
Model 5 – Model 1	21.7	1	< .01				
Model 6	147.0	15					
($b_{\text{old}} = b_{\text{young}}$)							
Older adults				.32	.18	*1.00*	.86
Younger adults				.33	*.18*	*1.00*	.62
Model 6 – Model 1	21.2	1	< .01				
Model 7	127.1	15					
($q_{\text{old}} = q_{\text{young}}$)							
Older adults				.28	.22	*1.00*	.76
Younger adults				.38	.13	*1.00*	.76
Model 7 – Model 1	1.3	1	> .25				

Note. Values set in italic were fixed or constrained for model estimation.

younger and older adults (e.g., Wickelgren, 1975). Thus, the higher frequency of intrusions and the stronger decline of intrusions across lists and across lags of intrusion origin within lists for older adults was linked primarily to age differences in the probabilities of constructing traces with and without list tags.

Discussion

The results of this experiment indicated an age-differential susceptibility to intrusion errors when recall was equated by means of criterion-referenced adjustment of presentation times. A necessary condition for intrusions to occur was the presence of traces without list tags, that is of landmark–noun combinations for which it was not clear in which list they were generated (see Table 1 for examples). The experimental design (i.e., [A–B, A–Br] paradigm, imaginable material, mnemonic instruction, and dynamic adjustment of presentation times) induced the construction of memory traces which, due to lack or loss of contextual cues, led to the buildup of a pool of potential intrusions across the lists of an experimental session. In a mathematical model mechanisms operating during construction and during storage were specified to account for correct recall as well as for intrusion errors resulting from earlier lists of a session. The mechanism that was most likely responsible for the age difference in intrusion errors appeared to be a deficit in the integration of context during the construction of memory traces; there were no significant age differences associated with the forgetting of trace content or the forgetting of context.

The model did not do very well in reconstructing the probabilities of correct responses across lists. Correct recall for the first list of an experimental session was especially

strongly underestimated in both age groups. Some decline in recall across lists was expected because of the addition of new traces without list tags in each list, which competed with those still available from previous lists. The predicted decline in recall across lists (see Figure 2; left panel) is due solely to response competition because Parameters *a* and *b* were held constant across lists. Moreover, because more traces without tags were added for older than for younger adults (and vice versa with respect to traces with tags) the model predicted a stronger decline in recall across lists for older adults. Our use of the term *response competition* is tied strictly to the rational selection of responses in the absence of relevant traces with tags and the presence of traces without tags as specified in the model.

The lack of fit between observed and expected recall probabilities suggests that response competition contributed some, but was not sufficient to account for all, of the proactive interference observed in correct responses. This result is consistent with results from paired-associate learning, where proactive interference was observed under conditions in which subjects were asked to emit all available responses after learning two lists (e.g., Houston, 1967; Koppenaal, 1963), that is, when response competition was minimized with "modified modified free recall" (MMFR; Barnes & Underwood, 1959). These results posed serious problems for the classical two-factor theory of interference, which assumed that proactive interference could be accounted for solely by response competition (Melton & Irwin, 1940; for a review see Postman & Underwood, 1973). In the present experiment we had hoped that the criterion-referenced adjustment of presentation time would counteract the difficulty of recalling later lists. Unfortunately, we ended up with differences in presentation time across lists that enhanced this difficulty: On average, presentation times were shorter, rather than longer, in the second and third lists than they were in the first list. Clearly, the assumption that the same proportion of traces with tags and without list tags would be generated in each list was not very plausible given these differences in presentation times. In summary, there were three reasons why later lists were more difficult in this experiment: (a) response competition due to the availability of potential intrusions from earlier lists, (b) shorter presentation times (at least in List 2 and List 3), and (c) possibly, sources of proactive interference other than response competition. Of these possibilities only the first one was considered in the formal model. To delineate the contributions of (a) and (b) one needs to keep presentation times constant across lists.

Alternative explanations of age differences in susceptibility to intrusions in terms of response bias, familiarity with the memory task, and knowledge about optimal response strategies were not very plausible in this experiment. Response bias was controlled by the forced-choice recognition procedure; participants were highly experienced in this task; and a careful instruction in guessing strategies was aimed at making optimal use of available information for maximizing recall. Of course, one cannot completely rule out the potential influence of age differences in such performance factors, but their relevance should be much reduced relative to the theoretical mechanisms postulated in the formal model.

Experiment 2

There were two issues we addressed in this follow-up study. First, we wanted to replicate the age-differential susceptibility to intrusions and the specificity of age differences with respect to the integration of contextual information about list membership of memory traces. Second, although presentation times were still individualized, they were kept constant across lists to avoid dependencies between serial position of list and presentation time and their compounded effect on recall. This procedural change was expected to delineate two possibly independent sources of proactive interference in recall: the increase in response competition and, possibly, a decrease in the probability with which traces with and without tags are generated across lists.

A decrease in the probability of generating traces with and without tags across lists could be due to persistence in encoding (Greeno, James, & DaPolito, 1971; see also Greeno, James, DaPolito, & Polson, 1978). Persistence in encoding is a tendency on the part of subjects to continue to use features of stimuli and responses that were generated in the first list (where they are likely to be appropriate for construction of an integrated trace) in later lists (where they are likely to be less appropriate). Thus, encoding persistence causes negative transfer because the first list sets up a bias toward sampling less than optimal stimulus and response features for trace construction in later lists. Greeno et al. (1978) saw no conceptual difference between their principle of encoding persistence and Martin's (1968, 1971) theory of encoding variability, which assumed a similar change in the sampling probabilities for stimulus features across lists. Also, an extension of classical interference theory along these lines was entertained by Postman and Underwood (1973) to overcome problems generated by patterns of proactive interference.

Proactive interference in recall of lists has been a tantalizing issue in cognitive aging. Although anecdotal evidence suggests a high likelihood of age-related differences, studies measuring release from proactive interference according to Wickens's (1970) paradigm yielded no evidence for adult age differences (for a review see Salthouse, 1982, p. 132). In two list learning experiments, however, large age differences were reported. Lair, Moon, and Kausler (1969) constructed lists with high-associate pairs but re-paired them in the list to be learned in the experiment. Older adults had considerably more difficulty in overcoming the preexperimental associations than did younger adults. Moscovitch and Winocur (1983) had subjects learn two lists. The first list consisted of high-associate pairs. In the second list the same A terms but an alternative high-associate B term had to be learned. Again, large age differences in favor of younger adults were observed. Kausler (1991, p. 285) concluded "that there are extreme transfer conditions in which elderly adults do seem to be more interference prone than young adults." Conditions appear to be extreme if preestablished relations must be overcome. Mnemonic construction of memory traces may establish extreme transfer conditions of this kind.

For each participant a presentation time was selected that allowed him or her an intermediate level of recall. To check the model validity across a wider range of accuracy, different levels of recall were induced in independent groups. Specifically, presentation times were determined for a 50% and an 80% criterion of accuracy, respectively. In all other respects, the experiment was identical to the previous one.

Method

Participants

About 6 months after Experiment 1, 15 younger and 18 older adults participated in a second training phase with the method of loci that comprised an additional 15 experimental sessions. The results of this training phase were reported in Baltes and Kliegl (1992). After this assessment, an additional three sessions were scheduled to replicate and extend the results reported here as Experiment 1. Three younger adults had moved away; 4 older adults were not available for reasons of health or extended vacation.

Procedure

The experimental set-up and procedure were very similar to those of Experiment 1. Again, in each of three sessions four lists of 30 landmark–noun combinations were presented; the combinations were different from those in Experiment 1 but were based on the same list of nouns and landmarks. Unlike Experiment 1, however, participants worked with the same presentation time in all sessions. Presentation times were fixed individually at values that had allowed one half of the subjects in each age group a recall of 15 of 30 words and the other half a recall of 24 of 30 words in the Baltes and Kliegl (1992) extension of the training. We did not expect to realize exactly these scores in the present experiment because in the training phase words had not been repeated across lists. Moreover, presentation times had been adjusted asymmetrically: A shorter presentation time had been selected whenever the criterion was met, but the time had been reset to the slow level only if the criterion was failed on two successive lists. Therefore, in the present experiment, means for correct recall should be below the nominal criterion levels.

Results

Presentation of results follows the schema used for Experiment 1. First we report results related to performance indicators and then the results pertaining to the mathematical modeling. As there was no calibration phase, data from all three sessions were used in the analysis.

Analyses of Performance Indicators

Presentation time. Presentation time was taken from the last three training sessions of the extension study reported in Baltes and Kliegl (1992). Means and standard deviations by age and criterion group are listed in Table 9. There were significant effects of age, $F(1, 26) = 28.2$, and criterion $F(1, 26) = 6.9$. Moreover, there was a significant interaction between these factors, $F(1, 26) = 4.4$ ($MS_e = 2.0$ for all F values). The age difference in need for presentation time was larger for the 80% than for the 50% groups.

Correct recall. Means and standard deviations of correct recall in the present experiment are also given in Table 9. Again, there were significant effects for age and criterion and a marginally significant interaction, $F(1, 26) = 18.3, 5.4$, and

Table 9

Means and Standard Deviations of Performance Indicators in Experiment 2

	Younger adults				Older adults			
	50% (n = 8)		80% (n = 7)		50% (n = 8)		80% (n = 7)	
Performance indicator	M	SD	M	SD	M	SD	M	SD
Presentation time[a]	1.5	0.5	1.8	0.8	3.2	1.3	5.7	2.4
Correct recall	15.6	4.5	21.0	2.4	12.8	2.6	13.2	3.6

[a] Individualized presentation times for this experiment were determined as part of participants' earlier mnemonic training reported in Baltes and Kliegl (1992).

3.9, respectively, $MS_e = 45.9$. Post hoc tests indicated that there was no difference for the 50% criterion groups, $t(14) = 1.6, p > .10$; there were, however, larger differences between scores of the 80% criterion groups in favor of younger adults, $t(14) = 4.8, p < .01$. Indeed, there was no difference between the two older adult groups. Thus, for older adults we cannot test the effects of criterion manipulation on model parameters; both groups should yield similar estimates. Similarly, there will be limitations on age comparisons for the 80% groups. With constant presentation times across lists, an Age × List interaction for correct recall would be indicative of age-differential susceptibility to proactive interference. The effect of list and the associated interactions were evaluated with three contrasts that tested the difference between successive lists. As expected, recall declined significantly from list to list, $F(1, 26) = 52.0, MS_e = 7.6, F(1, 26) = 31.2, MS_e = 3.9$ and $F(1, 26) = 5.5, MS_e = 3.9$, respectively. In addition, there was a significant interaction between age and the decline from List 1 to List 2, $F(1, 26) = 4.5, MS_e = 52.0$. None of the interactions involving criterion group was significant. Figure 3 (left panel) shows that older adults showed a stronger decline—that is, suffered more interference—than younger adults.

Intrusions per list. Age-differential susceptibility to pro-

Figure 3. Correct recall (left panel) and intrusions (right panel) as a function of list in experimental session and age group in Experiment 2. (There was significant age-differential proactive interference from List 1 to List 2.)

active interference was also shown in intrusions, $F(1, 26) = 41.2, MS_e = 4.1$ (see right panel of Figure 3). There were significant increases in intrusions from List 2 to List 3, $F((1, 26) = 72.1, MS_e = 1.4$, and from List 3 to List 4, $F(1, 26) = 8.4, MS_e = 1.7$. The main effect of criterion was not significant, but there was a trend toward more intrusions for 50% criterion groups. Finally, none of the interactions was significant.

Intrusions per list of origin: contrasts on lists. The right panels of Figure 4 (filled symbols) show the probability of intrusions as a function of list of origin (i.e., lag), age, and criterion. Overall, intrusions per list of origin declined significantly from List 2 to List 3, $F(1, 26) = 1.47, MS_e = 1.8$, List 3 to List 4, $F(1, 26) = 32.9, MS_e = 0.6$, and between List 4 and intrusions from the last session, $F(1, 26) = 86.6, MS_e = 0.3$. Also as in Experiment 1, the first and third interactions of these contrasts with age group were significant, $F(1, 26) = 3.9, MS_e = 1.8, p < .06$, and $F(1, 26) = 11.8, MS_e = 0.3$, respectively. The first interaction showed that the age difference in intrusions per list of origin was larger in List 2 than in List 3. The second interaction was due to a larger age difference for intrusions per list of origin in List 4 (younger: 1.2; older: 2.4), $t(28) = -5.5$, than for intrusions from the last session (younger: 0.5; older: 0.9), but the latter age difference was also significant, $t(28) = -2.6$. Contrasts for effects of lags within List 3 and List 4 yielded significantly more intrusions for shorter lags: That is, there were more intrusions of lag-1 than of lag-2 in List 3, $F(1, 26) = 13.8, MS_e = 0.8$; in List 4 lag-1 intrusions were more frequent than the average of lag-2 and lag-3 intrusions, $F(1, 26) = 16.5, MS_e = 0.7$; and there were more intrusions from lag-2 than from lag-3, $F(1, 26) = 8.8, MS_e = 0.5$. (Only the second of these three contrasts had been significant in Experiment 1.) There was one significant higher order interaction involving age, criterion, and the contrast between lag-1 and the average of lag-2 and lag-3 within List 4, $F(1, 26) = 5.3, MS_e = 0.7$. The primary source of this interaction was that the contrast was significant for younger adults with 50% criterion times and older adults with 80% criterion times. In view of the overall pattern of results we consider this effect spurious. As in Experiment 1, probably because of small expected effect size, none of the expected interactions between age and contrasts between lags within lists was significant. The main effects of intrusions per list of origin related to the decline of intrusions across lists and across lags within lists were predicted by the model and replicated the pattern of results reported for Experiment 1.
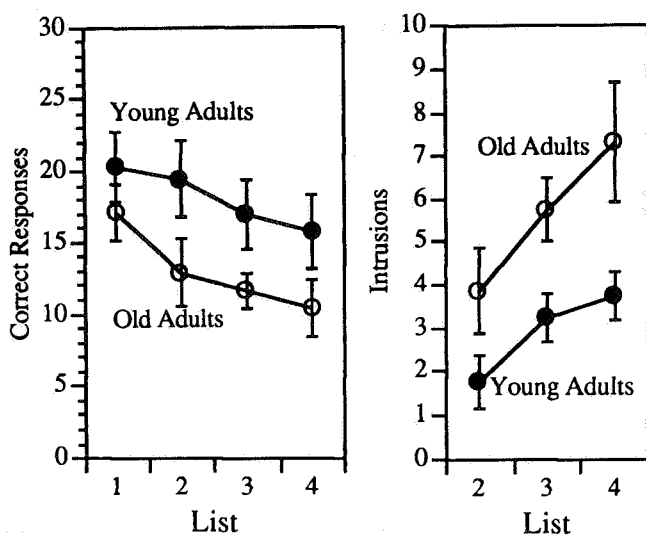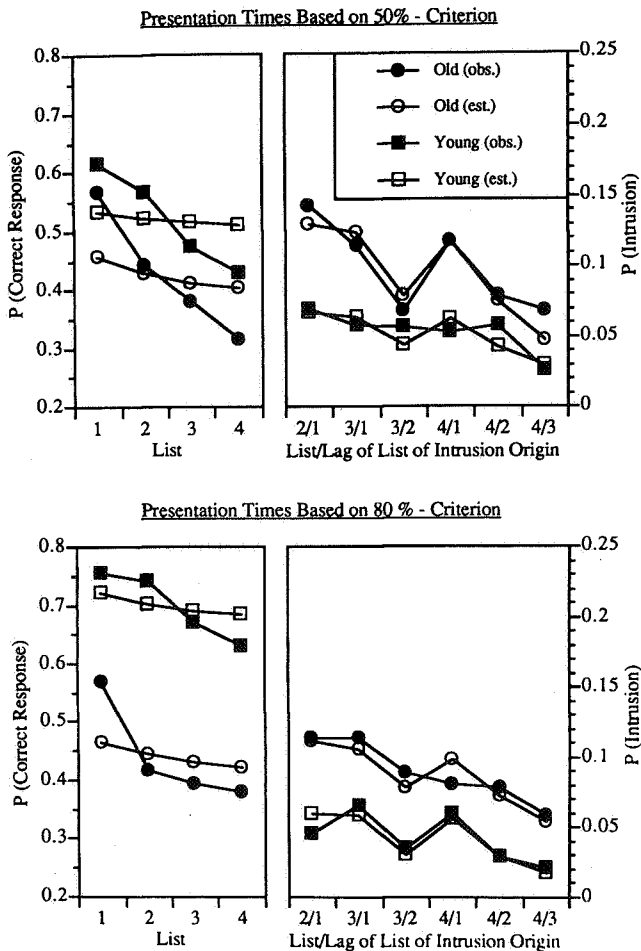
Figure 4. Observed (filled symbols) and estimated (open symbols) probabilities of correct and intrusion responses for younger and older adults with 50% and 80% criterion-referenced presentation times in Experiment 2. (Intrusions are given per lag of intrusion origin [e.g., 4/1 means intrusion that was observed in the fourth list and that the response was correct in the preceding List 3]).

Other false responses. Age groups differed in false responses that were not classified as intrusions, $F(1, 26) = 4.9$, $MS_e = 42.8$; older adults scored higher than younger adults. There was no significant main effect and no significant interaction involving criterion.

Response format. Data from the control task without memory demand (performed at the beginning of each session) yielded no age difference in error rate and no age difference in time taken to enter the response. As in Experiment 1, the pattern of results for correct answers and intrusions was not affected when this variable was included as a covariate.

Estimation of Model Parameters

Observed and expected probabilities for response categories. Figure 4 displays observed and estimated probabilities for nonredundant response categories. Again, the agreement was quite satisfactory for intrusions but much less

so for correct recall. For the latter, despite a constant presentation time, a divergence between the observed and estimate decline of recall across lists was notable in all four groups. In other words, proactive interference was more pronounced for observed than for estimated probabilities. The increase in response competition that was due to a growing pool of traces without tags across lists was not sufficient to account for the decline in correct recall across lists.

Model estimates. In Table 10, G^2 statistics and parameter estimates for the four groups are listed for the four-parameter model. As in Experiment 1, the overall fit of the model was poor. Checks of the relevance of model parameters replicated results reported for Experiment 1. The probability that traces with list tags do not alter their state across lists could be fixed at 1.0 without a significant loss of fit. Fixing each of the other parameters at an extreme value, however, led to significantly higher G^2 statistics (i.e., Models 2, 3, and 4).

Age differences in parameter estimates. Using Model 1 as a reference, we tested age differences in the three remaining parameters by evaluating the drop in goodness of fit when they were forced to be equal. As shown in Table 11, there were significant age differences in Parameters *a* and *b* but not in Parameter *q*. Thus, as in Experiment 1, the effect of age was limited to the generation of memory traces. Moreover, the pattern of parameter estimates was similar in both groups of older adults. For younger adults, the effect of the criterion manipulation resulted in an increase of Parameters *a* and *b*, whereas Parameter *q* was not affected. Because presentation time should influence primarily encoding, this result can be taken as further support for the model.

Extension of Model With Assumption of Encoding Persistence

Results of Experiment 2 were largely in agreement with those of Experiment 1. There were reliable age differences in the susceptibility to intrusions. This age difference could be reconstructed with differences in parameters of the construction process. Older adults apparently integrate context information to a lesser degree into their traces than do younger adults. The results also revealed a systematic discrepancy between observed and estimated probabilities for correct recall. In each of the four groups, observed probabilities were larger for the first list and smaller for the final two lists than was predicted by the model; this discrepancy was particularly pronounced for the first list of both groups of older adults (see Figure 4). The results imply that the intrusion of previously learned material can account for only a small portion of the observed proactive interference. A second mechanism, encoding persistence, can no longer be ignored. As described earlier, encoding persistence refers to the increase in difficulty of constructing a memory trace when the same cue has to be reused (Greeno et al., 1971, 1978). (An added difficulty in the [A–B, A–Br] paradigms is that this applies also to the item to be remembered.) A bias toward using less than optimal features of stimuli and responses in later lists could lead to lower construction probabilities for traces with and traces without list tags and, consequently, to a lower recall probability with increasing list number. We

Table 10
Goodness of Fit and Estimates of Four-Parameter-Model (Model 0) and Four Submodels in Experiment 2

Model	G^2	df	Probability	a	b	p	q
Model 0	198.1	24	< .01				
Older adults, 50%				.16	.29	.84	.59
Older adults, 80%				.23	.23	.88	.66
Younger adults, 50%				.36	.18	.75	.45
Younger adults, 80%				.41	.31	.79	.35
Model 1 (p = 1)	201.3	28					
Older adults, 50%				.17	.28	*1.00*	.67
Older adults, 80%				.24	.22	*1.00*	.77
Younger adults, 50				.36	.17	*1.00*	.71
Younger adults, 80%				.46	.26	*1.00*	.56
Model 1 – Model 0	3.2	4	> .25				
Model 2 (q = 1)	300.7	28					
Model 2 – Model 0	102.6	4	< .01				
Model 3 (a = 0, p = nd)	249.1	32					
Model 3 – Model 0	51.0	8	< .01				
Model 4 (b = 0)	282.9	28					
Model 4 – Model 0	84.7	4	< .01				

Note. Values set in italic were fixed for model estimation. nd = not defined.

allowed for this possibility by expanding the model with two Parameters i and j which represent the costs of encoding persistence; $a_{10} = a_{20} = a_{30} = a_{40}$ and $b_{10} = b_{20} = b_{30} = b_{40}$ were replaced with $a_{10} = a_{20} + 1*i = a_{30} + 2*i = a_{40} + 3*i$ and $b_{10} = b_{20} + 1*j = b_{30} + 2*j = b_{40} + 3*j$. With the constraints of Model 7 (i.e., $p = 1.0$ and $q_{young} = q_{old}$) enforced, goodness of fit and parameter estimates are listed in Table 12 under Model 8. The goodness of fit was acceptable—at least marginally—with $G^2(24) = 36.5, p > .04$. The age-differential pattern related to Parameters a and b was somewhat clearer than before, extending also to the 80% group of younger adults. Finally, encoding persistence

per list was estimated to be between 1% and 5% for both groups and both types of traces.

The small differences Parameters i and j suggested that they be replaced with a single parameter (i.e., $i = j$) and that this parameter be constrained to be equal across groups. This model fit equally well (see Model 9 in Table 12). Figure 5 displays expected probabilities for the 10 response categories that are based on Model 9. The agreement between data and model is acceptable. The model estimates for correct recall across lists reveal the expected Age × List interaction: Differences between age groups become larger—but the effect is very weak. The difficulty of establishing this interaction

Table 11
Age Differences in Parameters a, b, and q in Experiment 2

Model	G^2	df	Probability	a	b	p	q
Model 1 (p = 1)	201.3	28					
Model 5 ($a_{old;\ 50\%} = a_{old;\ 80\%} = a_{young;\ 50\%} = a_{young;\ 80\%}$)	250.9						
Model 5 – Model 1	49.6	3	< .01				
Model 6 ($b_{old;\ 50\%} = b_{old;\ 80\%} = b_{young;\ 50\%} = b_{young;\ 80\%}$)	212.2						
Model 6 – Model 1	9.9	3	< .01				
Model 7 ($q_{old;\ 50\%} = q_{old;\ 80\%} = q_{young;\ 50\%} = q_{young;\ 80\%}$)	206.8	31					
Older adults, 50%				.18	.27	*1.00*	.69
Older adults, 80%				.21	.25	*1.00*	.69
Younger adults, 50%				.36	.17	*1.00*	.69
Younger adults, 80%				.51	.21	*1.00*	.69
Model 7 – Model 1	5.5	3	> .15				

Note. Values set in italic were fixed or constrained for model estimation.

Table 12

Extension Models: Goodness of Fit and Estimates of a Six-Parameter Model and a Submodel in Experiment 2

Model	G^2	df	Probability	a	b	p	q	i	j
Model 8 ($q_{\text{old; 50\%}} = q_{\text{old; 80\%}} = q_{\text{young; 50\%}} = q_{\text{young; 80\%}}$)	36.5	23	> .04						
Older adults, 50%				.27	.28	*1.0*	*.69*	*.05*	.05
Older adults, 80%				.25	.29	*1.0*	*.69*	*.01*	.06
Younger adults, 50%				.43	.20	*1.0*	*.69*	*.03*	.06
Younger adults, 80%				.58	.19	*1.0*	*.69*	*.04*	.02
Model 9 ($qij_{\text{old; 50\%}} = qij_{\text{old; 80\%}} = qij_{\text{young; 50\%}} = qij_{\text{young; 80\%}}$)	44.6	30	> .04						
Older adults, 50%				.27	.27	*1.0*	*.71*	*.04*	*.04*
Older adults, 80%				.29	.26	*1.0*	*.71*	*.04*	*.04*
Younger adults, 50%				.43	.18	*1.0*	*.71*	*.04*	*.04*
Younger adults, 80%				.58	.19	*1.0*	*.71*	*.04*	*.04*
Model 9 − Model 8	8.1	7	> .25						

Note. Values set in italic were fixed or constrained for model estimation.

in earlier research could be related to insufficient statistical power. In contrast, age differences in susceptibility to intrusions appear to be quite robust and large.

General Discussion

The results of this article can be summarized in five points. First, given comparable levels of correct recall—achieved by adaptive adjustment of presentation times—there was a reliable age difference in the susceptibility for intrusions from earlier lists. Adult age differences in generic episodic memory were reflected in the amount of time needed to match a specified level of accuracy. Second, age differences in the susceptibility for intrusions were linked to theoretical parameters of a formal model. Specifically, younger adults were more likely to generate traces with information about the context of their generation (and less likely to generate traces without this information). Third, traces with tags were more likely to be retained across lists than traces without tags, but there were no age differences in these parameters. Fourth, decline in correct recall across lists (i.e., proactive interference) appears to be based on an increase in response competition that is due to traces without tags from previous lists and an additional factor, which could be encoding persistence (Greeno et al., 1971, 1978). Fifth, age differences were found for response competition but not for encoding persistence.

The present studies differ from most other experimental work in their reliance on highly practiced, laboratory-trained mnemonists. On the one hand, this is a constraint on the generalizability of the results. However, the results of age differences in intrusions in the absence of age differences in recall are consistent with a recent study by Schacter et al. (1991) on source forgetting that used conventional samples. Also, Mäntylä and Bäckman (1992, Experiment 3) found worse memory for contextual detail despite older adults' superior recognition of target objects displayed in a room. On the other hand, the expertise of participants in the memory skill may have helped the fit of the model because a high level of skill most likely reduces the influence of task-extraneous performance factors (such as test anxiety) and increases the ability to adhere to instructions.

Proactive Interference

Two sources of proactive interference in recall could be distinguished. The accumulation of traces without tags across lists led to an increase in response competition. The decline produced by response competition, however, was not sufficient to account for the observed decline in recall across lists. It was necessary, in addition, to assume that the probabilities associated with the generation of memory traces decreased across lists, an assumption compatible with the principle of encoding persistence (Greeno et al., 1971, 1978). The results also appear to be compatible with revisions of classical interference theory as outlined by Postman and Underwood (1973). Moreover, an age difference in proactive interference (i.e., in the decline in correct recall across lists) was observed in Experiment 2. In the formal model this age difference was restricted to the response-competition component (i.e., a consequence of the differential availability of traces with and without tags); age differences were not significant with respect to persistence of encoding.

The model expansion with linearly increasing effects of encoding persistence was motivated by the discrepancy between observed and estimated recall scores. With more lists per session, a negatively accelerated exponential function might be more appropriate than the one used here as a first approximation. Interestingly, the degree of estimated encoding persistence was not different for younger and older adults. Most likely, this was a consequence of the individualized adjustment of presentation times. The interpretation of encoding persistence as specific negative transfer is, however, in need of qualification. Explanations in terms of unspecific negative transfer (e.g., fatigue, decrease in cognitive effort) could be ruled out by using control groups with (A–B, A–C), and (A–B, C–D) designs (cf. Kausler, 1991). In this respect, the results presented here must be considered preliminary.
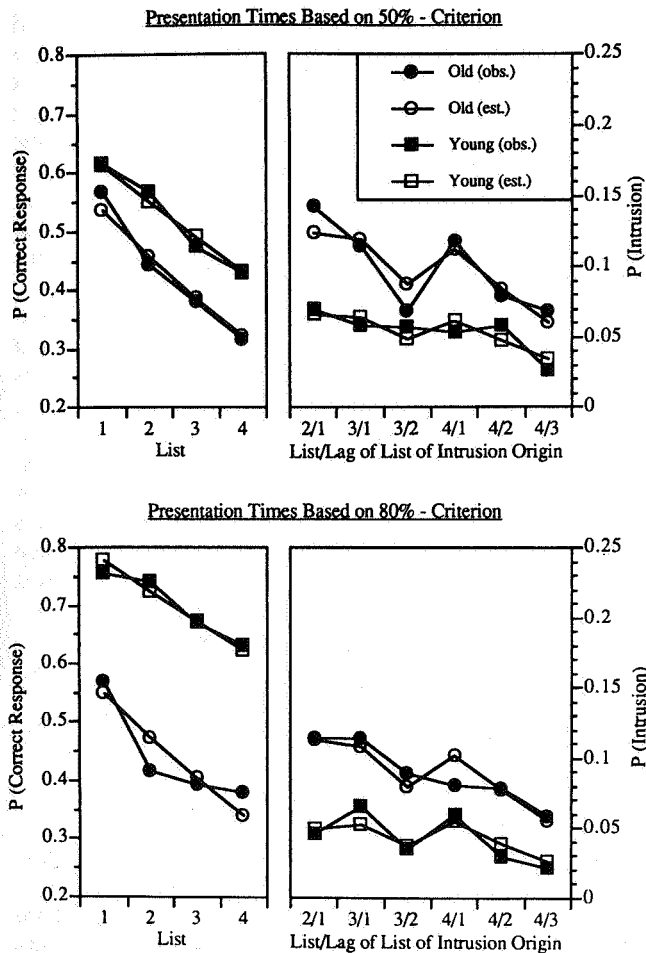
Figure 5. Observed (filled symbols) and estimated (open symbols) probabilities of correct and intrusion responses for younger and older adults with 50% and 80% criterion-referenced presentation times in Experiment 2. (Observed data are the same as those in Figure 4; estimates were based on Model 9.)

There has been considerable debate about whether interference theory predicts a negative dependency between first-list and second-list learning that is due to the need for unlearning of previous responses (e.g., Greeno et al., 1971; Hintzman, 1972; Martin, 1971; Postman & Underwood, 1973). Greeno et al. (1978) summarized a large body of research that suggested stochastic independence—a result consistent with the principle of encoding persistence but not with what Greeno et al. thought the unlearning factor of interference theory implied. Later reanalyses and simulations led to similar conclusions (Mensink & Raaijmakers, 1988; Riefer & Batchelder, 1988). The model presented here adds a little twist to this debate because it distinguishes between three potential sources implying independence, positive dependence, and negative dependence between first-list and second-list learning, respectively. First, encoding persistence leads to independence because of random sampling of stimulus features (i.e., Greeno et al. 1978). Second, a high probability for traces with tags (i.e., Parameter a) generates a positive dependence because it increases the chances of cor-

rectly recalling a trace without tag from the second list (i.e., due to list discrimination). Finally, a high probability for traces without tags (i.e., Parameter b) implies a negative dependence because of the associated increase in response competition in List 2. Given the relative prevalence of traces with and without tags in younger and older adults, correlations should be more positive in the younger than in the older adults. Obviously, the difference is likely to be too small to be detected with the usual statistical power.

Criterion-Referenced Testing

In the two experiments reported in this article, presentation times per word differed between individuals. Experimental equivalence in presentation time was traded for functional equivalence in overall level of recall. As a consequence, individual differences in memory ability were no longer reflected in accuracy but in the amount of time needed to meet the criterion of recall specified a priori. This was not only the case for differences between age groups but also for interindividual differences within age groups. The partial correlation (i.e., controlling for age group) between the adaptively determined presentation time of Experiment 1 and serial recall accuracy measured prior to this experiment in the posttest of the Kliegl et al. (1989) study was .70; accuracy in that study was based on the recall of six lists of 30 words administered with presentation times ranging between 1 s and 20 s per word. The same result was also obtained in an experiment by Thompson and Kliegl (1991). A rather different pattern was observed for intrusions: Susceptibility to intrusions was not related to serial recall; partial correlations were .13 with criterion-referenced presentation time and −.19 with the independent assessment of serial recall in the previous study—a result also consistent with the work of Keppel and Rauch (1966). Traditionally, the effect of age on recall has been equated by learning to criterion or by means of statistical control techniques. Criterion-referenced presentation times may be a useful alternative for pursuing questions of susceptibility to interference and cognitive development (Kliegl, Mayr, & Krampe, 1992; Mayr & Kliegl, 1992; Thompson & Kliegl, 1991).

As mentioned in the introduction, most research on age differences in memory for context is handicapped by the co-occurrence of age differences in memory for context and memory for target information (Light, 1991). Results of Experiment 1 yielded new information with respect to the recall of word lists: Criterion-referenced testing led to comparable probability of recall, but there was still a large age difference in susceptibility to intrusions. If a single mechanism were responsible for both types of age-related memory deficits then equating younger and older adults in target memory should also have eliminated age differences in memory for context. Also, the absence of significant partial correlations mentioned earlier suggests independent mechanisms. Finally, in Experiment 2, younger adults scored higher in memory for targets than older adults, but the profile of intrusion errors and associated age differences was almost identical to the one obtained in the first experiment. Thus, the data are supportive of the hypotheses (a) that memory for targets and

memory for contexts are based on different mechanisms, and (b) that age differences are more severe in memory for context. The next question to be discussed is whether the age difference in context memory can be linked to specific processing stages.

Processing Stages Revisited

The formal model was developed from an encoding perspective: Poor memory for context is assumed to be due to poor or deficient self-initiated constructive processing, which could be due to limited attentional resources (e.g., Craik & Byrd, 1982). Older adults are thought to engage in less processing of details, which leads to fuzzier traces overall. Adaptive adjustment of presentation time was sufficient to counteract the age deficit in target memory but did not eliminate differences associated with self-initiated constructive processing of contextual information.

Encoding

Integration and storage of target and context information was investigated in a four-parameter model. We conceptualized Parameters a and b as outcomes of the encoding process. This generation process can be conceptualized as a two-stage process of trace construction. In the first stage, traces are constructed without tags as mental images. In the second stage, list tags (contextual cues) are integrated into this trace with probability a. The estimation of transition probabilities for such a two-stage process requires multiple-trial learning. Because we administered each list only once, we can only derive conditional probabilities for integrating context given a trace. Obviously, the sum of Parameters a and b represents the probability of traces being generated; and $a/(a + b)$ represents the conditional probability that context was successfully integrated given a trace. For Experiment 1, the conditional probabilities were .56 and .75 for older and younger adults, respectively. In Experiment 2, the corresponding values were .50 and .70 for older and younger adults with a criterion of 50% recall and .53 and .75 for the 80% criterion groups. Because of the constraints on presentation time to achieve comparable accuracy, Parameters a and b can only be interpreted conjointly (i.e., older adults' higher values for Parameter b do not mean that they generated more traces without tags than younger adults but that more traces of that kind were left in the unfinished state). These results are compatible with an age deficit in integrating context information.

Storage

The validity of the distinction between traces with and without tags was corroborated by the differences in forgetting rates associated with them. Traces with tags were lost—if at all—to such a small degree that their probability of maintenance, p, could be fixed at 1.0 without loss in the quality of model fits. In contrast, traces without tags were maintained only with a probability of about .7. Given that lists were presented only once, the complete maintenance of traces with tags is a credit to the power of the mnemonic technique but would perhaps need to be corrected downward if delayed retention data had to be predicted as well. The lower probability, q, of maintaining traces without tags across lists was meant to indicate that these traces are lost or become unaccessible more readily. Note, however, that in the context of the present experiments a low value of q was advantageous because more potential intrusions were lost. Viewed from this perspective, $(1 - q)$ may reflect both the loss of traces without tags and the ability to select against temporally distant traces without tags. In a response competition situation subjects may not recall the exact list number in which a trace was generated but they may know that it was generated "way back" and decide not to select this trace as a response for this reason. This hypothesis could be tested by assessing the general availability of traces in the absence of response competition at the end of a session. Evidence that temporal distinctiveness can aid recall was provided by Bellezza (1982). The absence of age differences in parameters associated with the storage phase is consistent with previous research in which a similar conclusion was reached (cf., for a review, Kausler, 1991.)

Retrieval and Response Selection

Response selection was assumed to make use of available traces with and without list tags such that the probability of correct recall was maximized. Intrusion errors were a necessary consequence of optimal response selection. Note that the availability of responses is completely determined by the encoding and storage phases of the model. In other words, there are no retrieval processes beyond selecting a response from the available information. We also assumed that because of careful instruction and the expertise of our subjects there would be no age differences associated with response selection, that is, that the rational choice of responses represented in Tables 4 and 5 would apply equally to younger and older adults. Because the model could account for the data these assumptions can be maintained.

Alternative Accounts

Because the words to be remembered were invariant across lists, the subjects' task amounted to one of cued recognition with 30 well-known choices to each cue. One may doubt that subjects' response selection occurred in as pure a recollective manner as we assumed in the model. Although we instructed subjects very carefully, it seems likely that responses were also selected on the basis of familiarity, the second mechanism (besides deliberate recollection) contributing to scores in recognition tasks (e.g., Mandler, 1980). Indeed, the examples of traces without tags in Table 1 could be take as prototype examples of contributions by the familiarity mechanism. Of course, in principle, this could also be true for list tags of traces, but for them we probably would be more likely to give credit to recollection mechanisms. Jacoby (1991) argued that recollection and familiarity contribute to all memory tasks with task and subject characteristics regulating their relative importance. From this perspective, a somewhat different model can be postulated.

Suppose that during encoding only the first of the two processes involved in trace construction takes place—that is, traces are generated but no list tags. Then, these traces are lost with a certain probability. Finally, prior to response selection, the second stage is carried out: List tags for traces are not constructed but are reconstructed using both familiarity and recollection. Because age differences are large in recollection and small in familiarity (cf., Dywan & Jacoby, 1990; Light, 1991), older adults will rely comparatively more on familiarity than younger adults and will be more susceptible to interference. They are less able to select against traces from earlier lists and, consequently, commit more intrusion errors than younger adults. One argument in favor of this interpretation is that subjects were not instructed to integrate context (e.g., the list number) into their mental images or thoughts; rather, they were told to use all available information about prior locations of words during recall. In summary, age differences in conditional probabilities of integrating context can also be interpreted as age differences in conditional probabilities of reconstructing context. Note, however, that this account still begs the question of where the information comes from that allows younger adults to reconstruct context better than older adults.

One possibility is contextual fluctuation as postulated, for example, by Balota, Duchek, and Paullin (1989; for an encompassing model of interference and forgetting see Gillund & Shiffrin, 1984; Mensink & Raaijmakers, 1988). Balota et al. showed that older and younger adults' performance in a continuous cued-recall paradigm could be predicted on the assumption that older adults (a) encoded less contextual information and (b) have a slower rate of contextual fluctuation across time. There was a large difference between younger and older adults' recall level but, interestingly, none of the experimental manipulations interacted with age. It is unclear to what degree the general age difference was responsible for the results, but Balota et al.'s first finding was certainly compatible with the present results. As Balota et al. mentioned in their discussion, manipulations of presentation rate should be used to equate age groups in rate of contextual fluctuation to determine the dependency between overall level of performance and contextual fluctuation.

Age-specific decline in recollection of contextual information is also compatible with Hasher and Zacks's (1988; Hasher, Stoltzfus, Zacks, & Rypma, 1990) position that aging is associated with a decline in inhibitory processes that keep working memory clear of irrelevant information. Effective recollection may be required to counteract interfering contributions of familiarity. Evidence in support of this position derives from studies showing that older adults have great difficulty in keeping attention focused on a target task, as indicated, for example, in the frequency of unrelated content in verbal statements (Gold, Andres, Arbuckle, & Schwartzmann, 1988; Stine & Wingfield, 1987). Gerard, Zacks, Hasher, and Radvansky (1991) manipulated the amount of information overlap of the material to be learned in a recognition experiment. During recall, similar and unlearned items had to be distinguished. Older adults were much more impaired by this experimental manipulation of interference than were younger adults. Gerard et al. interpreted their re-

sults in terms of an age-differential retrieval deficit, but from the discussion in their article it appears that differences in the encoding phase could be the source as well.

Some of the ambiguities relating to the interpretation of parameters may be resolved by enriching the data and the model. The experimental approach could be extended in at least three ways: (a) Performance could be assessed across various retention intervals; we tested only immediate recall. (b) The availability of traces from earlier lists could be assessed with MMFR procedures; we tested only recall of the preceding list. With perfect maintenance of traces with tags across these two manipulations we would lean toward the reconstruction interpretation. (c) Training could be tailored toward the integration of context; our instructions focused only on response selection. We think the age-related encoding deficit reflects an irreversible fact of old age. Therefore we expect maintenance or magnification of age differences in intrusions. Finally, with respect to the model specification, the two-stage forgetting process could perhaps be expanded with a fluctuation process allowing for the recovery of lost tags and of lost traces. We would expect age invariance in the relevance of contextual fluctuation if overall level of recall is controlled.

Such extensions of the data base and the model itself would facilitate comparisons with other mathematical models that are concerned with the influence of the availability of context information on recall (e.g., Balota et al., 1989; Batchelder & Riefer, 1990; Mensink & Raaijmakers, 1988). We have little doubt that the present set of results could be accommodated within the general architecture of these models. Two differences to the approach presented here are their restriction to correct recall (i.e., there is no modeling of different types of errors) and the large number of model parameters (i.e., there are no tests of goodness of fit); the second difference, however, does not apply to the multinomial model of Batchelder and Riefer (1990; see also Riefer & Batchelder, 1988).

Conclusion

At a general level there is little disagreement that attempts need to be made to formalize adult age differences in cognitive functioning (cf., Charness, 1988; Salthouse, 1988). Formal models require the explicit specification of hypothetical constructs or mechanisms. As a consequence, age differences can be described in terms of a change in one or more parameters of the model (assuming that the model fits the data). If the parameters stand for hypothetical mechanisms of some generality, the model can be brought to bear on a variety of tasks in which these mechanisms are supposed to be relevant. Cognitive models are typically tailored toward the functioning of college students. The model introduced here had its origin in observations made on older adults, that is, in their difficulty to remember the context (i.e., the list) in which a particular memory trace had been generated. The low frequency of intrusions observed for young adults would hardly have inspired us to think about a formal model. Formal models such as the one presented in this article may be

636

helpful both in identifying specific processing deficits associated with aging and in exploring the generality of these mechanisms.

References

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging, 4*, 3–9.

Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology, 28*, 121–125.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97–105.

Baschek, I. L., Bredenkamp, J. Oehrle, B., & Wippich, W. (1977). Bestimmung der Bildhaftigkeit (I), Konkretheit (C) und der Bedeutungshaltigkeit (M') von 800 Substantiven [Determination of imagery, concreteness, and meaningfulness of 800 nouns.] *Zeitschrift für Experimentelle und Angewandte Psychologie, 24*, 353–396.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548–564.

Bellezza, F. S. (1982). Updating memory using mnemonic devices. *Cognitive Psychology, 14*, 301–327.

Bjork, R. A. (1978). The updating of human memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 235–259). San Diego, CA: Academic Press.

Bjork, R. A., & Landauer, T. K. (1978). On keeping track of the present status of people and things. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 52–60). San Diego, CA: Academic Press.

Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist, 58*, 496–510.

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin, 90*, 513–546.

Charness, N. (1988). The role of theories of cognitive aging: Comment on Salthouse. *Psychology and Aging, 3*, 17–21.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum Press.

Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: Differences in the susceptibility to false fame. *Psychology and Aging, 5*, 379–387.

Gerard, L., Zacks, R. T., Hasher, L., & Radvansky, G. A. (1991). Age deficits in retrieval: The fan effect. *Journal of Gerontology, 46*, P131–P136.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 62*, 145–154.

Gold, D., Andres, D., Arbuckle, T., & Schwartzmann, A. (1988). Measurements and correlates of verbosity in elderly adults. *Journal of Gerontology: Psychological Sciences, 43*, P27–P33.

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

Greeno, J. G., James, C. T., DaPolito, F., & Polson, P. G. (1978). *Associative learning; A cognitive analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Hartley, J. T., & Walsh, D. A. (1980). The effect of monetary incentive on amount and rate of free recall in older and younger adults. *Journal of Gerontology, 35*, 899–905.

Hasher, L., Stoltzfus, E., Zacks, R. T., & Rypma, B. (1990). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 163–169.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Hashtroudi, S., Johnson, M. K. & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4*, 106–112.

Hintzmann, D. L. (1972). On testing the independence of associations. *Psychological Review, 79*, 261–264.

Houston, J. P. (1967). Proactive inhibition and competition of recall. *Journal of Experimental Psychology, 75*, 118–121.

Hultsch, D. F., & Dixon, R. A. (1983). The role of pre-experimental knowledge in text processing in adulthood. *Experimental Aging Research, 9*, 17–22.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–514.

Kausler, D. H. (1982). *Experimental psychology and human aging.* New York: Wiley.

Kausler, D. H. (1991). *Experimental psychology, cognition, and human aging* (2nd ed.). New York: Springer.

Kausler, D. H., & Puckett, J. M. (1981). Adult age differences in memory for sex of voice. *Journal of Gerontology, 36*, 44–50.

Keppel, G. (1968). Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 172–213). Englewood Cliffs, NJ: Prentice-Hall.

Keppel, G., & Rauch, D. S. (1966). Unlearning as a function of second-list error interaction. *Journal of Verbal Learning and Verbal Behavior, 5*, 50–58.

Kliegl, R., Mayr, U., & Krampe, R. T. (1992). *Time-accuracy functions of process and person differences: An application to cognitive aging.* Manuscript submitted for publication.

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology, 25*, 247–256.

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of age-magnification in mnemonic training. *Developmental Psychology, 26*, 894–904.

Koppenaal, R. J. (1963). Time changes in strength of A–B, A–C lists: Spontaneous recovery. *Journal of Verbal Learning and Verbal Behavior, 2*, 310–319.

Lair, C. V., Moon, W. H., & Kausler, D. H. (1969). Associative interference in the paired associative learning of middle-aged and old subjects. *Developmental Psychology, 9*, 548–552.

Lehman, E. B., & Mellinger, J. C. (1984). Effects of aging on memory for presentation modality. *Developmental Psychology, 20*, 1210–1217.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333–376.

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

Mäntylä, T., & Bäckman, L. (1992). Aging and memory for expected and unexpected objects in real-word settings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1298–1309.

Martin, E. (1968). Stimulus meaningfulness and paired-associated transfer. *Psychological Review, 75*, 421–441.

Martin, E. (1971). Verbal learning theory and independent retrieval phenomena. *Psychological Review, 78*, 314–332.

Mayr, U., & Kliegl, R. (1992). *Sequential and coordinative complexity: Age-based processing limitations in figural transformations.* Manuscript submitted for publication.

McCormack, P. D. (1982). Coding of spatial information by young and elderly adults. *Journal of Gerontology, 37,* 80–86.

McCormack, P. D. (1984). Temporal cording by young and elderly in a list-discrimination setting. *Bulletin of the Psychonomic Society, 22,* 401–402.

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology, 41,* 175–192.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 53,* 173–203.

Mensink, G. J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455.

Moscovitch, M., & Winocur, G. (1983). Contextual cues and release from proactive inhibition in old and young people. *Canadian Journal of Psychology, 37,* 331–344.

Müller, G. E., & Pilzecker, A. (1990). Experimentelle Beiträge zur Lehre vom Gedächtnis [Experimental contributions to an understanding of memory]. *Zeitschrift für Psychologie,* Suppl. Vol. *1,* 1–300.

Paivio, A., Yuille, J. C. & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement, 76,* 1–25.

Park, D. C., & Puglisi, J. T. (1985). Older adults' memory for the color of pictures and words. *Journal of Gerontology, 40,* 198–204.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

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

Salthouse, T. A. (1982). *Adult cognition.* New York: Springer.

Salthouse, T. A. (1988). Initiating the formalization of theories of cognitive aging. *Psychology and Aging, 3,* 3–16.

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging, 6,* 559–568.

Simon, E. (1979). Depth and elaboration of processing in relation to age. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 115–124.

SPSS Inc. (1988). *SPSS-X user's guide* (3rd ed.). Chicago: Author.

Stine, E. L., & Wingfield, A. (1987). Process and strategy in memory for speed among younger and older adults. *Psychology and Aging, 2,* 272–279.

Thompson, L. A., & Kliegl, R. (1991). Adult age effects of plausibility on memory: The role of time constraints during encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 542–555.

Tun, P. A. (1989). Age differences in processing expository and narrative text. *Journal of Gerontology: Psychological Sciences, 44,* P9–P15.

Wickelgren, W. A. (1975). Age and storage dynamics in continuous recognition memory. *Developmental Psychology, 11,* 165–169.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Wolfram, S. (1988). *Mathematica.* Reading, MA: Addison-Wesley.