Relations among measures of autobiographical memory

JAMES R. ERICKSON and CAROL R. JEMISON
University of Texas at Arlington, Arlington, Texas

Students recorded one event from their lives each day for 12 weeks and rated events along several dimensions. About 5 months later, yes-no recognition, recognition-discrimination, cued-recall, and time-estimation tests were given. All tests included foil items from other subjects. Performance was high on both recognition tests, although about 11% of other students' events were accepted as "old." Time estimates were positively correlated with real time but were rather inaccurate. Cued-recall performance was fairly accurate for event cues that students recognized as old. Factor analysis of rating scales yielded three independent factors: outcome (memorable events were positive and satisfactory), frequency (memorable events were infrequent and atypical), and salience (memorable events were rated as memorable, surprising, exciting, and significant).

Research on autobiographical memory (memory for information related to one's self) has been reported with increasing frequency in the past few years; an overview of some of this research can be found in Rubin (1986). Early research in the area involved cued recall. For example, Galton (e.g., 1883) probed his memory by using a variety of words, a technique still in use (e.g., Crovitz & Quina-Holland, 1976). A fairly obvious problem with this method is that accuracy cannot be verified.

Another line of research tests memory of events known to have been experienced. For example, Linton (1975) and Wagenaar (1986) kept records of daily events for several years and tested themselves for memory of these events. Linton focused on the accuracy with which events could be dated, but because she did not attempt to date events for which she had no memory, she was able to report that event recognition remained high over several years. Wagenaar recorded "who, what, where, and when" information for each event and used these categories as recall cues. He found that a "what" cue yielded the best recall of other information, whereas a "when" cue produced the poorest, and that there was much better recall of "who" or "where" than "what" or "when" information. A problem with this kind of study is that the "subjects" know that all tested items actually occurred, which provides no control for potential response biases. If only to obtain base-rate information, it seems useful to obtain "memory" reports for foil items that did not occur (or were not recorded).

Brewer (1988), Barclay and DeCooke (1988), and Barclay and Wellman (1986) have used variants of the procedure introduced by Linton (1975), in which subjects record daily events for a period of time and are tested

for memory of these events. Brewer (1988) obtained records of events and thoughts at random times (by using "beepers"), whereas Barclay asked subjects to record memorable events each day. Brewer argued that foils are not necessary, and did not use them, whereas Barclay has used a variety of foils. For example, Barclay and DeCooke's foils (1988) involved meaning or style changes or were items from other subjects.

Both Brewer (1988) and Barclay and Wellman (1986) report very high recognition of old events over periods of up to 1 year, but, as in Linton (1975), they found relatively poor accuracy at dating events. Interestingly, Barclay and Wellman found that over 20% of items written by other subjects were judged as old. Barclay and DeCooke found much lower acceptance of "other" foils at shorter test intervals and found that "same content" foils were accepted as old at about the same rate as unchanged items, whereas "inconsistent" foils were almost never accepted as old.

In many of the above studies, ratings of items on several dimensions were correlated with performance. For example, Brewer (1986) found that action recognition correlated with infrequency, significance, and excitement; Wagenaar (1986) found that cued recall was related to salience (infrequency), emotional involvement, and pleasantness. Such correlations tend to be quite low ($r < .20$).

In the present study, students recorded one memorable event each day for 12 weeks and rated the events on eight scales. About 5 months later, four memory tests were administered. Items were rewritten to remove specific facts (e.g., names and dates) that would have allowed responses to be based on that information alone. Equal numbers of original items and foils (events written by other students) were presented on each test, providing base-rate information on some tests and response-bias data on others.

Two recognition tests were given: one in which the subjects made yes-no decisions and one in which they chose

Correspondence should be addressed to James R. Erickson, Department of Psychology, Box 19528, University of Texas at Arlington, Arlington, TX 76019-0528.

which item of a pair was theirs. In a cued-recall test, a phrase was deleted from each item, and the subjects were asked to complete the item. Finally, subjects estimated the week during which events occurred. The major purposes of the tests were to compare the various measures of autobiographical memory and to examine the extent to which they are related to the various event ratings.

METHOD

Subjects and Materials

Students in a cognitive psychology class at the University of Texas at Arlington were asked to record a personal event each day for a 12-week period. Events were rated on 7-point scales for memorability (1 = easily forgotten, 7 = memorable), personal significance (1 = irrelevant, 7 = significant), frequency (1 = infrequent, 7 = frequent), typicality (1 = unique, 7 = routine), satisfaction (1 = unfulfilling, 7 = satisfactory), surprise (1 = expected, 7 = surprising), emotionality (1 = boring, 7 = exciting), and outcome (1 = negative, 7 = positive). Nine of these students turned in records covering from 7 to 12 weeks and were asked to participate in the present study. Eight of them provided ratings for their events. Memory tests were given individually at the end of the next semester, an average of 22 weeks later.

Before testing, dates and names of people specific to a student were deleted from each item. Some items were slightly rewritten to produce test items that were similar in general style and length.

For example, one recorded event was: "Went to Zaks to get a couple of pictures framed (early Xmas presents)—they had 50% off all custom frames so I thought I better do it now or never. I saw my friend Susan there doing the same thing." This was modified to the following: "I went to Zak's to get a couple of pictures framed as their custom framing was 50% off. I saw my friend _____ there doing the same thing."

Procedure

Memory tests were given in this order: cued recall, yes-no recognition, recognition discrimination, and time estimation. For each subject, weekly events were randomly assigned to one of these tests. When more than four events were available for any given week, they were assigned to yes-no discrimination, recognition discrimination, cued recall, and time estimation, in that order, resulting in unequal numbers of items on the various tests. Events from the same week, but from other subjects, were randomly chosen as foils; each event was used once as an "old" item for the subject who wrote it and once as a "new" item for another subject. On each test, items, or pairs of items, were presented in random order.

On the cued-recall test, a phrase was deleted from each recorded event, and subjects were asked to complete the items. For example, "Monthly reports are due today and Lisa is out of town so I have to do them" became "Monthly reports are due today and _____." In addition to completing items, subjects rated them for recognition as "sure old," "probably old," "probably new," or "sure new."

On the yes-no recognition test, the subjects indicated whether or not each item was one of theirs and rated their confidence on a 3-point scale ranging from "know" (they could recall an event judged as old or were positive that an event judged to be new was not theirs) to "maybe" (an item judged to be old was consistent with their life-style, but they did not recognize it, or they did not think a "new" item was theirs but were not sure).

On the recognition-discrimination test, the subjects selected the item of a pair that they believed to be theirs and then rated *both* items by using the 3-point scale described above.

For the time-estimation test, the subjects were given a list of dates of the weeks of the semester—they chose a week for each event and rated their confidence on a 3-point scale (accurate within a week, within 2 to 3 weeks, or guessing).

RESULTS

Memory Tests

Recognition memory. Even though many items were modified stylistically, and specific information (e.g., names) was deleted, performance after 5-8 months was quite good on both recognition tests, almost at ceiling, and was essentially constant across the 12 input weeks.

Mean accuracy scores on each test were calculated for each subject and used for formal analyses. A Type I probability of .05 was used for all decisions regarding statistical significance. On the yes-no test, the subjects were equally accurate at recognizing their own items and at rejecting new items. The mean correct response probabilities for old and new items were .94 and .89, respectively, a nonsignificant difference [$t(8) = 1.15$] that corresponds to a d' of 2.78. Analysis of yes-no scores weighted for confidence showed similar results, with means of 2.33 and 2.20 for old and new items, respectively [$t(8) = .59$].

On the recognition-discrimination test, the mean probability of a correct response was .96, corresponding to a d' of 2.48. Scores weighted for confidence were almost identical on old (2.52) and new items [2.56; $t(8) = -.51$].

Cued recall. Recognition performance on cued-recall items was poorer than it was on the recognition tests, undoubtedly because items were incomplete. The correct recognition probabilities were .80 for old items and .76 for new items, corresponding to a d' of 1.54. Item completions were scored on a 4-point scale (4 = almost complete agreement between the completion response and the item as written, and 0 = no agreement). Mean cued-recall scores were above 0, but were not particularly high, and did not vary much with input time. Mean performance on old items (1.30) was significantly better than it was on new items [0.38; $t(8) = 5.97$]. This difference was entirely due to old items that students recognized as old, for which the mean completion score was 1.51; for old items not recognized as old, it was 0.36.

Time estimation. For both old and new items, estimated time was positively related to actual time, but the subjects were not particularly accurate at estimating the week during which events occurred. The mean difference between estimated and actual week was 3.20 for old items and 4.92 for new items; because of the high variance, this difference was not significant [$t(8) = .41$].

For old items where time estimates were not guesses (confidence ratings of 1 or 2), time estimation was more accurate, with a mean error of 2.64 weeks. For these items, the regression line predicting estimated week (y) from input week (x) was $y = 3.82 + .42x$; estimated time increased about 40% as rapidly as actual time. For these items, the slope of the regression line and the corresponding correlation of .82 between actual and estimated time were significant. For new items (and unrecognized old items), the regression slope ($b = .25$) and the correlation between actual and estimated times ($r = .33$) were

Table 1
Correlations Among Variables

Variable	Memorability	Significance	Frequency	Typicality	Satisfaction	Surprisingness	Emotionality	Outcome
Yes-no Recognition	.09	.03	-.14	-.22*	-.03	.03	.15	-.04
Recognition Discrimination	.31*	.22*	-.16	-.16	.15	.16	.30*	.09
Cued Recall	.14	-.01	-.32*	-.28*	.11	.07	.25*	.15
Time Estimation	-.23*	-.15	.03	-.05	.19*	-.22*	-.18*	.25*
Memorability		.30*	-.22*	-.29*	.17*	.19*	.49*	.16*
Significance			-.08	-.11*	.11*	.08	.26*	.08
Frequency				.81*	-.01	-.18*	-.17*	.00
Typicality					-.03	-.21*	-.20*	-.01
Satisfaction						-.14*	.35*	.85*
Surprisingness							.27*	-.16*
Emotionality								.28*
Outcome								

* $p < .05$.

positive but nonsignificant at the sample size of $n = 12$ weeks.

Relations among Item Ratings and Memory Tests

Students rated each item on the eight scales described above. Table 1 presents correlations among rating scales and among ratings and performance for old items on memory tests (weighted-recognition scores, cued-recall scores, and estimates of the week an item occurred were used). Correlations among rating scales are based on approximately 400 subject-items. Correlations between tests and ratings are based on 58 subject-items for time estimation, 114 for cued recall, 144 for recognition discrimination, and 153 for yes-no discrimination.

Memory scores were related rather weakly to the various item ratings, and item ratings were often strongly intercorrelated. Because of the lack of independence among scales, a principal-component factor analysis was performed on scale correlations. Three factors (with eigenvalues greater than 1.0) resulted. An orthogonal varimax rotation of these factors yielded a factor pattern approximating simple structure, with each scale loading .40 or more on only one factor. Factor 1 involved outcome and satisfaction, with loadings of .94 and .94. This factor is denoted *outcome*, with high scores indicating positive and satisfying events. Memorability, emotionality, significance, and surprisingness, with loadings of .78, .72, .68, and .49, respectively, formed Factor 2. This factor is denoted *salience*, with high scores indicating memorable, exciting, significant, and surprising events. Frequency and typicality, with loadings of .94 and .94, formed Factor 3, denoted *frequency*, with high scores indicating frequent and routine events.

Factor scores were calculated for each item for each subject and used, along with the week the event occurred, as predictors in regression analyses, with performance on memory tests as criterion variables.

Probably because of a ceiling effect, yes-no recognition scores weighted for confidence could not be predicted by the four-variable regression equation [$F(4,115) = .803$, $r < .027$]. In addition, none of the predictor vari-

ables were individually significant (the largest t value was -1.58 for frequency, with a weight of $-.11$).

Weighted recognition-discrimination scores were related to the regression equation [$F(4,107) = 4.95$, $r = .156$]. Individually, salience ($b = .23$, $t = 3.95$) and frequency ($b = -.13$, $t = -2.12$) were significant.

Cued-recall scores were also predictable from the full model [$F(4,84) = 3.93$, $r = .16$]. Individually, only frequency was significant ($b = -.36$, $t = -2.42$), although outcome ($b = .30$, $t = 1.87$) and input week ($b = -.09$, $t = -1.88$) approached significance ($p < .07$).

Finally, time estimates were related to the regression model [$F(4,53) = 3.48$, $r = .208$]. Individually, outcome ($b = 1.08$, $t = 2.39$) and salience ($b = -1.32$, $t = -2.55$) were significant. (Actual time and estimated time were positively correlated, but actual time was not significant in the regression equation that included the other variables.)

The correlations among memory scores and predictor variables are shown in Table 2 and are based on the sample sizes (subject-items) given above.

DISCUSSION

After 5 to 8 months, the students were quite accurate at recognizing events from their own lives and at discriminating them from events in other people's lives. The 11% false-alarm rate is not to surprising given the general similarity of events in college students' lives (going to class, taking tests, dating, going to movies and other events, etc.). As in other studies, time estimation was rather poor, even for items where students reported that they were not guessing. Performance on cued recall was intermediate and was fairly good for item cues that were recognized as old. Because it is usually desirable to avoid ceiling and floor effects,

Table 2
Correlations among Item Memory Scores, Factor Scores, and Input Time

Memory Test	Predictor Variables			
	Outcome	Salience	Frequency	Time
Yes-no recognition	.03	.07	-.14	.05
Recognition Discrimination	.08	.33*	-.16	.03
Cued Recall	.17	.11	-.31*	-.16
Time Estimation	.27*	-.33*	-.06	.25*

* $p < .05$.

it is recommended that cued recall be included in future studies of autobiographical memory. In the present experiment, item cues could not be systematically varied, as Wagenaar (1986) was able to do; in future research, it will be of value to continue to examine the efficacy of different types of cues.

The factor analysis yielded three item characteristics that were independent of one another. These were outcome (memorable events tend to have positive, satisfactory outcomes), frequency (memorable events tend to be infrequent and atypical), and salience (memorable events tend to seem memorable and to be exciting, significant, and surprising). An exception is the fact that high-salience events tended to be poorly dated, for unknown reasons. Rating scales based on these factors should be useful in future research, if for no other reason than for increasing the reliability of ratings by reducing subjects' confusion.

REFERENCES

- BARCLAY, C. R., & DECOOKE, P. A. (1988). Ordinary everyday memories: Some of the things of which selves are made. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 91-125). Cambridge, U.K.: Cambridge University Press.
- BARCLAY, C. R., & WELLMAN, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory & Language*, **25**, 93-103.
- BREWER, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 21-90). Cambridge, U.K.: Cambridge University Press.
- CROVITZ, H. F., & QUINA-HOLLAND, K. (1976). Proportion of episodic memories from early childhood by years of age. *Bulletin of Psychonomic Society*, **7**, 61-62.
- GALTON, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.
- LINTON, M. (1975). Memory for real world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376-404). San Francisco: W. Freeman.
- RUBIN, D. C. (Ed.). (1986). *Autobiographical memory*. Cambridge, U.K.: Cambridge University Press.
- WAGENAAR, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, **18**, 225-252.

(Manuscript received December 10, 1990.)