# Constancy of Clustering over the Course of Recall*

Devon D. Brewer

School of Social Sciences

University of California, Irvine

Paper presented at the 14th International Sunbelt Social Networks Conference in New Orleans, Louisiana, February 17 - 20, 1994.

*This research was supported by an Air Force Laboratory Graduate Fellowship awarded to the author by the U.S. Air Force Office of Scientific Research. I thank William H. Batchelder and A. Kimball Romney for their suggestions. I am grateful to Keiko Nakao for PSTAT, a program for measuring clustering. The nonparametric Monte Carlo clustering software is available from the author on request. Please direct all correspondence to: Devon Brewer, Social Development Research Group, School of Social Work, University of Washington, 146 N. Canal Street, Suite 211, Seattle, WA 98103.

Running head: CONSTANCY OF CLUSTERING OVER THE COURSE OF RECALL

## Abstract

In 1953, Bousfield showed that when subjects are presented with a randomized list of words from several different semantic domains (e.g., birds, occupations, and vegetables), they tend to cluster (mention adjacently) words in recall by domain. He also reported that subjects' tendency to cluster words by domain decreased over the course of recall to the point where recall was effectively random and not governed by semantic associations. This paper describes a re-examination of clustering over the course of recall. Data were analyzed from one study in which subjects recalled words from homogeneous semantic domains and from three studies in which subjects recalled the names of persons in socially bounded communities. The results indicate no decrease in subjects' clustering over the course of recall, demonstrating that an underlying cognitive structure influences associative patterns throughout the course of recall.

# Constancy of Clustering over the Course of Recall

In 1953, Bousfield documented the phenomenon of semantic clustering in the free recall of words from mixed category lists. He showed that subjects presented with randomized lists of words from several semantic domains (e.g., birds, occupations, and vegetables) tend to recall words from the same semantic domain successively in their recall protocols. This seminal article launched the systematic study of the organization of memory and is regarded as the classic demonstration of semantic clustering. However, in that article, Bousfield also reported another important result, namely that the degree of semantic clustering declined over the course of recall to the point where associations between consecutively recalled words appeared to be random. Except for replications by Bousfield and Cohen (1953) and Sakoda (1956), this result has been ignored by memory researchers in the intervening 40 years, despite its significant implications. This finding suggests that semantic structure and the organization of memory are not as clearly or directly related as memory researchers maintain (see, e.g., chapters in Puff, 1979).

In this paper, I follow up on an earlier suggestion by Romney, Brewer, and Batchelder (1993) to reexamine clustering over the course of recall using current methods for measuring semantic clustering. I describe analysis of Romney, et al.'s (1993) data on subjects' recall of words from homogeneous semantic domains as well as data from three studies on subjects' recall of the names of persons in socially bounded communities (Brewer, 1993a; Brewer & Yang, 1993; Brewer, 1993b). These studies include episodic and semantic memory recall tasks, and vary in terms of recall content and time allowed for recall.

Memory researchers apparently assume that clustering does not decrease over the course of recall. I have been unable to find a single reference in the published literature (aside from Bousfield and Cohen (1953) and Sakoda (1956)) to Bousfield's (1953) observation that clustering progressively diminishes over the course of recall.

However, there does seem to be some indirect evidence that clustering may not decline over the course of recall. Brewer and Yang (1993) and Brewer (1993b) studied how subjects in two socially bounded communities (a religious fellowship and a department in a formal organization) recalled the names of persons in those communities. In both cases, subjects' associative patterning was based on the social structure of the community, with subjects tending to mention adjacently persons who were socially close. In both studies, the inter-response times (log transformed) for adjacently recalled pairs of persons were moderately strongly correlated with proximity in the community's social structure. The strength of these relations did not decrease noticeably after output serial position was controlled for, indicating that associative patterning, as reflected in the inter-response times, did not decrease over the course of recall. On the basis of these results, Brewer (1993c) hypothesized that clustering in the recall of persons would remain constant over the course of recall. Since the recall of persons shows broad similarities to the recall of words (see Brewer, 1993a; Brewer & Yang, 1993; Brewer, 1993b), I extrapolate that semantic clustering in the recall of words might also hold steady throughout the course of recall.

The analysis of the Romney, et al. (1993) data is described first, followed by parallel reports for the three studies on the recall of persons. The results from all studies are considered in the discussion section.

## Study 1

### Method

Similarity judgments were collected for 17 lists of words from 11 domains (see Table 1). A different group of approximately 25 subjects (undergraduates at the University of California, Irvine) responded to each word list. Lists 1, 3, 5, 7, 9, 11, and 13 - 16 consisted of the 21 most typical items from Rosch's (1975) fruit, vegetable, furniture, vehicle, weapon, clothing, bird, toy, tool, and sport domains, respectively. Lists 2, 4, 6, 8, 10, and 12 were composed of 20 items (of varying typicality) from Rosch and Mervis' (1975) fruit, vegetable, furniture, vehicle, weapon, and clothing domains, respectively, plus the item from Rosch (1975) that had the greatest typicality per domain but was not already included in Rosch and Mervis' (1975) 20 items. List 17 consisted of the 21 most frequent fish terms from Battig and Montague's (1969) norms, excluding whale, rock, and shrimp.

Similarity judgments were collected on the 21 items for each word list with triads tests (Weller & Romney, 1988). Subjects were presented with sets of three items and asked to circle the item most different from the other two. Each triad test was individually randomized with the program ANTHROPAC (Borgatti, 1992) using a lambda-one, balanced, incomplete block design (Burton & Nerlove, 1976). This design produces 70 triadic comparisons for the test, with each pair of items occurring exactly once and each of the 21 items occurring 10 times. The triads test served the double purpose of measuring the judged similarity among items and presenting the items in a completely randomized fashion for the free-recall task.

Immediately after the triads tests were completed and collected from the subjects, and without previous warning, the subjects were asked to recall, in writing, as many of the items appearing on the triads test as possible within 75 seconds. These protocols constituted the free-recall data.

### Results

For each word list, the triads data were transformed into a 21-by-21 proximity matrix, with one unit of similarity scored for the two items in each triad that were not circled by a subject (Weller & Romney, 1988). A single aggregated proximity matrix was formed by adding the scores for all subjects. A three-dimensional spatial representation of the similarity of items for each word list was obtained through correspondence analysis (Gifi, 1990; Greenacre, 1984; Weller & Romney, 1990). This representation depicts the judged similarity structure of each word list in terms of euclidean distance, where more similar items are closer to each other than less similar items.

For measuring semantic clustering, I used the same basic procedures employed by Romney, et al. (1993), described here in brief. A subject's free-recall protocol may be conceptualized as a path through the spatial representation derived from the proximity matrix. A subject's observed path length may be defined as the sum of the euclidean distances between successively recalled items. Since subjects remember different and varying numbers of words (subjects in the study recalled on average a little more than half of the items; see Romney, et al., 1993, p. 29), the expected path lengths are computed separately for each subject. A method for computing the mean and variance of all possible path lengths for a given set of points in a euclidean distance matrix using quadratic assignment is elucidated in detail by Hubert (1987; Hubert & Levin, 1976; see also Carroll, Romney, Farner, & Delvac, 1976). Carroll, et al. (1976) and Romney, et al. (1993) both demonstrated that random path lengths among items in three-dimensional euclidean semantic space were approximately normally distributed.

Semantic clustering is indexed by calculating a *z* score for each observed path length for the set of words recalled by that subject. Negative *z* scores indicate clustering based on semantic similarity, while positive *z* scores indicate paths longer than chance. Romney, et al. (1993) provided additional details and illustrations of this clustering measure.

To allow testing whether the level of clustering changed over the course of recall, I divided each subject's recall path into two halves. These half-paths were treated *as if* they were separate paths from different subjects and were submitted to the analysis procedures just described. When the number of links ([number of words recalled, including repetitions] - 1) in a subject's path was even, I created two half-paths, splitting evenly. For instance, if a subject recalled five items (represented here by letters) in the order A - B - C - D - E, the whole path was divided into the A - B - C and C - D - E halves. When the number of links in a subject's whole recall path was odd, the second half-path contained one more link than the first half-path. The observed length of each half-path, then, was measured according to the euclidean distances between successive items recalled in that half. Similarly, the expected length of a half-path was computed on the basis of the euclidean distances among only those items in that half-path. When a subject's half-path included repetitions (which were very rare), the subject's observed half-path length was reduced by the number of repetitions times the mean observed distance between two successively recalled items in the subject's half-path.

Thus, the analysis yielded two semantic clustering *z* scores for each subject, one based on the first half of the subject's recall path and the other based on the second half of the subject's recall path. Table 1 presents the mean semantic clustering *z* scores for subjects' first and second half-paths for each word list. Consistent with Romney, et al.'s (1993) observations of semantic clustering in subjects' whole recall sequences, the results in the table show that subjects also displayed a tendency in both halves of their recall sequences to list semantically similar words adjacently. Since all mean *z* scores were negative, the absolute values of the *z* scores are reported in the table. The *z* scores shown in Table 1 are somewhat smaller in magnitude than those reported by Romney, et al. (1993) because they are based on subjects' half-paths which contain 50% fewer items than the whole paths. This produces, in effect, a substantial reduction in the degrees of freedom for the *z* scores. Hence, the lower *z* scores for the half-paths in comparison to the whole paths do not necessarily indicate a decrease in the actual strength of semantic clustering. The semantic clustering *z* scores for an individual subject's two half-paths, though, may be compared directly since they are based on the same (or very nearly the same) number of items.

To assess whether the degree of clustering varied over the course of recall, I carried out matched pair *t*-tests between subjects' first and second half semantic clustering *z* scores for each word list. No detectable pattern emerged, with subjects in some word lists tending to cluster a little more in the first half of their recall sequences and subjects in other word lists tending to cluster a little more in the second half of their recall sequences. The cumulative *Z* score based on aggregating the separate word lists' *t* values (Rosenthal, 1991; Winer, 1971) was 0.91, which indicates no significant difference in the level of clustering between first and second halves of recall sequences across word lists.

I conducted parallel analyses on sets of simulated recall paths. These simulated recall paths were generated according to the model of semantic clustering described in detail by Romney, et al. (1993). This two-stage model of clustering first determines which set of items will be "recalled" in a simulated path according to the marginal probability estimated by the observed proportion of subjects who recalled an item. Then the first item output in a simulated path is selected according to a weighted sampling rule based items' observed frequency of mention (to mimic the empirical finding that the first item recalled tends to have a high overall recall probability). The output recall sequence of subsequent items in a simulated path is determined probabilistically according to the following rule. The probability of selecting any given item next is its reciprocal semantic distance (in the 3 dimensional spatial representation obtained from correspondence analysis of subjects' similarity judgments) from the previous item divided by the sum of reciprocal distances from the previous item to all remaining items. This recall process proceeds to sample the remaining items, without replacement, until all items in the simulated path's recall set (as determined above) have been exhausted.

For each word list, the number of model-based simulated paths generated was equal to the number of subjects who had responded for that list. In the same manner as described earlier for the observed paths, I created half-paths from the whole simulated paths, computed the simulated half-paths' semantic clustering $z$ scores, and compared $z$ scores for the first and second halves of simulated paths. The simulated half-paths exhibited roughly comparable levels of clustering to subjects' observed half-paths (see Table 1). More importantly, the simulated paths displayed the same overall constancy of clustering over the course of recall that was found in subjects' recalls. (Analysis of other 12 word lists is in progress).

## Study 2

### Method

Subjects. Subjects were 15 graduate students from an interdisciplinary graduate program at the University of California, Irvine, including 12 females and 3 males, and students from each of the five most recent cohorts (from first to fifth year students) (see Brewer, 1993a for a complete description of this study). Aside from the absence of very senior students, this sample appeared to be quite representative in terms of students' demographic characteristics, academic interests, and office locations.

Procedure. Subjects were interviewed individually. The interview was administered via microcomputer with a standard keyboard layout. In the interview, the first screen presented to a subject consisted of an introduction to the experiment and instructions for how to respond on the keyboard. I then reviewed these instructions with subject orally. The next screen showed the question, "Who are all the graduate students in [name of the program] program? (You can use their first and/or last names and you do not need to put your name as a response)." No instructions were given regarding the order in which subjects were to list names. Several lines below the question, the cursor would be blinking and subjects could type names which would be displayed on the screen. After typing a name, subjects were to hit the Enter key, which would cause the name to disappear from the screen and the blinking cursor to return to its original position below the question. The subjects could type in additional names, each time hitting the Enter key after typing a person's name. Subjects were given up to 10 minutes to recall as many persons as they could, but all subjects finished with 7 minutes (mean time used by subjects = 3 minutes, 22 seconds), saying they could remember no more names.

### Results

On average, subjects recalled approximately half of the 41 students in the program. Brewer (1993a) demonstrated that subjects strongly clustered persons in recall by cohort, tending to mention consecutively persons in the same cohort.

Exactly as in Study 1, I created half-paths from each subject's whole recall sequence. The lengths of these half-paths were measured according to a binary persons-by-person matrix representing the program's cohort structure (1 = pair of persons in same cohort, 0 otherwise). Brewer (1993a) noted that the distribution of random paths among the set of persons recalled by a subject were not approximately normal. To measure clustering by cohort, then, Brewer (1993a) employed a nonparametric approach related to the method used by Romney, et al. (1993), and I used this approach in my analysis of clustering for subjects' half-paths. This technique consists of generating a large sample of random permutations of a subject's recall sequence (10,000 for the analyses reported in this paper for Studies 2-4) and noting the proportion of these random (half-)paths that were as or more clustered than the subject's observed (half-)path. This proportion is a one-tailed, nonparametric Monte Carlo $p$-value indicating the significance of clustering. (The clustering schemes for Studies 2-4 differ from those used in Study 1, because in Study 1 the clustering scheme was measured in distances, while the clustering schemes used in Studies 2-4 were measured in proximities). The same logic of comparing subjects' first and second half clustering $z$ scores in Study 1 holds for comparing subjects' first and second half clustering $p$-values in Studies 2-4. In addition, I calculated ARC scores (Roenker, Thompson, & Brown, 1971) for subjects' half-paths. This measure equals $(o - e)/(m - e)$, where $o$ is observed (half-)path length, $e$ is the expected (half-)path length, and $m$ is the maximum possible (half-)path length.

The median clustering $p$-value for subjects' first and second half-paths were .0604 and .0868, respectively. Hence, subjects' half-paths also clearly displayed clustering by cohort. Eight of the 15 subjects clustered more in the first half of their recall sequences than in the second (sign test $p$ = .78). Furthermore, the distributions of subjects' first half and second half clustering $p$-values were very similar. The minimum and maximum clustering $p$-values for subjects' first halves were .0005 and .9228, respectively, and for subjects' second halves were .0007 and 1.0, respectively. For subjects' first halves of their recall sequences, the fourth lowest clustering $p$-value was .0029 and the fourth highest clustering $p$-value was .3906. For subjects' second halves of their recall sequences, the fourth lowest clustering $p$-value was .0331 and the fourth highest clustering $p$-value was .4391. Moreover, a matched pair $t$-test for subjects' first and second half ARC scores showed no significant difference in clustering levels ($t(14) = 1.17$, $p$ = .26). (The ARC scores did not discriminate subjects' levels of clustering as well as Monte Carlo $p$-values, since a few subjects' first and second half ARC scores were tied even though their Monte Carlo $p$-values were noticeably different). In summary, subjects in Study 2 showed no significant changes in the level of clustering over the course of recall[1].

Study 3

Method

Subjects. Subjects were 25 college-aged members of a church-affiliated Christian fellowship of Taiwanese and Taiwanese-American young people in southern California, including 11 females and 14 males. (See Brewer & Yang, 1993, for a detailed description of this study).

Procedure. Subjects performed two tasks: a recall task and a pile sort of persons by social proximity. All 25 subjects did the recall task, while only 11 (5 females and 6 males) did the pile sort task. For both tasks, subjects were interviewed individually by Yang (who was a member of the fellowship), usually after fellowship meetings or church services in a private room or a quiet setting out of sight and earshot of other fellowship members. For most subjects who performed both tasks, there was a 2-3 week interval between the recall and pile sort tasks; for a few subjects, both tasks were performed during the same interview, with the pile sort task always following the recall task. Except for the instructions immediately preceding a task, subjects were not given any prior indication about the number or specific nature of the tasks to be performed. Subjects also were asked not to discuss the study with other fellowship members until data collection was finished.

Yang gave the following instructions orally and bilingually to subjects for the recall task:

> Who are all the people involved with the [name of the fellowship]? In giving your answers, please try to give first and last names, or as much of the person's name as possible. You do not need to mention your name or my name. List aloud the names of as many people involved in the [name of the fellowship] that you can think of.

No instructions were given regarding the order in which subjects were to list names. Subjects were given 10 minutes to mention persons (all subjects finished within 9 minutes, and the mean amount of time used by subjects was 2 minutes, 25 seconds) and their responses were recorded on audiotape.

After 20 subjects had done the recall task, the full name (or as much as was known) of each different person mentioned in the recall interviews was written on a separate 3" x 5" notecard in both English letters and Chinese characters. Cards with the names of the few persons first mentioned in the last five recall interviews were added to the set as they became available. The instructions for the unconstrained single pile sort by social proximity followed in large part those used by Freeman, Freeman, & Michaelson (1988). Subjects were first asked to separate out from the set of randomly shuffled cards those persons whom they did not recognize, i.e., could not match the name with a face. Then, subjects were instructed to sort the cards into piles of persons who tended to like, interact with, and hang around each other, both at fellowship meetings and elsewhere (see Weller & Romney, 1988 for other details on the single pile sort). Subjects' responses to this task constituted their <u>perceptions</u> of the social proximities among persons in the fellowship--i.e., perceptions of the fellowship's social network. Individuals' reports of social proximity in pile sort tasks have been shown to be highly accurate with respect to observed interaction patterns (Freeman, et al., 1988; Webster, 1992). For brevity, perceived social proximity will be referred to here simply as "social proximity."

Results

The mean number of persons recalled by subjects was 30, out of approximately 100 persons who had attended the fellowship in the year prior to data collection. Brewer and Yang (1993) demonstrated that subjects clustered persons in recall according to social proximity, and that other factors, such as persons' gender, kinship relations, or membership in high school or college-aged sections of the fellowship, could not account for this social proximity clustering.

As in Studies 1 and 2, I created two half-paths from each subject's whole recall sequence. One subject's recall sequence was not split into halves since this subject only recalled 5 persons, and therefore this subject's recall was excluded from analysis. I measured subjects' half-paths against a social proximity matrix aggregated from individual subjects' pile sort responses. Following Freeman, et al. (1988), the cells in this matrix contained proportions referring to the number of subjects who placed a pair of persons in the same social proximity pile in the pile sort task divided by the number of subjects who recognized both persons. Social proximity data were only available for 99 of the 105 persons recalled by subjects, since a few persons first mentioned in the last five recall interviews were inadvertently not included in the last few pile sort interviews. These six persons were subsequently omitted from subjects' half-paths (a total of 7 omissions). The same nonparametric Monte Carlo method used in Study 2 was used here for measuring social proximity clustering.

The median clustering $p$-values for subjects' first and second halves of their recall sequences were .0096 and .0103, respectively, indicating highly significant clustering by social proximity over the course of recall. Fourteen of 22 subjects clustered more during the first half than in the second half (sign test $p = .29$). The level of social proximity clustering for two subjects' first and second half-paths could not be compared since their $p$-values were all $<.0001$. In addition, the distributions of first and second half clustering $p$-values were similar. For subjects' first half-paths, the minimum $p$ was $<.0001$ and the maximum was .3984. For subjects' second half-paths, the minimum $p$ was $<.0001$ and the maximum was .7023. The sixth highest $p$'s for subjects' first and second half-paths were .0847 and .1224, respectively. The sixth lowest $p$'s for subjects' first and second half-paths were $<.0001$ and .0001, respectively. These results show, then, that the level of clustering did not vary significantly over the course of recall.

Study 4

Method

Subjects. Subjects were 13 employees (including 11 females and 2 males) of a department in the public affairs division of a research university in the southwestern U.S. (see Brewer, 1993b for a detailed description of this study). Subjects' mean age was 35.5 years (range: 19 to 55 years) and had worked in the department for a mean of 4.4 years (range: 3 months to 11 years, 6 months). Individuals from each departmental status level and main departmental function were represented in this sample.

Procedure. Ten subjects participated in two interviews, and 3 subjects participated in one interview. All interviews were conducted individually and privately in a vacant office in the department, except for one interview which was carried out in an office in another building on the university's campus.

The first interview (for the 10 subjects who were interviewed twice) consisted of a free recall task. I gave the following instructions orally to subjects for the free recall task:

> Who are all the people who work in the [department's name] Department? Please list aloud the names of all the people who work in the [department's name] Department. You do not need to mention your name.

No instructions were given regarding the order in which subjects were to list names and subjects were allowed as much time as needed to mention all the persons they could. When subjects appeared to be done or said they had listed everyone, I prompted them once by asking if there were any other persons in the department. Subjects' responses were recorded on audiotape. In all interviews, subjects were not given any prior indication about the number or specific nature of the tasks to be performed except for the instructions immediately preceding a task. Subjects were asked not to discuss the study with other departmental employees until data collection was finished.

The second interview (for those 10 subjects who were interviewed twice) occurred 2-3 weeks after the first interview. The second interview began with a recall task. Five subjects were assigned to a free recall task (as in the first interview) and 5 were assigned to an alphabetically directed recall task (see Brewer, 1993b for details on this assignment process). For the alphabetically directed recall task, I gave the same oral instructions as in the first interview, except for the second sentence, which was replaced with: "Please list aloud the names of all the people who work in the [department's name] Department in alphabetical order by their first names as best as you can."

After the recall task in the second interview, subjects performed two quasi-successive pile sort tasks (cf. Boster, 1987; Freeman, et al., 1988). The full name (or as much as was known) of each different person mentioned by subjects in the first interview was written on a separate 3" x 5" notecard. (No additional persons were mentioned in the second interview). Subjects sorted persons for two different social relations: how closely persons worked with one another (work proximity) and how much persons socialized with one another (socializing proximity). The order in which subjects performed the pile sort tasks was balanced across subjects. For each pile sort task, subjects were first asked to separate out from the set of randomly shuffled cards those persons whom they did not recognize, i.e., could not match the name with a face. For the work proximity pile sort task, subjects were instructed to:

> Sort these persons into different piles according to how much they work with each other on job-related activities. Put persons that work with one another into the same pile.

For the socializing proximity pile sort task, subjects were instructed to:

> Sort these persons into different piles according to how much they socialize with each other, such as going to lunch together, meeting outside of work after hours, and/or talking with each other about things unrelated to work or the [department's name] department. Put persons that socialize with one another into the same pile.

After the initial sort, a subject was asked to loosen her/his criterion for working together (socializing) and, if possible, join piles of persons into larger groupings on the basis of working together (socializing). This step was repeated with further loosening of the subject's criterion until the subject did not perceive larger groupings (other than the whole department as one pile). At this point, the cards were rearranged into the piles the subject made in the initial sort. Then the subject was asked to tighten her/his criterion for working together (socializing) and, if possible, split piles of persons into smaller groupings of persons who worked (socialized) more intensely with each other. This step was repeated until the subject did not perceive finer groupings (other than each person as a single pile).

Similar to Study 3, subjects' responses to these tasks constituted their perceptions of work and socializing proximities among persons in the department--i.e., perceptions of the department's work and socializing networks. The 3 subjects who participated in only one interview performed the free recall task and the pile sort tasks in the same session. The analysis of these subjects' recalls is presented with the other subjects' first interview recalls.

Results

Subjects recalled a mean of 16 persons (out of 21 persons then employed in the department) in both interviews. On average, subjects took approximately one minute to recall all the persons they could, although alphabetically directed subjects in the second interview took almost 2 minutes on average (all subjects in both interviews finished in less than 3 minutes). Brewer (1993b) showed that subjects clustered persons in recall according to work proximity, and that other factors, such as socializing proximity, gender, departmental status, and office location, could not explain this work proximity clustering.

Alphabetically directed subjects' associative patterns, too, were best described by work proximity.

As in Studies 1-3, I created two half-paths from each subject's whole recall sequence for the first and second interviews. Three subjects' recalls from each interview were not analyzed because their recalls were clearly locationally oriented (i.e., these subjects mentally "walked" around the department's office, listing persons as their offices were "encountered"). I measured subjects' half-paths against a work proximity matrix aggregated from individual subjects' pile sort responses. To construct this matrix, I ordered the groupings of persons sorted by a subject into levels from broadest (where the subject could not join any more piles) to narrowest (where the subject could not split any pile further). The work proximity of a pair of persons from the perspective of each subject was indexed by a proportion representing the number of levels the pair was placed in the same pile divided by the total number of levels that subject used in the task. The work proximity values for each pair of persons were averaged across all subjects who recognized both persons in that pair to arrive at the aggregated matrix. The same nonparametric Monte Carlo method used in Studies 2 & 3 was used here for measuring work proximity clustering.

For subjects' first interview recalls, the median clustering $p$-values for subjects' first and second halves of their recall sequences were .0257 and .0674, respectively, signifying significant clustering by work proximity in both halves. Eight of 10 subjects clustered more during the first half than in the second half (sign test $p = .11$). Additionally, the distribution of first and second half clustering $p$-values were similar. For subjects' first half-paths, the minimum $p$ was .0058 and the maximum was .4071. For subjects' second half-paths, the minimum $p$ was .0028 and the maximum was .2759. These results suggest, then, that the level of clustering did not change significantly over the course of recall.

The results from subjects' second interview recalls (including both free recall and alphabetically directed subjects' recalls) reinforce this finding. For subjects' second interview recalls, the median clustering $p$-values for subjects' first and second halves of their recall sequences were .0324 and .0252, respectively. Only 1 of 7 subjects clustered more during the first half than in the second half (sign test $p = .13$). For subjects' first half-paths, the minimum $p$ was .0052 and the maximum was .7148. For subjects' second half-paths, the minimum $p$ was .0002 and the maximum was .2997.

When the levels of clustering between the first and second halves of subjects' recall sequences were compared for all subjects in Studies 2-4 as a set (excluding second interview recalls for subjects in Study 4), 30 of 47 subjects clustered more in the first half than in the second half (normal approximation to the binomial, $z = 1.75$, two-tailed $p = .08$).

## Discussion

The findings from these four studies reveal no evidence for a decrease in clustering over the course of recall. On the contrary, subjects displayed a fairly constant level of clustering as recall progressed. The model-based simulated recalls in Study 1 mirrored subjects' observed stability of clustering over the course of recall, which increases confidence both in the empirical result as well as the Romney, et al. (1993) model of semantic clustering. The results from studies 2-4 demonstrate the generality of the constancy of clustering effect and confirm Brewer's (1993c) expectations. Thus, it appears that recall organization reflects an underlying cognitive structure during all stages of recall, and not just the earlier phases, as argued by Bousfield (1953).

It is not entirely clear why Bousfield (1953) found an apparent decline in clustering over the course of recall. One possible explanation is that his use of the ratio of repetition to measure clustering did not incorporate (at the time) an expected value of clustering for a given segment of a subject's recall. If *expected* levels of clustering for his subjects decreased over the course of recall, then observed levels of clustering, as measured by the ratio of repetition, might *appear* to diminish as well. To confirm the constancy observed here, further research in this area should examine clustering over the course of recall in episodic memory recall tasks with mixed category lists of words, as in Bousfield's (1953) original investigation.

References

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monographs, 80(3, Pt. 2).

Borgatti, S. P. (1992). ANTHROPAC 4.0. Columbia, SC: Analytic Technologies.

Boster, J. S. (1987). Agreement between biological classification systems is not dependent on cultural transmission. American Anthropologist, 89, 914-919.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 49, 229-249.

Bousfield, W. A., & Cohen, B. H. (1953). The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. Journal of Psychology, 36, 67-81.

Brewer, D. D. (1993a). Patterns in the recall of persons in a student community. Social Networks, 15, 335-359.

Brewer, D. D. (1993b). Patterns in the recall of persons in a department of a formal organization. Paper presented at the West Coast Conference for Small Group Research, UCLA, Los Angeles, CA, April.

Brewer, D. D. (1993c). The social structural basis of the organization of persons in memory. Paper presented at the 3rd European Conference on Social Network Analysis, Munich, Germany, June.

Brewer, D. D., & Yang, B. L. (1993). Patterns in the recall of persons in a religious community. Paper presented at the 13th International Sunbelt Social Network Conference, Tampa, FL, February.

Burton, M. L., & Nerlove, S. B. (1976). Balanced designs for triads tests: Two examples from English. Social Science Research, 5, 247-267.

Carroll, J. D., Romney, A. K., Farner, C., & Delvac, W. F. (1976). The distribution of a path length statistic with illustrative applications (Social Science Working Paper 92). Irvine, CA: University of California.

Freeman, L. C., Freeman, S. C., & Michaelson, A. G. (1988). On human social intelligence. Journal of Social and Biological Structures, 11, 415-425.

Gifi, A. (1990). Nonlinear multivariate analysis. Chichester, England: Wiley.

Greenacre, M. J. (1984). Theory and application of correspondence analysis. New York: Academic Press.

Hubert, L. J. (1987). Assignment methods in combinatorial data analysis. New York: Marcel Dekker.

Hubert, L. J., & Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. Psychological Bulletin, 83, 1072-1080.

Puff, C. R. (Ed.). (1979). Memory organization and structure. New York: Academic Press.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. Psychological Bulletin, 76, 45-48.
Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. Psychological Science, 4, 28-34.
Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.
Rosenthal, R. (1991). Meta-analytic procedures for social research, revised edition. Newbury Park, CA: Sage.
Sakoda, J. M. (1956). Individual differences in correlation between clustering and recall of meaningful words. Journal of General Psychology, 54, 183-190.
Webster, C. M. (1992). Data type: A comparison of cognitive and observational measures. Paper presented at the International Sunbelt Social Network Conference, San Diego, CA, February.
Weller, S. C., & Romney, A. K. (1988). Systematic data collection. Newbury Park, CA: Sage.
Weller, S. C., & Romney, A. K. (1990). Metric scaling: Correspondence analysis. Newbury Park, CA: Sage.
Winer, B. J. (1971). Statistical principles in experimental design, second edition. New York: McGraw-Hill.

## Notes

1. Incidentally, subjects' inter-response times (IRTs) displayed a cohort effect throughout the course of recall, similar to results reported by Brewer and Yang (1993) and Brewer (1993b). For 11 of 15 subjects, both raw and natural log-transformed IRTs (controlling for output serial position) were shorter on average for pairs of adjacently recalled persons who were in the same cohort than for pairs in different cohorts. The cumulative $Z$ scores from individual subjects' $t$ values comparing same vs. different cohort IRTs (controlling for output serial position) were -1.67 and -2.10 for subjects' raw and log-transformed IRTs, respectively.

Table 1. Clustering in the first and second halves of observed and simulated recall sequences in Study 1

| | Observed | | | Simulated | | |
|---|---|---|---|---|---|---|
| | 1st half | 2nd half | | 1st half | 2nd half | |
| Word list (n) | Mean *z* | Mean *z* | *t* | Mean *z* | Mean *z* | *t* |
| 1. fruit1 (26) | 1.16 | 0.57 | -1.94 | 1.38 | 0.71 | -2.14 |
| 2. fruit2 (25) | 0.62 | 0.59 | -0.08 | 1.09 | 0.97 | -0.45 |
| 3. vegetable1 (24) | 0.61 | 0.70 | 0.34 | 0.37 | 0.81 | 1.50 |
| 4. vegetable2 (28) | 0.26 | 0.76 | 1.69 | 1.15 | 0.65 | -1.97 |
| 5. vehicle2 (22) | 0.15 | 0.82 | 2.12 | 0.48 | 0.80 | 1.11 |
| Cumulative *Z* | | | 0.91 | | | -0.83 |

Note: All mean *z* scores reported as absolute values.