Age differences in hypothesis testing and frequency processing in concept learning

RONALD T. KELLOGG
University of Missouri, Rolla, Missouri 65401

Two types of processes, controlled hypothesis testing and automatic frequency processing, have been posited to explain concept learning. The present study employed a concept-identification task to compare fifth-graders, college students, and elderly adults in their ability to test hypotheses and compile feature frequency information. College students proved superior in both their efficiency of selecting the correct hypothesis and their accuracy of recalling sampled hypotheses, and fifth-graders and elderly adults performed about the same on these measures of hypothesis-testing ability. Ability to estimate feature frequencies followed the same developmental pattern. The results support the view that hypothesis testing is an effortful, controlled process, but they contradict the assumption that frequency processing is automatic.

Hypothesis testing and feature frequency-of-occurrence processing have historically been viewed as two competing explanations of concept learning (Levine, 1975). Recently, Kellogg (1982) proposed a dual process model assuming that frequency or associative strength processing is automatic and effortless, whereas hypothesis testing is controlled and attention demanding. According to this view, frequency processing should occur unconsciously and unintentionally whenever a person perceives instances of a concept. Hypothesis testing, on the other hand, should occur only when the person intentionally devotes conscious effort to sampling, testing, and storing hypotheses in short-term memory. The present study examined whether these two types of learning processes are subject to different developmental trends across fifth-grade children, college-aged adults, and elderly adults.

Zacks, Hasher, and Sanft (1982, p. 107) argued that compiling event frequency is "an automatic operation for which people may be genetically prepared." A high degree of skill in processing event frequency may develop during infancy and persist over the human life span. Supporting this, some studies have concluded that grade-school children, college students, and elderly adults estimate equally well the frequency of occurrence of words in a memory experiment (Attig & Hasher, 1980; Hasher & Zacks, 1979; Kausler & Puckett, 1980). In a concept learning task, college students can estimate the relative frequency of stimulus features among instances of the concept being learned (Kellogg & Dowdy, Note 1). But whether grade-school children and

elderly adults are as sensitive to feature frequency information in this task as are college-aged adults has apparently not been examined before.

Hasher and Zacks (1979) argued that controlled, effortful processes follow an inverted U-shaped developmental function because they are dependent on attention. Attentional capacity, they assumed, increases from infancy to adulthood and then decreases during old age. Some evidence suggests that the ability to test hypotheses follows a similar function. It is well documented that hypothesis-testing skills improve from childhood to early adulthood (Gholson, Levine, & Phillips, 1972; Phillips & Levine, 1975). Offenbach (1974) concluded that elderly subjects test hypotheses more poorly than even kindergarten children. He found that the older adults (mean age = 74.8 years) often seemed to forget the hypothesis that they had sampled on the previous trials of a concept-identification task. Although such a decline in hypothesis testing fits well with dual process theory, additional data confirming Offenbach's conclusion are needed.

The ability to estimate feature frequencies and to test hypotheses was assessed among fifth-graders, college students, and elderly adults in the present experiment. Subjects attempted to solve a concept-identification task in which a stimulus was classified and a current hypothesis was stated on each trial. The trial of last error (TLE) in hypothesizing the defining feature of the concept and the number of previously tested hypotheses that were correctly recalled measured the subject's skill at testing hypotheses (Kellogg, 1980; Levine, 1975). After a fixed number of learning trials, subjects estimated how often particular features had occurred among positive instances of the concept. Following this assessment of frequency estimation ability, subjects classified new positive and negative instances of the concept. Their ability to classify correctly new instances provided a general mea-

sure of concept learning, reflecting prior hypothesis testing (Levine, 1975), frequency processing (Bourne, Ekstrand, Lovallo, Kellogg, Hiew, & Yaroush, 1976), or a mixture of both (Kellogg, 1982).

If the developmental assumptions of the dual process model are correct, then the measures of hypothesis-testing ability should show an advantage for college-aged adults. In contrast, all age groups should provide identical frequency estimates that increase as actual frequency increases; no age differences are expected. Finally, classification performance, if it is mediated by both automatic and controlled processes, should reveal slight age differences favoring college students over fifth-graders and elderly adults.

METHOD

Subjects

The youngest subjects were fifth-graders (range = 11-12 years) in the Rolla Public School. The college-aged adults (range = 18-21 years) were undergraduate students enrolled at the University of Missouri–Rolla. The elderly adults (range = 65-81 years; mean age = 72.9 years) lived in a retirement apartment complex in the Rolla area. A brief screening questionnaire was used to insure that all elderly subjects were in good health and without uncorrected vision problems. There were 13 subjects in each age group, and all subjects received $4 as payment for their participation. Subjects were tested in small subgroups (n = 3-5) of people of the same age level.

Procedure

Subjects were instructed that the experiment involved a game in which they were to learn how to classify letter stimuli as positive or negative instances of a concept. Each stimulus consisted of six letters or dimensions; the features on each dimension were upper- and lowercase. The letters were arranged in the shape of a hexagon. A sample of stimuli was shown to subjects to familiarize them with the letters used (e.g., F, j, N, R, b, g).

The concept that they were to learn was defined by 1 of the 12 features. (For six of the subjects in each age group, G was relevant, and for the others, r defined the concept.) The relevant feature was called the winning letter. Subjects were instructed to learn how to classify stimuli correctly by discovering the winning letter. A positive instance of the concept contained the winning letter, whereas a negative instance did not. By attending to which stimuli the experimenter identified as positive and which negative, therefore, the subject could learn the concept. Examples of positive and negative instances of concepts defined by various winning letters were presented.

Next, the details of the study procedure were given. There were a total of 40 study trials. On each trial, a stimulus was presented using a slide projector. Half of the trials used positive instances of the concept to be learned and half negative, determined by a single random order used for all subjects. The stimulus was shown for 10 sec, and the experimenter informed subjects whether it was a positive or negative instance at the beginning of this interval. A blank slide then appeared, and the experimenter said, "Answer." Subjects were to write their best hypothesis or guess as to what the winning letter was on a slip of paper. Each subject had a stack of these slips that were ordered and numbered from 1 to 40 and a large envelope. After writing down the guess during a 10-sec response interval, he or she placed the slip into the envelope. This procedure prevented subjects from looking back at previous guesses in order to learn the concept or to remember their previous hypothesis.

For half of the trials, the hypothesis statement ended the trial. For the other half, a slide with the phrase "Remember Previous Guess" appeared next for 10 sec. This was the cue to write on the next slip of paper (labeled with the previous trial number and a prime superscript) exactly what they had written on the one just before and to place this recall slip into the envelope. When these hypothesis memory probes occurred was randomly determined (one order for all subjects), with the constraint that half followed positive- and half followed negative-instance trials.

Subjects were instructed to recall their previous guesses as well as they could, but to devote most of their attention to the main task of discovering the relevant letter. They were informed that they would later be asked to classify stimuli without the experimenter's providing feedback as to the correct classification. Hence, it was important that they attend carefully to their main task of learning the concept.

The next phase of the experiment was explained after the study phase was completed. Subjects were told to estimate how often each of the 12 features occurred among the positive instances that were presented on learning trials. Beside each feature on a test sheet was a scale with the following values: 0%, 20%, 40%, 60%, 80%, and 100%. Subjects circled the value corresponding to the percentage of positive study instances that exhibited the letter in question. No feedback was given. The winning feature of the relevant dimension (e.g., G) had actually occurred among 100% of these stimuli, the other feature of the relevant dimension (g) had occurred among none of them, and the other 10 irrelevant features had occurred among 50% of them. The frequency estimation test was self-paced.

Finally, subjects were asked to classify 20 stimuli without feedback according to whatever concept they had learned earlier. For each stimulus, they circled either positive or negative on a test sheet during a 10-sec response interval. A single random order of positive and negative instances was used for all subjects.

Design

Age was crossed with different within-subjects factors during three phases of the experiment. During the learning trials, subjects were asked to recall the hypothesis they gave on the immediately preceding trial. Half of the time a positive instance had been presented on the preceding trial, and half the time, a negative instance. On the frequency estimation test, features that had occurred among 0%, 50%, or 100% of the positive study instances were presented. Finally, on the classification test, half of the stimuli were positive instances and half were negative.

RESULTS

Hypothesis Testing

Hypothesis-testing ability was assessed in terms of the TLE in stating the relevant feature as the current hypothesis and in terms of hypothesis recall accuracy. The TLE means for the fifth-graders, college students, and elderly adults were 26.39, 18.77, and 31.39, respectively. The college students required the fewest trials before they consistently stated the correct hypothesis. An analysis of variance (ANOVA) revealed that the age effect was marginally significant [$F(2,36) = 2.40$, $p < .11$]. A subject was regarded as having learned the concept if his or her TLE was prior to the last trial. There were 9 fifth-graders, 12 college students, and 5 elderly adults who learned the concept, a pattern that again indicates superior performance by the college students.[1]

The hypothesis memory data are presented in Table 1. Regardless of the subject's age, it made little difference whether the hypothesis being recalled was initially stated

Table 1
Mean Number of Correctly Recalled Hypotheses Selected on Positive and Negative Study Trials

Trial Type	Fifth-Graders	College Students	Elderly Adults
Positive	7.92	9.69	8.00
Negative	7.46	9.69	7.92

Note—Perfect recall performance = 10.

Table 2
Means and Standard Deviations of Feature Frequency Estimates

	Fifth-Graders		College Students		Elderly Adults	
AF	Mean	SD	Mean	SD	Mean	SD
0	35.39	27.27	3.08	7.51	30.77	19.35
50	45.56	16.96	45.70	6.93	31.60	13.71
100	76.92	22.87	96.92	11.01	52.31	26.51

Note—AF = actual frequency. Both actual and estimated frequencies are given in terms of the percentage of study instances exhibiting a particular stimulus feature.

Table 3
Mean Number of Correctly Classified Stimuli on Positive and Negative Test Trials

Trial Type	Fifth-Graders	College Students	Elderly Adults
Positive	8.77	10.00	8.08
Negative	8.46	10.00	7.54

Note—Perfect classification performance = 10.

on a positive- or negative-instance trial. Averaged across trial types, the college students (mean = 9.69) recalled more hypotheses than did the fifth-graders (mean = 7.69) and the elderly adults (mean = 7.96). An ANOVA indicated a significant main effect of age [F(2,36) = 4.03, p < .05] and no other significant effects. Newman-Keuls tests showed that all means differed significantly (p < .05) from each other, with the exception of the fifth-grader/elderly adult comparison.

Frequency Processing

The data on frequency estimation are shown in Table 2. These cell means are based on the mean estimates given by each subject to features of the same frequency. Looking first at the mean estimates given by college students, it is clear that they processed the actual frequencies of occurrence of features among positive instances. The mean estimates of the fifth-graders increased as a function of actual frequency, but the 0% level of frequency was markedly overestimated and the 100% level was underestimated. Finally, estimates of the elderly subjects barely increased at all as a function of actual frequency. The ANOVA revealed a significant main effect of age [F(2,36) = 6.75, p < .01], a main effect of actual frequency [F(2,72) = 85.97, p < .001],

and an Age by Frequency interaction [F(4,72) = 14.29, p < .001].

The mean estimates suggest an age-related change in frequency processing. The standard deviations of these estimates, shown in Table 2, further demonstrate this change. At each level of actual frequency, there was much less variability in the estimates given by college students; the fifth-graders and elderly adults showed about the same degree of variability, averaging across all levels of actual frequency.

To clarify the nature of the Age by Frequency interaction, a separate ANOVA was conducted for each age group. The main effect of actual frequency was significant for the fifth-graders [F(2,26) = 14.35, p < .001]. Newman-Keuls tests indicated that the mean estimate for the 100% feature was significantly different (p < .05) from the means for the 0% and 50% features, but that these two were not reliably different. For the college students, the main effect was significant [F(2,26) = 343.66, p < .001], and all levels of actual frequency differed significantly from each other. Finally, the elderly adults showed the same pattern as did the fifth-graders. The main effect of actual frequency was significant [F(2,24) = 4.27, p < .05], the 100% feature was significantly different (p < .05) from the 0% and 50% features, and these two were statistically equal. In short, although all age groups produced reliable main effects of frequency, only the college-aged subjects showed a strong sensitivity to actual frequency.

Classification

Performance on the final classification test is shown in Table 3. The data followed the developmental pattern obtained on all other measures of the present experiment. Across all age groups, positive and negative instances were classified equally well. And, averaged across trial types, college students performed perfectly (mean = 10.00), whereas fifth-graders (mean = 8.62) and elderly adults (mean = 7.81) both correctly classified about 8 of 10 stimuli. The ANOVA revealed a significant main effect of age [F(2,36) = 4.30, p < .05] and no other reliable sources of variance. Newman-Keuls tests showed that only the means for the college students and the elderly adults differed significantly (p < .05).

DISCUSSION

Hypothesis testing, frequency processing, and classification ability followed the same developmental story. Performance improved between fifth-graders and college students and then declined. Fifth-graders and elderly adults yielded essentially the same levels of performance in the present concept-identification task. The age difference in hypothesis-testing ability supports the assumption that hypothesis testing is an effortful, controlled operation. The age difference in frequency processing, however, contradicts the assumption that frequency processing is an effortless, automatic operation.

The previously cited studies showing no age differences in frequency processing oppose the present conclusion (Attig & Hasher, 1980; Hasher & Zacks, 1979; Kausler & Puckett, 1980).

There were many procedural differences between these experiments and the present one. Yet, Warren and Mitchell (1980), using a memory task similar to the others cited here, also found that college-aged adults were significantly more accurate in their frequency judgments than were elderly adults. Thus, whether frequency processing is automatic and invariant across the life span requires further investigation.

REFERENCE NOTE

1. Kellogg, R. T., & Dowdy, J. C. *Controlled hypothesis testing and automatic frequency processing in schema acquisition.* Manuscript submitted for publication, 1983.

REFERENCES

ATTIG, M., & HASHER, L. The processing of frequency of occurrence information by adults. *Journal of Gerontology*, 1980, 35, 66-69.

BOURNE, L. E., JR., EKSTRAND, B. R., LOVALLO, W. R., KELLOGG, R. T., HIEW, C. C., & YAROUSH, R. A. Frequency analysis of attribute identification. *Journal of Experimental Psychology: General,* 1976, 105, 292-312.

GHOLSON, B., LEVINE, M., & PHILLIPS, S. Hypotheses, strategies, and stereotypes in discrimination learning. *Journal of Experimental Child Psychology,* 1972, 13, 423-446.

HASHER, L., & ZACKS, R. T. Automatic and effortful processes in memory. *Journal of Experimental Psychology: General,* 1979, 108, 356-388.

KAUSLER, D. H., & PUCKETT, J. M. Frequency judgments and correlated cognitive abilities in young and elderly adults. *Journal of Gerontology,* 1980, 35, 376-382.

KELLOGG, R. T. Feature frequency and hypothesis testing in the acquisition of rule-governed concepts. *Memory & Cognition,* 1980, 8, 297-303.

KELLOGG, R. T. When can we introspect accurately about mental processes? *Memory & Cognition,* 1982, 10, 141-144.

LEVINE, M. *A cognitive theory of learning.* New York: Erlbaum, 1975.

OFFENBACH, S. I. A developmental study of hypothesis testing and cue selection strategies. *Developmental Psychology,* 1974, 10, 484-490.

PHILLIPS, S., & LEVINE, M. Probing for hypotheses with adults and children: Blank trials and introtacts. *Journal of Experimental Psychology: General,* 1975, 104, 327-354.

WARREN, L. R., & MITCHELL, S. A. Age differences in judging the frequency of events. *Developmental Psychology,* 1980, 16, 116-120.

ZACKS, R. T., HASHER, L., & SANFT, H. Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 1982, 8, 106-116.

NOTE

1. The data of those who learned the concept, based on this TLE criterion, were analyzed separately. The hypothesis recall and frequency estimation measures followed the same pattern for these learners as they did for the analyses reported for all subjects. The classification data, as one might expect, did not follow the same pattern. The learners across all age groups achieved a ceiling level of performance on the final classification test. Once a subject identified the relevant feature, he or she had no trouble classifying both positive and negative instances of the concept.