

Accuracies and Inaccuracies in Autobiographical Memories

CRAIG R. BARCLAY

University of Rochester

AND

HENRY M. WELLMAN

University of Michigan

Records of everyday autobiographical events were gathered from a small group of adults during a 4-month period. This was followed by five memory tests extending over 2½ years. Recognition memory, temporal ordering, and dating accuracy declined as the events tested became more remote. Recognition accuracy on original items was high over the entire study; whereas the false recognition of nonevent, foil items increased after a 1- to 3-month delay. Confidence ratings of recognition accuracy remained consistently high over all tests, even though recognition accuracy deteriorated. Additional analyses of foil items indicated that false recognitions of nonevents as one's own memories were related positively to the semantic similarity between foils and the original records from which they were constructed. Taken together, the data support the hypothesis that the same autobiographical schemata account for the correct recognition of actual events, the false recognition of certain nonevents as one's own memories, the correct rejection of other nonevents, and an overconfidence in the "facts" of one's life. © 1986 Academic Press, Inc.

The importance of autobiographical recollections has long been acknowledged (Freud, 1905/1953, 1914/1960; Galton, 1879, 1911; Hall, 1899; Henri & Henri, 1895, 1898). Only recently, however, have psychologists attempted systematic studies of these memories (e.g., Bahrck, Bahrck, & Wittlinger, 1975; Fitzgerald, 1980; Linton, 1975, 1979; Robinson, 1976; Warrington & Sanders, 1971). Much of this interest stems from concerns regarding clinical phenomena like childhood amnesia (e.g.,

White & Pillemer, 1979), or the concordance between self-schemata and the processing of self-referenced information (e.g., Kuiper & Derry, 1980; Markus, 1977; Rogers, 1980).

A variety of differing methods have been used to study autobiographical memory (e.g., Crovitz & Cordon, 1982; Herrmann & Neisser, 1978; Linton, 1975; Rubin, 1982). Two somewhat contradictory themes emerge from this work. The first claims that people often forget or confuse everyday happenings, especially if similar episodes are encountered frequently. The second claims that autobiographical events seem to remain vivid for years; in particular, those events associated with strong affective cues (e.g., flashbulb memories, Brown & Kulik, 1977; Neisser, 1982; Winograd & Killinger, 1983). These claims are difficult to assess and the contradictions difficult to resolve because valid records of the "remembered" events themselves were usually un-

The first author was supported in part by NICHD Grant HD-07109 while at the University of Michigan, and the second author by NIMH Grant HD-00525. A portion of these data were presented at the American Psychological Association meetings, Washington, D.C., 1982. Appreciation is expressed to the participants and Nancy Abrams for her assistance. A special thanks to Marigold Linton and David Rubin for their critical reviews of earlier versions of this paper. Send requests for reprints to Dr. Barclay, Graduate School of Education and Human Development, University of Rochester, Rochester, NY 14627.

available to verify the accuracy of reported memories (exceptions include Gold & Neisser, 1980; Linton, 1975; Thompson, 1982). The general accuracy of autobiographical memories is thus unclear.

An influential study in which accurate records of daily events were available is that of Linton (1975). In a 6-year longitudinal investigation, Linton recorded from two to five events daily, rated each event on dimensions thought relevant to its memorability (e.g., emotionality and datability), and at the end of each month, tested her capacity to recognize, sequence, and date records selected randomly from previous months. Linton found that everyday, episodic events were *not* forgotten quickly—she recognized events as her own after lengthy delays. Since her results are based on verifiable earlier events, they are striking because forgetting seemed even less than suggested by previous research where the subjects' accuracy could not be independently checked. In addition, her findings suggested that the perceived affective qualities associated with some episodes were not consistently good predictors of memory. Of the events remembered, most were not associated with strong affective cues; and, conversely, some episodes rated high in emotionality were forgotten.

In interpreting Linton's results concerning recognition accuracy, it is important to note that the main focus of her work was on determining whether she could remember *when* some event occurred, not in deciding if an episode in fact happened. Our concern is primarily recognition memory accuracy. In this regard, it is informative to consider that Linton served not only as the observer of events in her study and therefore the subject, but also as her own examiner. Because of this, it was not plausible to present herself with "false memories" for recognition. Her recognition test thus involved judging actual events as remembered or not. A more typical and sensitive test of recognition memory requires two types of items: actual and foil events. In

this case, foils would be constructed records of nonevents—episodes that never happened.

Foils are needed to separate memory sensitivity from response bias since recognition judgments have been found to depend on these two related psychological processes (Banks, 1970; Green & Swets, 1966). Further, recognition decisions depend partly on the person's subjective familiarity with the test items (Mandler, 1980). Knowing that many (if not all as in Linton's case) of the items on a recognition test are taken from one's own records likely creates a bias toward a lenient response criterion, thereby leading to an overestimate of memory accuracy and increasing the probability of identifying foil items as true events (cf. Locksley, Stangor, Hepburn, Grosovsky, & Hochstrasser, 1984). From this perspective, memory accuracy can only be assessed by looking jointly at the relative differences in recognition performance for actual events and the false recognition of foils. The use of foil items also provides a method for determining which features of events are related to improved memory or heightened confusedness by examining false recognitions of specifically constructed nonevents. Thus, our aims were to chart the fate of everyday memories over long time periods with special emphasis on accuracy as assessed by comparing memory for actual events with that for foils.

METHOD

Design

Four consecutive 1-month data collection periods (February–May 1981) were followed by five recognition, ordering, and dating tests. The first four tests were given in mid-June, September, and December 1981, and February 1982. Items on the June and September tests were selected from the March, April, and May data collection periods, thus creating 1- to 3- and 4- to 6-month delay intervals, respectively. Items

on the December test were taken from the February, April, and May data collection periods yielding delays of 7, 8, and 10 months. On the February test, 9-, 11- and 12-month delay intervals were formed by choosing items from the February, March, and May data collection periods. The fifth test was not part of the original design but was given in mid-September 1983 as a follow-up to evaluate memory for autobiographical material after a further, extended delay. The records selected for this fifth test were drawn from February and March, 1981; thus, the delay interval was 30–31 months from when the memories were recorded and tested.

This design incorporates three important controls. First, all of the daily records were collected before testing began to ensure that future record keeping was not influenced by prior test knowledge. Second, items were selected for testing only once, to avoid the influence of repeated testing. Third, the items on each test were chosen so as to include records from a wide range of data collection periods. For example, the items used at each testing time were collected during three different periods. Our concern here was that the subjects might use specific temporal cues among items as a retrieval strategy to aid within-test dating performance.

Subjects

Six graduate student volunteers, four females and two males, were paid for their participation. Attrition claimed one student during the first month of data collection who found daily record keeping too time consuming. Two students were excluded from the second and three from the fourth test since they did not keep sufficient numbers of records. Three subjects completed the fifth test, the 2½ year follow-up.

Tasks and Materials

One set of materials was used for data collection, another for testing. In data collection, the materials were 240 possible

subject-generated daily records. Each day's records were kept on a separate 4 × 6 in. index card. Abbreviated instructions were included across the top of the card. Below the instructions was space for the subjects to write their records. The subjects were encouraged to use a simplified format for their episode record keeping that included three components: a context, a description of the event itself, and a reaction to or evaluation of the event. This format was derived from the elements of story grammars (cf. Mandler & Johnson, 1977). No attempt was made to define comprehensively the nature of everyday events; subjects defined for themselves what constituted the events they reported.

The tasks and materials used for each of the first four times of testing included a 45-item, "Yes/No"-type recognition test and a 12-item (six pairs) ordering and dating test. The 2½ year post-test was similar except 18 items were used for ordering and dating (nine pairs).

Each 45-item test included 18 original records (six records from three data collection periods) identical to those reported by the subjects. The remaining 27 items were foils. Three types of foils were used. Nine foil evaluations, three from each of three data collection periods, were rewritten from original records by changing the subjects' reactions to or evaluations of events. Nine foil descriptions, again based on three records chosen from three data collection periods, were constructed by rewriting some aspect of the subjects' description of the event itself. A simplified illustration would be if a subject had reported:

Event: I went to lunch at a new restaurant.

Original Description: The menu included a selection of deli sandwiches and fresh breads.

Original Evaluation: At last there is some variety in places to eat by my office.

Foil Description: The waiter asked if I wanted coffee before ordering.

Foil Evaluation: It looks like just another mediocre restaurant.

Nine foil-other items were taken from the

records of a person not participating in the study but who lived in the same university environment. A 7-point confidence rating scale was associated with each Yes/No item (1 = Not confident, 7 = Very confident the response was correct).

In developing the foil-evaluation items, we rewrote the evaluation of an event by (a) negating the actual reaction, and (b) fabricating an unlikely evaluative statement if the record lacked an evaluation. The foil-description items were manufactured by replacing or revising the details or elaborations of the event with a semantically consistent description of the event. Two judges working together constructed the entire pool of foil items for each recognition test. Total agreement was reached on the appropriateness of each foil before it was included on a test. A foil was judged appropriate if it was a plausible change given the judges' general knowledge of the person and a specific original record; thus, if a person was a known vegetarian, "hamburger" would not be substituted for "bean salad." It is noteworthy that we knew our subjects very well, allowing us to construct reasonable but still fundamentally altered foil items. It was not possible to have the subjects rate the plausibility of foils; however, all of the foils from the first four tests were rated by independent judges. Judgments were made regarding the degree of semantic difference between foil-evaluation and foil-description items, and the original records from which they were written.

The ordering and dating tasks for the first four tests consisted of six pairs (12 items) of original records. These originals were the same items as 12 of the 18 originals on the recognition test. The memories in each pair were selected and matched to ensure that they were recorded at least 1 month apart. Located to the right of each ordered pair was a 7-point confidence rating scale (1 = Not confident, 7 = Very confident that the records were ordered correctly in time). On the 2½ year post-test, nine pairs (18 items) of original records were matched

without regard for the delay interval between recording times because these items were selected from two adjacent data collection periods. This change in the construction of the fifth test was mandated by the limited number of items remaining from certain data collection periods.

Procedures

Data collection. The subjects were contacted individually and told that this was a study of their memory for everyday life events. They were informed that we were interested in the content of daily events self-selected as memorable, and with long-term memory for everyday happenings. In a set of general instructions, the subjects were asked to keep careful records of at least three memorable events a day, 5 days a week (Monday–Friday). These records were collected every Monday for the previous week and this arrangement continued for 4 months.

Specific record-keeping instructions were given to each subject. The essential features of these instructions were that on Monday of each week, the subject was to take five cards and date them with the corresponding day of the week it would be used. When recording events, subjects were asked to keep the following in mind: (1) the *content* of the events was entirely up to them, and (2) each event record should probably contain at least three elements—a context, an event, and a reaction. The subjects were encouraged but not required to use this format to ensure some uniformity in the records and to provide enough information so that appropriate foil items could be constructed. Subjects were told:

Context refers to the environment in which the memorable event is placed. Context may be general (for example, "This morning while I was walking across campus. . .") or specific (for example, "I was on the second floor of the library at 3:00 PM when. . ."). An *event* is something that happens or a noteworthy happening. That is, an event is the content of

what you think is memorable. Your *reaction* to the event includes any internal, emotional response elicited by the event and any external, overt physical or verbal response you make.

Note that subjects were instructed to record events that were memorable or noteworthy. This certainly restricted the pool of possible events. We adopted this procedure for several reasons. One was our need to get records of the original events for later accuracy comparisons. Like Linton, we used subjects' contemporary records of these events as our original data. So that these contemporary records would themselves reflect the original events accurately we asked for records of memorable, noteworthy events. Two concerns are obvious with this procedure. First, it may inflate accuracy for original events. But notice our analysis strategy rests on comparing memory for foils and originals, not simply examining accuracy on originals alone. Second, it is not entirely clear that our subjects' performance best reflects memory for events or for the records of those events. This ambiguity was necessary in order to collect a set of verifiable records against which to judge accuracy at all.

Testing. Subjects were tested individually. The recognition test was completed first, followed by the ordering and then dating tests. On the recognition test, subjects were instructed to positively identify only those items which were exact replicas of their original records, and then rate how confident they were that their answers were correct.

Ordering involved the sequencing of two original events in time. The subjects were told to decide which event in a pair was most remote and which most recent. Once the items in a pair were ordered, the subjects rated how confident they were that the items were sequenced correctly. They then tried to give the exact date each event had occurred—the dating test.

RESULTS AND DISCUSSION

In the data collection periods, the

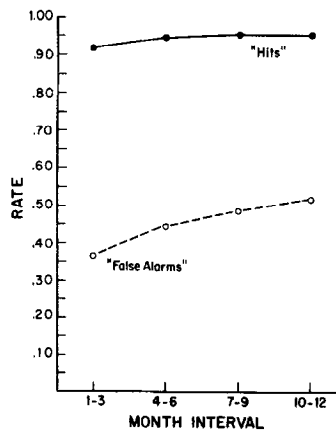


FIG. 1. Hit and false alarm rates as a function of month interval.

missing data rate for the five subjects completing the first test was .17 (range = .05–.49). That is, of the 240 possible records, on the average 199 were in fact collected.

The findings of primary interest include first, recognition performance, then ordering and dating accuracy. Figure 1 depicts the "hits" and "false alarms" for items on the recognition tests. Foil items were scored as false alarms if the subject responded "Yes" thereby falsely identifying a foil as an original. These data were collapsed across subjects and blocked into four, 3-month intervals to emphasize the performance trends. The hit rate was higher than the false alarm rate at every month interval (smallest mean difference, $t(31) = 32.74$, $p < .001$).¹ In Figure 2, the proportion of items correctly recognized is plotted separately for various foil types; the data on originals are included again for comparison. Accuracy on foil-other items was significantly less than that found for originals ($t(62) = 18.21$, $p < .001$) but greater than for foil-evaluation and foil-description items at every delay interval (smallest mean difference, $t(11) = 2.77$, $p < .05$). The overall mean for foil-other items was .76.

¹ These statistical tests are reported as rough estimates of Type I errors. Completely appropriate analyses are unavailable because of the limited number of subjects tested, the scores among item types were not independent, and repeated tests were given over delay intervals. The degrees of freedom reported represent number of data points in the analyses.

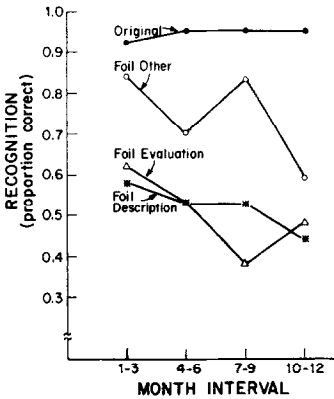


FIG. 2. Recognition accuracy as a function of month interval and item type.

These data show an accurate discrimination between original and foil memories, but also suggest a positive response bias. Thus, hits and false alarms significantly differ, but foils are often accepted as one's own. A signal detection analysis confirmed this impression. Our data, because of the small number of subjects, do not strictly conform to the statistical assumptions required for signal detection analysis. Nevertheless, estimates of sensitivity (A') or true ability to discriminate original items from foils, and response bias (B'')—computed using nonparametric signal detection measures (Grier, 1971)—corroborate the above findings. Based on scores summed across the first year delay intervals for original, foil-evaluation, and foil-description items, A' (sensitivity) = .84 and B'' (response bias) = $-.61$ (d' averaged 1.78). These results indicate a high degree of discriminability but one coupled with a bias toward identifying recognition test items as original records.

Because of the small number of measurement points, the data cannot support signal detection analyses at the different delay intervals. Nonetheless, it is clear from Figures 1 and 2 that accuracy declined with time: correct acceptance of original records stayed quite high, but false alarms increased. Specifically, the rate of false alarms increased from .37 at the 1- to 3-

month interval to .52 at 10–12 months, $t(56) = -8.35$, $p < .001$. The largest change between adjacent intervals was from .37 to .45 for intervals 1–3 versus 4–6, respectively, $t(61) = -6.15$, $p < .001$. Conversely, the correct rejection of foils decreased with time, and it did so for each of the foil types—foil other, foil evaluation and foil description (comparing 1- to 3- vs 10- to 12-month delays, the smallest mean difference across the three types yielded, $t(23) = 7.34$, $p < .001$). Subjects thus got worse at rejecting foil items with increasing delay. This was true even for foil-other items; subjects could usually, but by no means always, correctly reject the memories of someone else who was also engaged in familiar kinds of daily activities associated with college life. The means for foil-other items varied widely across time, with a range of .84 at 1–3 months and .59 at 10–12 months. This variation was probably due to differences in the specific nature of the foil-other items, i.e., the inclusion of varying numbers of distinctive events that were impossible as one's own. In line with this, variance across subjects was greater for foil-other items than for originals, foil-evaluations, or foil-descriptions at every delay except for the 7- to 9-month interval.

At the 2½-year test, 79% of the original records were identified accurately (hits). This compares with 95% at the 1-year test. At the same time, 74% of the foil items were correctly rejected. This high level of performance is clarified by examining the percentage of foil-evaluation, foil-description, and foil-other items correctly recognized; the means were 52, 48, and 96%, respectively. Two subjects were 100% correct in rejecting the foil-other items at the 2½-year post-test. Again, these particular foils were apparently so dissimilar from originals that they were easily identified as distractors. Excluding the foil-other items, the correct rejections of foil-evaluation and foil-description items did not differ from each other and were comparable to the rates reported in the 1-year data.

TABLE 1
MEAN CONFIDENCE RATINGS OVER DELAY INTERVAL
BY ITEM TYPE FOR CORRECT, MISSES, AND
FALSE ALARM RESPONSES

Item type	Months delay			
	1-3	4-6	7-9	10-12
Original				
Correct	5.76	5.89	5.76	5.92
Misses	4.89	6.33	5.75	5.25
Foil evaluation				
Correct	4.52	5.23	5.63	4.92
False alarms	4.39	4.59	5.29	5.21
Foil description				
Correct	3.00	5.69	4.48	5.53
False alarms	5.85	5.00	5.75	4.67
Foil other				
Correct	5.12	5.95	5.90	5.37
False alarms	5.08	4.58	2.75	3.76

Note. Maximum possible score = 7 (Very confident).

Inspection of the subject's confidence ratings (Table 1) for the first four delay intervals shows that overall, the ratings were only slightly higher for correct responses than for misses (i.e., saying "No" to original records) and false alarms. Most importantly, the ratings were relatively stable with increasing delay intervals. The only apparent decline in confidence was for false alarm responses on foil-other items. On the 2½-year post-test, the mean confidence ratings for hits and false alarms were 5.28 and 5.21, respectively. In comparison to decreasing accuracy (Figures 1, 2, & long-term post-test), the subjects remained fairly confident in their choices.

These findings show that recognition memory for autobiographical events is both strikingly accurate and inaccurate, when viewed from two different perspectives. Autobiographical memory is accurate in the special sense that people can correctly identify a great number of everyday events that actually occurred even after considerable periods of time. Indeed, if given an actual record from their past (as done here and by Linton), people are quite accurate at recognizing it as theirs. Accuracy of this

sort is intuitively consistent with many of our everyday recollections; we feel that those memories retrieved as "our life events" really happened as remembered. Our data support this intuition in that performance on original records was quite high overall and confidence in one's recognition of these items was also high and relatively stable across long time intervals. Thus, from the perspective of subjects' everyday experience, which focuses primarily on the correct acceptance of original memories, autobiographical memory is strikingly accurate.

However, from an experimental perspective, evaluation of memory accuracy cannot be based on data from original records alone; it requires the joint consideration of recognition performance on foil as well as original items. In fact, based on comparing both sorts of performances, autobiographical memory is accurate in that hits outweigh false alarms, but also clearly inaccurate because there is a strong bias toward identifying foil items as actual memories. Stated differently, if autobiographical memories were accurate in the sense of direct access to some stored memory trace we should have found few false recognitions of foils, as well as the reported high rate of correct responses to original records. Since the false alarm rate was so high, it is improbable that autobiographical memory is accurate in this sense. Note further that our subjects even had some difficulty identifying someone else's life events as not their own; the rejection of foil-other items never reached the same level of accuracy as the correct acceptance of originals.

The levels of false alarms in the present data are even more revealing of inaccuracy in everyday, autobiographical memories, given the fact that our subjects both knew at the time original records were collected that they would be tested on reported events and were instructed to record only reasonably "memorable" daily happenings (see also Thompson, 1982). These proce-

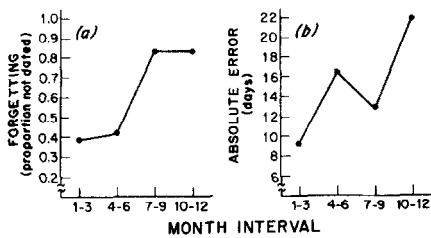


FIG. 3. Dating (a) and ordering (b) accuracy as a function of month interval.

dures should have served to increase recognition accuracy and may indeed be partly responsible for the high hit rate data. However, in spite of these procedures, it is still clear that memory for everyday autobiographical events is biased and inaccurate in the sense that people tend to incorrectly identify nonevents as being from their past.

This picture of accuracies and inaccuracies in recognition of autobiographical information also fits the results from the temporal ordering and dating tests. The proportion of item pairs correctly ordered (earlier versus later) when the two items represented delays of 1 month versus 3, 4 months versus 6, 8 months versus 10 and 9 months versus 11 were .73, .96, .67, and .72, respectively. Accuracy over all these tests of a 2-month comparison as above .50 and averaged .77 (smallest mean difference, $t(15) = 5.67$, $p < .001$).

Figure 3 presents the data on dating accuracy. Panel (a) shows the proportion of memories forgotten, that is, not dated or dated only in some broad time frame, for example, the season of the year the episode was recorded. These proportions are sizable, ranging from about .40 at the 1- to 3-month interval to over .80 at 10-12 months, and it is apparent that recall becomes more difficult with time. It can be seen in panel (b) that of the memories dated (60-20% as explained above), the absolute dating error increased from approximately 9 to 22 days in 12 months. This indicated that even if a record was dated, memory for the month and day the episode occurred declined as

the memory became more remote.² These data are consistent with comparable results reported by Baddeley, Lewis, and Nimmo-Smith (1978), Linton (1975), and Thompson (1982).

While ordering events separated by at least 1 month was reasonably accurate, dating the records, which required that the subjects recollect precise information about when some event occurred, was much more inaccurate. It is thus unlikely that memories of everyday life events are represented in a form isomorphic with the way in which they truly happened, including temporal and spatial features. Clearly many autobiographical memories are not strictly episodic in nature (Tulving, 1972).

Our findings, especially the particular mix of accuracy and inaccuracy uncovered, are conducive to thinking of autobiographical memories in schema-like terms. Although the data are consistent with other memory models (cf. Johnson & Raye, 1981), a schema view of autobiographical memory is appealing since everyday recollections may be integrated into personal knowledge structures that give meaning to seemingly isolated everyday occurrences. People know in general the sorts of events that have occurred in their life, even though most past episodes cannot be reproduced from memory in complete detail. Further, the term autobiography itself suggests a story-like schema. Similar to other story schemas, this schema would regulate the

² The ordering results from the long-term post-test indicated a modest decline in the accuracy of ordering two events from 1 ($X = .74$ for tests 1-4) to 2½ years ($X = .67$). This decrease probably reflects ordering errors attributable to the nearness of events separated in time of occurrence by less than a month in the latter but not the former tests. The dating test from the long-term follow-up showed that the absolute error in days for the records dated was 22.41 (range = 6-67) days. Thus, there was little increase in error for locating events precisely in time after about a year. This conclusion must be considered relative to the limited time span from which the events were selected and the fact that our subjects were no longer keeping daily records when tested.

encoding and retrieval of related material (cf. Alba & Hasher, 1983; Bower, Black, & Turner, 1979; Graesser, 1981; Spiro, 1977). While there is much change from day to day in most everyone's life and associated events and activities, there is also a corresponding invariance in one's sense or gist of their life. Here, gist at the most general level is, we believe, one's autobiography—the schematized story of one's life. What one remembers then is, at least in part, what could have happened or should have happened in one's life. People also, in part, remember actual past events and successfully monitor reality (Johnson & Raye, 1981; Neisser, 1976). Thus, as an autobiography, this sense of your life is also individualized and specialized according to the content and interpretation of your life events. From this point of view, the accuracy and inaccuracy of everyday memories as reported here is intimately related to reports that information consistent with one's prototypic self-concept is remembered more accurately than inconsistent information (Markus & Sentis, 1980; Rogers, 1980).

In line with an autobiography-schema interpretation of autobiographical memories, it is significant that comparing the accuracy versus confidence rating data from both the 1-year and long-term follow-up shows that while recognition performance declines over time, perceived certainty remains fairly stable. People are more confident about the exact nature of events occurring in their life than they should be. We believe that people claim confidence in autobiographical information, both in tests such as these and in simply recounting life events, because the events could have happened in "their life" and the gist of one's life remains fairly constant over a reasonable period of time. A stable perception of one's autobiography engenders a stable estimation of confidence in spite of the fact that memories become less accurate. At bottom, we hypothesize that the same au-

tobiographical schemata account for the correct recognition of actual autobiographical memories, the false recognition of certain nonevents as one's own memories, the accurate rejection of other nonevents, and an overconfidence in the "facts" of one's life.

Any strict test of this account of autobiographical memory is beyond the scope of our data; however, we can provide some further evidence consistent with it. We reasoned that people would falsely identify a foil item if it was typical of some abstracted representation of their own life events. If so, the probability of making a false alarm response should covary with the degree to which a foil item was similar or discrepant in general meaning, though not specific detail, to the original from which it was written. Degree of discrepancy between originals and foils is a subjective evaluation based on an assessment of whether or not the foil specifies essentially the same sort of experience conveyed in the original record. All of the foil-evaluation and foil-description items were paired with their respective original records. A total of 288 pairs were then judged by eight independent raters. Each pair was rated on a 7-point scale for the degree of "semantic discrepancy" between the foil and its original (1 = Not semantically different, 7 = Very semantically different). Semantic difference was explained to the raters as the extent to which the foil varied in truth value or meaning from the original. These ratings were averaged over judges for each subject and separated by item type. Point-biserial correlations were calculated between the mean ratings and the response outcomes (1 = "correct", 0 = "false alarm").

All the correlations were positive; the mean coefficient for foil-evaluation items was .51, for foil descriptions, .66. The greater the raters' estimated degree of semantic distance between foils and originals, the more likely the subjects were to correctly reject the foils as not their own mem-

ories. Similarly for both foil-evaluation and foil-description items the average semantic difference rating for false alarms (3.82 and 2.88, respectively) was less than that reported for correct rejections (4.39 and 3.65, respectively). On the one hand, these item analyses could have provided evidence against a schema-like account, but they did not; on the other hand, they are consistent with several interpretations, not only the one outlined here.

CONCLUSIONS

The view that autobiographical memory is largely the schematized reconstruction of previous life events is not new (e.g., Bartlett, 1932; Freud, 1914/1960). Our contribution to this claim is the evidence regarding the implicated inaccuracies that should occur in many autobiographical recollections from this perspective. Specifically, autobiographical memory ought to be inaccurate in certain predictable respects, and even ought to be generally inaccurate when considered in the context of what really happened or could have occurred. A critical estimate of accuracy in autobiographical memory has been difficult to achieve because the existing research findings are fragmentary; few studies have addressed this issue because of the difficulty collecting information about original events. Further, in those studies that have information about original events, memory tests appropriate to a comprehensive evaluation of accuracy have not been employed. Our study, in spite of its acknowledged limitations, approximates more closely than others conditions necessary for examining accuracy in autobiographical memory. Thus, it paints a needed, more balanced picture, capturing both the apparent accuracy of many autobiographical memories coupled with real inaccuracies and biases.

From our data it is clear that considering memory performance on original items alone, without comparison to foils, is misleading because people tend to inaccurately

accept nonevents which are consistently related to actual happenings. Neisser's (1981) description of "reepisodic" memory—remembering a supposed event but one in which the exact details are blended over repeatedly occurring and related episodes—would be one instance of this more general phenomenon. Therefore, memory functions of this sort showing little forgetting for personal, episodic events over long periods of time overestimate the amount of information accurately remembered (Linton, 1975, 1979, 1982).

REFERENCES

- ALBA, J. W., & HASHER, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203–231.
- BADDELEY, A. D., LEWIS, V., & NIMMO-SMITH, I. (1978). When did you last . . . ? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*. New York: Academic Press.
- BAHRICK, H. P., BAHRICK, P. O., & WITTLINGER, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*, 54–75.
- BANKS, W. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*, 81–99.
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. New York: Macmillan.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.
- BROWN, R., & KULIK, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–99.
- CROVITZ, H., & CORDONI, C. (1982). *Memory diaries for assessing real-world forgetting experiences*. Unpublished manuscript, Durham V.A. Hospital, Durham, NC, and Duke University.
- FIUZGERALD, J. M. (1980). Sampling autobiographical memory reports in adolescents. *Developmental Psychology*, *16*, 675–676.
- FREUD, S. (1905/1953). Three essays on the theory of sexuality. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 7). London: Hogarth Press.
- FREUD, S. (1914/1960). *The psychopathology of everyday life*. New York: Norton.
- GALTON, F. (1879). Psychometric experiments. *Brain*, *2*, 148–162.
- GALTON, F. (1911). *Inquiries into human faculty and its development*. New York: Dutton.
- GOLD, E., & NEISSER, U. (1980). Recollections of kin-

- dergarten. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 2, 77-80.
- GRAESSER, A. C. (1981). Prose comprehensive beyond the word. New York: Springer-Verlag.
- GREEN, D., & SWETS, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GRIER, J. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424-429.
- HALL, G. S. (1899). Note on early memories. *Pedagogical Seminary*, 6, 485-512.
- HENRI, V., & HENRI, C. (1895). On our earliest recollections of childhood. *Psychological Review*, 2, 215-216.
- HENRI, V., & HENRY, C. (1898). Earliest recollections. *Popular Science Monthly*, 53, 108-115.
- HERRMANN, D. J., & NEISSER, U. (1978). An inventory of everyday memory experiences. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*. New York/London: Academic Press.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- KUIPER, N. A., & DERRY, P. A. (1980). The self as a cognitive prototype: An application to person perception and depression. In N. Cantor & J. F. Kihlstrom (Eds.), *Personality, cognition and social interaction*. Hillsdale, NJ: Erlbaum.
- LINTON, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman.
- LINTON, M. (1979, August). *Cuing events in adult's and children's autobiographical memory*. Paper presented at the meeting of the American Psychological Association, New York.
- LINTON, M. (1982). Transformations of memory in everyday life. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts*. San Francisco: Freeman.
- LOCKSLEY, A., STANGOR, C., HEPBURN, C., GROSOVSKY, E., & HOCHSTRASSER, M. (1984). The ambiguity of recognition memory tests of schema theory. *Cognitive Psychology*, 16, 421-448.
- MANDLER, B. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252-271.
- MANDLER, J. M., & JOHNSON, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-151.
- MARKUS, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 35, 63-78.
- MARKUS, H., & SENTIS, K. (1980). The self in social information processing. In J. Suls (Ed.), *Social psychological perspectives on the self*. Hillsdale, NJ: Erlbaum.
- NEISSER, U. (1976). *Cognition and reality*. San Francisco: Freeman.
- NEISSER, U. (1981). John Dean's memory: A case study. *Cognition*, 9, 1-22.
- NEISSER, U. (1982). Snapshots or benchmarks? In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts*. San Francisco: Freeman.
- ROBINSON, J. A. (1976). Sampling autobiographical memory. *Cognitive Psychology*, 8, 578-595.
- ROGERS, T. B. (1980). A model of the self as an aspect of the human information processing system. In N. Cantor & J. H. Kihlstrom (Eds.), *Personality, cognition and social interaction*. Hillsdale, NJ: Erlbaum.
- RUBIN, D. (1982). On the retention function of autobiographical memory. *Journal of Verbal Learning and Verbal Behavior*, 21, 121-138.
- SPIRO, R. J. (1977). Remembering information from text: The "state of schema" approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum.
- THOMPSON, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, 10, 324-332.
- TULVING, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.
- WARRINGTON, E. K., & SANDERS, H. I. (1971). The fate of old memories. *Quarterly Journal of Experimental Psychology*, 23, 432-442.
- WHITE, S. H., & PILLEMER, D. B. (1979). Childhood amnesia and the development of a socially accessible memory system. In J. F. Kihlstrom & F. J. Evans (Eds.), *Functional disorders of memory*. Hillsdale, NJ: Erlbaum.
- WINOGRAD, E., & KILLINGER, W. A., JR. (1983). Relating age at encoding in early childhood to adult recall: Development of flashbulb memories. *Journal of Experimental Psychology: General*, 112, 413-422.

(Received April 10, 1985)

(Revision received August 12, 1985)