New Zealand Journal of Psychology, 1989, 18, 71–75

Knowing Your Limits: Expertise and the Feeling of Knowing*

Llinos Silyn Roberts and Gillian Rhodes
Department of Psychology,
University of Otago

Inability to answer a question or to recall some fact may be accompanied by the feeling that one knows the required information but is temporarily unable to retrieve it from memory. Such "feeling of knowing" (FOK) judgements are generally reliably above chance in predicting subsequent recognition performance (Blake, 1973). In the present experiment we investigated the relationship between expertise in a field and the strength and accuracy of FOK judgements. Self-rated experts on general knowledge questions correctly answered more general knowledge questions than did non-experts. However, neither FOK strength nor the accuracy of FOK judgements about unanswered questions were related to either self-rated or objectively assessed expertise level.

At one time or another, all of us will have experienced the frustration of being unable to recall some name, date or fact, that we feel we know. The information may seem to be on the tip-of-the-tongue, yet remain just out of reach. (Brown & McNeill, 1966; Koriat & Lieblich, 1977). The feeling that one knows the information despite being unable to retrieve it from memory is called the *feeling of knowing (FOK)* (Hart, 1965).

FOK research has focussed on the accuracy of FOK judgements for predicting subsequent memory performance on non-recalled items. Although far from perfect, FOK judgements are reliably above chance in predicting subsequent recognition (Blake, 1973), the effectiveness of retrieval cues such as the first letter of the item (Gruneberg & Monks, 1974), perceptual identification and relearning performance (Nelson et al., 1984). People, therefore, have some ability to assess what they do and don't know, even when they can't access the information concerned.

Less is known about how such metacognitive judgements (judgements about cognition) are made. Nelson, Gerler, and Narens (1984) note that, although it is premature to discuss detailed mechanisms, two classes of theroretical models exist. In the first class of models, FOK judgements are based on partial access to the non-recalled item. For example, the item may have some level of activation associated with it, which is too low to trigger recall, but is sufficient to generate a FOK. Alternatively, part of the item (e.g., its syllabic stress pattern) may be recalled. In the second class of models FOK judgements are inferences based on access to non-target information such as knowledge about or familiarity with the retrieval cue (the question asked). One type of inferential mechanism that has been suggested is that FOK is inferred from self-rated expertise with the general topic of the item, so that strong feelings of expertise lead to strong FOK (Nelson et al., 1984). This "inference-from-expertise" hypothesis, however, has yet to be tested.

The present experiment was designed to investigate the relationship between expertise and FOK and to test the inference-from-expertise hypothesis. If this hypothesis is correct, then greater expertise should be associated with stronger FOK. Changes in expertise might also affect FOK accuracy. Increasing expertise not only increases the amount of knowledge that a person has, but also causes restructuring of that knowledge: new relations among concepts may be represented and the core concepts themselves may change (for reviews see Carey, 1985; Chi, Glaser & Rees, 1982). It has also been claimed that experts have a fundamentally different, more intuitive (or at least more automatic and unconscious) cognitive style than novices (Chi, Glaser & Rees, 1982; Dreyfus & Dreyfus, 1986). How might these changes in the knowledge

* We thank Tom Nelson for supplying the FACTRETRIEVAL2 program and for his helpful advice. We also thank Jack Clarkson for comments on an earlier version of the manuscript, the Otago University Computer Services Centre and Department of Mathematics for use of their computers, Michael Hamel for technical advice, and Cadbury (NZ) Ltd for donating the chocolate fish.

Requests for reprints should be sent to Dr G. Rhodes, Department of Psychology, University of Canterbury, Private Bag, Christchurch, New Zealand.

base and access mechanisms affect FOK accuracy?

Two competing hypotheses seem plausible. First, FOK judgements might become more accurate as expertise increases, due to increased insight into one's knowledge or skills in a domain (the *insight hypothesis*)[1]. Indirect support for the insight hypothesis comes from evidence that young children, who have been regarded as the ultimate novices, have poor metacognitive abilities (Flavell, 1985; Wellman, 1977). Alternatively, as expertise increases, so too may the tendency to assume that inability to answer a question merely represents a temporary retrieval failure. This assumption may not solely reflect arrogance on the part of the expert, although experts are often unwilling to admit ignorance (Bradley, 1981). Rather, the presence of many memory structures in the domain of expertise, some of which would be activated by cues in the question (albeit too weakly to trigger an answer) could lead to strong FOK. However, if this strong FOK is not associated with superior recognition performance FOK accuracy may actually decrease with increasing expertise (the *overconfidence hypothesis*)[2].

To summarise, both hypotheses predict that FOK strength increases with increasing expertise. However, the two make opposite predictions concerning the relationship between FOK accuracy and level of expertise. According to the insight hypothesis FOK accuracy increases with increasing expertise, whereas according to the overconfidence hypothesis FOK accuracy decreases with increasing expertise. While it may seem a rather conservative strategy to predict that FOK accuracy either goes up or goes down with increasing expertise, in the absence of direct evidence on this question, both hypotheses seem plausible.

[1] Complete novices may also have accurate metacognitive judgements, because they don't know anything and they know it! However, in most areas of human cognition the complete novice is rare. According to the insight hypothesis, given some degree of knowledge/skill, greater expertise should be associated with more accurate metacognitive judgements.

[2] Slovic, Fischhoff and Lichtenstein (1977) reviewed studies showing that people are overconfident about the accuracy of their answers to questions and that experts are similarly overconfident. FOK judgements, however, are concerned with confidence in one's ability to recognize non-recalled information, and it is with this rather different type of judgement that the present experiment is concerned.

Method

Subjects

Forty-two students from the University of Otago participated in the experiment.

Equipment

Data were collected using the program FACTRETRIEVAL2 (Wilkinson & Nelson, 1984) run on an Apple IIe computer.

Procedure

At the beginning of the experiment subjects rated their performance at answering general knowledge questions on a 7-point scale (1 = poor, 7 = very good). The experiment consisted of four phases:

Preliminary Phase. Subjects were told that they would be asked general knowledge questions and that they would be rewarded (with chocolate fish) according to the proportion of questions answered correctly. Between one and three chocolate fish were earned depending on the proportion of questions answered correctly during the recall phase. Two practice questions were given, for example, "Who is the current President of the United States?"

Recall Phase. Questions on a variety of topics including geography, history, sport, art, science, literature and entertainment were presented on the computer monitor. Subjects responded by typing in the answer on the keyboard. To avoid mis-spelling errors, only the first three letters of each answer were used by the computer. Subjects also rated their confidence in each answer on a 6-point scale. If the subject had no idea of the correct answer they typed "NEXT". The recall phase was terminated when the subject had made 12 incorrect responses (not counting "Next" responses).

Feeling of Knowing Phase. Subjects made absolute FOK judgements followed by relative FOK judgements on the 12 incorrect items. For the absolute FOK judgements subjects rated FOK for each question on a 6-point scale indicating whether they would be able to recall the answer to the question at a later time (1 = "definitely no", 6 = "definitely yes"). For the relative FOK judgements subjects were presented with pairs of questions and asked to indicate which they would be most likely to recognize the answer to. The items were ranked from highest to lowest to give the relative FOK measure.

Recognition Phase. Subjects were presented with a 7-alternative forced-choice recognition test for each of the 12 non-recalled items. The program draws 6 distractors from a pool of 7 in the same general category as the correct answer, with the constraint that the subject's previous answer cannot be used (this prevents perseveration errors).

Results

Subjects were divided into three equal-sized expertise groups (N = 14 per group) based on their self-rated ability to answer general knowledge questions. Subjects in the high, medium and low expertise groups rated themselves greater than 4 (M = 5.3), 4, and less than 4 (M = 2.5), respectively on the 7-point rating scale. High, medium and low should be regarded as relative terms only, because there are no norms for these ratings.

FOK accuracy was measured by the (Goodman-Kruskal gamma) correlation between FOK and Recognition Performance (Absolute FOK vs. Recognition and Relative FOK vs. Recognition).[3] Correlations were also calculated between Absolute FOK and Relative FOK (to assess reliability), and between Confidence and Recall Accuracy. The mean correlation for each expertise group on each of the four measures is shown in Table 1. All correlations were significantly above zero for all expertise groups, all t's $>$ 3.84, indicating that FOK was significantly correlated with recognition performance, that Absolute and Relative FOK were significantly correlated and that confidence was significantly correlated with recall accuracy.

An ANOVA was carried out for each of the four correlations with expertise level as a between subjects factor. There was no significant main effect of expertise level on FOK accuracy (Relative FOK vs. Recognition or Absolute FOK vs. Recognition), on the correlation between Absolute and Relative FOK or on the correlation between Confidence and Recall, all F's $<$ 1.

In order to validate our assignment of subjects to expertise groups, we calculated the Proportion Correct in the Recall phase for each of the three groups (see Table 1). Table 1 also shows the Proportion Correct in the Recognition phase, the Proportion of Questions Passed in the Recall phase, and the mean Absolute FOK. High expertise subjects were expected to perform best in the recall phase, but not necessarily in the recognition phase, to pass on fewer questions and to have higher Absolute FOK ratings than less expert subjects, as predicted by Nelson et al. (1984).

ANOVAs with expertise level as a between subjects factor were carried out on these four dependent measures. There was a significant effect of expertise for Proportion Correct in the Recall phase, $F(2,39) = 4.68$, $p < .02$. As predicted, the high expertise group performed best. Planned t-tests showed that the high expertise group performed significantly better than the medium expertise group, $t(39) = 2.27$, $p < .025$ and the low expertise group, $t(39)$

[3] Nelson (1984) compared the mathematical properties of various measures of FOK accuracy and concluded that the Goodman-Kruskal gamma correlation between FOK and recognition performance was the best.

Table 1: *Dependent Measures as a function of Self-rated Expertise*

	Expertise		
Dependent Measures	High	Medium	Low
FOK Accuracy:			
Relative FOK vs. Recognition (gamma)	.43 (.11)	.55 (.07)	.56 (.10)
Absolute FOK vs. Recognition (gamma)	.49 (.10)	.62 (.07)	.57 (.10)
Absolute FOK vs. Relative FOK	.75 (.05)	.75 (.05)	.78 (.04)
Confidence vs. Recall	.91 (.02)	.94 (.02)	.91 (.02)
P(correct) for Recall	.56 (.03)	.44 (.04)	.41 (.04)
P(correct) for Recognition	.57 (.05)	.45 (.04)	.44 (.05)
Proportion Questions Passed for Recall	.19 (.03)	.22 (.03)	.31 (.05)
FOK Strength (Mean Absolute FOK Rating)	4.10 (.14)	3.75 (.13)	3.88 (.15)

Note. Standard errors in parentheses.

Table 2: *Dependent Measures as a function of Objectively Assessed Expertise*

Dependent Measures	Expertise	
	High	Low
FOK Accuracy:		
Relative FOK vs. Recognition (gamma)	.59 (.10)	.45 (.11)
Absolute FOK vs. Recognition (gamma)	.60 (.10)	.47 (.11)
Absolute FOK vs. Relative FOK	.78 (.04)	.73 (.04)
Confidence vs. Recall	.89 (.03)	.93 (.02)
P(correct) for Recall	.64 (.02)	.29 (.03)
P(correct) for Recognition	.55 (.05)	.40 (.06)
Proportion Questions Passed for Recall	.13 (.02)	.34 (.05)
FOK Strength (Mean Absolute FOK Rating)	3.97 (.16)	3.88 (.19)

Note. Standard errors in parentheses.

= 2.91, $p < .005$. There was no significant difference between the medium expertise group and the low expertise group, $t < 1$. There was a marginally significant effect of expertise on Proportion Correct in the Recognition phase, $F(2,39) = 2.58, p < .09$, with the high expertise group performing better than the other two groups, although none of the differences were significant by a post hoc Scheffé test, all t's < 2.05. There was no effect of expertise level on the Number of Questions Passed, $F(2,39)=2.70$, or on the mean absolute FOK ratings, $F(2,39) = 1.53$.

It is possible that the three expertise groups were not different enough or that self-rated expertise was not associated with actual expertise. A second set of analyses was therefore carried out using two extreme groups (see Table 2). A high and a low expertise group were selected on the basis of performance in the recall phase of the experiment. The top 12 subjects were assigned to a high expertise group and the bottom 12 subjects were assigned to a low expertise group. Of the 12 subjects in the high expertise group, 8 were in the high, 2 in the medium and 2 in the low expertise groups based on self assessed expertise. Of the 12 subjects in the low expertise group, 6 were in the low, 4 in the medium and 2 in the high expertise groups based on self-assessment.

Both FOK accuracy measures (Absolute FOK vs. Recognition and Relative FOK vs. Recognition) were significantly above zero, $t(11) = 4.33, p < .001$ (low expertise subjects), $t(11) = 6.27, p < .001$ (high expertise subjects), and $t(11) = 4.22$, $p < .001$ (low expertise subjects), $t(11) = 5.74, p < .001$ (high expertise subjects), for Absolute and Relative FOK accuracy, respectively.[4] There was no significant effect of expertise on either of the FOK accuracy measures, both F's < 1.

There was a significant effect of expertise for the Proportion Correct in the Recall phase, $F(1,22) = 98.57, p < .00001$ (see Table 2), which is not surprising given that the groups were selected on the basis of their Recall Performance. There was a marginally significant effect of expertise on Proportion Correct in the Recognition phase, $F(1,22) = 4.00, p < .06$ with high expertise subjects scoring higher than low expertise subjects (see Table 2). This result is consistent with the trend in this direction found in the first analysis. Experts passed on significantly fewer questions in the Recall phase than did non-experts, $F(1,22) = 13.62, p < .001$. As in the first analysis, there was no significant effect of expertise on the mean Absolute FOK ratings, $F < 1$.

Discussion

For all expertise levels, FOK was significantly correlated with recognition perfor-

[4] Table 2 also shows correlations between absolute and relative FOK, and between confidence and recall for each expertise group. As in the first analysis, all were significantly greater than zero, all t's > 17.16.

mance. Therefore, consistent with the results of other studies (Blake, 1973; Nelson, et al., 1984), FOK strength was a reliable, although imperfect, index of actual knowledge. High expertise subjects performed marginally better on the recognition test (for non-recalled items) than low expertise subjects, suggesting that recall failures were more likely to result from retrieval failure than absence of knowledge for high expertise compared with low expertise subjects. However, the high expertise subjects did not show greater FOK strength than did the low expertise subjects. The present results therefore offer no support for the inference-from-expertise mechanism (Nelson et al., 1984), in which FOK is assessed according to one's expertise in the general domain of the questions.

Increasing expertise in the domain of general knowledge was not associated with higher FOK accuracy. Experts were no better able to assess what they did and didn't know than were non-experts. Therefore the insight hypothesis was not supported. Nor was there evidence for the overconfidence hypothesis, according to which experts should have low FOK accuracy, due to strong FOK but no more accurate recognition performance than less expert subjects. FOK accuracy was apparently unrelated to expertise, either self-rated or objectively assessed.

The possibility remains open that expertise might be related to the strength and/or accuracy of FOK judgements if more extreme groups were used. However, the groups in the present experiment did show a marked difference in expertise as assessed by performance on the recall test, with the mean score for the high expertise group more than twice that for the low expertise group. The use of more extreme groups would be desirable in a replication of the present study, although it would be difficult to develop test materials suitable for groups with extremely different knowledge bases.

To conclude, the present results offer no support for the idea that a person's level of expertise affects FOK accuracy. Experts, despite their larger knowledge base, appear to have no priviledged access to what they do and don't know. Nor was there any support for the hypothesised inference-from-expertise mechanism for assessing FOK strength.

References

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.

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

Brown, R., & McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning & Verbal Behavior, 5,* 325–327.

Carey, S. (1985). *Conceptual Change During Childhood.* Cambridge, MA.: MIT Press.

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem-solving. In R. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence: Vol 1.* Hillsdale, NJ.: Erlbaum.

Dreyfus, H.L. & Dreyfus, S.E. (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer.* New York: MacMillan.

Flavell, J. (1985). *Cognitive Development* (2nd ed.). Englewood Cliffs, NJ.: Prentice-Hall.

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

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

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

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. (1980). A new technique for investigating the feeling of knowing. *Acta Psychologia, 46,* 69–80.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

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

Wilkinson, T. S., & Nelson, T.O. (1984). FACTRETRIEVAL2: A PASCAL program for assessing someone's recall of general information facts, confidence about recall correctness, feeling-of-knowing judgments for nonrecalled facts, and recognition of nonrecalled facts. *Behavior Research Methods, Instruments, & Computers, 16,* 486–488.