Age differences in encoding and retrieving details of a pediatric examination

RANDALL H. BENDER, THOMAS S. WALLSTEN, and PETER A. ORNSTEIN University of North Carolina, Chapel Hill, North Carolina

Baker-Ward, Gordon, Ornstein, Larus, and Clubb (1993) showed that recall improves over ages 3–7 for events experienced during a physical examination. We used a joint multinomial model to ask whether the improvement was due to encoding, to retrieval, or to likelihood to report. The model fit the Baker-Ward et al. data well and showed that (1) retrieval and reporting cannot be distinguished and (2) the observed effects were due primarily to age-related improvement in retrieval reporting rather than in encoding.

Research on young children's abilities to remember the details of salient, personally experienced events is important for at least two reasons: (1) it contributes to our understanding of the development of autobiographical memory, and (2) it aids in establishing methods for assessing the validity of children's reports about sexual and other forms of abuse. It is in this spirit that Ornstein and his colleagues (e.g., Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993; Ornstein, Gordon, Baker-Ward, & Merritt, in press; Ornstein, Gordon, & Larus, 1992) have been exploring young children's memories of various types of medical experiences.

In one such study, Baker-Ward et al. (1993) examined the abilities of 3-, 5-, and 7-year-olds to remember the details of a regularly scheduled physical examination by a pediatrician and nurse. Memory overall was good, although there were age differences in both initial and delayed recall. The younger children provided less information than did the older subjects and relied more strongly on specific probes. Moreover, the 7-year-olds demonstrated little, and the 3-year-olds considerable, forgetting over the course of a 6-week delay interval.

What factors might underlie this pattern of findings? Developmental studies of deliberate memorization suggest age-related improvement in strategies for both encoding and retrieval. For example, Ornstein and Naus (1978) have demonstrated that older children have superior rehearsal and organizational skills than younger chil-

This research was supported by a grant from the University Research Council of the University of North Carolina at Chapel Hill. The authors would like to thank William H. Batchelder, Charles J. Brainerd, Richard Schweickert, and an anonymous reviewer, whose many helpful comments on an earlier draft and suggestions markedly improved this paper. The authors also thank Patricia Clubb for help in working with the data from the Baker-Ward, Gordon, Ornstein, Larus, and Clubb (1993) and Clubb, Nida, Merritt, and Ornstein (1993) studies, both of which were supported by Grant MH 49304 from the U.S. Public Health Service. Correspondence should be addressed to R. H. Bender, Department of Psychology, Davie Hall CB 3270, University of North Carolina, Chapel Hill, NC 27599-3270 (e-mail: randy_bender@unc.edu).

dren, leading to better encoding. Ceci and Howe (1978) have shown that older children search memory better, resulting in higher retrieval rates. But, as Schneider and Pressley (1989) point out in their review of this literature, inferences about encoding processes generally are drawn from recall data, and children do not always retrieve all that they have encoded. Thus, it is sometimes difficult to disentangle the two processes. One approach to overcoming this problem is to represent memory theories by mathematical models with parameters that denote encoding and retrieval probabilities. The general conclusion from the application of a variety of such models is that encoding probabilities increase with age. Retrieval probabilities very likely do so as well, but that result has not been consistent over studies. Among the relevant modeling work are studies by Chechile, Richman, Topinka, and Ehrensbeck (1981), Brainerd (1985), Wilkinson, DeMarinis, and Riley (1983), Wilkinson and Koestler (1983, 1984), and Brainerd, Reyna, Howe, and Kingma (1990). A summary of most of these and related efforts has been provided by Schneider and Pressley (1989) and Brainerd (1985).

The Baker-Ward et al. (1993) study differs in important ways from most others that have demonstrated distinct age-related improvement in encoding and retrieval. For one thing, the bulk of the developmental memory literature deals with tasks in which children know in advance that their recall will be assessed. In contrast to these studies of deliberate memorization, the children in the Baker-Ward et al. experiment did not know that they were going to have to recall the details of their checkups, or if they did, they did not consciously attend to memorizing the experiences; therefore, their encoding of information was clearly incidental. The subsequent recall, however, was deliberate (see Brainerd & Ornstein, 1991, and Ornstein, Gordon, & Baker-Ward, 1992, for a discussion of issues concerning this combination of conditions).

Another difference between the Baker-Ward et al. study and many other explorations of memory is that their subjects experienced the "stimulus materials" (i.e., the physical examination) only once, whereas participants in most other experiments are presented with sev-

eral study/test trials. Finally, the bulk of the developmental literature has used verbal stimuli (most routinely, sets of words), whereas the Baker-Ward et al. investigation focused on an integrated, meaningful, and salient event. The children were at times in various states of undress, physically handled by an (often unknown) adult, and occasionally emotionally stressed. Given the potentially embarrassing circumstances of the physical examination, it cannot be assumed that features of the examination that were encoded and retrieved necessarily were reported in the recall sessions.

Because of these differences in procedure, context, and materials, it would not be surprising to find a pattern of age-related changes in encoding and retrieval processes different from the pattern that has been noted in other studies. For example, if the age-related encoding improvements observed in other studies were due to differential use of intentional encoding strategies and if the children of the ages tested by Baker-Ward et al. had sufficient prior knowledge about the physical examination to adequately encode its features, then encoding in the Baker-Ward et al. paradigm might be age-invariant. On the other hand, if the positive relation between age and encoding ability does not depend on the use of conscious strategies, or if prior knowledge is not equivalent across ages, then such effects might show up in the Baker-Ward et al. case. With respect to retrieval, age-related improvement has been found in some studies and not in others. Therefore, it is of particular interest to assess whether it was manifest in this context. Finally, the process of reporting an event, given that it was encoded and retrieved, must be considered and factored into our understanding of the memory system in this context.

To assess age effects on encoding and both age and delay effects on retrieval and reporting, we have reanalyzed the Baker-Ward et al. data using a joint multinomial model (Batchelder & Riefer, 1986; Greeno, James, DaPolito, & Polson, 1978; Kail, Hale, Leonard, & Nippold, 1984; Riefer & Batchelder, 1988, 1991). The multinomial approach has a number of advantages: (1) the model yields exact expressions that relate encoding and retrieval-reporting parameters to data, (2) parameters can be derived separately for the individual features of the physical examination, and (3) the model has a natural error theory that provides a statistical goodness-of-fit test. Point 2 is important because the features of the examination (e.g., ear check, heart assessment, genital exam, collection of a urine specimen, etc.) may be differentially encoded, depending on how they were experienced, and differentially salient or embarrassing at time of retrieval. Taking advantage of Points 1 and 3, hypotheses about agerelated differences in memory processes are naturally expressed as restrictions on the model parameter space and subject to statistical evaluation. Conclusions about processspecific age differences, therefore, are drawn only if the model is supported by the data and on the basis of statistical tests.

A related use of the model is first found in Greeno et al. (1978) applied to the modified modified free recall

task. They used the model to investigate the conditions under which the retention of B and C items, from AB and AC study pairs, were independent. Reifer and Batchelder (1988) used the model to look at storage and retrieval in retroactive inhibition. More recently, Reifer and Batchelder (1995) had a related application of the model to over 300 data sets from 44 separate studies of the recognition-failure paradigm, as an alternative to the Tulving—Wiseman function. They showed that experimental manipulations thought to affect recall or encoding were in fact reflected in changes in the respective parameter values. Thus, the model that we use here has received good support in other applications seeking to distinguish encoding and retrieval effects in memory.

The remainder of this paper is organized as follows. First, we will briefly describe the experimental procedures, referring readers to the original report (Baker-Ward et al., 1993) for details. Next, we will develop a family of models and use it to summarize the data in a novel way—namely, using features of the physical examination, rather than subjects, as the units of analysis. Nested models within the family will be compared with each other and with the data to test hypotheses about delay and age differences in the underlying processes. Finally, we will discuss implications, limitations, and extensions of the work.

METHOD

Subjects

The subjects were children who had undergone a "well-child" visit at either of two different clinics and had been seen by one of several different medical teams. They were from three age groups: 3 years (33–42 months), 5 years (57–66 months), and 7 years (81–90 months). The numbers of subjects in the respective groups were 47, 49, and 43.

Procedures

Parents were contacted prior to a child's regularly scheduled checkup to seek their permission and their child's approval for participation. Children of families who agreed to take part in the study received a regular checkup with one unusual event added—namely, after the examination was completed, the child's picture was taken by the nurse

Because of differences in the nature of individual physical examinations, checklists were used to establish what had actually occurred in each case. The examining physician and nurse indicated the procedures that were included in their segments of the examination, and the accompanying parent independently checked off procedures that the child had experienced. A procedure was assumed to have occurred if it was checked on either the medical team's or the parent's list.

Following the examination, the children underwent two videotaped interviews, one immediately and the other at a delay of either 1, 3, or 6 weeks. (A control group was interviewed only once, 3 weeks after the checkup, but is not included in these analyses.) The children, interviewed alone with parents generally nearby, were first prompted with general questions to provide a free recall account of what had happened in the checkup and were then cued in increasingly pointed ways about features of the examination that they had not mentioned. Trained interviewers were used, and the order of probes was varied across subjects.

The children's verbal protocols were coded from the videotapes by trained judges, using a system that focused on whether or not individual examination features were correctly remembered and, if so, the specificity of the prompt necessary to elicit recall. The present analyses use only the free-recall data, which were obtained initially in response to the general open-ended question about the examination or subsequently if the child spontaneously recalled an event not being addressed by a specific probe.

MODELING THE UNDERLYING PROCESSES

The Unit of Analysis and the Likelihood Function

Unlike most applications of multinomial models and, indeed, unlike most investigations of memory in general, our unit of analysis is the experienced event (e.g., a check of the ears), not the individual child. Because some procedures were not experienced by all children, we are limited to studying those events experienced by a sufficient number of subjects in all three age groups. Using the items as the unit of analysis allows different levels of encoding and retrieval to emerge for different events. For example, encoding and retrieval-reporting would not be expected to be the same for an ear and for a genital check. We allow these differences to manifest themselves by treating the events separately. It is probably also the case that these processes vary over individual children and depend on how the specific event was carried out by the health-care provider. We will consider such individual differences in the discussion, but, for the moment, ignore them.

To understand the data structure, consider a single procedure restricted to a single age group—for example, a chest examination of 3-year-olds. Every child who experienced the event subsequently had two opportunities to recall it; immediately and after a delay. (We are ignoring length of delay to increase statistical power, because the earlier analyses by Baker-Ward et al. showed only small unsystematic effects of this variable.) Each recall opportunity could result in a success (S) or a failure (F), leading to a four-event categorization for the outcomes: E1 = successfully recalled the exam both immediately and after the delay (SS); $E2 \equiv$ successfully recalled the exam immediately but not after the delay (SF); E3 \equiv failed to recall the exam immediately but succeeded after the delay (FS); and $E4 \equiv$ failed to recall the exam both immediately and after the delay (FF).

Let N_{ij} equal the number of children in age group j (j = A, B, and C for 3-, 5-, and 7-year-old groups, respectively) who experienced event i (i = 1,...,I) and let N_{ijk} index the number of those children who subsequently fell into cell E_k , with, of course, $N_{ij} = \sum_k N_{ijk}$. The data vector for event i and age group j, then, is $\mathbf{D_{ij}} = (N_{ij1},$ N_{ij2} , N_{ij3} , N_{ij4}). For example, 47 3-year-olds had their chests examined, of whom 5 freely recalled the event on both occasions, 7 did so at the first test only, 3 at the second test only, and 32 did so neither time. Thus, $N_{\text{chestA}} =$ 32). There are corresponding data vectors for the 5- and 7-year-olds, D_{chestB} and D_{chestC}. In addition, there are vectors, **D**_{ii}, for all three age groups for each of the other items used in the analysis. We might consider the single age group \times item vectors $\mathbf{D_{ii}}$ separately; however, generally, we will consider the vectors collectively, with $\mathbf{D} = (\mathbf{D_{1A}}, \dots, \mathbf{D_{IC}}).$

We assume that these data, **D**, meet the conditions for a joint (or product) multinomial distribution. That is,

considering a single age \times item data vector, $\mathbf{D_{ij}}$, we assume that each child independently falls by design into exactly one of the cells E1, E2, E3, and E4, with probabilities p_{ij1} , p_{ij2} , p_{ij3} , and p_{ij4} , respectively, such that $\sum_k p_{ijk} = 1$. The goal of the model is to express these probabilities in terms of psychologically meaningful parameters.

The Basic Model

Consistent with virtually all theorizing about free recall, we will assume that recall of an event on a given occasion depends on (1) its having been encoded sufficiently strongly or in suitable form to be available for retrieval and (2) its being successfully retrieved and reported.

We begin with the process of encoding an event. Although encoding may well be thought of as a continuous, rather than a discrete all-or-none process, we nevertheless will assume that a member of age group j either encoded item i with probability q_{ij} or did not encode it with probability $1 - q_{ij}$. If one prefers to view encoding as a more continuous process of building strength or laying down traces, then we can assume that with probability q_{ii} , he or she developed sufficient strength to support retrieval of the event and with probability $1 - q_{ij}$, he or she did not (see Batchelder & Riefer, 1986, or Chechile & Meyer, 1976, who make an identical assumption). The binary outcomes of (sufficiently) encoding or not are shown as the top two branches in Figure 1. An item that has not been encoded cannot subsequently be recalled, resulting in the FF event designated at the bottom of the

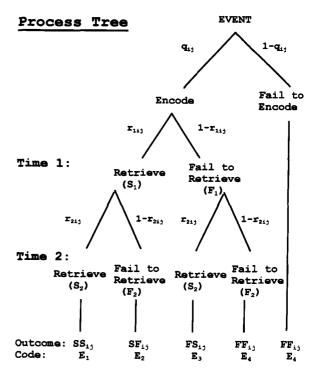


Figure 1. General process tree diagram of the full model for a child from group j and item i.

not-encode branch. We assume that if a member of group j has encoded item i, he or she will either retrieve and report it at the initial test with probability r_{1ij} or not with probability $1-r_{1ij}$. This assumption is represented by the second-level branch in Figure 1. The probability of retrieving and reporting the item at second test is r_{2ij} , independent of success at first test, as shown by the third-level branches in Figure 1. The code at the end of each branch shows the pattern of successes and failures for event i over group j's two retrieval opportunities. The outcome codes E_1 – E_4 are also shown, consistent with the notation we are using.

Three points remain to be clarified regarding the model. For reasons already mentioned, we would like to divide Step 2 above and include as Step 3, a decision to report the event given its retrieval. Unfortunately, it turns out that the parameters for the retrieval and reporting processes cannot be separately identified (see Appendix A for proof), and, therefore, the retrieval probability really must be interpreted as the probability of the two processes combined. However, in what follows, we generally will still simply refer to a retrieval parameter, but we recognize that its value depends multiplicatively on both processes. To the extent that the children report all they retrieve (which generally seems reasonable in this case), this practice will not result in misinterpretation.

Second, we are assuming that items are not being retrieved, for example, as part of a script. We can be reasonably assured that scripts are not producing the responses instead of memory search, because Baker-Ward et al. reported that intrusion errors were virtually nonexistent.

Finally, note, that the assumption that retrieval at Time 2 is independent of retrieval at Time 1 conditional on the event's having been encoded does not imply that retrievals at Times 1 and 2 are *unconditionally* independent. The two retrievals would be unconditionally independent if and only if the probabilities of retrieval at Time 2 were identical, given either retrieval or failure to retrieve at Time 1. That is, they would be independent if and only if $Pr(S_2|S_1) = Pr(S_2|F_1)$. But, in general, our assumptions imply that this equality does not hold. Instead, they imply that $Pr(S_2|S_1) > Pr(S_2|F_1)$, consistent with empirical observations (see Appendix B for a proof).

Returning to the model, the expected cell counts, M_{ijk} , for the outcomes E_1 – E_4 , for age group j and this event i are obtained as the sums of the products along the paths leading to the outcomes, multiplied by the sample size for that combination:

$$M_{ij1} = q_{ij} r_{1ij} r_{2ij} N_{ij}$$

$$M_{ij2} = q_{ij} r_{1ij} (1 - r_{2ij}) N_{ij}$$

$$M_{ij3} = q_{ij} (1 - r_{1ij}) r_{2ij} N_{ij}$$

$$M_{ij4} = [1 - q_{ij} (r_{1ij} + r_{2ij} - r_{1ij} r_{2ij})] N_{ij}.$$
(1)

The fit of the model with specific parameter values for event i and age j can be assessed using an appropriate goodness-of-fit statistic to compare the expected and the observed counts. The power divergence statistic (Read & Cressie, 1988) serves this purpose well (see Hu

& Batchelder, 1994). In the absence of a priori parameter values, we use minimum power divergence estimators by substituting Equations 1 into the following power divergence statistic (Equation 2 below) and minimizing it with respect to our model parameters:²

$$2P^{\lambda}(N:M)_{ij} = \frac{2}{\lambda(\lambda+1)} \sum_{k=1}^{4} N_{ijk} \left[\left(\frac{N_{ijk}}{M_{ijk}} \right)^{\lambda} - 1 \right], \quad (2)$$

where N_{ijk} is the observed count in cell k. Read and Cressie (1988) discuss this family of goodness-of-fit statistics, which include the Pearson X^2 and likelihood ratio statistics as particular cases ($\lambda=1$ and $\lambda=0$, respectively). They recommend using $\lambda=2/3$ when the observed counts are low, as they are here, and the expected discrepancy between observed cell counts and the expected number under the model are unknown. The resulting fit measure is asymptotically distributed as χ^2 , with degrees of freedom, η_{ij} , equal to the difference between the number of independent cells in the data and the number of estimated parameters in the model.

Equation 2 is the power divergent statistic for estimating parameters and testing the model on a single event, i, and a single age group, j. A given model can be tested over all events and ages simultaneously by assuming event and group independence and summing the separate power divergence statistics from Equation 2 into an overall goodness-of-fit measure,

$$2P^{\lambda} = \sum_{i=1}^{I} \sum_{j=A}^{C} 2P^{\lambda}(N:M)_{ij}.$$
 (3)

Under the null hypothesis of a perfect model fit, $2P_0^{\lambda}$ is asymptotically χ^2 distributed with degrees of freedom,

$$\eta = \sum_{i=1}^{I} \sum_{j=A}^{C} \eta_{if}. \tag{4}$$

The numerical value for the minimized power divergence statistic in Equation 3 with degrees of freedom in Equation 4 is then the goodness-of-fit measure for comparing that model with the data. It can also be used to as. All calculations were done using a computer program written by Hu (1991) to estimate parameters for general processing tree models by minimizing Equation 2 using an EM (expectation-maximization) algorithm. This approach is discussed in detail by Hu and Batchelder (1994).

Nested Models

In the case of our full model, the number of independent counts in the data equals the number of parameters needed to specify the model, thus leaving the model untestable (i.e., there are 4 counts, or 3 degrees of freedom in the data, for each age—item combination and 3 parameters in the full model for that combination). The fit of the full model is perfect, then, because it is only a rewrite of the data in terms of the model. Testable models are achieved by imposing restrictions on the parame-

ters so that their number is less than the degrees of freedom in the data. For example, one might test the hypothesis that the subjects are equally as likely to retrieve the memory of an event at Time 1 as at Time 2 by setting $r_{1ij} = r_{2ij} = r_{ij}$ and substituting in Equations 1. For each age \times item combination, this model has 2 parameters, q_{ij} and r_{ij} , to predict data with 3 degrees of freedom, leaving 3I degrees of freedom (1 degree of freedom for each of the 3 age group \times I item combinations) for the statistical test.

The hierarchical family of models. A hierarchy of testable models can be defined by imposing increasingly restrictive constraints on the parameter space through setting certain parameters equal, corresponding to particular psychological hypotheses. These hypotheses can be investigated by testing these nested models against the data or against each other by means of χ^2 difference tests.

A hierarchy of models for a given event, i, obtained by imposing successive constraints on the parameters is shown in Figure 2. The subscript i has been omitted for clarity in the figure. At the top is the full model, allowing all 9 parameters per item over the 3 age groups in the model. Model 1 represents the hypothesis that encoding probability is age-invariant—that is, $q_{iA} = q_{iB} = q_{iC} = q_i$. Model 2 represents the hypothesis that retrieval reporting is unaffected by delay—that is, $r_{1ij} = r_{2ij} = r_{ij}$, for j = A, B, and C, respectively. Model 3 assumes that retrieval reporting at each delay is independent of age—that is, $r_{1iA} = r_{1iB} = r_{1iC} = r_{i1}$, and similarly for the r_{2ij} .

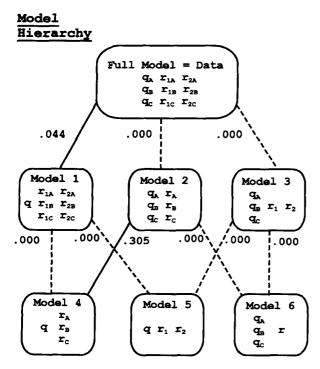


Figure 2. Diagram of the model hierarchy for a single item. Lines indicate nested models, with best fits connected by bold lines and others by dashed lines. The numbers along the lines connecting the various models represent the p values for the χ^2 difference test between the respective two model fits.

Given the nonrejection of one or more of these models, we can test the imposition of further assumptions. Models 4, 5, and 6 are obtained by combining the constraints from two of the three models above them. Model 4 represents the joint hypothesis that encoding is age-invariant and retrieval reporting is delay-independent. Model 5 represents the joint hypothesis that encoding and retrieval reporting are both constant over age. Model 6 represents the joint hypothesis that retrieval reporting is independent of both age and delay.

RESULTS

Twenty-eight features were available to be checked by the health-care providers or the child's parent as having occurred during the physical examination. Of these, analyses were limited to the I = 15 events, for which outcomes E₁-E₄ occurred with nonzero frequencies and with sufficient numbers within all three age groups (see below for a discussion of possible confounds). Seven items were dropped due to insufficient usage, and six items were dropped due to zero frequencies. The frequency distributions for the items included are shown in Table 1. A few points can be noted from the distributions prior to applying the model-based analyses. First, performance varied substantially over the 15 events, indicating that the features were indeed processed differentially at one or more levels in the memory system. For example, most 7-year-olds reported the weight check and failed to report the genital check on both recall occasions. This fact can be seen by contrasting the predominance of outcome SS and the single occurrence of outcome FF for the 7-year-olds' recall of the weight check with the reverse pattern for their recall of the genital check. Other such contrasts can be noted as well. A second feature to see in Table 1 is the general performance improvement with age: for most (but not all) events, SS frequency increases and FF frequency decreases with age. Note next the substantial occurrence of outcomes SF and FS within each age group, suggesting that at least some failures to freely recall are due to failures to retrieve and report encoded information. Finally, note that FS is less frequent than SF within each age group, suggesting that retrieval at Time 2 is less likely than at Time 1.

A precise interpretation of the data in terms of encoding and retrieval-reporting probabilities can be achieved only within the context of the models. Accordingly, minimum power divergence estimates of the parameters along with model fits for each of Models 1–6 were obtained for each event by means of Hu's (1991) program, mentioned above.

The model fits to the data, along with their p values for each of the six models broken down by item, are found in Table 2. Higher p values indicate better fitting models. Note that, for Model 1, only two items—chest exam and knee reflex—have p values less than .05. Over half of the items have p values greater than .30. The number of items with p values less than .05 for Models 2–6 are 4, 5, 5, 10, and 6, respectively. The last row of the

Table 1	
Number of Observations per Exam Item in Each Response Pattern by Ag	e

		Ag	ge 3			Ag	ge 5 Age 7				ge 7	
Item	SS	SF	FS	FF	SS	SF	FS	FF	SS	SF	FS	FF
Weight check	1	5	6	35	17	16	3	13	30	8	4	1
Sticker given	18	14	4	9	22	12	9	3	17	6	5	6
Ear exam	3	12	4	28	16	5	15	13	29	7	4	3
TB test	9	12	3	7	8	16	2	13	12	9	2	3
Mouth exam	2	11	3	31	12	15	3	19	24	9	5	5
Chest exam	5	7	3	32	10	13	6	20	21	9	7	6
Hearing test	2	3	3	16	5	16	5	23	17	6	3	5
Vision test	1	1	1	18	13	11	6	17	25	6	4	6
Blood pressure	1	3	2	40	2	6	3	38	5	16	5	17
Abdominal exam	3	9	5	30	9	13	8	19	14	7	8	14
Knee (reflex)	9	11	4	16	22	4	3	17	28	6	4	3
Walk on toes	1	1	1	8	5	4	1	7	6	1	2	3
Photo taken	1	5	3	23	1	5	1	34	2	3	3	19
Genital exam	6	6	5	28	8	6	3	32	6	5	8	23
Urine sample	1	1	2	33	1	2	4	34	8	5	3	25

Note—SS, successfully recalled the exam both immediately and after the delay. SF, successfully recalled the exam immediately but not after the delay. FS, failed to recall the exam immediately but succeeded after the delay. FF, failed to recall the exam both immediately and after the delay.

table, which has all 15 items combined for each model and the p values for the combined power divergence statistics, provides the basis for the model comparisons.

The model comparison results are represented in Figure 2. The tail p value for the difference χ^2 resulting from the comparison of each model to one above it in the hierarchy is shown alongside the line connecting the two models in the figure. The p values are given, instead of the power divergence statistics, because the values of the latter are not comparable when the degrees of freedom differ for the χ^2 distribution to which they belong.

Note first that Model 1, which assumes no age-related encoding differences, gives fits substantially better than does Model 2, which assumes no delay-related differences in retrieval, and than does Model 3, which assumes

no age-related differences in retrieval. On these grounds, we conclude that age-related encoding effects, if any, are of less consequence than are age or delay effects on retrieval. Moving to the next level of models, we see that Model 4, which allows age differences in retrieval with no other effects, differs nonsignificantly from Model 2. All the other model comparisons show significant differences. Note that Model 4 differs from Model 2 by the same constraint as does Model 1 from the full model. The same constraint also distinguishes Model 5 from Model 3. Each of these immediately nested comparisons presents a test of the hypothesis of no age effects on encoding, but in the presence of different presuppositions. The hypothesis accrues reasonable support in two out of the three tests (Model 1 vs. the full model, Model 4 vs.

Table 2
Power Divergence Statistics (PDS) and p Values, With $\lambda = 2/3$, by Item and Combined Over Items for Each of the Six Models

	Mod	el I	Mod	lel 2	Mode	el 3	Mode	14	Mode	1 5	Mode	el 6
Items	PDS	p	PDS	\overline{p}	PDS	p	PDS	p	PDS	p	PDS	p
Weight check	4.58	.10	10.29	.02	23.71	.00	11.52	.04	72.16	.00	29.29	.00
Sticker given	3.92	.14	6.56	.09	3.52	.47	9.15	.10	6.99	.32	7.70	.17
Ear exam	1.58	.45	9.76	.02	29.57	.00	10.12	.07	56.93	.00	29.63	.00
TB test	1.07	.59	20.70	.00	2.97	.56	20.70	.00	7.18	.30	25.13	.00
Mouth exam	3.51	.17	13.58	.00	13.59	.01	14.42	.01	40.81	.00	27.85	.00
Chest exam	6.22	.04	4.41	.22	4.87	.30	9.25	.10	32.69	.00	9.10	.11
Hearing test	0.87	.65	6.77	.08	14.10	.01	8.07	.15	29.86	.00	20.14	.00
Vision test	3.58	.17	1.87	.60	6.17	.19	5.17	.40	36.20	.00	8.00	.16
Blood pressure	3.26	.20	7.39	.06	0.72	.95	10.84	.05	25.95	.00	7.64	.18
Abdominal exam	0.33	.85	2.39	.49	5.90	.21	2.77	.74	15.92	.01	7.45	.19
Knee (reflex)	10.78	.00	3.79	.29	11.22	.02	14.53	.01	24.43	.00	15.43	.01
Walk on toes	1.22	.54	2.11	.55	2.86	.58	3.13	.68	8.46	.21	3.33	.65
Photo taken	0.78	.68	3.11	.37	2.10	.72	3.54	.62	3.87	.69	3.81	.58
Genital exam	1.96	.37	1.78	.62	2.69	.61	3.66	.60	3.77	.71	2.69	.75
Urine sample	1.21	.55	1.49	.68	4.38	.36	2.53	.77	13.87	.03	4.54	.47
Combined items	44.87	.04	95.98	.00	128.3	.00	129.4	.00	379.11	.00	201.75	.00

Note—The degrees of freedom for Models 1, 2, 3, 4, 5, and 6 were 2, 3, 4, 5, 6, and 5, respectively. Each of these degrees of freedom refers to its application to each individual item. The degrees of freedom for the combined items are 15 times those of the individual items for each model.

Model 2, and Model 5 vs. Model 3), while all other constraints were rejected regardless of the assumptions under which they were tested. We conclude from this entire pattern of results that the data best support the absence of age-related encoding effects. It is likely that retrieval depends on both delay and age. Thus, Model 1 provides the most satisfactory representation of these data.

It should be noted that nothing in the data structure or model structure forces one model to fit better than another. To check this, we constructed reasonable data sets that were best fit by alternative models. Take, for example, the following three fairly similar data sets, each about the same size as those encountered in our analysis, which are best fit by different models: $D_{\text{ex.1}} = (18,8,4,15,20,9,5,11,23,10,6,7)$ is best fit by Model 3; $D_{\text{ex.2}} = (11,8,8,19,18,8,8,12,26,6,6,6)$ is best fit by Model 2; and $D_{\text{ex.3}} = (15,10,6,13,20,9,5,11,26,6,3,10)$ is best fit by Model 1.

In all three hypothetical cases, recall increases with age, with the proportion of SS events (Components 1, 5, and 9 of the data vector) increasing and the proportion of FF events (Components 4, 8, and 12) decreasing. Thus, different data sets, each reasonable, can be fit better by different models, showing that our results are not artifactual. One can also note in Table 2 that certain items are better fit by one model than by another: hearing test by Model 1, knee reflex by Model 2, and sticker by Model 3.

To check whether the findings in this study were due to low power, we conducted post hoc power analyses of our model comparisons. To look at the power for testing Model 1 against the full model, 1,000 data sets were generated in accord with the full model parameter estimates for the 15 items. The error variability was introduced using the multinomial distribution. Model 1 was fit for each sample, and power was calculated from the 1,000 fit statistics combined over the 15 items. This procedure was repeated for several different sample sizes: 30, 45,

60, and 90 per age group, with the 45 sample being closest to the size of our actual data set. The results showed very good power: .874 for n = 30, .997 for n = 45, .999 for n = 60, and .9995 for n = 90.

Similarly, we checked the Type I error rates for the same sample sizes and found them all to be in line. To carry out this check, we generated random data with Model 1 parameters and then fit the vectors with the same model. The Type I error rates with α set to .05 were .01 for n = 30, .031 for n = 45, .030 for n = 60, and .033 for n = 90.

Moving to the parameter estimates, Table 3 displays the Model 1 estimates (with 95% confidence bounds) for the 15 events, listed in order of decreasing probability of encoding. In light of the lack of experimental control with which the features of the physical examination were administered, any interpretation of the parameter estimates must be cautious. Nevertheless, the generally very high encoding probabilities are remarkable. For 12 of the 15 events, they are .69 or greater. The exceptions are the unusual event of having one's picture taken during the examination, the genital exam, and the collection of a urine specimen. Model 1 allowed age differences in retrieval, but it did not constrain them to be in a particular order. Nevertheless, for 9 of the 15 events, retrieval probabilities at both Time 1 and Time 2 increased monotonically with age. The exceptions are the events of receiving a sticker (as a reward), the TB test, the hearing test, walking on one's toes, having one's picture taken, and the genital exam. Given the size of the confidence intervals, the retrieval probabilities across the three age groups might be described as flat in many of these cases; even so, the point estimates were, with only two exceptions (receiving the sticker and genital exam), always greater at age 7 than at age 3. We will consider the parameter estimates further in the Discussion.

Table 3

Model 1 Parameter Estimates With 95% Confidence Intervals
(Estimated With Power Divergence Statistic, Lambda = 2/3)
Ordered by the Encoding Parameter Estimates

		Retrieval Reporting									
		Ag	ge 3	Ag	ge 5	Age 7					
Item	Encoding	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2				
Weight check	.98±.08	.13±.08	.15±.09	.68±.14	.43±.12	.89±.08	.80±.10				
Sticker given	$.94 \pm .08$.77±.14	$.53 \pm .14$	$.76 \pm .11$	$.69 \pm .11$	$.73 \pm .16$	$.70 \pm .16$				
Ear exam	$.93 \pm .08$	$.35 \pm .12$	$.16 \pm .09$	$.47 \pm .13$	$.69 \pm .14$	$.89 \pm .09$	$.81 \pm .10$				
TB test	$.90 \pm .12$	$.75 \pm .17$	$.43.\pm 16$	$.69 \pm .17$	$.29 \pm .13$	$.87 \pm .13$	$.58 \pm .17$				
Mouth exam	.88±.11	$.32 \pm .13$	$.12 \pm .09$	$.64 \pm .17$	$.36 \pm .13$	$.84 \pm .10$	$.74 \pm .12$				
Chest exam	$.87 \pm .12$	$.31 \pm .13$	$.22 \pm .11$	$.54 \pm .16$	$.38 \pm .14$	$.77 \pm .12$	$.72 \pm .13$				
Hearing test	$.86 \pm .12$	$.25 \pm .16$.25±.16	$.50 \pm .15$	$.24 \pm .11$	$.85 \pm .13$	$.74 \pm .15$				
Vision test	.84±.10	$.14 \pm .14$.14±.14	$.62 \pm .16$	$.49 \pm .15$	$.87 \pm .10$	$.81 \pm .11$				
Blood pressure	$.83 \pm .28$	$.11 \pm .09$	$.09 \pm .08$	$.20 \pm .13$	$.13 \pm .10$	$.58 \pm .22$.28±.15				
Abdominal exam	.79±.13	$.33 \pm .14$	$.22 \pm .12$	$.56 \pm .16$	$.44 \pm .15$	$.63 \pm .17$	$.66 \pm .17$				
Knee (reflex)	.7 6 ±.08	$.66 \pm .18$	$.43 \pm .16$	$.85 \pm .15$.82±.15	$.89 \pm .08$	$.83 \pm .09$				
Walk on toes	.69±.18	$.29 \pm .30$	$.29 \pm .30$	$.80 \pm .29$	$.54 \pm .28$	$.76 \pm .23$	$.87 \pm .19$				
Photo taken	$.50 \pm .29$	$.37 \pm .29$	$.25 \pm .23$	$.30 \pm .25$	$.11 \pm .13$	$.38 \pm .31$	$.38 \pm .31$				
Genital exam	.49±.11	$.55 \pm .22$	$.50 \pm .21$	$.65 \pm .25$	$.51 \pm .23$	$.49 \pm .19$	$.62 \pm .19$				
Urine sample	$.41 \pm .14$	$.15 \pm .17$	$.22 \pm .20$	$.18 \pm .17$	$.30 \pm .23$	$.74 \pm .21$	$.62 \pm .21$				

DISCUSSION

Both technical and substantive features of this study warrant discussion. The technical features concern aspects of our analyses and model specification, whereas the substantive ones focus on interpretations of the results.

Technical Aspects

Before discussing the interpretations of this study, we consider two aspects of the research that may raise concern. First, with regard to the data, as noted, we could only use items with all nonzero frequencies for the four events. Eliminating some items in this way may introduce a bias. Perhaps those not used were all less memorable and would have yielded substantially different parameters estimates from the items included in the study. We believe that is not the case. The frequency distributions of events E_1 – E_4 for items with all nonzero cells were similar to those excluded except that where an included item's cell count was 1, 2, or 3, the excluded item's was a zero. (An ad hoc adjustment, such as substituting a one for the zero, has a potential to skew the results, which we did not want to risk.)

The second area of concern is the possible misspecifications of the model. One can imagine a number of more general memory models of which this one would be a constrained version. Perhaps ours is a misspecified constraint of one such model. Although we cannot cover every imaginable formulation under which this model could fall, two examples are worth noting.

First, suppose one modeled the retrieval at the second interview as conditionally dependent on retrieval at the first interview (a model that we could not estimate, due to a shortage of degrees of freedom in the data). We have done some simulations with this model in which we lowered the retrieval probabilities (by .1), conditional on a failure at the first interview. In these simulations, we found that when encoding levels are fairly high (.8 and up) and retrieval-reporting levels are greater than .5, then the parameter bias introduced is not greater than ±.05 for any of the parameters. This bias tends to lower the encoding probability a little, while inflating the first retrievalreporting probability, leaving the second retrieval-reporting probability little changed. Thus, for example, for a single age group, if the actual population model and parameters are q = .80, $r_1 = .70$, $r_{2|S} = .70$, and $r_{2|F} = .60$ (where $r_{2|S}$ is now defined as retrieval at the second interview conditional on a successful retrieval at the first interview, and $r_{2|F}$ is the retrieval probability conditional on a prior failure), the resulting estimates using our misspecified model are q = .77, $r_1 = .73$, and $r_2 = .70$.

As noted, our full model is saturated, not allowing us to directly fit a model with such a dependency as above. An anonymous reviewer suggested a more complex version of our model with the same number of parameters but additional paths to perhaps capture some of the possible dependency of the second retrieval on the first. This version adds a second encoding possibility (using the same encoding parameter a second time) following a success-

ful retrieval at the first interview. The second retrieval then consists of independent attempts to retrieve either the original or the new trace. We fit this model and its constrained versions analogous to our Models 1 and 3 (including suggested Submodels 3a and 3b, which equate retrieval parameters across age groups separately for the first and second interviews). The alternative Model 1 fit quite well (p=.49), but none of the other alternative models fit well at all. Thus, this somewhat more complicated version of our model points to exactly the same results—age-related differences in retrieval but not in encoding.

A second more general model is one that allows loss of an encoded item prior to the second retrieval with probability t. We again ran simulations to see the extent to which our model yields biased estimates if, in fact, such a model is correct. If the probability of storage loss is fairly small, then our model estimates the parameters well. However, if the probability is great, our model tends to underestimate the second retrieval parameter. For example if t = .2, q = .9, $r_1 = .8$, and $r_2 = .7$, then our model estimates are q = .9, $r_1 = .8$, and $r_2 = .56$. If t = .1, then q = .9, $r_1 = .8$, and $r_2 = .63$. These simulations are instructive and reassuring. They suggest that our model is fairly robust with respect to deviations from possibly suspect assumptions. In our opinion, this fact strengthens confidence in our primary conclusions.

We also reiterate that the simulations (reported in the Results section), examining the asymptotic properties of the model and model testing, show our results to be robust at reasonable sample sizes. Power and Type I error rates, particularly for the sample sizes encountered in this study, were quite satisfactory. Reifer and Batchelder (1988) also studied the statistical properties of this model and found the parameter estimates to be unbiased and power to be satisfactory for samples of 50 or greater when individual differences were small to moderate. With these concerns about the model and data addressed, we can now attend with more confidence to the substantive findings of this study.

Substantive Aspects

The results of this study are surprising in light of previous modeling efforts, which, as we mentioned earlier, revealed clear developmental differences in encoding and likely differences in retrieval. In contrast, we found the best fitting model to be one in which there were age differences in retrieval only. As we discussed in the introduction, the Baker-Ward et al. (1993) study differed from others in numerous ways, any of which may be responsible for the different pattern of results.

First, perhaps we obtained age differences in retrieval but not in encoding because, unlike other studies, the encoding was incidental and only the recall task was deliberate. That is, the children were asked to recount their experiences, but they were not instructed to remember them at the time they occurred. Indeed, one might argue that the paradigm employed here is the appropriate one for studying children's autobiographical memories for personally experienced events, especially if the conclusions are to be applied to issues of courtroom testimony.

Second, regardless of age, the children had prior experience and knowledge of physical examinations. This fact has implications for encoding in that it is very likely that, though the older children had richer experiences to draw on, the children in all three age groups understood and easily integrated features of the examination into their memory structures. Therefore, the ability to encode features adequately for the recall task was uniformly high across the groups and did not result in an increase in encoding probability with age, as might be expected.

Third, the nature of the events being remembered may have implications for retrieval estimates. Recall that our retrieval parameter is really a multiplicative combination of retrieval probability and the probability of reporting given a retrieval (see Appendix A). Thus, retrieval, as usually thought of in memory research, may in some cases have been more successful than indicated by the parameter estimates. That is, one might speculate that the children censored their reports of potentially embarrassing items that they in fact had retrieved. If so, such censoring might be age-dependent. This, of course, could go in either direction, but it seems reasonable to suppose, in light of most children's experiences, that the probability of censoring increases with age for some items. Thus, we conjecture that if the older children censored their reports of the genitalia examination more frequently than the younger children did, this effect would counterbalance to some degree an age-related increase in actual retrieval. And, in fact, the retrieval parameter for the genital check showed no overall age trend.

Finally, this study differs from many others in that the children are from younger age groups. It may be that the encoding differences others have found have not yet developed in this age range or are too small to detect.

It is interesting to see what specific interpretations can be put on the best fitting model's parameter estimates. However, we would caution the reader against making too much of the precise parameter estimates we have reported. There are several reasons for this. First, because the sample sizes are small, the 95% confidence intervals around the point estimates are large. Second, the children may have encoded features of the physical examination into configured events different from the ones we used for purposes of analysis. Third, the data on which the parameter estimates are based were obtained over a large number of circumstances—two different clinics, 12 physicians, and 12 nurses—any of which can affect the values. And finally, the parameters represent performance averages for which individual differences might be great.

These cautions notwithstanding, there may yet be some cases where specific parameter values are broadly interpretable. Consider, for example, the items with low versus high encoding probability. The high encoding probability items are all better known or more salient aspects of a physical exam, whereas the low encoding items are

more peripheral or not generally a part of a physical checkup. Encoding seems to be tied to familiarity.

As for the retrieval-reporting parameters, 9 out of 15 show age-related increases for both retrieval opportunities. This result is readily interpretable in terms of developmental changes in facility with retrieval operations. Many of the features with the greatest spread in parameter values from the 3- to 7-year-olds (examination of the chest, ears, hearing, mouth, and vision, and the weight check) are very common in a well-child checkup and may form parts of a script (Nelson, 1986) that older children have learned better than younger children. Therefore, their recall performance is better on those events. While we saw that the children's experience of common items led to a uniformly high encoding probability for all three ages, the retrieval of these same commonly used procedures showed age-related differences. This may indicate the relative difficulty in retrieval versus encoding. It seems to be the case that the benefits of age are more crucial in retrieval than in encoding for good performance.

In looking at the model's parameter estimates for various features, the blood pressure item presents some anomalous results. It has a high probability of encoding (.83) but low retrieval-reporting probabilities (.11, .20, .58, for 3-, 5-, and 7-year-olds, respectively, at first interview, and .09, .13, and .28, respectively, at the second interview). Is this pattern sensible? Although a blood pressure test is quite involved and extended, presumably facilitating the encoding process, it may be quite difficult for a young child to retrieve and organize the component parts of this little-understood procedure or to put it into words. While not wanting to put too much into these parameter values, it is important to see that they are, for the most part, sensible, thus lending credibility to the model.

All these points about the parameter estimates lead to the possibility that some of the component features of the physical examination fit into children's developing scripts for visits to the doctor. How could this type of prior knowledge about checkups affect the children's performance? In the extreme case, of course, the children could rely on their knowledge about office visits to fill in the gaps (possibly unconsciously) about what happened at their physical examination. Nonetheless, the lack of intrusion errors and the children's effective responses to questions about events that did not take place in the checkup led Baker-Ward et al. (1993) to argue that there was little evidence for such confabulation. But there are other ways in which children's scripts could influence memory performance, and, in the context of the present model, we can ask if prior knowledge enhances encoding operations, retrieval processes, or both.

To examine the possible linkage between underlying knowledge and the encoding and retrieval parameters of Model 1, we used normative data on young children's knowledge of physical examinations. These data were obtained by Clubb, Nida, Merritt, and Ornstein (1993), who asked 5-year-olds to provide information about what usually happens when they go to the doctor for a checkup. For each feature of the physical examination,

Clubb et al. calculated a "knowledge score," which was the proportion of the children in the normative sample who nominated that item. We correlated the encoding and retrieval parameters displayed in Table 2 with Clubb et al.'s knowledge scores for those 15 features of the checkup. The results were that the knowledge scores correlated significantly with the age-invariant encoding probabilities (r=.65, p<.01) and with the retrieval-reporting probabilities for the 5-year-olds at the second interview (r=.52, p<.05). None of the other correlations approached significance. Thus, it seems likely that the children's underlying knowledge substantially influenced their encoding and perhaps, in some cases, their retrieval and reporting of the physical examination.

One final note about the parameter estimates: Because the retrieval-reporting parameter is a combination of two processes, it should not be compared with other model estimates of retrieval alone. Moreover, though we believe the reporting probability is generally quite high in this situation and that age differences in the retrieval-reporting parameter are due primarily to retrieval differences, we cannot rule out the possibility that these differences are due simply to differences in the probability of reporting.

To summarize, the model provided important insight into children's memory processes by locating the source of age-related differences in recall in retrieval processes. While our results seem surprising on the surface, there may be unique aspects of this study that explain its divergence from others.

REFERENCES

- BAKER-WARD, L., GORDON, B. N., ORNSTEIN, P. A., LARUS, D. M., & CLUBB, P. A. (1993). Young children's long-term retention of a pediatric examination. *Child Development*, **64**, 1519-1533.
- BATCHELDER, W. H., & RIEFER, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. British Journal of Mathematical & Statistical Psychology, 39, 129-149.
- BRAINERD, C. J. (1985). Model-based approaches to storage and retrieval development. In C. J. Brainerd & M. Pressley (Eds.), Basic processes in memory development: Progress in cognitive development research (pp. 143-207). New York: Springer-Verlag.
- Brainerd, C. J., & Ornstein, P. A. (1991). Children's memory for witnessed events: The developmental backdrop. In J. Doris (Ed.), The suggestibility of children's memory (pp. 10-20). Washington, DC: American Psychological Association.
- BRAINERD, C. J., REYNA, V. F., Howe, M. L., & KINGMA, J. (1990). The development of forgetting and reminiscence. Monographs of the Society for Research in Child Development, 55(3-4, Serial No. 222).
- CECI, S. J., & Howe, M. J. A. (1978). Age-related differences in free recall as a function of retrieval flexibility. *Journal of Experimental Child Psychology*, 26, 432-442.
- CHECHILE, R., & MEYER, D. L. (1976). A Baysian procedure for separately estimating storage and retrieval components of forgetting. Journal of Mathematical Psychology, 13, 269-295.
- CHECHILE, R. A., RICHMAN, C. L., TOPINKA, C., & EHRENSBECK, K. (1981). A developmental study of the storage and retrieval of information. *Child Development*, 52, 251-259.
- CLUBB, P. A., NIDA, R. E., MERRITT, K. A., & ORNSTEIN, P. A. (1993).
 Visiting the doctor: Children's knowledge and memory. Cognitive Development, 8, 361-372.

- GREENO, J. G., JAMES, C. T., DAPOLITO, F., & POLSON, P. G. (1978). Associate learning: A cognitive analysis. Englewood Cliffs, NJ: Prentice-Hall.
- Hu, X. (1991). General program for processing tree models, Version 1.0 [Computer program]. Irvine: University of California.
- Hu, X., & BATCHELDER, W. H. (1994). Statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59, 21-47.
- KAIL, R., HALE, C. A., LEONARD, L. B., & NIPPOLD, M. A. (1984). Lexical storage and retrieval in language-impaired children. Applied Psycholinguistics, 5, 37-49.
- NELSON, K. A. (1986). Event knowledge: Structure and function in development. Hillsdale, NJ: Erlbaum.
- ORNSTEIN, P. A., GORDON, B. N., & BAKER-WARD, L. (1992). Children's memory for salient events: Implications for testimony. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), Development of longterm retention (pp. 135-158). New York: Springer-Verlag.
- Ornstein, P. A., Gordon, B. N., Baker-Ward, L. E., & Merritt, K. A. (in press). Children's memory for medical experiences: Implications for testimony. In D. P. Peters (Ed.), *The child witness in context: Cognitive, social, and legal perspectives*. Dordrecht, The Netherlands: Kluwer.
- Ornstein, P. A., Gordon, B. N., & Larus, D. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology*, 6, 49-60.
- ORNSTEIN, P. A., & NAUS, M. J. (1978). Rehearsal processes in children's memory. In P. A. Ornstein (Ed.), Memory development in children (pp. 69-99). Hillsdale, NJ: Erlbaum.
- READ, T. T. C., & Cressie, N. A. C. (1988). Goodness-of-fit statistics for discrete multivariate data. New York: Springer-Verlag.
- RIEFER, D. M., & BATCHELDER, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95, 318-339.
- RIEFER, D. M., & BATCHELDER, W. H. (1991). Age differences in storage and retrieval: A multinomial modeling analysis. *Bulletin of the Psychonomic Society*, 29, 415-418.
- RIEFER, D. M., & BATCHELDER, W. H. (1995). A multinomial modeling analysis of the recognition-failure paradigm. *Memory & Cognition*, 23, 611-630.
- Schneider, W., & Pressley, M. (1989). Memory development between 2 and 20. New York: Springer-Verlag.
- WILKINSON, A. C., DEMARINIS, M., & RILEY, S. J. (1983). Developmental and individual differences in rapid remembering. Child Development. 54, 898-911.
- WILKINSON, A. C., & KOESTLER, R. (1983). Repeated recall: A new model and tests of its generality from childhood to old age. *Journal* of Experimental Psychology: General, 112, 423-451.
- WILKINSON, A. C., & KOESTLER, R. (1984). Generality of a Markov model for repeated recall. *Journal of Mathematical Psychology*, 28, 43-72.

NOTES

- 1. We assume that the recall behavior of one child is not affected by that of another, given the private checkups and interviews. The only case where independence might be questioned is in the case of a very few families with more than one child in the study.
- 2. Note that Read and Cressie (1988) denote this statistic with $2I^{\lambda}(N:M)$, signifying its connection with information theory. We substituted a P for the I in this equation to avoid confusion with our use of I to represent the total number of events.
- 3. See Read and Cressie (1988, p. 15) for the technical details involved in defining the statistic for $\lambda = 0$.

APPENDIX A

To understand why distinct retrieval and reporting probabilities cannot be identified for this paradigm, consider the model for one age group and a single feature. If we assume a separate

retrieval (r) and reporting (s) parameter and a single recall opportunity, then the probability of successful recall is

$$P(S) = qrs. (A1)$$

The probability of a recall failure, P(F), is the complement of P(S), or P(F) = 1 - qrs.

Note that, in this simple example, the three parameters, q, r, and s, are not separately identifiable. The maximum likelihood estimate of their product in Equation A1 is the proportion of items recalled, but nothing in the data allows estimates of the individual components.

In the case of two recall opportunities, as in this study, the model equations are:

$$P(SS) = q(r_1s_1)(r_2s_2)$$

$$P(SF) = q(r_1s_1)(1 - r_2s_2)$$

$$P(FS) = q(r_2s_2)(1 - r_1s_1)$$

$$P(FF) = 1 - P(SS) - P(SF) - P(FS).$$
(A2)

Note that the encoding parameter can now be identified, but the retrieval and reporting parameters still cannot. These two parameters continue to appear together as a product in Equation A2 and therefore cannot be separately estimated. Thus, they are not separately identifiable.

APPENDIX B

To understand why $Pr(S_2|S_1)$ does not generally equal $Pr(S_2|F_1)$, note that

$$\Pr(S_2|S_1)=r_2,$$

and

$$\Pr(S_2|F_1) = [q(1-r_1)r_2]/[q(1-r_1)+(1-q)].$$

These two expressions are equal if and only if

$$\frac{q(1-r_1)}{q(1-r_1)+(1-q)}=1,$$

that is, if and only if q=1, the situation where encoding is perfect. In other words, if q<1, we expect $\Pr(S_2|S_1)>\Pr(S_2|F_1)$. This assumption of conditional independence therefore implies that dependencies should be observed in the data unless encoding is perfect.

(Manuscript received February 21, 1995; revision accepted for publication December 30, 1995.)

Nominations for the Editorship of Memory & Cognition

Nominations are solicited for the editorship of *Memory & Cognition*. The term of the present editor, Geoffrey R. Loftus, expires at the end of 1997. The new editor will begin an official 4-year term on January 1, 1998, and will begin to receive manuscripts early in January 1997. The Publications Committee of the Psychonomic Society expects to appoint the new editor by November, 1996.

Nominations (including self-nominations) should be submitted by August 1, 1996, to:

Keith Rayner Chair, Memory & Cognition Search Committee Psychology, Tobin Hall University of Massachusetts Amherst, MA 01003