Tulving, E., & Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1963, *1*, 321–334.
Tulving, E., & Colotla, V. Free recall of trilingual lists. *Cognitive Psychology*, 1970, *1*, 86–98.
Warrington, E. K., & Shallice, T. The selective impairment of auditory verbal short-term memory. *Brain*, 1969, *92*, 885–896.
Watkins, M. J. The concept and measurement of primary memory. *Psychological Bulletin*, 1974, *81*, 695–711.
Watkins, M. J., & Watkins, O. C. Processing of recency items for free recall. *Journal of Experimental Psychology*, 1974, *102*, 488–493.
Watkins, M. J., Watkins, O. C., & Crowder, R. G. The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 430–447.
Watkins, O. C., & Watkins, M. J. The modality effect and echoic persistence. *Journal of Experimental Psychology: General*, 1980, *109*, 251–278.
Waugh, N, C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, *72*, 89–104.
Wickelgren, W. Size of rehearsal group and short-term memory. *Journal of Experimental Psychology*, 1964, *68*, 413–419.
Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, *70*, 102–108.
Wickelgren, W. A. Auditory or articulatory coding in verbal short-term memory. *Psychological Review*, 1969, *76*, 232–235.
Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, *77*, 1–15.
Wickens, D. D., Born, D. G., & Allen, C. K. Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 440–445.
Woodward, A. E., Jr., Bjork, R. A., & Jongeward, R. H., Jr. Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 608–617.

CHAPTER 4

MARTIN D. MURPHY
C. RICHARD PUFF

Free Recall: Basic Methodology and Analyses

INTRODUCTION

The basic procedure of the typical free-recall experiment is probably easier to grasp than that of any other task normally used in the study of human memory and cognition. In its prototypical form, a list of words is presented to subjects who are instructed that, following the presentation, they will be asked to reproduce the items in any order that they choose.

One appeal of the method stems from the ability to perform analyses of input–output discrepancies, using the evidence of transformations imposed by the subject as a basis for making inferences about the nature of intervening mental processes. To increase our understanding of the nature of the intervening transformational, or organizational, processes, the influence of a great variety of input manipulations on the output has been studied. Furthermore, several types of output analyses have been developed in order to specify the nature of mental transformations.

Two of the major organizational phenomena investigated in free-recall studies are referred to as *clustering* and *subjective organization*. Taken together, these two phenomena have been the object of the great preponderance of free-recall research over the past 25 years. A brief introduction to the defining operations for these phenomena is necessary for understanding much of the methodological effort in this area.

HANDBOOK OF RESEARCH METHODS IN HUMAN MEMORY AND COGNITION

ISBN 0-12-566760-4

The standard operations for the study of clustering follow from the work of Bousfield (1953). The typical study of clustering begins with the investigator selecting stimulus materials that comprise items from a number of different taxonomic categories, such as *animals, vegetables, pieces of furniture,* or *occupations*. Items from the different categories are scrambled into a randomized sequence for presentation to the subjects who are given the usual instructions that they can recall the items in any order. The sequence of items recalled by each subject is then scored to determine the extent to which he or she has recalled items from the same categories together in runs, or clusters. The discrepancy between the haphazardly arranged input and the categorically organized output is taken as strong evidence that the organizing process of clustering has intervened.

The basic defining operations for the phenomenon of subjective organization follow primarily from Tulving's (1962) study. In this case, the experimenter usually begins by selecting a list of items that are operationally unrelated in the sense that they do not come from the same categories and do not elicit each other in free association. Multiple randomizations of the list are prepared for presentation over a series of free-recall trials. The subject's recall protocols are scored in successive pairs (i.e., Trials 1 and 2, 2 and 3, etc.) to determine the extent to which items recalled in adjacent positions on one trial are again recalled in adjacent positions on the next trial. The inference of an intervening organizing process is made when the subject draws together items that were presented in haphazardly varying positions over trials in the input list and repeatedly recalls them in adjacent positions in the output list. The repeated adjacency of the items in recall allows the inference that they reflect an organized, or integrated, unit. The phenomenon is referred to as subjective organization because the experimenter, having created a list of operationally unrelated items, is normally quite unaware of the basis on which the subject has formed the unit; all the experimenter may know is that the subject has organized consistently on some basis.

Because of great interest in subjective organization and clustering, much of the methodology in the free-recall area has involved (*a*) the selection of materials to be explicitly related or unrelated, (*b*) the manipulation of parameters of the situation to influence the opportunity for the subject to detect and use the structure built into the list or to impose his or her own structure, and (*c*) the exploration of alternative techniques for assessing and specifying the extent to which the output reflects the operation of several different organizing and strategic activities.

In the remainder of this chapter, we will review some of the important methodological considerations involved in the collection of free-recall data, as well as a few of the more traditional ways of measuring the characteristics of the output. We will pay special attention to some of the particular problems associated with life-span developmental research in free recall, because there are a number of those, and because this is one of the most active areas of free-recall research today. This chapter is prefatory to the next one by Pellegrino and Hubert. Their chapter picks up where this one leaves off, in the sense that it presents a new systematic approach to the analysis of many facets of organization and structure.

MANIPULATION AND CONTROL OF TASK PARAMETERS

Many of the potentially important task parameters that need to be taken into account in free-recall research are discussed in this section in order to help with the kinds of decisions just mentioned. This brief analysis cannot cover all of the relevant variables or their interactions. A few explicit suggestions will be offered, but the major goal of this section is to provide some background for thinking about what levels of what variables are appropriate for what purposes in free-recall studies. This kind of framework should presumably generalize to other variables and combinations of variables.

Input Parameters

LIST LENGTH

Because secondary organizational processes, such as clustering or subjective organization, are frequently of interest, list length is usually chosen to be well above the immediate memory span of approximately seven items. Otherwise, perfect recall in the order of presentation may result, leaving no room for such organizational effects.

The finding that recall level dramatically increases with age in children (Cole, Frankel, & Sharp, 1971) has led to some concern in developmental studies that the presentation of the same number of items may represent subjectively different tasks for subjects of different ages. One solution has been to present more items to the older children (e.g., Moely, Olson, Halwes, & Flavell, 1969). However, whereas the concern about subjective difficulty is reasonable, the problem is probably most severe with the short lists commonly used in serial-recall studies. Generally, free-recall studies employing the same number of items across

ages have replicated those in which number of items and age covary (Neimark, Slotnick, & Ulrich, 1971).

ITEM CHARACTERISTICS

Selection of items for free-recall studies involves decisions about a number of normative characteristics, including intraitem properties such as frequency, meaningfulness, and imagery values, as well as interitem association, or relatedness.

The major principle in selecting items for studies of subjective organization is to avoid interitem associations and categorical relationships (Tulving, 1962; Bousfield, Puff, & Cowan, 1964). The items are selected to be operationally unrelated, so that the subjects can impose their own organizational schemes rather than adopting some salient structure built into the list by the experimenter.

For studies of clustering, on the other hand, some basis for organization is intentionally built into the list by choosing groups of two or more items that are related in some way. Both the type and the strength of the relatedness are important factors. Most often, the items are chosen to represent conceptual, or taxonomic, categories such as *animals, pieces of furniture,* or *occupations* (e.g., Bousfield, 1953). However, interitem relatedness can also be defined on the basis of direct association strength (Jenkins, Mink, & Russell, 1958), physical characteristics, such as shape (Frost, 1971), or functional relationships, such as *foot–sock* (Denney & Ziobrowski, 1972). In addition, the relatedness, or categorization, can be induced situationally on the basis of spatial locations (Stukuls, 1975).

The strength with which the instances represent the categories is also an important parameter of recall and clustering, as demonstrated, for example, by Bousfield, Cohen, and Whitmarsh (1958). With taxonomic categories, the extent to which the instances represent the category can be readily specified through the use of a set of category norms such as those constructed by Battig and Montague (1969) for college students, Posnansky (1978) for children of different ages, and Howard (1980) for adults of different ages. Work on exemplar typicality can also serve as a basis for item selection.

Clearly, if one is interested in getting a good look at the operation of organizing processes, it makes little sense to select items that are only weakly representative of the categories. However, there are also problems associated with picking the items that are the highest in strength or typicality. Such items can be guessed by simply remembering the category name, and they are often directly associated with each other. Such problems have led Lange (1978) to argue that clustering of such highly related items might not reflect categorical organization at all,

especially with young children. Differences in the normative properties of the materials may thus be responsible for discrepant findings about the age of occurrence of significant clustering in young children. A reasonable solution in many situations is to choose items that are moderately related to the category name and minimally associated with each other. If high strength items need to be used, some from each category can be included in the list and other high strength exemplars omitted. Guessing rates can then be estimated from the frequency of intrusion in recall of the items that were not in the list.

An item selection issue of increasing concern in developmental and cross-cultural studies involves possible differences among subject populations in both the preferred bases for organizing and the within-category organization of exemplars. For example, investigators such as Nelson (1977) have discussed the possibility that young children may prefer functional rather than taxonomic organizations. Chi (1978) has strongly argued for the importance of semantic knowledge in accounting for developmental differences. Furthermore, Myers and Perlmutter (1978) have found developmental differences in knowledge structures among preschool children. An outcome of this work should be a better description of developing category knowledge and more useful theoretical accounts of the development of knowledge and how it may interact with episodic free recall.

NUMBER OF CATEGORIES AND ITEMS PER CATEGORY

The impact of variations in the number of categories and the number of items per category has not been the subject of much systematic investigation. Some investigators have shown that recall and clustering vary as an inverted-U function of the number of categories (or number of items per category) with list length held constant (e.g., Murphy, 1979). With a small number of large categories, little can be learned about differential category accessibility; with many small categories, a high proportion of the categories may not appear in recall without explicit cuing. Unless there is a specific reason to do otherwise, the best procedure is probably to follow previous work that has generally used 3–12 categories of 3–5 items with children and up to 10 items per category with adults.

PRESENTATION ORDER

Once the items have been selected, they have to be ordered for presentation. In the study of clustering, the degree of contiguity among the members of the same category in the presentation order must be

decided upon. The most common types of presentation orders are *random, explicitly unblocked,* and *blocked.* A randomized sequence allows the chance occurrence of runs of items from the same category, though a restriction is usually added to limit the length of such runs to no more than two or three consecutive words from a category. The explicitly unblocked presentation order is constructed with the restriction that an item from a given category cannot be followed by another item from the same category. In the blocked, or completely organized presentation, all items from the same category are presented contiguously. Some studies have included still other degrees of stimulus list organization. See Puff (1974) for a review of much of the earlier list organization research and Batchelder and Riefer (1980) for an extensive recent investigation.

The explicitly unblocked presentation has often been used to insure that subjects recalling in serial order do not show clustering that could be confused with that produced by the operation of an organizing process. However, with clustering measures for which a chance value can be computed, perfect serial recall of unblocked lists would lead to below-chance values. Since below-chance values of many of these measures are difficult to interpret, the random (or random with restrictions) ordering would seem to be preferable. Blocked presentation almost invariably leads to greater clustering and frequently to greater recall (Puff, 1974), especially with children. These effects of blocked presentation may follow from a number of factors, including the optimal opportunity for same-category items to be rehearsed together and the increased salience of the structure of the list. These possibilities make blocked presentation a particularly good procedure for promoting organization when the category structure is otherwise fairly weak. However, if the list includes more than a few categories, blocked presentation may result in fewer of the categories being represented in recall.

Regardless of the type of stimulus list organization that is chosen, the use of multiple input orders for presentation to different subjects is desirable. This controls for any effects due to the particular item adjacencies as well as for effects arising from which categories are represented (and by which items) in the primacy and recency portions of the input sequence. If the situation will allow it, a separate input sequence should be used for every subject.

When unrelated lists are used, as in studies of subjective organization, some of the same considerations are relevant. Since each subject receives multiple trials, a series of presentation sequences is required. The major issue in this case is whether to randomize the orders on each trial, thereby allowing a chance number of items to appear contiguously on successive trials, or to order the items so that no two items appear

together on consecutive lists. Since the latter would lead to below-chance organization scores if recall is serial, the randomization alternative seems preferable. However, as suggested for categorized lists, the same items should not be allowed to appear consistently in the primacy or recency portions of the sequences across trials. Furthermore, if every subject does not receive a unique set of presentation orders, then the fixed set of orders should be presented in different sequences to different subjects.

PRESENTATION MANNER AND MODALITY

Decisions about whether to present the materials serially or simultaneously, visually or auditorially, and as words or pictures in the visual modality are most often based on convenience or convention in a particular line of research unless these variables are of direct interest to the investigator. A serial, or successive, presentation helps to insure that the subjects are exposed to every item and equates the nominal amount of study time for all of the items. The simultaneous presentation of all items, on the other hand, allows the subject more flexibility in making interitem comparisons, rehearsing same category items together, and so on. Simultaneous presentation should, therefore, be expected to facilitate organization and recall of unrelated or weakly categorized materials, especially by subjects who may not be highly efficient in holding previously presented items in memory while implementing rehearsal strategies. Finally, it should be noted that the potential advantage of the simultaneous presentation might be augmented by combining it with the opportunity for the subjects to move items into groups (e.g., Mandler, 1967).

Several factors are relevant to the choice of the modality of presentation. First, it is well known that auditory presentation of words leads to a slightly increased recency effect on an immediate test (Cole *et al.,* 1971; Murdock & Walker, 1969). Second, the visual presentation of printed words may introduce a decoding (reading) problem that is obvious with prereading children but may also be a factor with elderly subjects, low socioeconomic class, or low IQ groups. The auditory presentation of words reduces this concern, although the investigator has to be careful about unintentionally including items that have homonyms. The visual presentation of pictures also circumvents the reading problem, but a picture may often have several possible labels. Therefore, with pictorial presentation, subjects are usually asked to label the items as they are presented. Labels that differ from those of the researcher are then either corrected at presentation or counted as correct at recall.

RATE OF PRESENTATION

In general, a slower rate leads to better performance, but the main consideration about the rate of presentation (or the total study time with a simultaneous presentation) is that there be sufficient time for the subjects to do whatever the experimenter expects them to be able to do during the input phase of the task. For example, a rate as fast as one item per sec does not afford much opportunity to engage in rehearsal or organization at input. Furthermore, a rapid rate of input might put younger or older subjects at a disadvantage in comparison with college students, even if memory ability per se does not differ. If precise control of the rate by the experimenter is not essential, some of these concerns can be overcome by allowing the subjects to control their own presentation rates. The distribution of times during study can then become an informative component of the results (Belmont & Butterfield, 1971).

INSTRUCTIONS

There are several elements that can be considered for inclusion in the instructions given to the subjects. If the task is intentional, the subjects are asked to study the items and informed that a memory test will follow. Also, subjects are generally told that they may recall the items in any order. In the case of categorized lists, if there is any concern about whether or when different subjects may detect the categorical structure, the structure can be described in the instructions. Finally, if written recall is to be obtained, it is helpful to use a lined data sheet and to instruct the subjects neither to skip lines as they are writing nor to insert any items between previously filled lines so that the experimenter has an unambiguous record of the order of recall.

The Retention Interval

The recall test usually begins immediately after the presentation is finished. The use of an immediate test can be expected to result in a high probability of recall of items presented at the end of the list. These items are generally recalled early and may preclude the occurrence of much organization until somewhat later in recall. If the experimenter is interested in minimizing the primary organizational components in the output, longer retention intervals can be used that are filled with some rehearsal inhibiting activity such as arithmetic problems (Puff, Murphy,

& Ferrara, 1977) or backward counting (Glanzer & Cunitz, 1966). Distractor tasks are normally chosen for the specific population under study so that young children might be asked to count forward by ones, older children to count backward by ones, and college students to count backward by threes or sevens. Distractor tasks are normally seen as reducing, but not completely eliminating, the possibility for rehearsal. In general, effective distractor tasks tend to decrease the recency effect, but they may thereby lead to increased clustering or subjective organization. Finally, the effects of rehearsal activities between presentation and recall can be assessed by comparing performance following filled and unfilled intervals.

Output Parameters

OUTPUT MODE

Whether the subjects are asked for written or spoken recall is usually not much of an issue. Spoken recall sequences have to be transcribed and cannot be obtained in group testing situations. However, written recall is slower and may lead to decreased recency effects. In addition, with written recall the subjects can (unless a mask is used) look back at previously remembered items. Looking back may provide retrieval cues for the recall of additional items and might well facilitate the editing process so that fewer items are recalled a second time. Of course, in comparing different subject populations capable of different writing speeds (i.e., college students versus just about anyone else), spoken recall would be the preferred mode of output.

OUTPUT DURATION

A fixed period of between 1 and 5 min is usually allowed for recall, with the exact length depending upon the number of items in the list. Population differences may be important here, also. It has been argued that older adults' recall performance may be disproportionately hampered if they are pushed for speed at recall (Botwinick, 1978). On the other hand, if young children are given too long, they tend to get bored with the task. The use of a dual criterion is a possible solution to some of these problems. The maximum recall period is set, but if the subject recalls no items in a given time period (about 15 sec), the experimenter asks "Can you remember some more?" If not, recall is terminated.