Journal of Experimental Psychology:
Learning, Memory, and Cognition
1986, Vol. 12, No. 2, 279–287

0278-7393/86/$00.75

A Comparison of Three Predictors of an Individual's Memory Performance: The Individual's Feeling of Knowing Versus the Normative Feeling of Knowing Versus Base-Rate Item Difficulty

Thomas O. Nelson, R. Jacob Leonesio, and Robert S. Landwehr
University of Washington

Louis Narens
University of California, Irvine

Distinctions are drawn between different predictors of an individual's memory performance, with emphasis on the notion of privileged access to idiosyncratic knowledge. Research is reported in which undergraduates attempted to recall the answers to general-information questions, then made feeling-of-knowing judgments on nonrecalled items, and subsequently had a criterion test (relearning, perceptual identification, or one of two versions of recognition). For predicting an individual's criterion performance, the individual's own feeling-of-knowing predictions were intermediate between two kinds of normative predictions: The individual's feeling-of-knowing predictions were more accurate than predictions derived from normative feeling-of-knowing ratings but were less accurate than predictions derived from base-rate item difficulty (normative probabilities of correct recall). Subsidiary analyses showed that factors other than unreliability are responsible for the partial inaccuracy of the individual's feeling of knowing. Ramifications are discussed for possible ways to improve the accuracy of an individual's feeling-of-knowing predictions.

> "I know what it's like to be me!" Probably you do, at least about some matters, but that's something to be checked in performance. Maybe, just maybe, you'll discover that you really don't know as much as you thought you did about what it is like to be you. (Dennett, 1981, p. 429)

A key aspect of consciousness in both psychology and philosophy is the concept of privileged access to idiosyncratic knowledge (Nisbett & Wilson, 1977), which has also been referred to as "private experience" (Wittgenstein, 1953), "point of view" (Nagel, 1974), and the "idiosyncratic component" (Underwood, 1966). This concept has been sharpened recently by philosophers (e.g., Nagel, 1974; Dennett, 1978), and Dennett (1978) remarked about the problem of justifying the concept: "The problem begs for a cooperative solution; my attempt trespasses deep in psychologists' territory, and I would hope to stimulate assistance, not a boundary dispute from that quarter" (p. 154).

Psychologists seek empirical justification for the concept[1] by comparing predictions derived from a given individual against predictions derived from other people. The latter predictions, here called "normative," have also been referred to as "actuarial" or "consensus" or "base rate" (Kahneman & Tversky, 1979).

Three different kinds of normative predictions that have sometimes been blurred in the literature need to be distinguished. One kind of normative prediction consists of expectations about someone else's criterion performance (e.g., Vesonder & Voss, in press). Another kind of normative prediction stems from averaging people's expectations about their own criterion performance (Underwood, 1966; Lovelace, 1984). Still another kind of normative prediction stems from the criterion performance of other people on the task that the individual is about to undertake (Brown, 1923; Gruneberg, Monks, & Sykes, 1977; Gruneberg & Sykes, 1981).

Suppose an individual has failed to recall the correct answer to a number of general-information questions such as "What is the capital of Delaware?" and is trying to predict which of those items he or she (hereinafter referred to for brevity as "he") subsequently will recognize the answer to—this prediction is called the *individual's feeling of knowing.* Suppose further that normative item-difficulty predictions, defined in terms of the percentage of people who initially recalled the correct answer to each question, are available.

How accurate will the individual's feeling of knowing be for predicting his subsequent performance on nonrecalled items,

This research was supported by National Institute of Mental Health Grant MH32205.

We thank A. Singh for his computer programming and E. Butterfield, A. Jameson, E. Lovelace, and A. Glenberg for comments on the manuscript.

Correspondence concerning this article should be addressed to Thomas O. Nelson, Psychology Department (NI-25), University of Washington, Seattle, Washington 98195.

[1] At least two possible meanings of "privileged access" can be distinguished. The first meaning refers to an individual's personal knowledge that someone else *might not happen to have,* and the second meaning refers to an individual's personal knowledge that someone else *necessarily does not have* (i.e., the second meaning is a proper subset of the first). For instance, someone else might not happen to have some particular information that you have (e.g., whether the person who burned his hand was you or your friend) but perhaps they could have had that information if they had accompanied you; by contrast, there are other kinds of information that someone else necessarily does not have (e.g., how pain feels to you). Throughout the remainder of this article, "privileged access" will always have the first of these two meanings.

relative to predictions about his subsequent performance that are derived from normative item difficulty? One hypothesis, stemming from the fundamental idea that memory varies across people, is that the individual's feeling of knowing will be more accurate. For instance, to say that the normative percentage correct for 100 people on a given item such as "What is the capital of Delaware?" equaled 70% is tantamount to saying that everyone of 70 people was correct whereas everyone of 30 people was incorrect (i.e., there are variations in memory across the 100 people). Suppose in this example that all 70 people who were correct on the item had previously taken a high-school geography course, whereas all 30 people who were incorrect had not previously taken a high-school geography course; to the degree that the individual incorporates into his predictions the fact that he did/didn't previously take a high-school geography course, the individual can increase his predictive accuracy beyond that of normative item-difficulty predictions that reflect only the average of collapsing over different individuals' past histories. Put differently, the individual can use idiosyncratic information that is lost by the normative item-difficulty predictions due to averaging.

Finally, suppose that normative feeling-of-knowing ratings by everyone who failed to recall the correct answer are also available. A hypothesis similar to the one in the previous paragraph occurs for comparisons of the predictive accuracy of the individual's feeling of knowing versus the normative feeling of knowing. That is, idiosyncratic-background information can be reflected in the individual's predictions but not in the predictions about the individual's performance that are derived by averaging other people's feelings of knowing.

The present research compared the individual's predictions with normative item-difficulty predictions and with normative feeling-of-knowing predictions, both in terms of which particular items are predicted to be easiest (i.e., the degree of relationship between the predictors) and in terms of the relative accuracy at predicting the individual's subsequent memory performance on nonrecalled items.

Method

The data-collection technique for feeling-of-knowing research has recently been improved (Nelson & Narens, 1980a) and incorporated into a computer program called FACTRETRIEVAL (Shimamura, Landwehr, & Nelson, 1981). The FACTRETRIEVAL data-collection program was used unmodified, except as noted below.

Procedure

FACTRETRIEVAL first has the subject go through a recall test for the answers to general-information questions from the norms of Nelson and Narens (1980b) until the subject incorrectly answers (or fails to provide any answer to) a preset number of items. This isolates a set of nonrecalled items, which ranged in size from 12 to 21 items across the present experiments. Then the subject makes feeling-of-knowing predictions by ranking the nonrecalled items in terms of his relative feeling of knowing for every item. For instance, given nonrecall on two questions such as "What is the capital of Chile?" and "What was the name of Tarzan's girlfriend?" the subject indicates whether his feeling of knowing is greater for the answer to the first item or for the answer to the second item. When repeated for every pair of nonrecalled items, this procedure yields a rank order of those items in terms of the subject's relative feeling of knowing for each item. Subsequently, a criterion test occurs on the nonrecalled items to assess the accuracy of the feeling of knowing; the criterion test employed in the standard FACTRETRIEVAL program is a forced-choice recognition test with seven distractors available per item. To the degree that the subject's feeling of knowing is accurate, criterion performance should be better for items nearer the top of the rank order than for items nearer the bottom.

Subjects and Design

The subjects were 168 undergraduate students from the University of Washington who participated for psychology course credit. All four groups made feeling-of-knowing predictions in terms of the question, "Which item are you more likely to recognize the answer to?" The groups differed in terms of the criterion task that followed the feeling-of-knowing predictions. The reason for examining performance on different criterion tasks was to assess the qualitative generality of the results, so as to determine whether the findings are limited to a particular type of criterion task. The data from the first three groups consisted of an analysis of different aspects of the performance observed by Nelson, Gerler, and Narens (1984).

The criterion task for the first group (n = 44) was relearning. For the study phase of relearning, each of the 12 questions that had been answered incorrectly was presented for 5 s, followed by the correct answer for 1 s. After the study phase, there was a self-paced relearning test phase, during which the subject tried to recall the correct answer to each question, with the questions being presented randomly in a different order than during the study phase.

The criterion task for the second group (n = 32) was a perceptual-identification task in which each incorrectly answered question was presented, and the correct answer was flashed for increasing-duration flashes (beginning with an average flash time of 27.4 ms and increasing this duration by an average of 4.1 ms on each flash) until the subject identified the answer. The dependent variable was how quickly the answer was identified (i.e., reciprocal of the number of flashes before perceptual identification occurred). The procedural details of the relearning and perceptual-identification tasks, along with other data not pertinent here, can be found in Nelson et al. (1984).

The criterion task for the third group (n = 43) was a four-alternative forced-choice (4-AFC) recognition test in which each incorrectly answered question was presented along with three distractors from the FACTRETRIEVAL program; each distractor was plausible in that it was from the same category as the correct answer. For instance, the four recognition alternatives for the question "What is the capital of Chile?" were "Santiago," "Quito," "Lima," and "Bogota."

The fourth group (n = 50) was a replication and extension of the third group. To increase the reliability of each person's feeling-of-knowing accuracy score, two changes were made. First, and most important, the recognition test was an eight-alternative forced-choice (8-AFC) recognition test in which each incorrectly answered question was presented with all seven of the distractors stored for that question in the FACTRETRIEVAL program. The greater number of distractors in this 8-AFC test (relative to the aforementioned 4-AFC test) helped to reduce the noise that necessarily occurs in a forced-choice recognition test; that is, the greater the number of distractors, the less likely an item is to be correct by chance alone. Second, the number of nonrecalled items entering into each subject's feeling-of-knowing predictions was raised from 12 to 15. However, because FACTRETRIEVAL's paired-comparison procedure becomes unwieldy with 15 items (i.e., 12 items necessitates 66 paired comparisons, but 15 items necessitates 105), a third change was necessary: FACTRETRIEVAL was modified so that instead of the feeling-of-knowing judgments being made on two nonrecalled items at a time, they were made on three nonrecalled items at a time. These judgments occurred on a computer-determined random sample of all possible unordered triples of nonrecalled items, with the restriction that across the entire set of triples a given nonrecalled item appeared exactly once with every other nonrecalled item—that is, for $K = 15$ items the randomly drawn sample consisted of $K(K - 1)/6 = 35$ triples. For each triple, the person indicated which of the three items he had the greatest feeling of knowing for, and then he indicated which of the remaining two items he had a greater feeling of knowing for. In all other ways, the procedure for this recognition group was identical to that for the other recognition group.

Results

The report of the results is organized into three sections. The first section describes the overall analyses of the accuracy at predicting an individual's subsequent performance on nonrecalled items. The second section describes fine-grained subsidiary analyses that were conducted to clarify the major findings from the first section, with emphasis on the accuracy at predicting the individual's recognition from the individual's feeling of knowing versus from normative item difficulty. The third section examines the relationship between the predictions from the three predictors. Throughout, statistical significance is defined as $p < .05$.

Overall Analyses of the Accuracy at Predicting an Individual's Criterion Performance

For each individual's set of nonrecalled items, we compared the accuracy of three rankings as predictors of the individual's criterion performance. Letting the degree of predictive accuracy, V, be defined as $P(I_c > J_c | I_p > J_p)$—that is, the probability that Item I is higher than Item J in criterion performance, given that Item I is higher than Item J in the predictor ranking—the most appropriate of the available measures of feeling-of-knowing accuracy is the Goodman-Kruskal gamma correlation G (Nelson, 1984). Moreover, because of the linear relationship $V = .5G + .5$ (Nelson, 1984, Eq. 7), all conclusions (e.g., from parametric statistics) involving interval-scale aspects of G will generalize meaningfully to V.

The first predictor was the individual's own feeling of knowing. The items were ranked from highest to lowest in terms of the individual's feeling-of-knowing judgments. Then predictive accuracy was assessed by computing a gamma correlation between the individual's feeling-of-knowing ranking and his criterion performance, where the criterion performance for a given item was correct versus wrong in relearning and recognition, and was the reciprocal of the number of flashes required for correct identification in perceptual identification.

The second predictor was the normative feeling of knowing. The individual's items were ranked in terms of the normative feeling-of-knowing ratings from the Nelson–Narens (1980b) norms; those ratings had been obtained by having subjects who did not recall the correct answer to a given item make a feeling-of-knowing rating about how likely they were to recognize the correct answer. A gamma correlation was computed between the normative feeling-of-knowing ranking and the individual's criterion performance.

The third predictor was normative item difficulty. The individual's items were ranked in terms of the normative probability of recall from the Nelson–Narens (1980b) norms. A gamma correlation was computed between the normative item-difficulty ranking and the individual's criterion performance.

The mean predictive accuracy for each of the three rankings is shown separately for each group in Figure 1. (*Note:* The correlations for the 8-AFC recognition group are based on $n = 49$ rather than $n = 50$ because 1 subject correctly recognized every item and therefore had indeterminate correlations.) Of the 12 correlations shown in Figure 1, all were significantly greater than zero except for the correlations of normative feeling of knowing with 4-AFC recognition and with perceptual identification.

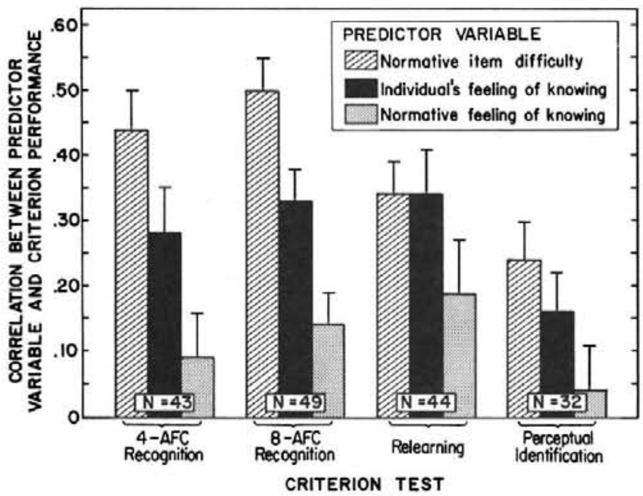

Figure 1. Correlation between each predictor variable and the criterion tests of 4-AFC recognition, 8-AFC recognition, relearning, and perceptual identification. (Height of each bar indicates the mean of the individual-subject correlations, bracket indicates the standard error, and number inside the bar indicates the number of individual-subject protocols upon which the mean is based. AFC = alternative forced choice.)

The most important findings concern how the individual's feeling-of-knowing predictions compare in accuracy with the two kinds of normative predictions. Collapsed across these 168 subjects, the mean correlation between the individual's criterion performance and each of the three predictor variables was +.38 for normative item difficulty, +.28 for the individual's feeling of knowing, and +.12 for the normative feeling of knowing. The difference between the first two means is significant, paired $t(167) = 2.65$, as is the difference between the last two, paired $t(167) = 4.79$. However, because the statistical significance of these differences varied across the four versions of criterion performance, the results shown separately for each group in Figure 1 may be more informative.

First, for every group the mean correlation between the individual's feeling of knowing and criterion performance was at least as high as the mean correlation between the normative feeling of knowing and criterion performance. In three of the groups (4-AFC recognition, 8-AFC recognition, and relearning), the individual's feeling of knowing was a significantly more accurate predictor of his criterion performance than was the normative feeling of knowing. For 4-AFC recognition, paired $t(42) = 2.89$; for 8-AFC recognition, paired $t(48) = 3.33$; for relearning, paired $t(43) = 2.14$; and for perceptual identification, paired $t(31) = 1.33$, $p > .10$ (all ts two-tailed). These results confirm the hypothesis that the individual uses idiosyncratic information that improves the prediction of his subsequent memory performance.

Second, for every group the mean correlation between normative item difficulty and criterion performance was at least as high as the mean correlation between the individual's feeling of knowing and criterion performance. In the two recognition groups, the predictive accuracy from the individual's feeling of knowing was significantly lower than that from normative item difficulty. For 4-AFC recognition, paired $t(42) = 2.14$; for 8-AFC recognition, paired $t(48) = 2.90$; for relearning, paired $t(43) = .01$; and for perceptual identification, paired $t(31) = 1.10$. This finding is surprising because if the individual has a kind of privileged access that is accurate in terms of predicting his recognition performance, then his feeling-of-knowing judgments might be expected to be more accurate than normative item difficulty for making those predictions (i.e., the individual has information about idiosyncratic aspects of his own history that potentially could modulate the predictions derived from normative item difficulty).

Fine-Grained Analyses of Accuracy at Predicting Recognition

Next we explored possible reasons for why the individual's feeling of knowing was not more accurate than normative item difficulty as a predictor of his recognition.

Reliability of the individual's feeling-of-knowing judgments. One possible explanation is that the individual's feeling-of-knowing judgments have low reliability (at least lower than the predictions from normative item difficulty, which have binomial variability in terms of the proportion of subjects who correctly recalled each item). An estimate of reliability is available from two groups (relearning and 4-AFC recognition) because after the 66 paired comparisons were completed, the subject made retest feeling-of-knowing judgments on 12 paired comparisons randomly chosen from the first 50 paired comparisons. For each subject the proportion of reliable feeling-of-knowing paired comparisons, designated P(reliability), was determined by computing the proportion of times that the retest judgments were the same as the original judgments. Across all retested paired comparisons, the mean P(reliability) was .83 (Nelson et al., 1984). Although fairly high, this may have been imperfect enough to account for the reduced predictive accuracy from the individual's feeling of knowing, relative to that from the normative item-difficulty predictions.

To explore this possibility, we partitioned the retested pairs in terms of the number of other items in the individual's feeling-of-knowing ranking that intervened between the two items in a given retested pair. For instance, if the observed feeling-of-knowing ranking for some items was J, R, G, K, M, then two items intervene between items J and K (namely, items R and G), no items intervene between items R and G (i.e., those two items are adjacent), three items intervene between items J and M, and so on.[2] Each subject's P(reliability) was computed for each partition of retested pairs. The results are summarized in Figure 2.

Given that the items with tied ranks are not discriminably different in terms of the individual's feeling of knowing, it is not surprising that the retest reliability was at chance for pairs consisting of tied items. (*Note:* The unreliability here is probably due to the procedure rather than to the people, because the procedure did not allow a response of "indifferent" as a feeling-of-knowing judgment on any pair of items.) Of greater importance, the reliability of the individual's feeling-of-knowing judgments

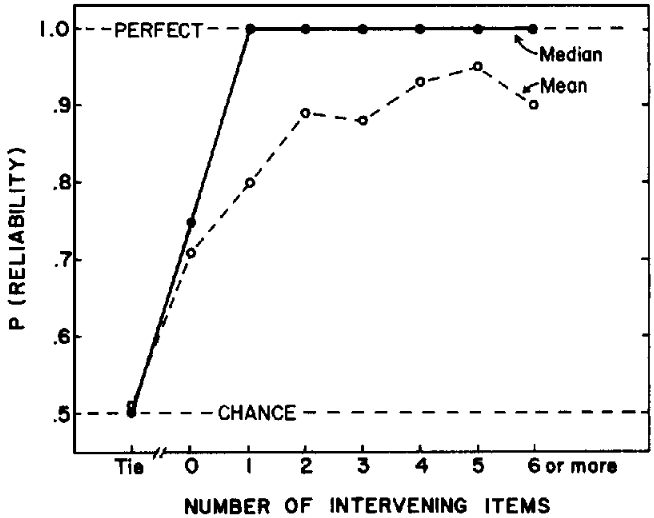

Figure 2. Proportion of reliable feeling-of-knowing paired comparisons as a function of the number of other items that intervene in the individual's feeling-of-knowing ranking between the two items being compared, ranging from pairs of items tied in the feeling-of-knowing ranking to pairs with six or more other items intervening. (Solid curve indicates the median of the individual-subject proportions, and dashed curve indicates the mean.)

increased dramatically to a high level for pairs with even a small number of intervening items. For instance, Figure 2 shows that the median P(reliability) of the feeling-of-knowing judgments was *perfect* for pairs with one or more intervening items. Thus, there appears to be little or no unreliability in the feeling-of-knowing judgments, except for items that are tied or adjacent in the feeling-of-knowing ranking (and adjacency in the observed ranking may sometimes be due to the paired comparison on two items tied in the underlying feeling of knowing).

Predictive accuracy as a function of the number of intervening items. We next determined whether the lower predictive accuracy for the individual's feeling of knowing than for normative item difficulty is related to the number of items intervening between the two items in a given pair. Given that the unreliability of an individual's feeling of knowing is confined primarily to pairs with zero items intervening between the two items comprising the pair, perhaps all of the reduced predictive accuracy of the individual's feeling of knowing is also confined to those pairs. This would be expected if all of the reduced predictive accuracy of the individual's feeling of knowing is due to unreliability.

The accuracy of the predictor variables was analyzed as a function of the number of items intervening between the two items in each pair (tied items are excluded because the feeling-of-knowing ranking of them is indeterminate).

[2] Aggregating the retested pairs into partitions in this way is somewhat crude because the placement of a given pair depends upon the other items in the feeling-of-knowing ranking, not just upon the given pair. However, until an alternative aggregation procedure with a better scaling foundation is developed, this scheme may be useful for providing a rough approximation of the spread between the two items in a given pair.

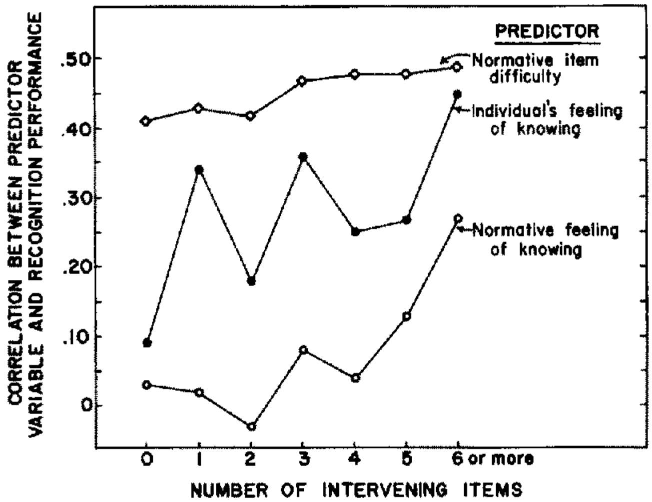

Figure 3. Mean of the individual-subject correlations between predictor variable and performance on the recognition test for the predictor variables of normative item difficulty, the individual's feeling of knowing, and the normative feeling of knowing as a function of the number of other items that intervene in the individual's feeling-of-knowing ranking between the two items being compared, ranging from pairs with no other items intervening (i.e., items adjacent in the feeling-of-knowing ranking) to pairs with six or more other items intervening.

The results from the 93 subjects in the recognition groups, summarized in Figure 3, show that the extra predictive accuracy for normative item difficulty is not confined to pairs with zero intervening items. The difference in predictive accuracy for normative item difficulty versus the individual's feeling of knowing is significant even by a sign test on the means from the partitions of one through six or more intervening items (*Note:* Every partition consists of different subsets of items). Also, Figure 3 shows that all three predictors tend to increase in accuracy as the number of intervening items increases, probably not because of the greater number of intervening items per se but rather because a greater number of intervening items is likely to reflect a greater spread in whatever does underlie the predictions (cf. the measurement concept of a semiorder, e.g., Coombs, Dawes, & Tversky, 1970). Finally, Figure 3 indicates that as the number of intervening items increases, the predictive accuracy of the individual's feeling of knowing begins to approach that of normative item difficulty. Although the individual's feeling of knowing is less accurate than normative item difficulty as a predictor here, in other situations the opposite may occur.[3]

Predictive accuracy of the individual's feeling of knowing versus normative item difficulty for extreme items. Data were also examined from only the most extreme items in the individual's feeling-of-knowing ranking—that is, the two items for which the individual judged his feeling of knowing to be the highest versus lowest. There were 48 subjects whose recognition performance differed on these extreme items and therefore whose gamma correlation on extreme items was not indeterminate. The mean gamma correlation (and the standard error of the mean) on these items was +.77 (±.09) for the individual's feeling of knowing versus +.69 (±.09) for normative item difficulty, but this difference is not statistically significant, paired $t(47) = 1.00$.[4] Notice that the mean gamma correlation of +.77 for the individual's feeling-of-knowing accuracy here is remarkably higher than any of the corresponding values in Figures 1 and 3, demonstrating that the observed degree of feeling-of-knowing accuracy for discriminating between various nonrecalled items is strongly dependent upon the particular items being judged (cf. psychophysical discrimination between two tones being dependent upon the particular tones being discriminated).

Accuracy of prediction for items that another predictor does not discriminate. Besides analyzing predictive accuracy on items that are the *most* discriminable in the individual's feeling-of-knowing ranking, we also examined items that are the *least* discriminable for one of the predictors. Two questions were investigated: (a) Can normative item difficulty make accurate predictions about items that are not discriminated by the individual's feeling of knowing? and (b) Can the individual's feeling of knowing make accurate predictions about items that are not discriminated by normative item difficulty?

First we isolated items that were not discriminably different in the individual's feeling-of-knowing ranking (i.e., tied items) but for which the individual's subsequent recognition differed (i.e., correct recognition on one item but incorrect recognition on another). The mean gamma correlation (and standard error of the mean) between normative item difficulty and the individual's recognition on those items was +.28 (±.11), demonstrating that normative item difficulty has significant accuracy at predicting the individual's recognition performance on items that the individual's feeling of knowing does not discriminate (also

[3] Although normative item difficulty is more accurate than the individual's feeling of knowing at relative prediction (i.e., one item relative to another item), we would not expect the same to hold for all aspects of absolute prediction. In absolute prediction (e.g., Lichtenstein & Fischhoff, 1977), the individual predicts his probability of correctly recognizing a given item; then the items with the same predicted probability are aggregated to determine the actual probability of correct recognition, which is compared against the individual's predicted probability. If an investigator so desired, the individual's actual probabilities of recognition could also be compared against predictions from normative item difficulty. A given individual might know that his memory is generally poor and therefore be accurate in estimating probabilities that are uniformly lower than those from normative item difficulty; however, this is knowledge about the general individual rather than knowledge about specific items in memory and therefore is different from the aspect being investigated here. In the terminology of Lichtenstein and Fischhoff (1977), our focus is on "resolution" rather than "calibration"; resolution, in comparison with calibration, "is a more fundamental aspect of probabilistic functioning" (p. 181). If an investigator conducting an absolute-prediction experiment were to rank the items in terms of the probabilities predicted by the individual, and the accuracy at predicting recognition is compared for that ranking versus the ranking from normative item difficulty, then we expect the outcome to be the same as here.

[4] By contrast, when extreme items are defined in terms of normative item difficulty (i.e., the two items for which normative item difficulty was the highest vs. lowest), the mean gamma correlation for predicting the individual's performance was +.50 (±.11) for the individual's feeling of knowing and +.87 (±.06) for normative item difficulty, paired $t(59) = 3.64$. However, defining extreme items in terms of the individual's feeling of knowing is more pertinent to the issue of whether there are any items for which the predictive accuracy is greater for the individual's feeling of knowing than for normative item difficulty.

found by Gruneberg et al., 1977, but note that they used a rating scale that forced ties in the feeling-of-knowing judgments).

Second, we isolated items that were not discriminably different in normative item difficulty but that differed during the individual's subsequent recognition. However, when these were required to be identical in normative item difficulty, there were too few items for stable results; such items occurred for only 14 subjects, with the mean gamma correlation between the individual's feeling of knowing and subsequent recognition having a positive value of +.21 (but with an unacceptably large standard error of ±.26). Therefore, the algorithm defining "not discriminably different" was broadened to include all pairs of items in which either of the two items' normative proportions was within one standard deviation of the other's. This algorithm yielded data from 75 subjects, with the mean gamma correlation between the individual's feeling of knowing and subsequent recognition having a significantly positive value of +.34 (±.08). To ensure that this did not reinstate the predictive accuracy of normative item difficulty, the mean gamma correlation between normative item difficulty and the individual's recognition was computed for those items; the obtained value was +.08 (±.09), which is not significantly different from zero. In a direct comparison between the two predictors, a paired t test showed that the individual's recognition on those items was predicted significantly more accurately by the individual's feeling of knowing than by normative item difficulty, $t(73) = 2.15$. This is an important demonstration that the individual's feeling of knowing can, under some circumstances, yield more accurate predictions than those from normative item difficulty.

Accuracy of predicting recognition for nonrecalled items that were commission errors versus omission errors. Two previous studies of the feeling of knowing (Krinsky & Nelson, 1985; Nelson et al., 1984, Figure 3) found effects that differed depending upon whether the recall errors had been commissions (i.e., incorrect answers) versus omissions (i.e., no answer guessed during recall). We explored whether the pattern of relationships between the present three predictors and the individual's recognition (cf. Figure 1) is different for commissions versus omissions by partitioning the data from the recognition groups into commissions versus omissions and then recomputing the correlations. The mean correlations are shown in Figure 4 for each of the three predictors on each of the two types of recall error. Although the overall pattern tends to be accentuated for commissions, there is no significant interaction between the predictor variable and the type of recall error. A two-way repeated-measures ANOVA yielded $F(2, 108) = 1.67$ for the interaction; the main effect of predictor variable was highly significant, $F(2, 108) = 12.95$, and the main effect of type of recall error was nonsignificant, $F(1, 54) < 1$. Thus, in contrast to other aspects of the feeling of knowing (Krinsky & Nelson, 1985; Nelson et al., 1984, Figure 3), the pattern of predictive accuracy for the present three predictors is not modulated significantly by the type of recall error.

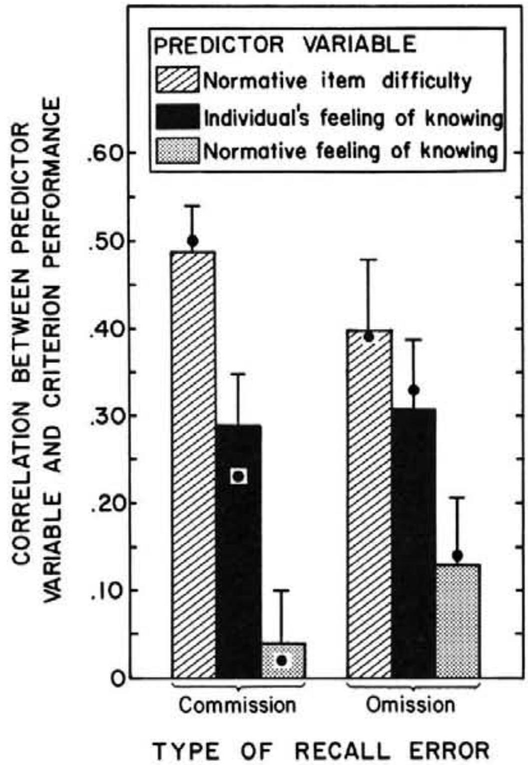

Figure 4. Correlation between each predictor variable and recognition performance for items that had been commission errors versus omission errors during recall. (Height of each bar indicates the mean of the individual-subject correlations [for the 83 subjects yielding complete commission data and the 64 subjects yielding complete omission data], bracket indicates the standard error, and dots indicate the corresponding means for the 55 subjects yielding complete commission and omission data entered into the repeated-measures ANOVA.)

Relationships Between the Three Predictors

In addition to examining the relative accuracy of the three predictors, we also investigated the relationships between those three predictors. The correlations between the three predictors' rankings of the individual's nonrecalled items are shown in Table 1. (*Note:* In contrast to Figure 1, where criterion performance could affect the relationships, criterion performance was not involved in the relationships summarized in Table 1, where each group is essentially a separate replication.) Each of the mean correlations in the table is significantly greater than zero.

The first column shows a positive relationship between the individual's feeling of knowing and the normative feeling of knowing, confirming the idea that people are somewhat similar in terms of what they believe they know about items they cannot recall (Koriat & Lieblich, 1977). The intermediate degree of relationship suggests that although there may be some common bases for different individuals' feeling-of-knowing judgments, there are also idiosyncratic factors that make one person's feeling-of-knowing ranking different from someone else's (cf. Lovelace, 1984).

The second column shows a small positive relationship between what an individual believes he knows about nonrecalled items and what people in general do know. The third column shows a positive relationship between what people who fail to recall items believe they know about those items and what people in general do know about the items.

Table 1
Relationship Between the Predictions From Three Predictor Variables for Each Criterion-Performance Group

Group	Predictor variables being correlated					
	IFK–NFK	SE_M	IFK–NID	SE_M	NFK–NID	SE_M
4-AFC recognition (n = 43)	.44	±.04	.25	±.03	.35	±.04
8-AFC recognition (n = 50)	.43	±.03	.14	±.02	.21	±.03
Relearning (n = 44)	.41	±.03	.19	±.03	.32	±.04
Perc. ident. (n = 32)	.33	±.06	.18	±.05	.34	±.04
All groups (n = 169)	.41	±.02	.19	±.02	.30	±.02

Note. Cell entry is mean gamma correlation, given along with standard error of that mean (SE_M).
IFK = individual's feeling of knowing; NFK = normative feeling of knowing; NID = normative item difficulty; Perc. ident. = perceptual identification.

A comparison of the first two columns showed that the individual's feeling of knowing is more highly correlated with what people who fail to recall the item believe they know than with what people in general do know. This difference is significant for every group as well as for the combined data, paired *t*s (with *df* of $n - 1$) = 4.12, 8.93, 5.74, 2.90, and 10.08, respectively.

A comparison of the last two columns showed that normative item difficulty is more highly correlated with the normative feeling of knowing than with the individual's feeling of knowing (significant for every group as well as for the combined data, paired *t*s = 2.55, 5.82, 3.25, 2.35, and 5.75, respectively). This suggests that across individuals the idiosyncratic factors may cancel each other, making overall estimates of item difficulty more prominent in the normative feeling-of-knowing ratings than in the individual's feeling of knowing.

A comparison of the first and third columns showed that the normative feeling of knowing tends to be more highly correlated with the individual's feeling of knowing than with normative item difficulty. This difference is significant for the combined data and for the 8-AFC recognition group, is marginally significant for the 4-AFC recognition group and the relearning group, and does not approach significance for the perceptual-identification group, paired *t*s = 3.91, 2.27, 1.68 (p = .10), 1.74 (p = .09), and .20, respectively.

Finally, what an individual feels he knows is more related to what other people who also fail to recall the items feel they know (mean correlation = +.41, first column of Table 1) than to what he actually does know, when "knowing" is operationalized in terms of subsequent memory performance (mean correlation = +.28, reported above). This difference is significant, paired t (167) = 3.41, and is consistent with previous findings (Koriat, 1976).

Discussion

The Individual's Feeling of Knowing Versus the Normative Feeling of Knowing

What if every individual's point of view (Nagel, 1974) is unique? Sometimes the individual's point of view would yield a product differing from the normative product (i.e., an idiosyncratic outcome), and at other times the individual's product would be the same as the normative product (i.e., a nonidiosyncratic outcome).[5] The reliably-below-unity correlations between the individual's feeling of knowing and the normative feeling of knowing (Table 1), especially given the high reliability of the individual's feeling of knowing (Figure 2), imply that every individual's point of view differs from at least someone else's. If points of view did not differ across people, everyone's rankings of a given set of nonrecalled items would, within the bounds of sampling error, be identical.

A critical determinant of whether an individual's point of view yields an idiosyncratic outcome is the heterogeneity/homogeneity of the comparison population. Another critical determinant, if that population is heterogeneous, is the similarity between the particular individual's product and the normative product.

What theoretical structure underlies the observed idiosyncracies? Multidimensional aspects are presumably combined into a point of view that produces a greater feeling-of-knowing rank for one item than another. Accordingly, people's points of view could differ either because they contain different components of information (e.g., the particular courses taken in high school), because they have different weights for the various components (e.g., greater weight on high school courses than on recency), or because they use different combinatorial rules (e.g., regression rules vs. lexicographic rules). That is, the idiosyncracies could stem either from monitoring different components or/and from drawing different inferences from the components that are monitored.

At least some of the idiosyncratic aspects of the individual's feeling of knowing are useful for enhancing predictive accuracy, as demonstrated by the recurring finding of greater predictive accuracy for the individual's feeling of knowing than for the normative feeling of knowing (Figures 1, 3, and 4). Our explanation for this finding is straightforward: The normative feeling of knowing eliminates (by averaging) individuating information that is predictive of subsequent memory performance and that is reflected in the individual's feeling of knowing.

The Individual's Feeling of Knowing Versus Normative Item Difficulty

The low correlation between the individual's feeling of knowing and normative item difficulty (Table 1), especially given the high

[5] Hereafter, we will use the term "point of view" to refer to the nonobservable psychological process that produces the observable feeling-of-knowing judgments, and we will use the term "idiosyncratic" to refer to an observable outcome in which there is a difference between something (e.g., feeling-of-knowing judgments) produced by a particular individual versus by other people. Thus the occurrence of a unique point of view is hypothetical, whereas the occurrence of an idiosyncratic outcome is empirical.

reliability of the individual's feeling of knowing (Figure 2), suggests that the individual's feeling of knowing is based on different subsets of components and/or ways of combining components than normative item difficulty. The different predictions from normative item difficulty and the individual's feeling of knowing yielded two outcomes. First, there was the relatively infrequent outcome of greater predictive accuracy from the individual's feeling of knowing than from normative item difficulty (e.g., on items that were not discriminably different in normative item difficulty).

Second, there was the relatively frequent outcome of greater predictive accuracy from normative item difficulty than from the individual's feeling of knowing (e.g., Figures 1, 3, and 4). This is a theoretical puzzle that the present research raises but does not solve. One plausible explanation, namely, that the individual has access to accurate information coupled with noise, was disconfirmed by the high level of reliability of the individual's feeling of knowing (Figure 2). Some mechanisms that might instead yield systematic distortions in the feeling of knowing and thereby produce inaccuracy with little or no noise are described in Nelson et al. (1984) and in Krinsky and Nelson (1985, Figure 1).

Finally, it can be noted that the lower accuracy from the individual's feeling of knowing than from normative item difficulty helps provide a scientific explanation for the supposedly nonscientific phenomenon of "mind reading": A medium can use the power of base rates to predict or retrodict someone's psychological performance without having any access to that person's mind—by knowing the group's performance, one can predict the individual's performance.

Improving the Predictive Accuracy of an Individual's Feeling-of-Knowing Judgments

Because normative item difficulty is frequently a better predictor than the individual's own feeling of knowing, future research should explore the possibility of having the individual incorporate normative item difficulty (or at least his estimates of it and/or whatever underlies it) into his feeling-of-knowing judgments. However, the low correlation between the individual's feeling of knowing and normative item difficulty does not necessitate that the individual could make his predictions more in accord with normative item difficulty. The extent to which people have insight into normative item difficulty that they are not already using in their feeling-of-knowing judgments is unknown.

Does the individual have reasonably accurate estimates of normative item difficulty? Nickerson, Baddeley, and Freeman (1984) asked University of Massachusetts undergraduates to estimate the normative probability of recall for 90 questions from the Nelson–Narens norms. For items that the undergraduates did not themselves recall, the slope of a linear regression line was +.24 when the estimated probabilities were plotted as a function of the normative probabilities ($Y = .24X + .29$).

When making feeling-of-knowing judgments, does the individual underutilize his estimates of normative information? Such underutilization is well known in other areas of psychology (e.g., Kahneman & Tversky, 1973; Nisbett & Borgida, 1975). Reviews of those literatures have concluded that the underutilization of normative information "is perhaps the major error of intuitive prediction" (Kahneman & Tversky, 1982, p. 416), and "it would appear that reasonably accurate base rate information is available to subjects but that they simply do not make appropriate use of it" (Borgida & Brekke, 1981, p. 72).

Perhaps during feeling-of-knowing judgments the individual underutilizes his estimates of normative item difficulty and overemphasizes individuating factors (cf. Bradley, 1981; Koriat & Lieblich, 1977). Persuading the person to de-emphasize individuating factors might increase the emphasis on normative factors (Ginosar & Trope, 1980), and strategies are available for increasing people's use of normative information (Borgida & Brekke, 1981; Kahneman, Slovic, & Tversky, 1982, Part VIII; Kassin, 1979). An interesting question for future research is whether individual idiosyncracies can be combined with estimates of normative item difficulty to yield feeling-of-knowing predictions whose accuracy exceeds that from normative item difficulty alone.

The present article is the first to suggest that the individual's feeling of knowing can be made more accurate. Developing techniques for improving the individual's feeling-of-knowing accuracy may be difficult but seems worthwhile because as feeling-of-knowing accuracy improves, so might other kinds of memory performance that potentially are mediated by the feeling of knowing, such as memory-search termination (Gruneberg et al., 1977; Nelson et al., 1984) and inferences during decision making (Gentner & Collins, 1981). Meanwhile, however, if someone wants to guess which nonrecalled items an individual is most likely to recognize, the present results demonstrate that the normative probability of recall can be a more accurate predictor than the individual's own opinion—empirically confirming the speculation by Dennett (1981) that began this article.

References

Borgida, E., & Brekke, N. (1981). The base-rate fallacy in attribution and prediction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research.* Hillsdale, NJ: Erlbaum.

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

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

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

Dennett, D. C. (1978). Toward a cognitive theory of consciousness. In D. C. Dennett (Ed.), *Brainstorms: Philosophical essays on mind and psychology* (pp. 149–173). Montgomery, VT: Bradford Books.

Dennett, D. C. (1981). Reflections. In D. R. Hofstadter & D. C. Dennett (Eds.), *The mind's I* (pp. 427–429). New York: Basic Books.

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

Ginosar, Z., & Trope, Y. (1980). The effects of base rates and individuating information of judgments about another person. *Journal of Experimental Social Psychology, 16,* 228–242.

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., & Sykes, R. N. (1981). Eyewitness accuracy and feelings of certainty. *IRCS Medical Science: Psychology and Psychiatry; Social and Occupational Medicine, 9,* 935–936.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science, 12,* 313–327.

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge, England: Cambridge University Press.

Kassin, S. M. (1979). Consensus information, prediction and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology, 37,* 1966–1981.

Koriat, A. (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition, 4,* 244–248.

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

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica, 58,* 141–158.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20,* 159–183.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 756–766.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83,* 435–450.

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., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

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.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1984). *Are people's estimates of what other people know influenced by what they themselves know?* (Technical Report). Cambridge, MA: Bolt, Beranek, & Newman, Inc.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32,* 932–943.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

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.

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

Vesonder, G. T., & Voss, J. F. (in press). On the ability to predict one's own responses while learning. *Journal of Memory and Language.*

Wittgenstein, L. (1953). *Philosophical investigations.* New York: Macmillan.

Received September 24, 1984
Revision received July 10, 1985 ■