Journal of Experimental Psychology: General
1984, Vol. 113, No. 2, 282–300

Accuracy of Feeling-of-Knowing Judgments for Predicting Perceptual Identification and Relearning

Thomas O. Nelson and Daniel Gerler
University of Washington

Louis Narens
University of California, Irvine

SUMMARY

The feeling of knowing refers to predictions about subsequent memory performance on previously nonrecalled items. The most frequently investigated type of subsequent performance has been recognition. The present research explored predictive accuracy with two new feeling-of-knowing criterion tests (in addition to recognition): relearning and perceptual identification. In two experiments, people attempted to recall the answers to general-information questions such as, "What is the capital of Australia?", then made feeling-of-knowing predictions for all nonrecalled answers, and finally had a criterion test to assess the accuracy of the feeling-of-knowing predictions. Experiment 1 demonstrated that perceptual identification can be employed successfully as a criterion test for the feeling of knowing. This opens a new way for metamemory research via perception. Moreover, the feeling-of-knowing accuracy for predicting perceptual identification was not significantly correlated with the feeling-of-knowing accuracy for predicting recognition, in accord with the idea that these two tests assess memory differently. Experiment 2 demonstrated that relearning performance can also be predicted by feeling-of-knowing judgments.

Both experiments showed that there is a positive relationship between the feeling of knowing and the amount of time elapsing before a memory search is terminated during recall. Further analyses showed that this relationship is substantial for nonrecalled items for which the person did not guess an answer (omission errors), but the relationship is null or negative for nonrecalled items that the person guessed incorrectly (commission errors).

Several theoretical mechanisms that may underlie the feeling of knowing are proposed.

The first empirical research on the feeling of knowing was the recall–judgment–recognition study by Hart (1965a). First, the subject attempted to recall the answers to general-information questions (e.g., "Which planet is the largest in our solar system?"). This served to provide a subset of questions for which the subject did not recall the correct answers. Next, the subject made feeling-of-knowing judgments on these questions by predicting whether he or she would be able to recognize the correct answers (hereinafter, "he or she" is referred to more briefly as "he"). Finally, to assess the accuracy of the feeling-of-knowing judgments, Hart gave a multiple-choice recognition test on each question (e.g., for the question, "Which planet is the largest in our solar system?" the alternatives were Pluto, Venus, Earth, and Jupiter). A recognition test is useful as a criterion test because it generally is more sensitive than recall for assessing memory (in the sense of differentiating between items that the recall test treats homogeneously in terms of all being nonrecalled), and thus it may be correlated with the feeling of knowing, which presumably is also more sensitive than recall (Hart, 1967a).

This paradigm, or modifications of it, has been used in dozens of feeling-of-knowing experiments. Various subject populations have been examined, including psychology undergraduates (Hart, 1965a), children (Wellman, 1977), the mentally retarded (Brown & Lawton, 1977), geography undergraduates (Gruneberg & Monks, 1974), and older adults (Lachman, Lachman, & Thronesbery, 1979). Various kinds of items have been examined, including general-information questions (Hart, 1965a), laboratory paired associates (Hart, 1967a), nonsense syllables (Blake, 1973), labels for pictured objects (Wellman, 1977), names of famous people (Gruneberg & Sykes, 1978), definitions of words (Gardiner, Craik, & Bleasdale, 1973), names of entertainers (Read & Bruce, 1982), capitals of countries (Gruneberg & Monks, 1974), and names of personal acquaintances (Gruneberg, Smith, & Winfrow, 1973).

The standard finding is that the accuracy of the feeling of knowing for predicting subsequent recognition on nonrecalled items is *intermediate,* being reliably above chance but far from perfect (Blake, 1973). One reason to expect imperfect accuracy is that the criterion test of recognition is "noisy" in at least two ways:

1. Although the difficulty of a recognition test increases with the degree of target-distractor similarity (Bernbach, 1967), functional (as opposed to nominal) target-distractor similarity is probably a multidimensional construct for which some of the potential dimensions may be unknown. Also, the perceived "similarity" is likely to be idiosyncratic and to vary across subjects. Thus, it is possible that sometimes the distractors will be good (in the sense of being similar to the target) for an item judged high in the feeling of knowing and bad for an item judged low in the feeling of knowing and sometimes vice versa. Moreover, correct recognition can occur either because the subject recognizes the target or because the subject eliminates the distractors. The feeling of knowing is conceptually related only to the former; the latter is a source of noise.

2. The practical limit on the number of distractors per multiple-choice item necessarily produces still more noise due to guessing. For instance, even if an item judged low in the feeling of knowing is not known at all, the subject may fortuitously choose the correct answer. This yields an even greater reduction in the observed accuracy of the feeling of knowing.

The present research explored the relation between the feeling of knowing and criterion tests that do not require distractors. Also, because a major goal of our research program is to determine the overall relation between the feeling of knowing (as an instance of metamemory; see Flavell & Wellman, 1977) and the target-memory system, a natural research strategy is to substitute criterion-memory tests that reflect processes different from those reflected by recognition tests or other previously examined criterion tests.

CRITERION TESTS FOR FEELING-OF-KNOWING RESEARCH

Three criterion tests have been used in previous studies: (a) Most studies used recognition tests (e.g., Hart, 1965a); (b) a few used cued-recall tests in which the first letter of the answer is provided by the experimenter (Freedman & Landauer, 1966; Gruneberg & Monks, 1974; Gruneberg, Monks, & Sykes, 1977); and (c) a few have examined reminiscence during additional uncued recall tests (Gardiner et al., 1973; Gruneberg et al., 1973; Gruneberg & Sykes, 1978; Hart, 1967b; Read & Bruce, 1982). As Schacter (1983) recently pointed out:

> It would be desirable to explore the relation of the feeling of knowing to modes of expressing knowledge other than recall and recognition. . . . Future studies that attempt to elucidate the nature of and relations between the various modes of expressing knowledge are likely to sharpen our insight into many facets of human memory. (p. 53)

To select new criterion tests for the feeling of knowing, it is helpful to consider how different tests might assess memory:

A forced-choice recognition memory test is essentially a test of *discriminating* between the target and the distractors (Bernbach, 1967).

A cued-recall test in which the experimenter provides some letters of the target is a test of

Portions of these results were presented at the annual meeting of the Psychonomic Society in Minneapolis, Minnesota, 1982. This research was supported by National Institute of Mental Health Grant MH-32205.

Thanks go to A. Shimamura for his computer programming and other contributions to early phases of the project; to R. Landwehr and A. Singh for their computer programming; to M. McCloskey and M. Watkins for their comments on the manuscript; and to J. Leonesio, J. Raley, A. Weisman, and D. Malcolm for their help in the data collection and data analysis.

Requests for reprints should be sent to Thomas O. Nelson, Psychology Department (NI-25), University of Washington, Seattle, Washington 98195.

redintegration of the whole by one or more of its parts (Horowitz & Prytulak, 1969). It can be conceptualized in terms of successively more specific cues for recall because the experimenter initially provides no letters of the target, then provides the first letter, and perhaps then provides the first two letters (Gruneberg & Monks, 1974).

Reminiscence across a second uncued-recall test is less well understood but may be a simultaneous test of an item's *variability* combined with its initial closeness to the threshold of recall. Still more speculatively, perhaps this variability is related to the distribution of "memory strength" for the item, where memory strength is conceptualized as a hypothetical unidimensional entity (cf. Eich, 1982; Hayes-Roth, 1977; Hull, 1943; Ratcliff, 1978; Wickelgren, 1977) or as one dimension of an underlying multidimensional entity.

The two new criterion tests for the feeling of knowing explored here—perceptual identification and relearning—can be considered *concatenation-like* tests in which the target itself is presented and so can add to whatever is already in memory about the target (Bahrick, 1967; Nelson, 1978).[1] The perceptual-identification test consists of successively longer tachistoscopic exposures that are assumed to add small increments to whatever is already in memory about the target; this test employs an immediate test of identification that occurs as soon as each exposure terminates. The relearning test differs from the perceptual-identification test in at least three ways: Relative to the perceptual-identification test, the relearning test (a) employs a longer retention interval between the presentation and test of a given item, (b) is thought to add larger increments to whatever is already in memory about the target (because of the greater presentation time relative to the perceptual-identification test), and (c) presents all of the items for study before any item is tested. Aside from these differences, the perceptual-identification test and the relearning test assess memory in the same basic way, via a concatenation-like mechanism that is fundamentally different from the mechanisms of the previous feeling-of-knowing criterion tests.

What hypotheses should be proposed about the relations between the feeling of knowing and relearning and between the feeling of knowing and perceptual identification? Because relearning is more sensitive than recognition (Nelson, 1978; replicated in Groninger & Groninger, 1980) and because recognition is correlated with the feeling of knowing, the feeling of knowing should be correlated with relearning. For the relation between the feeling of knowing and perceptual identification, an a priori hypothesis is not proposed because (a) the relative sensitivities of perceptual identification and recognition are unclear from the previous literature—consequently, perceptual identification is not necessarily expected to be successful as a criterion test for the feeling of knowing—and (b) some critical aspects of the interpretation of perceptual identification are highly controversial (discussed next).

EXPERIMENT 1: PERCEPTUAL IDENTIFICATION

We had two reasons for exploring perceptual identification as a criterion test for the feeling of knowing. First, its concatenation-like properties were of interest. Second, perceptual identification has been suggested by previous investigators (Hart, 1965b, p. 73; Schacter, 1983, p. 53) as a potentially useful criterion test for the feeling of knowing.

Our paradigm built on previous findings that prior exposure to words enhances subsequent perceptual identification of those words (Jacoby & Dallas, 1981; Murrell & Morton, 1974; Neisser, 1954). In our experiment, however, any prior exposure to the targets came from the subject's preexperimental history. Whether those prior exposures will enhance perceptual identification depends on

[1] Concatenation refers to linking one thing together with another, such as when a 2-g weight is added to a 3-g weight to produce a 5-g weight (Coombs, Dawes, & Tversky, 1970, p. 8; Krantz, Luce, Suppes, & Tversky, 1971). We use the term *concatenation-like* because the process cannot strictly be said to be concatenation because of nonadditivity on the observed performance scale (e.g., Haber & Hershenson, 1965). However, concatenation may be occurring on the underlying strength dimension and not be evident in performance because the relation between strength and performance is nonlinear. Until more is known about strength and how it relates to performance, the term concatenation-like seems more appropriate for the supposition of the target's presentation being linked together with whatever is already in memory about the target.

the type of information to which perceptual identification is sensitive, and this currently is controversial (e.g., see Jacoby & Dallas, 1981, p. 306, for one view; for another view, see Allport, 1977; Fowler, Wolford, Slade, & Tassinary, 1981, p. 345; Marcel, 1983). Moreover, even if perceptual identification uncovers some prior information in memory about nonrecalled items, the feeling of knowing may not tap that particular information (see Fowler et al., 1981, p. 345; but also see Merikle, 1982).

Method

Design and Subjects

A within-subjects design was employed. The criterion test for each of nine nonrecalled items was a perceptual-identification test for which the correct answer was repeatedly presented tachistoscopically; for a given item, the exposure time was increased until the subject identified the correct answer (cf. "method of ascending limits" in Winnick & Daniel, 1970). For nine other items, the criterion test was a four-alternative-forced-choice (four-AFC) recognition test. The subjects were 32 University of Washington undergraduates who participated for course credit. They were fluent in English and had normal or corrected-to-normal vision.

Procedure

The procedure consisted of four discrete phases. First, the increment time and start time were established for the eventual perceptual-identification test (preliminary phase). Second, a recall test was given for general-information items until the subject incorrectly answered 21 questions. Third, the subject made relative feeling-of-knowing judgments on those 21 incorrect items. Fourth, the two criterion tests assessed the accuracy of the subject's feeling-of-knowing judgments.

Preliminary Phase

The high variability in subjects' thresholds (Marcel, 1983) necessitated that the start and increment times be determined individually for each subject. The algorithm for this is reported elsewhere (Nelson, Gerler, & Narens, in press). It yields a start time and increment time such that (a) the start time is not so slow that the subject tends to identify the answer on the first exposure, and (b) the start time is not so fast and the increment time is not so small that the subject tends not to identify the answer even after 10 incremented exposures. Thus, the start and increment times were determined for each subject so as to have sufficient fineness to yield perceptual identification on an intermediate number of exposures (i.e., neither a ceiling effect nor a floor effect). These start and increment times were used later during the perceptual-identification phase.

Recall Phase

The items were 145 general-information questions from the 240 that are available in the FACTRETRIEVAL computer program (Shimamura, Landwehr, & Nelson, 1981) and that originally came from the Nelson and Narens norms (1980b). These questions span a wide variety of topics that include history, sports, art, geography, science, literature, and entertainment. The 145 selected items each had an answer that was five to seven letters in length (e.g., "What is the name of Socrates' most famous student?" Answer = Plato).

The questions were presented in a different random order to each subject. Each question was presented individually on a video monitor, and the subject typed his or her response on the keyboard of an Apple II computer. The subject was asked to search "memory hard in an attempt to find the answer." If the subject had no guess at all concerning the answer to a particular question then he typed "next." Response latencies were recorded for all responses. To determine correctness, the computer examined only the first three letters of the subject's response. This yielded a purer pool of nonrecalled items by not insisting on perfect spelling. After the subject missed the answer to 21 questions (9 of which eventually were used in the recognition test, 9 of which eventually were used in the perceptual-identification test, and 3 of which were catch-trial items as discussed later), the computer generated a printout so that slides of the correct answers could be selected for the eventual perceptual-identification phase.

Feeling-of-Knowing Phase

Subjects made feeling-of-knowing judgments on the 21 incorrectly answered items. Three questions at a time were displayed on the video monitor. Subjects were instructed to indicate their relative feeling of knowing by selecting whichever of those 3 questions they believed they were most likely to recognize the answer to. Then that question disappeared from the screen, and subjects made a relative feeling-of-knowing judgment for the remaining 2 questions by selecting whichever question they believed they were more likely to recognize the answer to. These feeling-of-knowing judgments occurred for every triad of items. The particular triads presented to each subject consisted of a random choice of all possible triads with the restriction that every possible unordered pair from the 21 items appeared exactly once. For N items, this yields $N(N-1)/6$ triads, or a total of 70 triads for the 21 items (Burton & Nerlove, 1976). This procedure yields a feeling-of-knowing rank order (with the possibility of tied ranks) in terms of how many times a given item is chosen over all other items (Nelson & Narens, 1980a), where the item chosen most frequently is assigned the highest feeling-of-knowing rank. To minimize an inappropriately high feeling of knowing for items that the subject had guessed incorrectly during the recall phase, the subject was informed that all of the items in the feeling-of-knowing judgments either had a response of "next" during the recall phase or had been answered incorrectly (Krinsky & Nelson, in press).

Criterion-Test Phase

This consisted of the perceptual-identification test and the four-AFC recognition test. Excluding the top, bottom,

Table 1
Mean and 95% Confidence Interval for Each Measure in Experiment 1

Measure	*M*	95% Confidence interval
P(recall)	.41	.35 ↔ .46
P(recognition)	.49	.43 ↔ .55
Perceptual-identification start time	27.4 ms	23.8 ↔ 30.9
Perceptual-identification increment time	4.1 ms	3.6 ↔ 4.6
Median number of exposures	5.8	5.1 ↔ 6.6
P(correct on perceptual identification)	.80	.75 ↔ .86
Gamma: Feeling of knowing and recognition	.29*	.15 ↔ .43
Gamma: Feeling of knowing and number of exposures	−.16*	−.04 ↔ −.27
Gamma: *L*(Incorrect recall) and recognition	.10	−.06 ↔ .25
Gamma: *L*(Incorrect recall) and number of exposures	−.07	−.19 ↔ .04
Gamma: *L*(Incorrect recall) and feeling of knowing	.17*	.07 ↔ .26

Note. Each measure is based on the scores from 32 subjects.
* $p < .05$, correlation is reliably different from zero.

and middle items in the feeling-of-knowing rank order, the remaining 18 items were divided into two sets: the items with even-numbered feeling-of-knowing rank positions (i.e., 2, 4, 6, etc.) and the items with odd-numbered feeling-of-knowing rank positions. The perceptual-identification test occurred on one set of items, whereas the recognition test occurred on the other set. Whether the perceptual-identification test occurred on the even-numbered versus odd-numbered items was counterbalanced across subjects. Order of the perceptual-identification test and the recognition test was also counterbalanced across subjects. The procedure for these criterion tests was as follows.

Perceptual identification. Two display screens were used. One screen was a video monitor for the nontachistoscopic aspects of the experiment. Adjacent to the video monitor was the back-projection screen for the display from a tachistoscopic slide projector that was controlled by the computer.

The question portion of each item appeared individually on the video monitor and remained there until that item was no longer tested. The subject was instructed to look at the tachistoscopic screen and to press the space bar when ready to view the answer portion of the item. Immediately after the exposure, a second slide projector displayed for 256 ms a visual-noise mask consisting of randomly overlapping letters. Then the subject made a best guess at the correct answer. He was told not to be concerned with any guesses during previous exposures but instead to focus on the correct answer for the current exposure; this was done to discourage a strategy of repeating the same answer. The subject was also informed that some questions might have to be answered correctly more than once; this discouraged a strategy of guessing a different answer after each exposure. One practice item was presented at the start time and increment time established during the preliminary phase. The answer to this item was presented either until the subject identified it correctly twice or until 10 exposures had elapsed (whichever came first). The computer then advanced to the test items for which each answer was presented until it was identified correctly once or until 10 exposures had elapsed, whereupon the computer advanced to the next item. Additional details of the perceptual-identification test are reported in Nelson et al. (in press).

The test items also contained three other items for catch trials. These consisted of questions for which the correct answer was never presented. These were the top, middle, and bottom items in the feeling-of-knowing rank order. They were utilized to assess the possibility that some items might be identified correctly even without the correct answer being presented, due either to reminiscence or to a strategy of switching guesses across the trials on a given item. The slides that replaced the correct answers for the catch-trial items contained seven nonletter characters from the top of the typewriter keyboard (e.g., ampersand, percent sign, etc.). These catch-trial items occupied the 4th, 7th, and 11th positions in the perceptual-identification presentation order. The order of the items in these 3 catch-trial positions was counterbalanced across subjects (the remaining 9 perceptual-identification presentation positions were randomly assigned to the nine items that had the correct answers presented). The exposure duration for each of the three catch trials was 5 ms on the first exposure, with a 1-ms increment on each subsequent exposure. Thus the longest possible exposure for a catch trial was 14 ms. This short exposure duration was used so the subject would not realize that no correct answer was being presented for the catch-trial question.

Recognition. The recognition test was self-paced, and the order of the nine items was random. For each item, the video monitor displayed the question and a random ordering of the four alternatives. The distractors were from the same category as the correct answer (e.g., for the question, "What is the last name of the inventor of the wireless radio?", the alternatives were Marconi, Morse, Bell, and Faraday) and came from the recognition phase of FACT-RETRIEVAL (Shimamura et al., 1981).

Results and Discussion

The primary results are summarized in Table 1. The first column shows the 11 measures

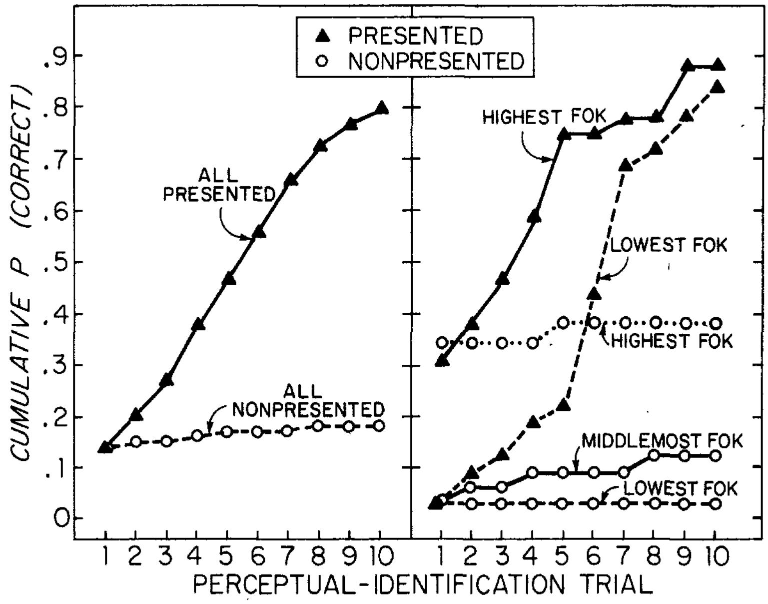

Figure 1. Mean cumulative *P*(correct) on each perceptual-identification trial for presented items and nonpresented items. (Left panel shows performance on all items. Right panel shows performance on highest and lowest feeling-of-knowing presented items and on highest, middlemost, and lowest feeling-of-knowing nonpresented items.)

obtained from each subject, the second column shows the mean value of the 32 individual-subject values, and the third column shows the 95% confidence interval for the mean. Some of these measures are primarily descriptive and will be mentioned only briefly. The discussion of the remaining findings is organized around focal theoretical issues. (The alpha level for reliability was set at $p < .05$, unless noted otherwise.)

Recall and Recognition

During the recall phase, the subjects correctly recalled the answers to approximately 41% of the questions. During the recognition phase, performance was at an intermediate level (mean proportion correct = .49), which was desirable to minimize floor and ceiling effects.

Perceptual Identification

During the preliminary phase, the mean values of the parameters for the eventual perceptual-identification phase were 27.4 ms for the start time and 4.1 ms for the increment time. These values successfully produced nonfloor and nonceiling performance during the subsequent perceptual-identification phase: The mean across subjects of each subject's median number of exposures was 5.8, which is intermediate between the floor value of 1 and the ceiling value of 10. The predetermined upper limit of 10 exposures per item was sufficient to allow most of the items (i.e., 80%) to be identified correctly.

The left panel of Figure 1 shows a more detailed examination of perceptual identification in terms of the rate of increase in correct performance as a function of the number of exposures. The data for presented answers show that the mean cumulative *P*(correct) increased fairly steadily across perceptual-identification trials, which suggests that the increment size established in the preliminary phase added approximately 10% correct for each additional exposure beyond the first exposure. Substantially more perceptual identification occurred for presented answers (sixth row of Table 1) than for catch-trial items [whose overall mean *P*(correct) = .18, with a 95% binomial confidence interval of .10 ↔ .25]. This demonstrates that perceptual identification on presented answers is due to something

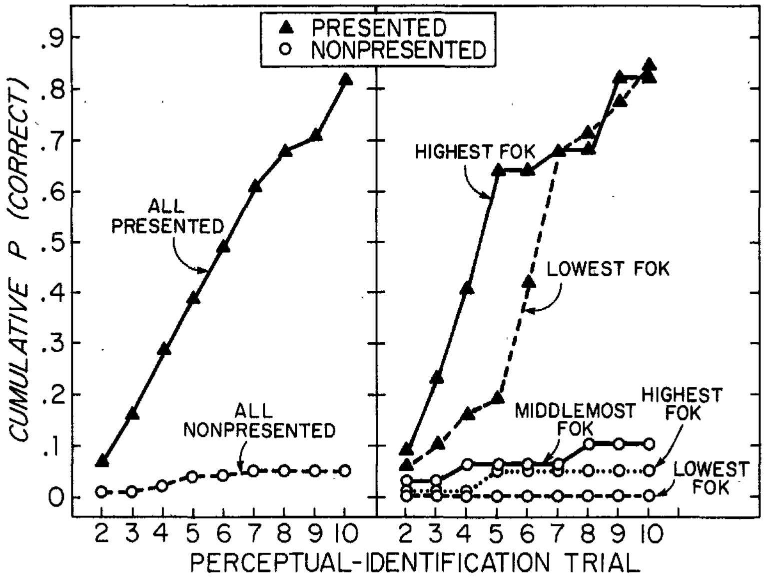

Figure 2. Mean cumulative *P*(correct) on each perceptual-identification trial for presented and nonpresented items that were not identified on the first perceptual-identification trial.

other than mere guessing or reminiscence across repeated test trials. The data for catch-trial items, which show no reliable increase across perceptual-identification trials, are markedly different from the data for presented answers.

Reminiscence

The finding of reliably greater-than-chance performance on catch-trial items is in accord with conclusions from previous research on the feeling of knowing and reminiscence (e.g., Gruneberg et al., 1973; Hart, 1967b; Read & Bruce, 1982). Reminiscence, operationalized as correct recall of catch-trial items during the perceptual-identification phase after nonrecall during the recall phase, varied reliably with the feeling of knowing (right panel of Figure 1). The item with the highest feeling of knowing had reliably greater reminiscence [*P*(correct) = .38] than did either the item with the middlemost feeling of knowing [*P*(correct) = .13] or the item with the lowest feeling of knowing [*P*(correct) = .03], by a sign test for each comparison. Reminiscence did not differ reliably for the item with the middlemost feeling of knowing versus the item with the lowest feeling of knowing, but the amount of reminiscence is near the floor for these items.

Reminiscence may be due either to (a) spontaneous recovery of the answer during the interval between the recall phase and the perceptual-identification phase (cf. Read & Bruce, 1982) or/and to (b) the person initially retrieving two answers to a question, giving one (the incorrect answer) during recall and giving the other during the first trial of perceptual identification (cf. Hart, 1966). The latter predicts more reminiscence on commission-error items (incorrect guess during recall) than on omission-error items (no guess during recall). This prediction was confirmed. The mean *P*(correct) during perceptual-identification catch trials was reliably greater for commission-error items than for omission-error items, $t(19) = 3.04$, mean *P*(correct) = .29, and mean *P*(correct) = .05, respectively.

Reminiscence occurred primarily on the first perceptual-identification trial. This is obvious both in Figure 1 and in Figure 2, which shows performance on each perceptual-identification trial for only those items not correctly identified on the first perceptual-identification trial.

Feeling of Knowing and Criterion Performance

For each subject, the Goodman–Kruskal gamma correlation was computed between the feeling-of-knowing rank and criterion performance. (The rationale for using the gamma correlation is given by Nelson, 1984.) The feeling of knowing is reliably correlated with subsequent recognition (see Table 1).

The new finding here is that the feeling of knowing is also reliably correlated with subsequent perceptual identification. One example of this is shown in the right panel of Figure 1 where the presented item with the highest feeling of knowing has a higher cumulative *P*(correct) than the presented item with the lowest feeling of knowing. The difference is most striking midway through the perceptual-identification trials (e.g., on Trial 5) and decreases as performance approaches asymptote.

The differential perceptual identification shown for the presented items in Figure 1 may be partly due to reminiscence, especially on the first perceptual-identification trial. However, the right panel of Figure 2 shows that perceptual identification was greater for the highest feeling-of-knowing presented item than for the lowest feeling-of-knowing presented item, even when reminiscence was minimal. This manifests itself in two ways. First, a sign test showed that the cumulative *P*(correct) is reliably higher for the highest feeling-of-knowing presented item than for the lowest feeling-of-knowing presented item on Trials 4 and 5 and is marginally higher ($.05 < p < .10$) on Trial 3. We have no explanation for why the effect disappears so abruptly by Trial 7. Second, the number of perceptual-identification trials required to obtain a given level of performance can be compared for the highest feeling-of-knowing item versus the lowest feeling-of-knowing item; for example, for a cumulative *P*(correct) of .4 in Figure 2, the highest feeling-of-knowing item required approximately four perceptual-identification trials whereas the lowest feeling-of-knowing item required six.

Table 1 shows that the correlation between the feeling of knowing and the number of exposures required for perceptual identification is significant and negative, indicating that items with a higher feeling of knowing require fewer exposures to be identified. However, the absolute value of this correlation was not large and was not higher than the absolute value of the correlation between the feeling of knowing and recognition; the difference between these two correlations was not significant, paired $t(31) = 1.36$.[2]

If there is a stable unidimensional feeling-of-knowing ability that underlies accurate predictions of criterion performance, then there should be a positive relation between the absolute value of the correlation for the feeling of knowing and recognition versus the absolute value of the correlation for the feeling of knowing and perceptual identification. The Spearman correlation computed across subjects on this pair of gamma correlations from each subject yielded $\rho = -.06$, which is not reliably different from zero. This lack of relation, especially given the ranges of $-.53$ to $+1.00$ for the feeling-of-knowing/recognition correlations and $-.65$ to $+.50$ for the feeling-of-knowing/perceptual-identification correlations, indicates either that the feeling of know-

[2] Because the design of this experiment yielded recognition that was dichotomous ("correct" or "wrong") whereas perceptual identification was more fine-grained (1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 exposures) and because the absolute magnitude of the gamma correlation is known to decrease as the fineness of the criterion scores increases (Blalock, 1974), we recomputed each subject's feeling-of-knowing/perceptual-identification gamma correlation by dichotomizing that subject's perceptual-identification performance. This dichotomization was accomplished by determining a cut point in the distribution of each subject's number of exposures, so as to minimize the difference between the number of items above the cut point versus the number of items below the cut point. This procedure is similar to a median split on the number of exposures; the two procedures would be identical if no two items had the same number of exposures. However, because ties in the number of exposures were common and because our goal was to dichotomize the number of items rather than the number of exposures, we employed the reported procedure rather than a median split. Hence there became only two possible values for the number of exposures on each item, namely "above" versus "below" that subject's cut-point number of exposures. This reanalysis raised the absolute value of the gamma correlation between the feeling of knowing and perceptual identification: the mean of the recomputed individual-subject correlations was $-.21$, with a 95% confidence interval of $-.05 \leftrightarrow -.37$. However, this recomputed correlation also was not significantly different from the correlation between the feeling of knowing and recognition, $t(31) = 0.73$.

ing does not map unidimensionally into different criterion tests such as recognition versus perceptual identification or/and that the wide range in individual-subject correlations is due to some kind of noise. The former possibility is in accord with previous findings that the effects of a given variable sometimes differ for recognition versus perceptual identification (Jacoby & Dallas, 1981). Before any firm conclusions are drawn, however, the latter possibility should be assessed (e.g., by determining the test–retest stability of individual-subject feeling-of-knowing correlations).

Search Termination

Previous studies have reported that people terminate a memory search fastest when no relevant information is found (Glucksberg & McCloskey, 1981; Kolers & Palef, 1976). The amount of time that a person searches memory for a nonrecalled answer could, in principle, be determined either by whether the person knows the answer or by whether the person feels that he knows the answer (or both). An ideal search-termination person would continue searching for items that he knows and would discontinue searching for items that he doesn't know, regardless of what he feels that he knows.

For each subject, the gamma correlation was computed between the latency of incorrect recall and various other measures. The latency of incorrect recall (see Table 1) is not reliably correlated with either recognition or perceptual identification (two operational definitions of what the subject knows), but is reliably correlated with the feeling of knowing (i.e., with what the subject feels that he knows). This pattern has been reported elsewhere for the criterion test of recognition (Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982) and suggests that people are far from being ideal search-termination devices.

The individual-subject correlations between the latency of incorrect recall and the feeling of knowing ranged from −.31 to +.59, which is a wide enough range to suggest the possibility of individual differences. Because half of each subject's items were tested via recognition and half were tested via perceptual identification, separate gamma correlations between the latency of incorrect recall and the feeling of knowing were computed on each of these two subsets of items. A Spearman rho correlation was then computed across subjects on the pair of gamma correlations from each subject, with the expectation being a rho of zero if there are no reliable individual differences in the relation between search termination and the feeling of knowing (i.e., if the aforementioned range is due only to noise). The obtained result was $\rho = +.41$, which indicated that the subjects are reliably different in the degree to which their latency of incorrect recall and their feeling of knowing are related.

EXPERIMENT 2: RELEARNING

Experiment 2 was designed to explore the accuracy of the feeling of knowing for predicting relearning performance. Hart (1967a) has speculated the following:

> The threshold for activation of a FK [feeling of knowing] signal from the MEMO [memory-monitoring] process is thought to lie between the recall and savings [during relearning] thresholds. The model implies that Ss would be unable to predict those unrecalled items for which they would show a savings score. In other words, memory traces can exist that are too weak to be monitored. This implication has not been tested. (p. 690)

However, the feeling of knowing might have above-chance accuracy at predicting relearning. Like perceptual identification, relearning has concatenation-like properties (Bahrick, 1967; Nelson, 1978; Nelson, Fehling, & Moore–Glascock, 1979). Moreover, previous research found that relearning detects extra information that recognition does not detect (Nelson, 1978). If feeling-of-knowing judgments tap this extra information, they may even be more accurate at predicting relearning than at predicting recognition.

Method

Design and Subjects

A between-subjects design was employed. For one group of subjects ($N = 43$), the criterion test was recognition, and FACTRETRIEVAL (Shimamura et al., 1981) was used without modification. For the second group ($N = 44$), the criterion test was relearning. The subjects were from the same population as in Experiment 1.

Stimuli and Apparatus

The items consisted of the 240 general-information questions from FACTRETRIEVAL. These included all of the items from Experiment 1 plus additional items whose correct answers were longer or shorter than five to seven letters in length. Items were presented on a video monitor connected to an Apple II computer, and the subject typed his responses on the computer keyboard. Instructions to subjects, data collection, and statistical analysis were computer-controlled (Shimamura et al., 1981).

Procedure

The procedure consisted of three discrete phases. The first was a recall test on general-information items, which lasted until the subject incorrectly answered 12 questions. Second, the subject made relative feeling-of-knowing judgments on the 12 incorrectly answered items. This was accomplished via paired comparisons (Nelson & Narens, 1980a). Third, the criterion test assessed the accuracy of the feeling-of-knowing judgments.

Recall Phase

This was identical to the recall phase in Experiment 1 except for terminating after 12 incorrectly answered questions rather than after 21.

Feeling-of-Knowing Phase

Following the recall phase, pairs of questions from the set of 12 incorrectly answered questions were displayed on the video monitor. For each pair, the subject selected whichever question he felt he was more likely to recognize the correct answer to. All possible $N(N-1)/2 = 66$ paired comparisons of the 12 questions occurred, and the presentation sequence of the pairs was random. The 2 questions in each pair were displayed 1 above the other, and the display order of the 2 questions was random.

After the feeling-of-knowing judgments, 12 of the 66 pairs were presented again for a second feeling-of-knowing judgment to allow for an estimate of retest reliability (Nelson & Narens, 1980a). The choice of which 12 pairs would be retested was random except that none of the retested pairs came from the last 16 of the original 66 pairs. Six of the retested pairs were presented with the display order reversed such that, for each pair, the question previously presented on top was now presented on the bottom. The 2 questions within each of the remaining 6 retested pairs were presented in their original display order.

The feeling-of-knowing rank order was derived by tallying the number of times that a given item was chosen over all others in the paired comparisons, not counting the retest phase. This yields (a) a relative feeling-of-knowing rank order of the items, (b) an index of retest reliability, and (c) the number of intransitivities in the subject's feeling-of-knowing judgments (i.e., if Item A is chosen over Item B, and Item B is chosen over Item C, then an intransitivity occurs if Item C is chosen over Item A).

Criterion-Test Phase

Half of the subjects had a relearning study-test trial, and half had a four-AFC recognition test.

Relearning. The subject viewed one of the 12 incorrectly answered questions for 5 s. Then the question disappeared and was followed by a 1-s "READY" signal. The correct answer was then displayed for 1 s and was immediately followed by a three-digit number. When the number appeared, the subject counted aloud backward by threes for 30 s; this procedure was employed to minimize short-term memory effects and rehearsal between successive items. Then the next question was presented, and so on until each of the 12 questions and answers had been presented. Finally the subject had a relearning recall test for the answers to the 12 questions, using the same procedure as in the recall phase. Both the study order and test order of the 12 questions were random except for the restriction that none of the last 3 questions from the relearning study phase was among the first 3 relearning test questions.

Pilot research had shown that a relatively short display duration for the relearning study trial was necessary to avoid a ceiling effect on relearning performance. To familiarize the subject with the timing and procedure, a relearning practice trial occurred prior to relearning the test items. The practice item was not one of the FACTRETRIEVAL questions.

Recognition. This was identical to the recognition phase in Experiment 1 except that there were 12 recognition items rather than 9.

Results and Discussion

The primary results are summarized in Table 2. Measures pertaining to recall and the feeling of knowing are based on $N = 87$ because all subjects were treated identically during those phases. Measures involving recognition or relearning are based on $N = 43$ or $N = 44$, respectively. As in Experiment 1, the findings are organized around focal theoretical issues.

Overall Performance

Recall (first row of Table 2) was similar to that in Experiment 1 (first row of Table 1).

The retest reliability during the feeling-of-knowing judgments was substantially better than chance [the chance P(mismatch) = .50], although not perfect at P(mismatch) = 0. Also, the P(mismatch) was reliably greater [$t(86) = 2.62$] when the two items in a retested paired comparison were presented in reversed order ($M = .21$) than in the same order as the original paired comparison ($M = .13$). This appears to be attributable to some subjects who had a position bias (e.g., tending to choose the uppermost item on the video screen). The main ramification is methodological: To avoid underestimating or overestimating the magnitude of P(mismatch), retests should use either a random order of the items within each paired comparison or a counterbalanced order with

Table 2
Mean and 95% Confidence Interval for Each Measure in Experiment 2

N	Measure	*M*	95% Confidence interval
87	*P*(recall)	.40	.35 ↔ .44
87	*P*(mismatch)	.17	.14 ↔ .20
87	*P*(intransitivity)	.08	.06 ↔ .10
43	*P*(recognition)	.44	.39 ↔ .49
44	*P*(relearning)	.66	.59 ↔ .72
43	Gamma: Feeling of knowing and recognition	.28*	.14 ↔ .42
44	Gamma: Feeling of knowing and relearning	.34*	.19 ↔ .48
43	Gamma: *L*(incorrect recall) and recognition	.02	−.09 ↔ .14
44	Gamma: *L*(incorrect recall) and relearning	.02	−.10 ↔ .13
87	Gamma: *L*(incorrect recall) and feeling of knowing	.12*	.06 ↔ .18

Note. N = number of subjects on which each measure is based.
* $p < .05$, correlation is reliably different from zero.

equal numbers of same and reversed presentations.

The *P*(intransitivity) was substantially below chance [the chance *P*(intransitivity) = .25 if judgments are made randomly], although not at *P*(intransitivity) = 0. Across subjects, the magnitudes of *P*(intransitivity) and *P*(mismatch) were reliably related, Spearman $\rho = +.35$. This correlation is not surprising because one source of intransitivities is unreliability (Krantz, Luce, Suppes, & Tversky, 1971), such that those subjects who are more unreliable would be expected to have more intransitivities (although not necessarily vice versa because unreliability is not the only source of intransitivities; see Krantz et al., 1971).

Both recognition and relearning were at an intermediate level. This was desirable to avoid floor and ceiling effects.

Feeling of Knowing and Criterion Performance

The feeling of knowing is reliably correlated with subsequent recognition, similar to the corresponding finding in Experiment 1 (Table 1). The new finding here is that the feeling of knowing is also a reliable predictor of relearning (see seventh row of Table 2). Although the sample mean is higher for the correlation between the feeling of knowing and relearning than for the correlation between the feeling of knowing and recognition, the difference between these two correlations is not statistically significant, $t(85) = .61$.[3]

Search Termination

Similar to Experiment 1, the present findings suggest that the amount of time that people search for nonrecalled answers is determined not by what they know but rather by what they feel that they know. Table 2 shows that the latency of incorrect recall is not reliably correlated with either recognition or relearning. However, the correlation between the latency of incorrect recall and the feeling of knowing is reliably positive.

Although statistically significant, this correlation and the analogous one in Experiment 1 (last row of Table 1) are small. More detailed analyses were conducted by partitioning the data to examine this relation separately for different types of recall failures. The results are striking and are summarized in Figure 3. When the recall failure was an omission error, the correlation was reliably positive and substantial for all three groups of subjects. That is, for omission-error items whose labels have not been retrieved, people search longer when the feeling of knowing is high rather than low.

[3] The direction of this difference may be maintained and become statistically significant with a larger sample of subjects, but the substantial size of the present confidence intervals suggests that the number of subjects needed to attain conventional levels of statistical significance would preclude such an experiment. Given the same standard deviations and difference between the means found here, the number of subjects that would be needed for significance at the .05 level is 455 subjects per group (Hays, 1973, Equations 10.13.1 and 10.15.2).

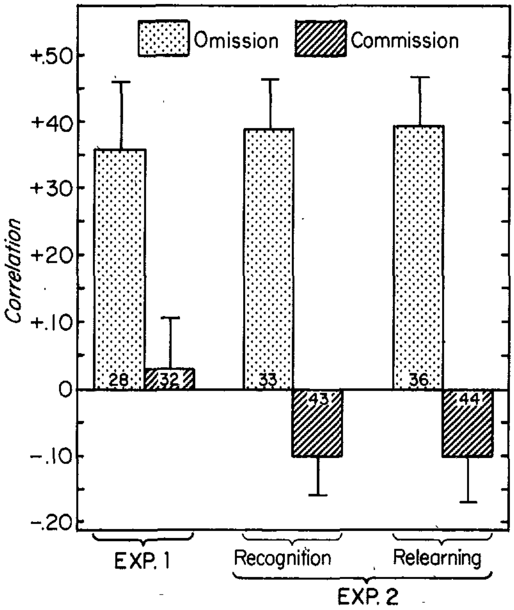

Figure 3. Mean (and standard error) of the individual-subject gamma correlations between the latency of incorrect recall and the feeling of knowing. (Entry inside each bar shows the number of subjects on which the mean is based; indeterminant correlations are excluded. Bars show data for two kinds of recall errors, omission versus commission, in three groups of subjects: Experiment 1 subjects, recognition group in Experiment 2, and relearning group in Experiment 2.)

In sharp contrast, when the recall failure is a commission error, the correlation is not reliably different from zero for any of the three groups. That is, when people retrieve the wrong answer but believe they are correct, their latency will be no different (or perhaps may even be shorter) than when they are less sure that they are correct. Thus, the correlation between the latency of incorrect recall and the feeling of knowing is qualitatively and reliably different for omission-error recall failures than for comission-error recall failures, $t(27) = 2.94$ for Experiment 1, $t(32) = 4.09$ for the recognition group in Experiment 2, and $t(35) = 4.33$ for the relearning group. Implications of this finding are described later.

Given this difference between omission errors and commission errors concerning the relation between the latency of incorrect recall and the feeling of knowing, one may expect that these two kinds of recall failure will also differ in latency of incorrect recall. In particular, the latency of incorrect recall may be longer for omission errors than for commission errors, which in turn may be similar to the latency of correct recall because the person may process commission errors and correct recalls in a similar way (cf. Krinsky & Nelson, in press).

Contrary to this expectation, the latencies for commission errors were more like the latencies for omission errors than like the latencies for correct recall. For the same three groups of subjects as in Figure 3, Table 3 shows the mean (and 95% confidence interval) of the individual-subject median latencies for each type of recall outcome. Although the latencies tended to be shorter in Experiment 1 than in Experiment 2, every group of subjects displayed the same pattern of latencies for omission errors versus commission errors; the paired t values from subjects contributing both kinds of latencies were $t(29) = 0.06$ for Ex-

Table 3
Mean (and 95% Confidence Interval) for Individual-Subject Median Latencies of Correct Recall, Incorrect Commission-Error Recall, and Incorrect Omission-Error Recall

	Subjects		
		Experiment 2	
Latency	Experiment 1	Recognition group	Relearning group
Correct recall	9.8 (8.8 ↔ 10.7)	13.4 (11.9 ↔ 14.9)	15.2 (13.3 ↔ 17.1)
Commission error	20.4 (17.5 ↔ 23.3)	25.9 (23.2 ↔ 28.6)	26.4 (22.8 ↔ 30.0)
Omission error	20.5 (15.0 ↔ 26.0)	26.0 (20.4 ↔ 31.7)	25.1 (19.3 ↔ 30.9)

periment 1, $t(35) = 0.14$ for the recognition group, and $t(35) = 0.36$ for the relearning group. In contrast, the paired t values for the comparison of correct-recall latencies with commission-error latencies were $t(31) = 8.93$ for Experiment 1, $t(41) = 9.19$ for the recognition group, and $t(40) = 6.31$ for the relearning group; the paired t values for the comparison of correct-recall latencies with omission-error latencies were $t(29) = 3.94$ for Experiment 1, $t(34) = 4.18$ for the recognition group, and $t(32) = 3.13$ for the relearning group (all six ts indicate reliable differences). Moreover, as Table 3 shows for every group of subjects, the 95% confidence interval for latency of correct recall did not even overlap with that of commission errors or omission errors.

Thus, in both experiments the relation between search termination and the feeling of knowing was different for commission errors versus omission errors. At the same time, however, people searched for approximately the same amount of time whether their incorrect recall was a commission error or omission error, and they searched longer when they were incorrect than when they were correct.

GENERAL DISCUSSION

The feeling of knowing may determine how long a memory search is continued, especially when the recall failure is an omission error (Experiments 1 and 2). Because search termination is related to the feeling of knowing rather than to subsequent criterion performance (Experiments 1 and 2), if the accuracy of the feeling of knowing can be improved, then a corresponding improvement may occur in the efficiency of memory search. That is, the person would continue searching for items that are in memory, thereby avoiding unnecessary memory failures due to premature search termination (a problem in Korsakoff patients; see Hirst, 1982, p. 454), and would stop searching for items not in memory. An improvement in feeling-of-knowing accuracy may be particularly relevant for individuals whose correlation between the feeling of knowing and search termination is high (reliable individual differences in the magnitude of these correlations occurred in Experiment 1).

The magnitude of feeling-of-knowing accuracy did not differ reliably for perceptual identification versus recognition (Experiment 1) or for relearning versus recognition (Experiment 2). Thus, the information in memory that is assessed during feeling-of-knowing judgments does not seem to be assessed any better (or worse) by a recognition test than by concatenation-like tests such as relearning and perceptual identification. Caution should be exercised, however, when generalizing this conclusion to other versions of these criterion tests. For instance, feeling-of-knowing accuracy probably would be greater for eight–AFC recognition than for four–AFC recognition because the former is less noisy than the latter (Nelson, 1984). Different versions of relearning also may yield different feeling-of-knowing accuracy; for example, Bahrick (1982) reported unusually high feeling-of-knowing accuracy when the delay between a relearning study phase and a relearning test phase was 1 month. Feeling-of-knowing accuracy may also vary across different versions of perceptual identification, such as repeated exposures at a constant exposure duration (Haber & Hershenson, 1965).

The feeling-of-knowing accuracy here was above chance, but far from perfect. However, a low correlation between the feeling of knowing and criterion performance does not necessarily reflect an inherently poor metacognitive ability. For instance, although feeling-of-knowing accuracy was somewhat low for predicting perceptual identification on all items, it was substantial for predicting perceptual identification on the highest and lowest feeling-of-knowing items (Figure 1). The nature of to-be-judged items affects feeling-of-knowing accuracy, such that differentiating among items is more difficult when the underlying information in memory is similar for the items. Acquisition parameters also can affect the degree of feeling-of-knowing accuracy. For instance, Nelson et al. (1982) found that the correlation between the feeling of knowing and recognition was .00 for items originally learned to a criterion of one correct acquisition response but +.40 for items originally learned to a criterion of four correct acquisition responses.

No reliable relation occurred between the feeling-of-knowing accuracy for predicting perceptual identification and the feeling-of-knowing accuracy for predicting recognition (Experiment 1). This could be due to noise (a possibility that can be explored by examining the reliability of individual subjects' feeling-of-knowing accuracy) and/or to the possibility that recognition and perceptual identification assess memory in fundamentally different ways. Other research also has found a nonsignificant correlation between criterion tests similar to the ones examined here (see Jacoby & Dallas, 1981). Predictions about the relation between a given criterion test and the feeling of knowing should improve as more is known about (a) how various criterion tests assess information in memory and (b) how information in memory influences the feeling of knowing.

Table 4
Some Plausible Theoretical Mechanisms Underlying the Feeling of Knowing

Trace-access mechanisms
Subthreshold strength
Forward-backward associations
Correct semantic referent but no label
Partial recall of label
Wrong referent
Multidimensional target
Inferential mechanisms
Related episodic information
Claimed related episodic information
Recognition of cue
Expertise on topics
Actuarial information
Social desirability

Theoretical Mechanisms Underlying the Feeling of Knowing

On what information do people base their feeling-of-knowing judgments? The data here and in the literature do not yet provide enough constraints to determine the mechanisms that underlie the feeling of knowing. Nevertheless, a discussion of the theoretical mechanisms that currently seem reasonable as potential bases for the feeling of knowing may be timely and may help to guide future research. The present discussion is not a critical review of the evidence in the literature, but rather is a brief description of some plausible mechanisms that are summarized in Table 4. They frequently overlap—one mechanism can involve aspects of another—but their emphases are different.

Trace-Access Mechanisms

The trace-access mechanisms share the characteristic that the person is presumed to have access to the nonrecalled item during feeling-of-knowing judgments. Different mechanisms presumably monitor different aspects of the nonrecalled item.

Subthreshold Strength

Here the focus is on the amount of associative strength between the cue and target (Hart, 1967a; Read & Bruce, 1982, p. 296), for example, between the cue "What is the capital of Australia?" and the target "Canberra." Supposedly there is a threshold amount of strength, R, for recall and another threshold amount of strength, F, for the feeling of knowing (with $F < R$). When the amount of strength in the soughtafter item is greater than R, the person correctly recalls the target. When the amount of strength is less than R but greater than F, incorrect recall occurs, but the person feels that he knows the target. When the amount of strength is less than F, incorrect recall occurs, and the person feels that he does not know the target (see Hart 1967a, Figure 1). More than two gradations of the feeling of knowing could be postulated by assuming additional thresholds, each of which would correspond to a different feeling-of-knowing rating (Hart, 1967a, Experiment 2; Read & Bruce, 1982). The primary problem with the subthreshold-strength mechanism is that the amount of strength is hypothetical and not directly measurable, making tests of the mechanism difficult. Nevertheless, this mechanism is popular, as Gruneberg (1983) mentioned: "Most accounts of the nature of feeling of knowing assume that the individual in some way makes a preliminary estimate of the likelihood of an item being in store by assessing either the strength or. . . ." (p. 20)

Forward-Backward Associations

This is an extension of the subthreshold-strength mechanism. Instead of one association between the cue and target, this mechanism proposes two associations: a forward association from the cue to the target and a backward association from the target to the cue (Hart, 1967b, p. 196). For instance, there may be an association from "What is the capital of Australia?" to "Canberra" and another association in the opposite direction. In one version of this mechanism, the feeling of knowing would presumably be monitoring the strength of the forward association, whereas recognition would be based on both the forward association and the backward association (cf. Wolford, 1971). Consequently, an item with associations that are asymmetric in strength could, for instance, produce a feeling of not knowing (due to a weak forward association), but a high likelihood of recognition (due to a strong backward association).

An assumed model of criterion performance in terms of forward–backward associations may be required to assess the validity of this mechanism. For example, if one assumes that (a) the feeling of knowing monitors only the forward association, (b) relearning is based on only the forward association, and (c) recognition is based on both the forward and backward associations, then one may expect the feeling of knowing to be a more accurate predictor of relearning than of recognition, particularly when the items have asymmetric associations. Experiment 2 did not yield such results, which (given the independent support of Assumption C by Wolford, 1971) suggests either that Assumption A or B is incorrect or that the associations were not asymmetric enough to produce a detectable difference. A more direct test of Assumption A has not been conducted, but would be desirable. Perhaps the methodology of Nelson et al. (1982) could be employed to manipulate overlearning in the forward versus backward direction of paired-associate acquisition prior to assessing the feeling of knowing and criterion performance.

Correct Semantic Referent but No Label

This mechanism assumes that the target of an item consists of two portions: the semantic referent and a label for the referent (Eysenck, 1979; Gardiner et al., 1973). The cue elicits the correct semantic referent (or what has been referred to as "contextual information" by Read & Bruce, 1982), but not the label for that referent. For instance, the cue "What is the capital of Australia?" may produce the semantic referent of the target (which could include images of Canberra from photographs—cf. May & Clayton, 1973—or descriptions from personal experience), but not the label "Canberra." Such an occurrence would yield an omission error, perhaps one associated with a high feeling of knowing (see *Search Termination* section in Experiment 2).

Partial Recall of Label

Here some (but not all) attributes of the label are recalled. In their study of the tip-of-the-tongue phenomenon, Brown and McNeill (1966) called this "generic recall" and wrote that it involves "letters (or phonemes), affixes, syllables, and stress location" (p. 335). The feeling of knowing appears to be greater for items whose labels are partially recalled (Blake, 1973; Koriat & Lieblich, 1974; Wellman, 1977).

Wrong Referent

Here the cue elicits the wrong semantic referent (Koriat & Lieblich, 1974, 1977; also see Eysenck, 1979, p. 245; Yarmey, 1973, p. 288). For instance, the cue "What is the capital of Australia?" may elicit the semantic referent of Sydney. In turn, the corresponding label of "Sydney" may be elicited (which would yield a commission error) or may not be elicited (which would yield an omission error). When the cue elicits a semantic referent that the person believes is correct (even though the experimenter considers it to be wrong), the person tends to report a high feeling of knowing (Koriat & Lieblich, 1974, 1977; Krinsky & Nelson, in press). Investigators who want to examine this mechanism could either develop word-definition items with the properties described by Koriat and Lieblich (1977, p. 156f) or use general-information items that have both a low probability of recall and a high feeling of knowing. For instance, the cue "What is the last name of the Union general who defeated the Confederate army at the Civil War battle of Gettysburg?" (see Nelson & Narens, 1980b) tends to elicit a high feeling of knowing and the wrong answer (e.g., Grant, Sherman, or Sheridan), rather than the correct answer (Meade).

Multidimensional Target

This mechanism is more abstract than the aforementioned mechanisms (although it could include aspects of them) and emphasizes the multidimensionality of a target (Koriat & Lieblich, 1977; Underwood, 1969; also see Bower, 1967). A person may not retrieve information from enough dimensions for recall to be correct, but the information that is retrieved could serve as the basis for a high feeling of knowing (e.g., information from dimensions of the semantic differential; Eysenck, 1979). Such a mechanism is consistent with the finding of no relationship between the feeling-of-knowing accuracy for predicting perceptual identification and the feeling-of-knowing accuracy for predicting recognition (Experiment 1). These different criterion tests may tap different dimensions of the underlying memory structure. Moreover, the feeling of knowing may tap a richer combination of dimensions than does any one criterion test. That is, each criterion test may tap a different *subset* of whatever information in memory is tapped by the feeling of knowing; hence, even though the feeling of knowing may be completely valid as a monitor of nonrecalled information in memory, no criterion test would be correlated perfectly with the feeling of knowing. Consistent with this possibility, Seamon, Brody, and Kauff (1983) recently reported evidence of a multidimensional memory structure that recognition taps only partially.

Inferential Mechanisms

For these mechanisms,the feeling of knowing does not monitor the nonrecalled target item. Instead, other information in memory is monitored and serves as the basis for an inference about the likelihood of correct performance on the criterion test.

Related Episodic Information

The concept of "related information" is sometimes vague and misleading. For example, suppose that someone does not recall the answer to the question, "What is the name of the man who assassinated Abraham Lincoln?" and has to make a feeling-of-knowing judgment about whether he will recognize the correct answer. If the person recalled that the assassination occurred in Ford's theatre during April of 1865 and that the assassin broke a leg leaping to the stage, we might expect this related information about the assassination episode to a trigger a high feeling of knowing. Accordingly, the conclusion drawn from this example might be that related information is an important basis for the feeling of knowing. However, suppose instead that the question had been, "What is the shoe size of the man who assassinated Abraham Lincoln?" and that the person did not recall the answer, but recalled the same related information about the assassination episode as mentioned above. Now we would not expect the person to have a high feeling of knowing, even though the "related information about the assassination episode" is identical to that from the first situation!

Perhaps the related episodic information that is critical pertains less to the episode addressed in the question than to prior episodes in the answerer's life during which the soughtafter item was encountered. If the person had no prior encounters with the item (e.g., "What is President Reagan's telephone number?"), then he probably would have a feeling of not knowing (Kolers & Palef, 1976). By contrast, if information about prior encounters with the soughtafter item is present in memory, a high feeling of knowing is more likely (cf. *availability* in Tversky & Kahneman, 1974). For instance, a person who is unable to recall the answer to the question "What is the capital of Australia?" might have been in a geography class where everyone learned the capitals of all major countries, and this might serve as the basis for a high feeling of knowing. Consistent with this, Nelson et al. (1982) found that the feeling of knowing for nonrecalled items was a direct function of the prior amount of overlearning. This mechanism also might involve other aspects of prior encounters, such as how difficult the item was to learn.

Claimed Related Episodic Information

This mechanism is similar to the previous one, except here the emphasis is on what the person recalls about prior encounters with the item rather than on the prior encounters per se (cf. Gentner & Collins, 1981). For instance, suppose that there are two nonrecalled items, *A* and *B*, and that there had been more overlearning on Item *A* than on Item *B*, but the person claims to have had more overlearning on Item *B* than on Item *A*. Which item will receive the higher feeling of knowing? By having people make various judgments about their prior encounters with nonrecalled items, future investigators may be able to separate the effect of prior encounters per se from the effect of what the person claims those prior encounters were.

Recognition of Cue

Here the feeling of knowing for the nonrecalled target is based on the person's degree of recognition for the cue (cf. Hoffding, 1891). Even kindergartners, when making a feeling-of-knowing judgment about a nonrecalled target, utilize information about whether they have seen the cue previously (Wellman, 1977). Moreover, the feeling of knowing for a nonrecalled target can be increased by merely adding redundancy to the cue (e.g., by repeating it or by adding alternative wording; Koriat & Lieblich, 1977). Thus, memory factors concerning the cue produce the inference about whether the nonrecalled target is known.

Expertise on Topics

Here the person draws an inference about the correctness of his criterion performance based on his expertise with the general topic of the item (Bradley, 1981; also see Gentner & Collins, 1981). Koriat and Lieblich (1974, 1977) initiated research on the relation between the feeling of knowing and information common to a category of items versus information specific to a particular item. The former information, even in the absence of the latter information, may trigger a high feeling of knowing. There has not, however, been research on how a person's rated expertise for a category is related to his feeling of knowing for nonrecalled items in that category. For instance, relative to other people, a person who has (or at least claims to have) expertise about sports may have an unrealistically high feeling of knowing for an extremely difficult (i.e., low normative probability) sports item such as the name of the player who had the most hits in the 1937 World Series. Put differently, the person may underutilize base-rate information about actuarial difficulty because it may not seem representative of his expertise (cf. *representativeness* in Tversky & Kahneman, 1974). However, given that correct recognition is sometimes mediated by recognition of general-category information instead of specific-item information (MacLeod & Nelson, 1976), a high feeling of knowing based on expertise about the category sometimes may be appropriate.

Actuarial Information

In contrast to the individual basing his feeling of knowing on idiosyncratic information (e.g., such as his own expertise on the topic), he may utilize actuarial information about the normative difficulty of the item. Nelson, Leonesio, Landwehr, and Narens (1983) found correlations of approximately +.25 between the individual's feeling-of-knowing ranking of items and the ranking obtained from the *P*(recall) in the Nelson and Narens norms (1980b). This mechanism would tend to increase the accuracy of feeling-of-knowing judgments because an individual's criterion performance on nonrecalled items tends to be correlated with the group's initial performance (Brown, 1923; Gruneberg et al., 1977, p. 369). Furthermore, Nelson et al. (1983) found that actuarial predictions derived from the *P*(recall) in the Nelson and Narens norms (1980b) were more accurate for predicting an individual's criterion performance than were the individual's own feeling-of-knowing judgments. Because people are accurate at estimating actuarial memory performance (Arbuckle & Cuddy, 1969; Seamon & Virostek, 1978; Underwood, 1966), perhaps they should give actuarial information more weight during feeling-of-knowing judgments.

Social Desirability

Here the person reports a high feeling of knowing based on what he thinks he ought to know "because of a desire not to be thought stupid" (Gruneberg et al., 1977, p. 370). Although such demand characteristics may serve as a basis for absolute feeling-of-knowing judgments from the Hart methodology (e.g., by lowering the person's threshold for reporting a high feeling of knowing), they may be less influential during the relative feeling-of-knowing judgments from the Nelson–Narens methodology.

The aforementioned mechanisms are primarily static insofar as they do not emphasize the possibility of changes in the feeling of knowing. Research on the dynamic aspects of the feeling of knowing (e.g., effects of feedback; Krinsky & Nelson, in press) is just beginning. One potentially useful direction for future research may be to explore dynamic processes such as changes in the feeling of knowing during the course of an unsuccessful memory search.

References

Allport, A. (1977). On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.), *Attention and performance VI.* Hillsdale, NJ: Erlbaum.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81,* 126–131.

Bahrick, H. P. (1967). Relearning and the measurement of retention. *Journal of Verbal Learning and Verbal Behavior, 6,* 89–94.

Bahrick, H. P. (1982, August). *Maintenance of knowledge.* Invited address at the meeting of the American Psychological Association, Washington, DC.

Bernbach, H. A. (1967). Decision processes in memory. *Psychological Review. 74,* 462–480.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior, 12,* 311–319.

Blalock, H. M., Jr. (1974). Beyond ordinal measurement: Weak tests of stronger theories. In H. M. Blalock, Jr. (Ed.), *Measurement in the social sciences: Theories and strategies* (pp. 424–455). Chicago: Aldine.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press.

Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society, 17,* 82–84.

Brown, A. L., & Lawton, S. C. (1977). The feeling of knowing experience in educable retarded children. *Developmental Psychology, 13,* 364–370.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Brown, W. (1923). To what extent is memory measured by a single recall? *Journal of Experimental Psychology, 6,* 377–382.

Burton, M. L., & Nerlove, S. B. (1976). Balanced designs for triads tests: Two examples from English. *Social Science Research, 5,* 247–267.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction.* Englewood Cliffs, NJ: Prentice-Hall.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Eysenck, M. W. (1979). The feeling of knowing a word's meaning. *British Journal of Psychology, 70,* 243–251.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail, Jr., & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition.* Hillsdale, NJ: Erlbaum.

Fowler, C. A., Wolford, G., Slade, R., & Tassinary, L. (1981). Lexical access with and without awareness. *Journal of Experimental Psychology: General, 110,* 341–362.

Freedman, J. L., & Landauer, T. K. (1966). Retrieval of long-term memory: "Tip-of-the-tongue" phenomenon. *Psychonomic Science, 4,* 309–310.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1,* 213–216.

Gentner, D., & Collins, A. (1981). Studies of inference from lack of knowledge. *Memory & Cognition, 9,* 434–443.

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 311–325.

Groninger, L. K., & Groninger, L. P. (1980). A comparison of recognition and savings as retrieval measures: A reexamination. *Bulletin of the Psychonomic Society, 15,* 263–266.

Gruneberg, M. M. (1983). Strategies in human memory. In A. Mayes (Ed.), *Memory in animals and man.* London: Van Nostrand.

Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing and cued recall. *Acta Psychologica, 38,* 257–265.

Gruneberg, M. M., Monks, J., & Sykes, R. N. (1977). Some methodological problems with feeling of knowing studies. *Acta Psychologica, 41,* 365–371.

Gruneberg, M. M., Smith, R. L., & Winfrow, P. (1973). An investigation into response blockaging. *Acta Psychologica, 37,* 187–196.

Gruneberg, M. M., & Sykes, R. N. (1978). Knowledge and retention: The feeling of knowing and reminiscence. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 189–196). New York: Academic Press.

Haber, R. N., & Hershenson, M. (1965). Effects of repeated brief exposures on the growth of a percept. *Journal of Experimental Psychology, 69,* 40–46.

Hart, J. T. (1965a). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56,* 208–216.

Hart, J. T. (1965b). *Recall, recognition, and the memory monitoring process.* Doctoral dissertation, Stanford University. (University Microfilms No. 66-2565).

Hart, J. T. (1966). Methodological note on feeling-of-knowing experiments. *Journal of Educational Psychology, 57,* 347–349.

Hart, J. T. (1967a). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6,* 685–691.

Hart, J. T. (1967b). Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology, 58,* 193–197.

Hayes-Roth, B. (1977). Evolution of cognitive structures and processes. *Psychological Review, 84,* 260–278.

Hays, W. L. (1973). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart, & Winston.

Hirst, W. (1982). The amnesic syndrome: Descriptions and explanations. *Psychological Bulletin, 91,* 435–460.

Hoffding, H. (1891). *Outlines of psychology* (M. E. Lowndes, Trans.). London: Macmillan.

Horowitz, L. M., & Prytulak, L. S. (1969). Redintegrative memory. *Psychological Review, 76,* 519–531.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Kolers, P. A., & Palef, F. R. (1976). Knowing not. *Memory & Cognition, 4,* 553–558.

Koriat, A., & Lieblich, I. (1974). What does a person in a "TOT" state know that a person in a "don't know" state doesn't know. *Memory & Cognition, 2,* 647–655.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica, 41,* 151–164.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.

Krinsky, R., & Nelson, T. O. (in press). The feeling of knowing for different types of retrieval failures. *Acta Psychologica.*

Lachman, J. L., Lachman, R., & Thronesbery, C. (1979). Metamemory through the adult lifespan. *Developmental Psychology, 15,* 543–551.

MacLeod, C. M., & Nelson, T. O. (1976). A nonmonotonic lag function for false alarms to associates. *American Journal of Psychology, 89,* 127–135.

Marcel, A. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology, 15,* 238–300.

May, J. E., & Clayton, K. N. (1973). Imaginal processes during the attempt to recall names. *Journal of Verbal Learning and Verbal Behavior, 12,* 683–688.

Merikle, P. M. (1982). Unconscious perception revisited. *Perception & Psychophysics, 31,* 298–301.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology, 102,* 963–968.

Neisser, U. (1954). An experimental distinction between perceptual process and verbal response. *Journal of Experimental Psychology, 47,* 399–402.

Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 453–468.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133.

Nelson, T. O., Fehling, M. R., & Moore-Glascock, J. (1979). The nature of semantic savings for items forgotten from long-term memory. *Journal of Experimental Psychology: General, 108,* 225–258.

Nelson, T. O., Gerler, D., & Narens, L. (in press). A technique for determining perceptual-identification parameters for use in feeling-of-knowing research. *Psychological Documents.*

Nelson, T. O., Leonesio, R. J., Landwehr, R. F., & Narens, L. (1983). *Accuracy of feeling-of-knowing predictions vs. predictions from normative recall.* Unpublished manuscript, University of Washington, Seattle.

Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 279–288.

Nelson, T. O., & Narens, L. (1980a). A new technique for investigating the feeling of knowing. *Acta Psychologica, 46,* 69–80.

Nelson, T. O., & Narens, L. (1980b). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19,* 338–368.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Read, J. D., & Bruce, D. (1982). Longitudinal tracking of difficult memory retrievals. *Cognitive Psychology, 14,* 280–300.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 39–54.

Seamon, J. G., Brody, N., & Kauff, D. M. (1983). Affective discrimination of stimuli that are not recognized: Effects of shadowing, masking, and cerebral laterality. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 9,* 544–555.

Seamon, J. G., & Virostek, S. (1978). Memory performance and subject-defined depth of processing. *Memory & Cognition, 6,* 283–287.

Shimamura, A. P., Landwehr, R. F., & Nelson, T. O. (1981). FACTRETRIEVAL: A program for assessing someone's recall of general-information facts, feeling-of-knowing judgments for nonrecalled facts, and recognition of nonrecalled facts. *Behavior Research Methods & Instrumentation, 13,* 691–692.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology, 71,* 673–679.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76,* 559–573.

Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child Development, 48,* 13–21.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

Winnick, W. A., & Daniel, S. A. (1970). Two kinds of response priming in tachistoscopic recognition. *Journal of Experimental Psychology, 84,* 74–81.

Wolford, G. (1971). Function of distinct associations for paired-associate performance. *Psychological Review, 78,* 303–313.

Yarmey, A. D. (1973). I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory & Cognition, 1,* 287–290.

Received September 6, 1983
Revision received November 1, 1983 ■