Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AD-A211 629

REPORT DOCUMENTATION PAGE

DTIC ELECTE AUG 1989 D B

DTIC FILE COPY

1b. RESTRICTIVE MARKINGS	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 89-1113
6a. NAME OF PERFORMING ORGANIZATION Department of Psychology 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research/NL
6c. ADDRESS (City, State, and ZIP Code) University of Washington Seattle, WA 98195	7b. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, D.C. 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-88-0226
8c. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, D.C. 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F PROJECT NO. 2313 TASK NO. A4 WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Metacognition and retrieval from long-term memory at Mount Everest

12. PERSONAL AUTHOR(S)
Thomas O. Nelson

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Technical Report	FROM 5/13/88 TO 6/23/89	6/27/89	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
05	10		Cognition, memory, metacognition, judgment, hypoxia, fear, danger, mountaineering, retrieval, confidence, altitude.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Behavioral data were collected from climbers at various altitudes on Mount Everest. In contrast to earlier findings that altitude impairs the acquisition of information into memory, we found no changes in the accuracy or latency of retrieving information from memory, even at extreme altitudes above 21,000' (6,400 m). This lack of effect on retrieval occurred for both the recall and recognition of answers to general-information questions (e.g., "What is the capital of Finland?"). Self-confidence about the accuracy of recent retrieval was also not affected by altitude. However, the feeling of knowing (i.e., self-confidence about upcoming retrieval) declined at extreme altitudes and remained lower even after return to Kathmandu. This pattern of results is close to opposite of the pattern obtained when the independent variable is alcohol intoxication and the same test battery is employed. These and related results are described in an attempt to give a relatively comprehensive picture of the climbers' performance, and suggestions are offered for future research.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Alfred R. Fregly	22b. TELEPHONE (Include Area Code) (202) 767-5021 22c. OFFICE SYMBOL NL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

89 8 1 032

Rationale

In his classic textbook, McFarland (1953) reviewed data on World War II fliers who had been diagnosed as having an impairment of consciousness. The most frequent cause was "anxiety state" (as assessed for 117 of the 483 RAF pilots and air-crew men showing impaired consciousness). He stated that "the most common cause of impairment in consciousness, confusion, and amnesia in flight was a neurosis, usually an anxiety state, *basically caused by fear*" (p. 253). The way in which this impairment--from fear and hypoxia during high-altitude missions--is manifest in cognitive behavior was that "the higher mental functions, such as memory, judgment, insight, and reasoning, are the first to deteriorate....The deterioration is usually insidious and unnoticed by the person experiencing it" (p. 270). Thus, effects may occur both (1) in cognition and (2) in the monitoring of one's own cognition (called *metacognition*).

A 1981 expedition to Mount Everest, known as the American Medical Research Expedition to Everest (AMREE), was sponsored by the Army. (The Army Research Institute of Environmental Medicine also sponsored two altitude-chamber research projects: Operation Everest, and Operation Everest II, neither of which involved the same kind of dangers that occur in-vivo on an actual Everest expedition, where the average fatality rate is two fatalities per expedition, as tallied by West, 1986.) The Soviets also have conducted in-vivo research at Mount Everest (data published in ***Everest 82. Report of the Soviet Expedition to Everest.*** Moscow: 1983).

Theoretical Background

One distinction for memory theory that essentially all cognitive psychologists agree on is that memory should be divided into at least two subcategories: (1) *acquisition* of information into memory (also called storage or learning or encoding) and (2) *retrieval* of previously acquired information from memory. In any given situation, a fundamental two-part question is how do the variables of interest affect acquisition and how do they affect retrieval.

For the variable of extreme altitude, research on Mount Everest during AMREE established that the acquisition of new information is impaired as altitude increases (Townes et al., 1984). However, those researchers pointed out that they had investigated only half of the memory situation and that future research should be done by "separating learning from the retrieval of information" (p. 35). Moreover, for most applied situations, the retrieval phase of memory may be even more important than the acquisition phase, because the typical applied situation is not that people have to learn information at high altitude, but rather that people learn information near sea level and subsequently have to retrieve the information at another altitude (e.g., retrieval of previously learned first-aid information). One goal of the present research was to explore the effects of extreme altitude on the retrieval of information from memory.

As indicated in the opening paragraph of this report, a question related to the retrieval of information is, to what degree do people accurately monitor their degree of success at retrieving information from memory? These self-awareness judgments are part of the new area called *metacognition*, which refers to the monitoring and control of one's own cognitive processes (for a brief review, see Nelson et al. 1986). In contrast to traditional memory *processes such as the acquisition of information*, which appears to be related to physiological activity in the hippocampus (Townes et al., 1984, p. 35), metacognitive processes appear to be related to physiological activity in the frontal lobe. Therefore, an independent variable may affect retrieval without affecting the person's self-monitoring of that retrieval, or vice versa.

One independent variable that has qualitatively different effects on cognition versus metacognition is acute alcohol intoxication. Nelson et al. (1986) found that acute alcohol intoxication disrupts retrieval without affecting metacognitive judgments about that retrieval. Although mountain climbers report anecdotal evidence that their judgment and self-monitoring are disrupted at altitude, no systematic research had previously been conducted on that topic. The present research explored the effects of altitude on metacognition by using the same methodology as in Nelson et al. (1986). The present use of that same methodology is of more than passing interest, given the speculation by one expert that "hypoxia resembles overindulgence in alcohol" (Houston, 1987, p. 176).

Finally, in addition to exploring the effects of *being at* high altitude, the present research explored the *aftereffects of having been* at high altitude. That is, the participants were examined not only as they went to increasingly higher altitudes, but also when they returned to the lower altitudes from which they began. The previous research by Townes et al. (1984) had shown that the disruptive effects of altitude on acquisition persisted even upon return to Kathmandu (i.e., the climbers' *ability* to acquire information had been impaired, and this impairment continued even after descending to low altitude).

Experiment 1: Retrieval and metacognition in climbers

METHOD

Subjects and experimenters

The subjects were the twelve climbers from the 1988 Northwest American Everest Expedition, all of whom lived in the United States and all of whom expected to climb above 6000 m on Mount Everest. Although they were not professional climbers (i.e., did not make their living via mountaineering), they were highly experienced climbers (e.g., median number of years of climbing experience was 16 years). They were nine males and three females, of which during this expedition one female became the first American woman to reach the summit of Mount Everest. They were above-average in formal education (all of them had at least two years of college education, with five having attained the M.D. degree, one having a J.D. degree, two having M.A. degrees, and three having B.A. degrees).

At each testing location, every participant served as both a subject (being tested by another climber) and as an experimenter (testing another climber). The plan was for each climber to be tested once at each of six locations (see next paragraph). The testing of climbers was counterbalanced according to a prearranged schedule: (1) each experimenter tested someone else on only those items that the experimenter himself/herself had already been tested on; (2) an attempt was made to test each climber by a different person at each location and to have all climbers serve approximately equally often as experimenters (namely, six times--once at each of the six locations). To accomplish this, each climber received an individualized Mastercard to keep in their pack throughout the expedition; the Mastercard listed which set of items would be tested at each location, who should be sought for the testing at each location, and which other people could do the testing if the soughtafter testor was not available because of being at a different place on the mountain at that time. As an emergency backup, (1) the basecamp manager had a duplicate of all Mastercards and could be contacted from any high camps by radioing to basecamp, and (2) one of the members of the research team could do the testing (e.g., the PI at Camp 2, another researcher at basecamp), although the goal was to have climbers testing climbers whenever possible (note: this goal was 100% successful). The research team consisted of the following people. During data collection in Nepal: T. Nelson (PI), D. White, B. Townes, and V. Dewey; during data analysis in the USA: T. Nelson, J. Dunlosky, D. White, J. Steinberg, and D. Anderson.

Locations for testing

The locations originally targetted for testing were: (1) 48 hrs after arrival in Kathmandu (el. 1200 m) but before departure on the two-week trek to basecamp, (2) 48 hrs after arrival at basecamp (5400 m), (3) 48 hrs after arrival at Camp 2 (6400 m), (4) a second time at 6400 m or higher (either at Camp 2, 6400 m, or at Camp 3, 7000 mm) near the time of summit attempts (approximately two weeks after Camp 2 had been established), (5) at basecamp after having been at the highest point that a given climber attained during this expedition, (6) at Kathmandu at the end of the expedition (which for everyone was at least one week after returning from the high camps). Three unplanned--but not unexpected for a Himalayan expedition--changes occurred that affected the data matrix: First, one climber decided to remain at basecamp (therefore no high-altitude data were obtained from that climber, and accordingly he is not included in the final data matrix). Second, because three of the remaining eleven climbers were delayed at the high camps in an attempt to save the life of a climber from another expedition, they did not return to basecamp until the evening before the expedition departed basecamp (therefore they did not have time to administer/receive the second test at basecamp, and accordingly the second test at basecamp is not included in the final data matrix). Third, one of the eleven climbers who contributed high-altitude data had a pulmonary embolism at 7000 m and therefore returned immediately to basecamp before going through the second high-altitude test. Thus, the final data matrix consisted of data from 54 testing sessions, i.e., five sessions from each of ten climbers and four sessions from the eleventh climber.

Stimulus materials, training of climbers, and apparatus

The items were 238 of the 240 general-information questions contained in FACTRETRIEVAL2 (Wilkinson & Nelson, 1984). For instance, one question was "What is the capital of Finland?" (answer = Helsinki). The normative difficulty of these questions had been determined previously (Nelson & Narens, 1980), and they were arranged into seven subsets (34 items per subset), with each subset having approximately the same normative difficulty. Every stimulus card showed one general-information question and 8 randomly-ordered recognition alternatives; these were printed in 24-point Geneva type, so as to minimize any difficulty of reading, and the card was laminated in plastic so as to be sturdy and waterproof. The order of the seven subsets varied across climbers, such that a given climber was tested on a different subset of items during every testing session and across climbers a given subset was used approximately equally often at each testing location. One subset was used for training the climbers in the procedure.

During training, which occurred in Seattle shortly before the expedition left for Nepal, the PI trained each climber individually by first running the climber through that subset in the role of subject (hereafter, abbreviated as "S") and then by having the climber serving in the role of experimenter (hereafter, "E") and running the PI through the same subset.

The major apparatus, which necessarily had to be minimal because of the Everest environment, consisted of two Nagra-brand tape recorders. This brand of tape recorder had previously been used successfully at 8000 m on Mount Everest and on K2. It was known to be reliable at extreme temperatures and provided a tape-recording of each session, both as a backup to the climber's recording of the subject's responses (for purposes of assessing any possible recording errors by E) and as a source for later scoring to determine the subject's response latencies (this potential use of the tapes was not mentioned prior to completion of the research).

The only other apparatus, aside from pencils and the testing materials, was a blindfold. The blindfold was employed during testing so that S could not receive any feedback concerning his answer to a given question (i.e., seeing the E's facial reaction after guessing an answer might affect the S's confidence judgment about the accuracy of that guess).

Procedure

In advance of testing. The E put a DO NOT DISTURB sign outside the door of the hotel room (Kathmandu) or tent (all other testing locations). Then E removed the testing materials for that location (these were stored in waterproof plastic envelopes, all of which were kept in a waterproof canvas bag when not in use) and set up the tape recorder.

During testing. First, E gave S an "Instructions to Subject" card to review. Second, E read aloud from the top of the blank data sheet the following: (1) S's name, (2) E's name, (3) the date, (4) the testing location, and (5) which set of items was being tested. This information was needed for later identification of the tapes.

Third, S read aloud in his/her normal speaking voice a one-page speech sample, whose primary purpose for the present research was to get the S settled in the testing environment.

Fourth, the memory experiment began. S put on the blindfold, which remained on during the recall phase and the feeling-of-knowing phase to eliminate all visual feedback from E that otherwise might have affected S's judgments of confidence or feeling-of-knowing judgments. Then E shuffled the deck of 34 memory cards. Next E slowly read aloud the question on the first card. The S took as long as he/she wanted to try to recall the answer to the question. S was encouraged to guess whenever possible and to take as much time as necessary to try to think of an answer. When S made a guess, E wrote it on the data sheet (or wrote "RIGHT" or wrote "NO GUESS" if S could not think of any guess at all). Immediately after making a guess--and without any feedback from E--the S said a confidence judgement of either 0%, 20%, 40%, 60%, 80%, or 100% to indicate his/her confidence that the guess was the correct answer. Then E recorded S's confidence rating. This procedure was repeated for each of the 34 questions.

Following the recall phase, E assembled the deck of stimulus cards for the feeling-of-knowing phase, which contained only those questions that S had not answered correctly. Then the deck was shuffled briefly.

For the feeling-of-knowing phase, E read aloud a question, and S said 1, 2, 3, 4, 5, or 6, depending upon S's subjective likelihood of recognizing the correct answer to the question if shown a pool of eight plausible answers. S said 1 if such recognition would be completely by chance, 6 if such recognition were certain to be correct, or 2, 3, 4, or 5 to reflect in-between states of confidence about the likelihood of recognition. After S had made feeling-of-knowing judgments on all of the previously nonrecalled items, E shuffled that deck of cards again, and S made another feeling-of-knowing judgment on each of those items again. For each item, S was asked to concentrate on only the current feeling-of-knowing judgment. Following the feeling-of-knowing phase, S was told to remove the blindfold.

Finally, S had an 8-alternative multiple-choice recognition test on each previously nonrecalled item. E shuffled the deck of nonrecalled items and passed them one-at-a-time to S. S read aloud the item number (so that E could record S's response at the correct location on the data sheet) and then read aloud the answer that he/she believed was most likely to be correct for that question. E recorded S's answer. S was told in advance that the correct answer would never be the one that he/she gave during recall, because the only items being tested for recognition were those that S had not correctly recalled the answers to. Upon completion of the recognition test, the memory experiment ended.

RESULTS AND DISCUSSION

Unless otherwise mentioned, all differences that are significant have $p < .05$, and all differences that are nonsignificant have $p > .10$. In an attempt to give a somewhat comprehensive picture of the retrieval/metacognitive performance of the climbers, we have organized the results into four broad

Latency of recall

During recall, three kinds of responses were possible on each item: (1) correct recall, (2) commission error (i.e., incorrect guess), and (3) omission error (i.e., no guess). The median latency for each kind of response was determined for each S at each location, and the means (across Ss) of those medians are reported in Table 1.

See Table 1 on next page

Before considering each category of latency separately, the three categories were compared. A 3x5 ANOVA showed that the three categories of latency were reliably different, $F(2,12) = 17.54$. Two follow-up orthogonal comparisons showed that (1) the latencies were significantly longer for commissions and omissions than for correct responses, $F(1, 12) = 33.52$, but (2) there were no significant differences between commissions and omissions. Neither the main effect of altitude nor the interaction was significant, both $Fs < 1$. Next, each category of latency is considered separately for a brief interpretation.

Latency of correct recall. The first row of Table 1 shows the latency of correct recall, which sometimes is regarded as a more sensitive measure of retrieval than the aforementioned accuracy measures (for discussion, see MacLeod & Nelson, 1984). This measure also did not differ significantly across locations when assessed via a nonparametric Friedman test (Friedman chi-square = 2.6, $p = .63$).

Latency of commission error. The second row of Table 1 shows the latency of commission errors, which also did not differ significantly across locations when assessed via a nonparametric Friedman test (Friedman chi-square = 1.23, $p = .87$).

Latency of omission error. The third row of Table 1 shows the latency of omission errors. This latency is interpreted as indicating how long the person is willing to search on a particular item before giving up (MacLeod & Nelson, 1984; Nelson, Gerler, & Narens, 1984). This measure also did not differ significantly across locations when assessed via a nonparametric Friedman test (Friedman chi-square = 6.03, $p = .63$).

Metacognitive monitoring judgments

Three kinds of *metacognitive judgments* that differ in terms of temporal considerations were obtained: (1) The postdicted likelihood of correct recall is *retrospective* (i.e., a confidence judgment about *past* performance); (2) the proportion of incorrect recall responses that were commission errors vs omission errors may be regarded as an *on-going* metamemory judgment (discussed below); and (3) the feeling of knowing is *prospective* and is a confidence judgment about *future* performance on currently nonrecalled items. Each metacognitive judgment is discussed in turn.

Postdicted percent correct recall. Following each recall response, the S

Table 1

Response Latency (sec.) during Recall at Each Test Location

Recall response	Location					
	1st Kathmandu	Basecamp	1st 6400m	2nd 6400m	2nd Kathmandu	$\overline{X}$
Correct	2.6	2.8	2.8	3.4	2.8	2.9
Commission error	8.3	10.8	11.7	11.0	11.8	10.7
Omission error	12.0	10.8	10.2	6.9	9.4	9.9
$\overline{X}$	7.6	8.1	8.2	7.1	8.0	

Note. Entry is the mean of the individual subjects' median latencies in seconds.

categories: (1) results about retrieval, (2) results about metacognitive monitoring judgments (i.e., characteristics of those judgments, regardless of their accuracy), (3) results about the accuracy of metacognitive monitoring judgments, and (4) results about metacognitive control.

Retrieval

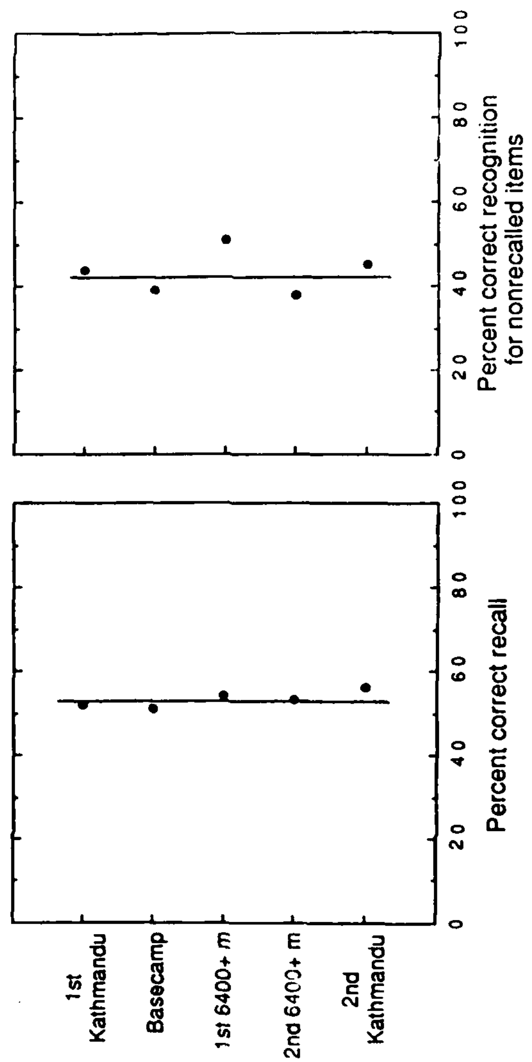
Three measures of retrieval--two focussing on accuracy and one focussing on latency--were obtained: percent correct recall of answers to general-information questions, percent correct recognition of nonrecalled answers, and latency of correct recall. *Each is discussed in turn, and the data* are summarized in dot charts (Cleveland & McGill, 1985), with a zero-slope line of best fit shown wherever the null hypothesis cannot be rejected.

Percent correct recall. As indicated in the left panel of Figure 1, there were no significant changes in the climbers' mean percent correct recall across the five testing locations, $F(4, 36) = .36$. This finding is in marked contrast to the finding of an effect of altitude on acquisition that was found by Townes et al. (1984) at these same altitudes. Moreover, to the extent that any state-dependency effects are operating, the robustness of this retrieval from long-term memory is all the more impressive because the general-information facts that our climbers retrieved at 6500 m had not been originally learned at that altitude. Also, this finding is not due to an insensitivity of the testing procedure, because this same procedure yielded significant differences in recall of these same items when the independent variable was alcohol intoxication (Nelson et al., 1986). Contrary to Houston's speculation mentioned in the Introduction, the hypoxia at the present altitudes did not resemble overindulgence in alcohol, in terms of effects on recall.

See Figure 1 on next page

Percent correct recognition on nonrecalled items. Another measure of retrieval is shown in the right panel of Figure 1. There were no significant changes in the climbers' mean percent correct recognition on nonrecalled items across the five testing locations, $F(4, 36) = 1.24$. This measure is somewhat noisier than the recall measure in the left panel of Figure 1, probably due to (1) recognition performance being based only on approximately half as many items (namely, the incorrectly recalled items) and (2) chance performance due to correct guesses on the 8-AFC recognition test.

Another aspect of recognition performance that should be mentioned is the occurrence of perseveration errors during recognition performance. That is, sometimes people choose the same recognition alternative that they gave as an incorrect recall response (Krinsky & Nelson, 1985). However, in the present situation, there were--across all Ss, tests, and items--only 3 perseverations out of more than 450 recognition responses on items for which commission errors had occurred during recall. Apparently the climbers were quite good at remembering the incorrect response they had made during recall and avoided making it again during recognition.



made a postdiction about the likelihood of that response being correct. All items given the same likelihood were aggregated and the actual percent correct was computed (ala Lichtenstein & Fischhoff, 1977; Nelson et al., 1986). The mean confidence (across Ss) at each location is shown in the left panel of Figure 2.

See Figure 2 on next page

There was no significant effect of location, $F(4, 36) = .29$. Thus the climbers overall confidence about their recent past performance was unaffected by altitude.

To examine this more analytically, we segregated the two categories of recall responses to which the aforementioned confidence judgments pertained, namely, correct responses versus commission-error responses. The results are shown separately for those two categories in the right panel of Figure 2 (where solid data points are from commission-error items, and open data points are from correctly recalled items). A 2-way ANOVA showed that confidence was significantly greater for correctly recalled items than for commission-error items, $F(1, 9) = 155.62$. However, as in the overall analysis reported above, the effect of location was not significant, $F(4, 36) = 1.40$. The interaction of location and correct versus commission was also nonsignificant, $F(4, 36) = .42$.

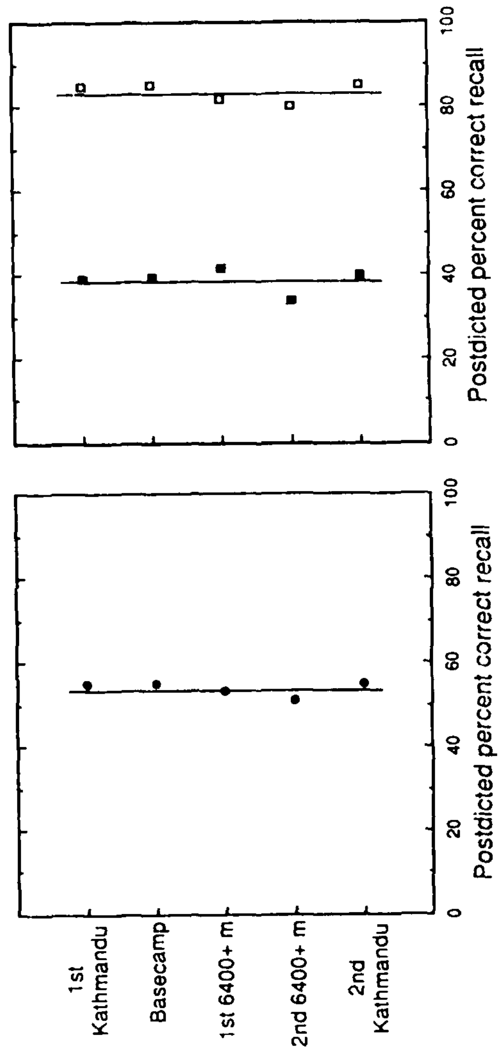
What underlies retrospective confidence? To help answer this question, we computed several correlations between the aforementioned retrospective confidence judgments and various potential factors that might underlie retrospective confidence. All correlations are Goodman-Kruskal gamma correlations, which are the best correlations available for these kinds of data (for rationale, see Nelson, 1984), and were computed individually for each S; what is reported next is the mean across Ss.

Table 2 shows the mean correlation between retrospective confidence and the latency of recall. A 2-way ANOVA showed that the absolute magnitude of this correlation is marginally greater for correctly recalled items (first row of Table 2) than for commission-error items (second row of Table 2), $F(1, 10) = 3.40$, $p = .095$. However, neither the effect of location nor the interaction effect was significant, both Fs < 1.

See Table 2 on next page

The correlations in Table 2 suggest two major conclusions: First, the *absolute magnitude* of the correlations is substantial, and *their direction* is in accord with the hypothesis that quickly recalled responses are associated with greater confidence. Second, the magnitude of these correlations--i.e., the degree to which the underlying aspects of latency might influence subsequent confidence--is not affected by altitude.

P(commission | incorrect recall). One potential index of on-going confidence during search is the likelihood of a commission error, given incorrect recall (i.e., the likelihood that the person will output a guess when the



correct answer is not available). The mean across Ss is shown in Figure 3 for each location.

See Figure 3 on next page

As indicated by the lines of best fit in Figure 3, three orthogonal comparisons yielded the following conclusions about the effects of altitude on this dependent variable: (1) There was *no* significant difference between the first Kathmandu test and the basecamp test, $F < 1$; (2) the likelihood of a commission error declined significantly from the aforementioned two tests to the last three tests (i.e., was lower during the high-altitude tests and the subsequent Kathmandu test), $F(1, 36) = 7.82$; and (3) there was no significant difference among the last three tests, $F(2, 36) < 1$.

The second outcome in the above pattern can be interpreted to mean either (or both) of the following: (1) On-going confidence becomes lower at high altitude (6400 m), such that the Ss are more reluctant to output their guesses, or (2) the on-going confidence remains the same but there is a decrease in the likelihood of retrieving a candidate to output as a guess (although if this were the case, the aforementioned lack of effect of altitude on the probability of correct recall is surprising). In any case, it is noteworthy that although the P(commission | incorrect recall) is known to be affected by drugs such as marijuana and lithium (Weingartner, Rudorfer, & Linnoila, 1985)--and now, by high altitude--it is not affected by acute alcohol intoxication (Nelson et al., 1986).

The third outcome in the above pattern suggests that whatever the interpretation of P(commission | incorrect recall), there is an *aftereffect* of going to high altitude that is still present a week later upon return to Kathmandu.

Feeling of knowing (FOK). First, we consider some characteristics of the structure underlying the climbers' feeling of knowing (FOK), and then the effects of altitude on FOK are examined.

The structure underlying the FOK appears to be quite stable, as indicated by the remarkably high retest reliability. The first row in Table 3 shows that the mean correlation between the first set of FOK judgments and the second set of FOK judgments on those same items was +.95 (i.e., nearly perfect reliability). Hereafter, therefore, only the first set of FOK judgments will be used when referring to "FOK."

See Table 3 on next page

The second row of Table 3 shows that the correlation between FOK and the postdicted likelihood of recall was substantially lower (mean = +.41) than the aforementioned retest reliability correlation. A 2-way ANOVA showed that these two correlations were significantly different, $F(1, 9) = 90.35$, in accord with the idea that the underlying structure for each of these correlations is somewhat different (discussed below). Moreover, the magnitude of these correlations does not change with altitude, nor was the interaction significant,

Table 2

Relationship between Postdicted Likelihood of Being Correct and Two Kinds of Response Latency during Recall at Each Test Location

Recall response	Location: 1st Kathmandu	Basecamp	1st 6400m	2nd 6400m	2nd Kathmandu	$\bar{X}$
Correct	-.57	-.51	-.59	-.51	-.56	-.55
Commission error	-.43	-.43	-.31	-.47	-.36	-.40
$\bar{X}$	-.50	-.47	-.45	-.49	-.46	

Note. Entry is the mean of the individual subjects' gamma correlation between response latency and postdicted likelihood of being correct.

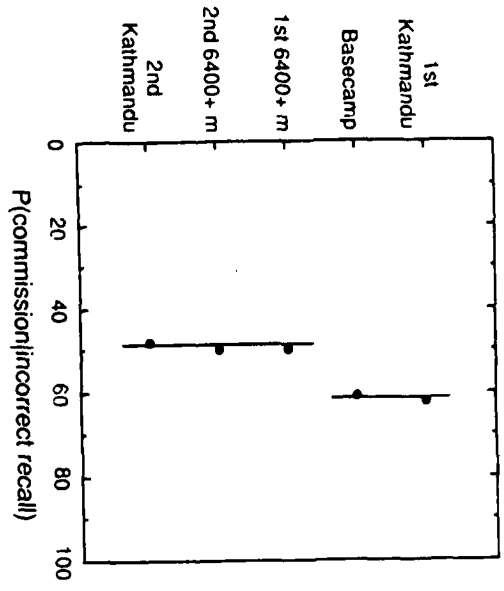


Fig. 3

Table 3

Relationship between Two Kinds of Metamemory Judgments: Feeling of Knowing and Postdicted Likelihood of Correct Recall at Each Test Location

Two variables being correlated	Location: 1st Kathmandu	Basecamp	1st 6400m	2nd 6400m	2nd Kathmandu	$\overline{X}$
1st FOK and 2nd FOK	+.96	+.95	+.98	+.98	+.89	+.95
1st FOK and postdicted likelihood of recall	+.36	+.50	+.31	+.46	+.40	+.41
$\overline{X}$	+.66	+.73	+.65	+.72	+.65	

Note. FOK = feeling-of-knowing rating. Entry is the mean of the individual subjects' gamma correlation between the two variables being correlated.

both $Fs < 1$. This disconfirms the a priori possibility that climbers would be less reliable in their judgments when *at high altitude than at lower altitudes* (*e.g.*, because of shifting dimensions and/or shifting weights on those dimensions; Butterfield, Nelson, & Peck, 1988)--if anything, the trend is for retest reliability to increase at altitude!

To determine the effect of altitude on FOK per se, we computed the median FOK for each S at each location. The left panel of Figure 4 shows the mean (across Ss) of the individual-S medians at each location.

See Figure 4 on next page

As indicated by the lines of best fit in Figure 4, three orthogonal comparisons *yielded the following conclusions* about the effects of altitude on FOK: (1) There was no significant difference between the first Kathmandu test and the basecamp test, $F < 1$; (2) the FOK declined significantly from the aforementioned two tests to the last three tests (i.e., was lower during the high-altitude tests and the subsequent Kathmandu test), $F(1, 36) = 16.84$; and (3) there was no significant difference among the last three tests, $F(2, 36) < 1$. (Note: The effect of altitude on FOK was also confirmed by a nonparametric Friedman test, Friedman chi-square = 14.26, $p < .01$.)

Because of the importance of this finding, we examined it further in two ways: In one followup analysis, we examined the individual S's median FOK to see the data upon which the mean across Ss is based (this seemed especially *appropriate because of* the possibility of less-than-interval-scale aspects of FOK). Figure 5 is a scatterplot matrix (Cleveland, 1985) showing bivariate data in which each data point is from one S and indicates that S's median FOK for the row location plotted against that S's median FOK for the column location (note: the number of petals on each sunflower indicates the number of Ss when more than one S had that bivariate entry; e.g., the three-petal sunflower in the subplot for column = "Basecamp" and row = "2nd Kathmandu" indicates that three Ss had a median FOK of 4 at basecamp and a median FOK of 3 on the 2nd Kathmandu test).

See Figure 5 on next page

The results shown in Figure 5, which may require extra effort to see, are illuminating: For all six of the subplots comparing any of the first two tests with any of the last three tests--these subplots are enclosed inside the dashed box--the preponderance of Ss who had a lower median FOK after going to high altitude (versus before going to high altitude) is striking. For instance, *every* climber had a lower median FOK at the 2nd high-altitude test than at basecamp. Also, there were only two inversions in those six subplots (namely, one in each of the upper two of those six subplots). Thus the group effect that was shown in Figure 4 is also obvious at the level of individual Ss in Figure 5.

In a second followup analysis, we did separate analyses for omissions and commissions because of possible mediation from P(commission | nonrecall); i.e., Krinsky and Nelson (1985) found greater FOK for commissions than

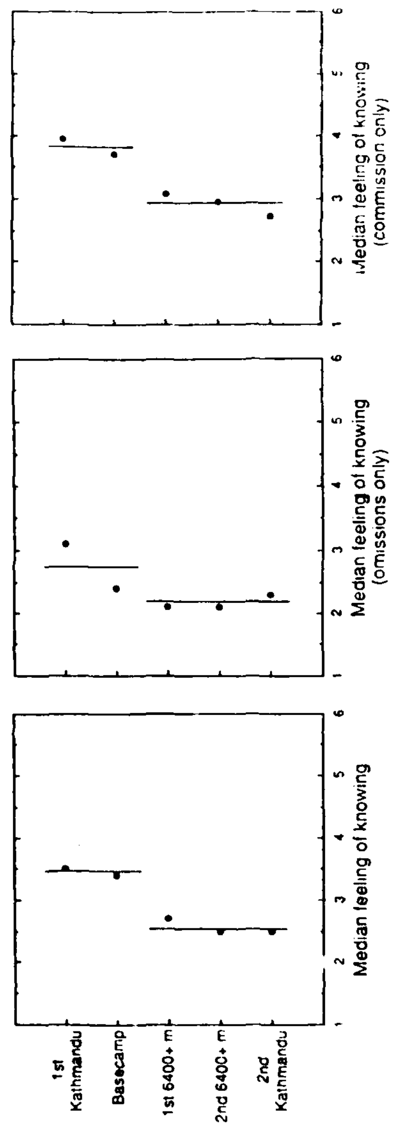
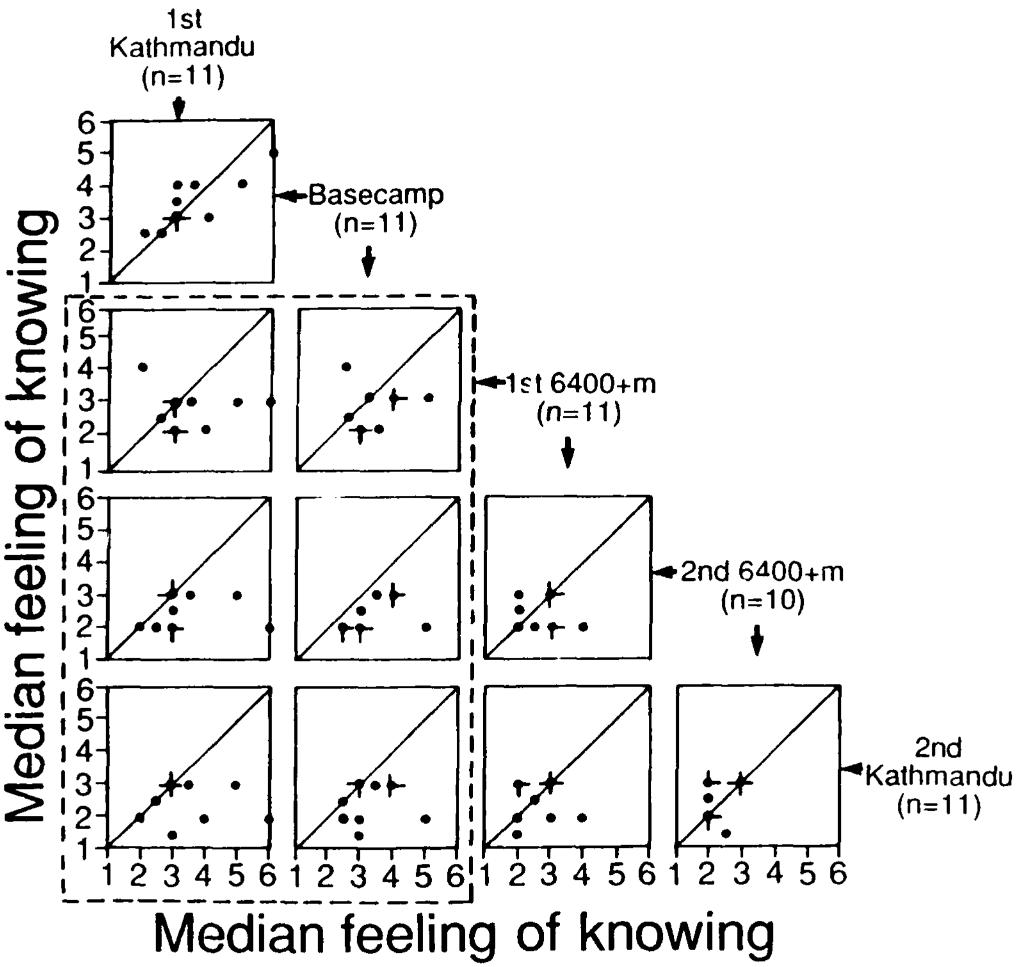


Fig 4



omissions, so the reduced P(commission | nonrecall) in Figure 3 might be the sole source of the reduced FOK in the left panel of Figure 4. This mediation hypothesis would predict no effect of location when the items are segregated into commissions and omissions because such segregation would neutralize any mediation from P(commission | nonrecall). The results, shown in the middle and rightmost panels of Figure 4, are that the overall pattern is less in accord with the aforementioned mediation hypothesis than with the direct-effect hypothesis indicated by the two lines of best fit in each panel. Treating those two panels as independent tests of the same hypothesis (W ... r, 1962)--in particular, the null hypothesis versus the pattern of lines of best fit shown in the leftmost panel of Figure 4--we found that the segregated data at yielded a significant decline in FOK at high altitude: FOK is lower at the last three locations than at the first two, chi-square = 26.28, $p < .005$.

<u>Accuracy of metacognitive monitoring judgments</u>

The accuracy of metacognitive monitoring is assessed via the relationship between metacognitive monitoring judgments and the retrieval that supposedly is being monitored, with that relationship being assessed either in relative or cardinal (absolute) form. For instance, to have *relative* accuracy, the actual percentage of correct recall should be *greater* for items judged to have an 80% chance of being recalled than for items judged to have a 60% chance; to have *cardinal* accuracy, the actual percentage of correct recall for items judged to have an 80% likelihood of being recalled should be 80%, *for items judged to have* a 60% chance should be 60%, etc. In the terminology of Lichtenstein & Fischhoff, 1977, the latter is called "calibration" whereas the former is called "resolution" (and perfect calibration requires perfect resolution, but not vice versa). Several kinds of metacognitive accuracy were assessed, both for retrospective confidence judgments (concerning recall in the recent past) and for prospective FOK judgments (concerning upcoming recognition).

<u>Relative accuracy of retrospective confidence judgments about recent recall</u>. We measured the degree of relative accuracy by the Goodman-Kruskal gamma correlation, G, which is the best available measure of the relative accuracy of metacognitive judgments (Nelson, 1984). For any pair of nontied items, G gives (via Eq. 7 in Nelson, 1984) the likelihood that the item receiving the greater confidence judgment will also be the item that is retrieved better (or more quickly), with $G = +1.0$ representing perfect metacognitive accuracy.

The first row of Table 4 shows the mean of the individual-S Gs at each location for the relative accuracy of confidence judgments about recent recall. Although this metacognitive accuracy is near the ceiling (mean $G = +.93$), the *SEMs* were all .02 or .03 and only 10 of the 54 Gs were +1.0. Nevertheless, there is no apparent effect of altitude, $F(4, 36) < 1$.

See Table 4 on next page

<u>Calibration curves during recall for absolute (cardinal) accuracy</u>. Figure 6 shows calibration curves, separately for each location, for the climbers'

cardinal accuracy at postdicting the likelihood of correct recall, with each curve being the mean of the climbers' performance at a given location. The main diagonal (heavy line) represents perfect accuracy. Three things are evident in these data: (1) The data are somewhat noisy, but acceptable. (2) Consistent with the dot charts for degree of confidence in Figures 1 and 2, there is no systematic effect of *altitude to make people overconfident or* underconfident (note: the former is indicated when the data points are below the main diagonal, and the latter is indicated when the data points are above the main diagonal. (3) The general tendency is for extremely good calibration, as indicated at the top of Figure 6 by the parameter values for the lines of best fit; perfect calibration would yield $Y = X$ (i.e., $Y = 1X + 0$), and all five curves have slopes and intercepts remarkably close to 1 and 0, respectively.

See Figure 6 on next page

In contrast to the aforementioned analysis of the climbers as a group, another way of examining the climbers' retrospective confidence judgments is to look at the postdictions and recall from individual climbers (i.e., individual differences). Figure 7 shows a separate pair of panels for each location. In each pair, the upper panel shows each individual climber's mean postdictions and recall, and the lower panel (cf. Tukey sum-difference graphs, as described in Cleveland, 1985) shows each climber's deviation from perfect overall calibration (note: the dotted line represents zero deviation, and the solid line summarizes the mean deviation for all climbers at that location). In every pair of panels, regardless of altitude, two findings emerge: (1) The individual climbers' mean judgments are well-calibrated regardless of whether they recalled few or many items (upper panels); and (2) there is no systematic tendency for the climbers who had worse recall to be overconfident or underconfident about their recall performance (lower panels). In particular, in the lower panels where each person's actual percent recall is subtracted from his or her postdicted percent recall, the means at the five locations are +2.9%, +3.5%, -0.6%, -2.2%, and -0.8% (i.e., all are ±3.5%, which indicates almost nil overall bias for the climbers at any altitude).

See Figure 7 on next page

Relative accuracy of FOK judgments about upcoming recognition. The mean (across Ss) of the individual-S gammas for relative accuracy of the item-by-item FOK judgments is shown for each location in the second row of Table 4. The overall mean across locations was +.45, indicating substantial FOK accuracy and similar to the mean of +.41 in Nelson et al. (1986). If anything, the trend across locations was for *higher* FOK accuracy after going to extreme altitude. However, there was no *significant* effect of altitude, $F(4, 36) < 1$, but this measure appeared to be noisier than most of the other measures, probably because (1) it is affected by chance recognition (e.g., two Ss with identical FOK ability and identical recognition ability might receive very different FOK-accuracy gammas if one of those Ss recognized by chance

Table 4

Relative Accuracy of Three Kinds of Metamemory Judgments at Each Test Location

Predictor variable and criterion variable	Location					
	1st Kathmandu	Basecamp	1st 6400m	2nd 6400m	2nd Kathmandu	$\bar{X}$
Postdicted recall and actual recall	+.90	+.93	+.93	+.95	+.94	+.93
Feeling of knowing and recognition	+.36	+.32	+.64	+.45	+.47	+.45
Postdicted recall and recognition	-.08	+.12	+.08	+.19	+.37	+.14

Note. Entry is the mean of the individual subjects' gamma correlation between predictor variable and criterion variable.

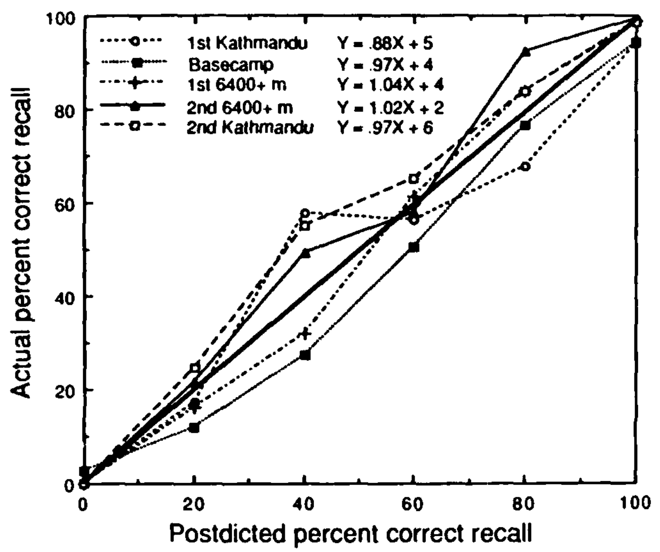
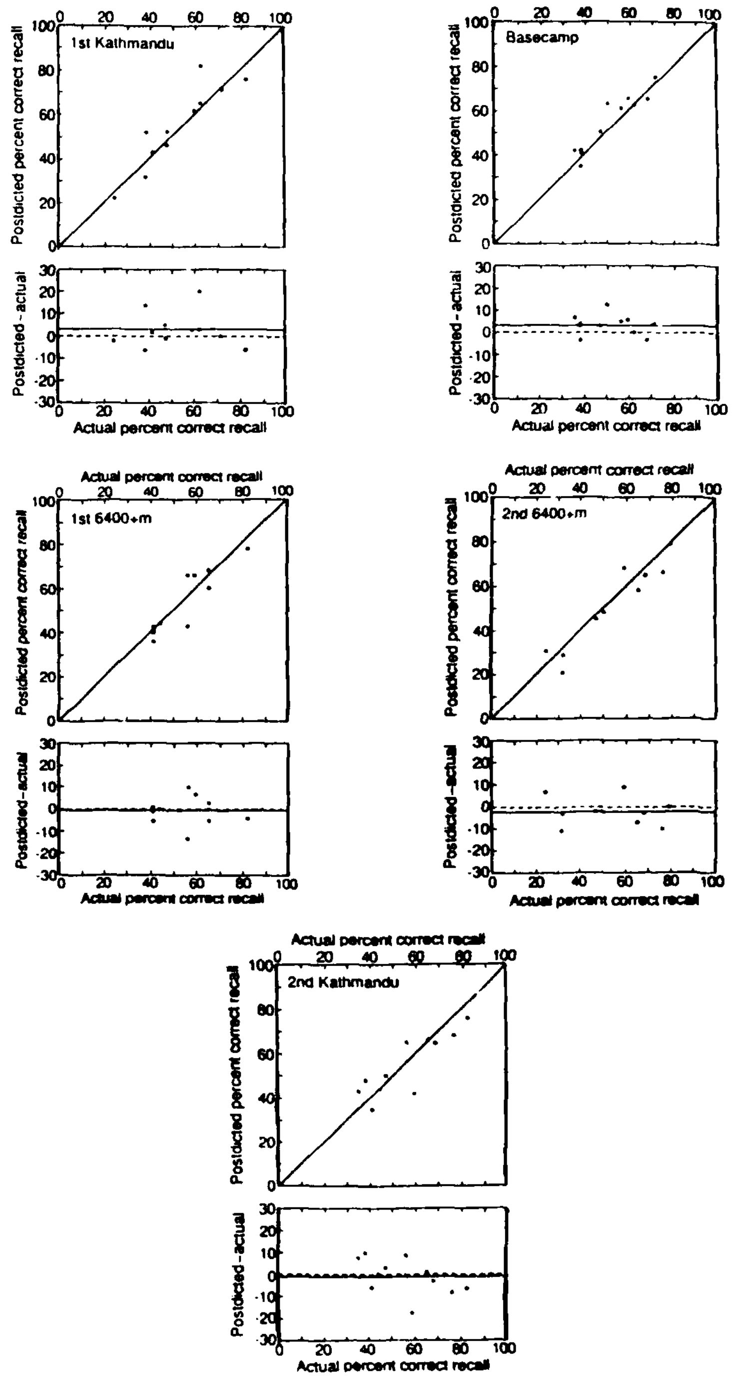

1st Kathmandu Y = .88X + 5
Basecamp Y = .97X + 4
1st 6400+ m Y = 1.04X + 4
2nd 6400+ m Y = 1.02X + 2
2nd Kathmandu Y = .97X + 6
Actual percent correct recall
Postdicted percent correct recall

fig 6


1st Kathmandu
Basecamp
1st 6400+m
2nd 6400+m
2nd Kathmandu
Postdicted percent correct recall
Postdicted - actual
Actual percent correct recall

the item that had the lowest FOK rating), and (2) it is based on relatively few observations, namely, only those nonrecalled items that weren't tied (which eliminates all dyads involving any recalled items, all dyads involving two recognized items or two nonrecognized items, and all dyads involving two items with same FOK rating). The *noisiness* of this particular measure is in accord with previous research showing that individual-S FOK-accuracy gammas for predicting recognition are typically noisy (Nelson, 1988). The large increase from basecamp to the first high-altitude test was marginally significant, $t(10) = 1.84, p = .096$.

Measures related to metacognitive control

Two measures were examined that pertain to the possibility that the termination of memory searching is controlled by FOK (i.e., people search longer for items that they believe they know than for items they believe they don't know). The correlation of FOK and the latency of incorrect recall is shown in Table 5.

See Table 5 on next page

Correlation of FOK and incorrect-recall latency for omission errors. The first row of Table 5 shows that there is a *strong* correlation (mean = +.60) of FOK and the latency of omission errors, in accord with previous research (Nelson et al., 1984) and with idea that people terminate search faster on items for which they have a lower FOK. The magnitude of this relationship was not significantly affected by altitude, $F(4, 36) < 1$.

Correlation of FOK and incorrect-recall latency for commission errors. The second row of Table 5 shows the correlation between FOK and the latency of commission errors. This relationship was not significantly different from zero (mean = +.03), in accord with previous research (Nelson et al., 1984), and was not significantly affected by altitude, $F(4, 36) < 1$. Thus, the latency of commission errors appears to be related to retrospective confidence judgments about recall (second row of Table 2) but not related to FOK (second row of Table 5), which confirms the aforementioned idea that the two kinds of metacognitive judgments are monitoring different aspects of memory.

Experimenter recording errors

One a priori concern was that just as the climber who is serving as the *S* might be affected by altitude, so might the climber who is serving as the *E*, in turn giving rise to the possibility of frequent experimenter errors during data collection. Our tape recordings of the testing sessions allowed us to assess this possible problem. We listened to each tape recording and compared the *S*'s spoken responses with *E*'s written records on the data sheets. Of the more than 5800 responses recorded by the climbers in the role of *E*, there were only 22 experimenter recording errors, i.e., an error rate of less than 0.4%. (Note: 8 of these 22 errors occurred during the first tests in Kathmandu.) Thus the climbers were remarkably accurate (namely, 99.6%) in their role as *E*, and the a priori concern of frequent experimenter errors was not borne out.

Table 5

Relationship between Feeling of Knowing and Two Kinds of Response Latency during Recall at Each Test Location

Recall response	Location: 1st Kathmandu	Basecamp	1st 6400m	2nd 6400m	2nd Kathmandu	$\overline{X}$
Omission error	+.62	+.50	+.67	+.68	+.52	+.60
Commission error	+.07	-.06	+.05	-.03	+.14	+.03
$\overline{X}$	+.35	+.22	+.36	+.33	+.33	

Note. Entry is the mean of the individual subjects' gamma correlation between response latency and feeling of knowing.

Experiment 2: Retrieval and metacognition in trekkers

METHOD

Subjects

The subjects were eleven trekkers who volunteered to be in the experiment and who were tested at all three testing locations. These were Americans who trekked (i.e., *hiked*) *from* Kathmandu *to* Everest basecamp and back to Kathmandu.

Locations for testing

The locations were the first, second, and last at which the tests occurred for climbers (see above).

Procedure

The procedure was the same as for the climbers in Experiment 1 (see above), except for the following modifications: (1) because only three tests occurred, there were more items available for testing at each location, so each of the three decks contained 79 items (i.e., total = 237 items); (2) sessions were not tape-recorded; (3) there was no speech sample; (4) each trekker was tested by one of the four members of the research team in Nepal.

RESULTS AND DISCUSSION

The main results from the trekkers are summarized in Table 6. With one exception, all *p*s for the effects of altitude are greater than .10 (i.e., negligible effect of altitude when the highest altitude is 5400 m). The one exception may be a small change in "G: FOK1 and FOK2" (i.e., retest reliability of FOK), $F(2, 20) = 3.65$, $p = .04$. Post-hoc followup tests showed that the reliability of FOK was greater at basecamp than at the first Kathmandu test, $t(10) = 2.69$, but did not differ significantly for either the two Kathmandu tests, $t(10) = 1.50$, or for basecamp versus the second Kathmandu test, $t(10) = 1.20$ (note: such a three-way outcome is impossible for the corresponding population means, indicating a likely fluke on one of those three inequalities).

See Table 6 on next page

The primary reason for testing the trekkers was to provide a control comparison to evaluate any significant differences between the first and second Kathmandu tests on the climbers. Although the absolute level of performance of the climbers versus trekkers should not be compared directly (because individuals were not randomly assigned to those two groups), the *pattern across locations* in each of the two groups can be compared. In particular, a significant shift on the second Kathmandu test *for the climbers*, but not for the trekkers, would indicate an aftereffect of the climbers' foray to high altitude. The critical comparison data from the trekkers, therefore, pertain to the P(commission | nonrecall) and the median FOK, both of which declined significantly from basecamp to the second Kathmandu test for the climbers (see above). For the trekkers, however, there was no significant change in the

Table 6

Mean Performance on Each Measure at Each Location for Trekkers

Measure	Location: 1st Kathmandu	Basecamp	2nd Kathmandu	$\bar{X}$
Percent correct recall	51	47	49	49
Percent correct recognition	43	42	46	44
Postdicted percent correct recall	52	51	49	51
P (commission/nonrecall)	.49	.52	.48	.50
G: FOK 1 and FOK 2	+.90	+.96	+.93	+.93
G: FOK and confidence	+.31	+.31	+.36	+.33
Median FOK	3.2	2.9	2.9	3.0
G: Postdicted recall and actual recall	+.94	+.92	+.92	+.93
G: FOK and recognition	+.37	+.48	+.52	+.46
G: Postdicted recall and recognition	+.15	+.10	+.19	+.15

Note. G = Goodman-Kruskal gamma correlation. FOK = feeling-of-knowing rating.

either of those variables across testing locations, $F(2, 20) < 1$ for each. These different patterns for the climbers and trekkers implies that the climbers' decreases in P(commission | nonrecall) and median FOK on the second Kathmandu test is due to an aftereffect of going to high altitude. A similar aftereffect occurred in previous research (Townes et al., 1984) for variables concerning acquisition, which decreased at high altitude and was still impaired upon return to Kathmandu.

The final aspect of trekkers' performance to be considered here is the calibration accuracy of the metacognitive judgments during recall. Figure 8 shows the three calibration curves for trekkers, corresponding to their three testing locations. The same general conclusions drawn previously for climbers' calibration (see above) also apply to the trekkers' calibration. A major conclusion is that there was a high degree of overall calibration for both groups.

See Figure 8 on next page

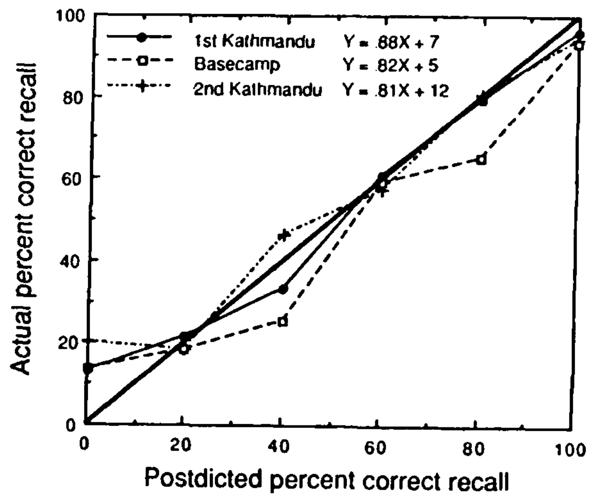
This can be seen more easily in Figure 9, in which the dashed lines are lines of best fit to the means of the climbers' and trekkers' performance across all testing locations (i.e., the data points in Figure 9 are the means of the data points in Figures 6 and 8). This overall calibration of the climbers is so accurate that it is difficult to see the small difference between the major diagonal (representing perfect calibration) and the line of best fit for the climbers' data points.

See Figure 9 on next page

Subjective impressions

Climbers' subjective impressions. Contrary to the results showing a lack of effect on the climbers' recall performance (see above), the climbers' peripheral remarks during the taped sessions at high altitude often indicated that they thought their recall accuracy and recall latency were being negatively affected by altitude. Also contrary to the results showing an effect of altitude on the climbers' metacognitions, the climbers made no mention of any noticeable effect on their metacognitive judgments (e.g., feeling of knowing). Thus the climbers' ability to monitor and describe these molar aspects of their behavior indicated a less-than-accurate awareness of that behavior.

PI's subjective impressions. The PI, who went through the Khumbu Icefall and up to Camp 2 (6400 m), recorded his subjective impressions in a diary. There were several noteworthy impressions, in addition to the general feeling of (as climbers frequently describe it) "spaciness": One in particular, that lasted several months after return to Seattle, was a recurring failure of memory for intention (sometimes called "prospective memory"). For instance, one day while climbing, the PI felt too warm and decided to remove his pile vest. He stopped, took off his pack and then his windbreaker, had a drink of



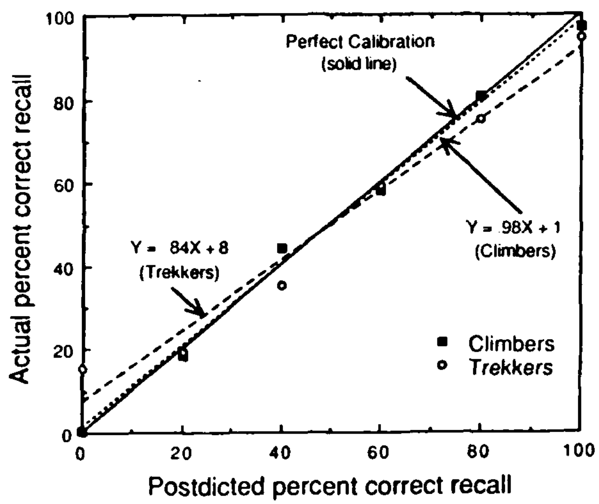


fig 9

water from the waterbottle in his pack, put the windbreaker and pack on again, and continued climbing; several hundred feet later, he still felt too hot and decided to remove his windbreaker, so he stopped, took off his pack, and upon taking off his windbreaker was shocked to discover that he had never removed his pile vest at the earlier stop! Upon return in Seattle, such failures were so frequent that he began to carry a portable dictaphone to give himself reminders *of what he intended to do during the day.*

In addition to the aforementioned failure of memory for intention, several other impressions suggested that several other topics should be explored systematically in future resear h. One impression (but see the inaccurate impressions from the climbers above!) was that retrieval of information from memory did seem more difficult. Perhaps the present general-information task provided such excellent cues that retrieval deficits were not evident in the climbers' memory performance, but would have occurred with a memory task consisting of less specific cues. This and other possibilities for natural followups to the present research are described next.

Directions for future research

FUTURE RESEARCH IN THE HIGH-ALTITUDE ENVIRONMENT:

1. *Memory for intention should be explored* (cf. the PI's subjective impressions mentioned above).

2. The aforementioned memory for intention (and the pattern of metacognitive deficits reported above) are in accord with the hypothesis that high altitude affects the frontal lobe. For instance, Shimamura & Squire (1986, 1988) found that Korsakoff patients had deficits in FOK but were no different than normals on confidence judgments about previous recall (cf. our findings above), and Shimamura, Jernigan, & Squire (1987) found that Korsakoffs have impaired frontal lobes (also see Janowsky, Shimamura, & Squire, 1987, for related behavioral deficits in frontal patients). To the degree that this hypothesis is valid, *memory for recency (aka "judgments of recency") should be explored* because that task is also sensitive to frontal-lobe impairment and therefore may be affected by high altitude. Other researchers (Oelz & Regard, 1988) recently reported that climbers who repeatedly had ventured above 8000 m without supplementary oxygen now have impaired concentration at sea level, and those researchers speculated that a probably permanent dysfunction has occurred in the fronto-temporal basal brain areas.

3. Another type of prospective metacognition (besides the FOK that we found to be affected by high altitude) is called "judgments of learning"(JOL), which are judgments made soon after learning and pertain to the likelihood of future recall. Previous research has found that JOL are impaired in Korsakoff patients (Bauer et al., 1984). *JOL should be explored* because they, like FOK, might be affected by high altitude. Also, as an extension of the different effects on retrospective vs prospective self-confidence, after the person has completed the phase of recall on K general-information questions (attempting recall and

making confidence judgments), request an overall estimate of how many items would be recalled correctly if K new items were presented.

4. *Two other kinds of retrieval from long-term memory should be explored*. Although the recall of general information (that had been learned long ago) was robust across the danger and hypoxia of high altitude in the present research, what would happen to retrieval when the information had been learned relatively recently before beginning the expedition (e.g., similar to the kind of information that a pilot might learn prior to a mission)? And what would happen to retrieval when such excellent and specific cues are not given to S (e.g., to get someone to recall "Helsinki", there are probably few cues better than "what is the capital of Finland?"). Subsequent research at high altitude should *(a) examine the retrieval of information that had been learned not long before the expedition leaves the USA* and *(b) examine retrieval from long-term memory when the cueing conditions are less specific* (e.g., categorized free recall with and without category cues given during retrieval). The "spaciness" that seems so obvious introspectively at high altitude might make access more difficult to a given category but once in the category, the access to a given item might be unaffected.

FUTURE RESEARCH IN OTHER ENVIRONMENTS:

The effects of high altitude on P(commission | nonrecall) and FOK should be teased apart into effects from danger vs hypoxia. Thus, two other situations should be examined:

5. The FACTRETRIEVAL battery should be administered in dangerous but nonhypoxic environments (e.g., undersea, Antarctica, space station, etc.).

6. The FACTRETRIEVAL battery should be administered in hypoxic but nondangerous environments (e.g., altitude chamber).

Actual Danger During This Season on Mount Everest

During the 1988 post-monsoon season, there were expeditions from several countries climbing on Mount Everest: the United States (our expedition), Korea, France, Spain, and a joint Czechoslovakia/New Zealand expedition. Our expedition and the Korean expedition had no fatalities; there were nine fatalities altogether on the expeditions from the other countries.

References Cited

Bauer et al. (1984) *Society for Neuroscience Abstracts*, 10, 318.

Butterfield, E. C., Nelson, T. O., & Peck, G. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, in press.

Cleveland, W. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.

Cleveland, W., & McGill, R. (1985). *Science*, 229, 828.

Houston, C. (1987). *Going higher: The story of man and altitude*. Boston: Little, Brown, & Co.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1987). Cognitive deficits in patients with frontal lobe lesions: Significance for the amnesia of Korsakoff's syndrome. *Society for Neuroscience Abstracts*, 13, 1454.

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.

MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, 57, 215-235.

McFarland, R. A. (1953). *Human factors in air transportation*. New York: McGraw-Hill.

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

Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (vol. 2). New York: Wiley.

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., McSpadden, M., Fromme, K., & Marlatt, G.A. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General*, 115, 247-254.

Nelson, T.O., & Narens, L. (1980). 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.

Oelz, O., & Regard, M. (1988). Physiological and neuropsychological characteristics of world-class extreme-altitude climbers. American Alpine Journal, 83-86.

Riefer, D. M. (1987). Is 'backwards reading' an effective proofreading strategy? (Report No. CS 008 786). *Resources in Education*. (ERIC Document Reproduction Service No. ED 281 175).

Shimamura, A. P., Jernigan, T. L., & Squire, L. R. (1987). Radiological (CT) findings in patients with Korsakoff's syndrome and their relationship to memory impairment. *Society for Neuroscience Abstracts*, 13, 1453.

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 452-460.

Shimamura, A. P., & Squire, L. R. (1988). Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 763-771.

Townes, B.D., Hornbein, T.F., Schoene, R.B., Sarnquist, F.H., & Grant, I. (1984). Human cerebral function at extreme altitude. *High Altitude and Man*, 31-36.

Weingartner, H., Rudorfer, M., & Linnoila, M. (1985). Cognitive effects of lithium treatment in normal volunteers. *Psychopharmacology*, 86, 472-474.

West, J. (1986). Do climbs to extreme altitude cause brain damage? *Lancet*, 2, 387.

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.

Winer, B.J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.