Transfer in Part-whole and Whole-part Free Recall: 
A Comparative Evaluation of Theories

ROBERT J. Sternberg and GORDON H. BOWER
Stanford University

Two experiments, involving part-whole and whole-part transfer respectively, investigated the adequacy of two organizational and two list discrimination accounts of transfer. The organizational accounts were those of Tulving (1966) and of Novinski (1972), while the list discrimination accounts were those of Slamecka, Moore, and Carey (1972) and of Anderson and Bower (1972). In each transfer task, 45 subjects were instructed to recall all words from both initial and transfer lists, indicating a list membership for each word and a confidence rating for each list assignment. A list-identification task was administered at the end. Recall and list-identification were scored for overall proportions of types and tokens, and for proportions of tokens recalled from each of three list partitions. Analyses of recall and organizational measures failed to support either organizational interpretation of transfer. Patterns of recall, list-identification, and confidence ratings were consistent with the theory of Anderson and Bower, but not with the response-withholding theory of Slamecka, Moore, and Carey.

Tulving (1966) reported negative part-to-whole list transfer in free recall of unrelated words, and Tulving and Osler (1967) reported the same for a whole-to-part list transfer paradigm. While the latter paradigm has received relatively little attention, the former has been subjected to extensive examination, appraisal, reexamination (Novinski, 1972) and reappraisal (Hicks & Young, 1972).

The finding of negative transfer was originally of interest first because it is counter-intuitive, and second because it is incompatible with conventional frequency theories, apparently requiring postulation of persisting subjective organizations which can impede transfer of training. We believe it is important to distinguish between the two reasons for interest in negative transfer, since recent research has progressively further dissociated them.

With respect to the first reason for interest, the negative transfer finding was clearly an unexpected one; indeed, the sheer novelty of the phenomenon seems at times to have attracted more attention than the theoretical issue it was meant to elucidate. The consequence has been a sort of “functional autonomy” of the empirical phenomenon from theoretical issues. This autonomy has led to the generation of a disproportionately high ratio of data to theory. Thus, while we are now aware of some of the limiting conditions of negative transfer in part-to-whole recall (e.g., Birnbaum, 1969; Wood & Clark, 1969), it is not altogether clear what general relevance these limitations have for a theory of memory.

With respect to the second reason for interest in the negative transfer phenomenon, the frequency theorists have yet to concede the error of their ways: The nature of the processes underlying negative transfer, indeed, the very existence of such transfer, remains in dispute. Let us consider evidence for the negative transfer phenomenon before considering explanations of it.

Slamecka, Moore, and Carey (1972) noted that while significant differences have been reported in the slopes of the whole-list learning curves for experimental and control subjects.

---

1 This research was supported by research grant MH-13950-06 to the second author from the National Institutes of Mental Health.

Copyright © 1974 by Academic Press, Inc.
All rights of reproduction in any form reserved.
Printed in Great Britain
(the former having previously learned a part-list and the latter an unrelated list), no significant differences have been reported in the subjects' final recall scores. If the obtained crossovers reflected chance sampling errors, then non-significant differences would be consistent with a frequency hypothesis, which would predict asymptotic convergence of the learning curves of experimental and control subjects.

Slamecka et al. (1972) raised the important issue of how negative transfer should be defined. One alternative defines it in terms of a difference in final recall of experimental and control subjects. A second alternative defines negative transfer in terms of a slower acquisition rate for experimental subjects, “correcting” in some way for group differences in starting values or asymptotes of the learning curves. Investigators have been hard-pressed to demonstrate differences of the former kind, and while they have demonstrated differences of the latter kind, they have usually done so inappropriately. The standard technique has been to apply a two-way analysis of variance (groups by trials), and test for a groups-by-trials interaction. This technique is unsatisfactory, first because it rests upon the false assumption that learning curves are linear, and second because the results can be affected by unimportant differences in initial levels of recall. An alternative and preferable approach would be to assume a standard exponential model of mathematical learning theory,

\[ r_{n+1} = r_n + \theta(\lambda - r_n), \]

where \( r_n \) is the proportion of items recalled on the \( n \)th trial. This equation may be fit to data for individual subjects to yield estimates of \( \theta \) (slope or rate) and \( \lambda \) (asymptote). Between-group comparisons of mean slope and asymptote are each informative in their own right, and together tell us more about transfer phenomena than either one alone.

While part-to-whole and whole-to-part negative transfer are not always conclusively demonstrated, the crossover between the acquisition curves of experimental and control subjects (which suggests differences in \( \theta, \lambda \), or both) has been replicated enough times by enough investigators to warrant its tentative acceptance as a stable result. We shall therefore proceed to compare explanations of the phenomenon.

Explanations of part-to-whole and whole-to-part negative transfer are of two general types, subjective organization and list discrimination. Explanations of the former kind interpret the result in terms of word-to-word relations, while explanations of the latter kind attribute the result to word-to-list tag relations. Two different and mutually incompatible interpretations of each kind have been advanced.

Tulving (1966) and Tulving and Osler (1967) suggested that their experimental subjects had tried to accommodate the second list to the subjective organization they had formed for the first list; while this persisting organization had been appropriate to the first list, it was not optimal for learning the second list, so that eventually it interfered with rather than facilitated recall of the second list by experimental subjects (relative to control subjects).

Novinski (1972) has proposed the reverse hypothesis. She has argued that the negative transfer effect is caused by subjects' failure to retain their prior (part-list) organization. Were experimental subjects to retain their prior organization, she conjectured, positive rather than negative transfer would ensue. One way to encourage subjects to retain their prior organization would be to inform them of the relationship between part and whole lists; and indeed, when subjects are informed of this relationship, positive rather than negative transfer appears (Novinski, 1972; Wood & Clark, 1969). It should be noted that Novinski offered her theory in the context of a part-to-whole transfer experiment. She has not extended her theory (as did Tulving) to the whole-to-part paradigm, and it is not clear how her theory would apply to it. However, it need not be the case that the same processes...
underlie the two conditions of negative transfer.

List discrimination hypotheses have been offered by Slamecka et al. (1972), by Anderson and Bower (1972), and by Schwartz and Humphreys (1973). Slamecka and his colleagues focused upon strategies which subjects might adopt in recalling words, while Anderson and Bower and Schwartz and Humphreys focused upon less deliberate mechanisms associated with poor learning and subsequent recall.

Slamecka et al. (1972) suggested that on the transfer task, subjects in the experimental condition will recognize that some of the part-list (old) items appear on the second list, but they may not be certain that all of the old items have been included. Their uncertainty, coupled with a desire to avoid intrusions, may lead experimental subjects to set a "high criterion" for emitting responses, resulting in the withholding of responses which in fact would be correct if recalled. Indeed, when instructions encouraged subjects to guess, part-to-whole negative transfer disappeared.

Anderson and Bower (1972) and Schwartz and Humphreys (1973) have argued for a list discrimination hypothesis, with the former authors describing a possible mechanism in some detail. These authors suggested that negative transfer effects derive from subjects' difficulties in associating List-2 markers (list tags) to items previously associated to several List-1 markers. As a result, the learning of new list-tags for old words repeated on List 2 is retarded; consequently, when subjects are asked to recall List-2 words only, old words not yet tagged as appearing on List 2 are edited out and not recalled. The Anderson-Bower theory is compatible with the finding (Wood & Clark, 1969) that informing subjects of the relationship between lists removes negative transfer (since a rule automatically assigns a List-2 tag to all words bearing a List-1 tag). The theory also explains the curious finding (Schwartz & Humphreys, 1973) that if subjects are not informed of the relationship between lists, negative transfer results even when the "two lists" are identical.

The present experiments were intended to discriminate among the explanations described above. Experimental procedures were designed to render observable certain consequences of processing which have been previously either presupposed or overlooked.

**EXPERIMENT I: PART-WHOLE TRANSFER**

**Method**

**Design.** Forty-five subjects were equally assigned to one control and two experimental groups. In the control group, subjects received successive presentations of two lists containing no common (overlapping) words. In the experimental group, subjects received two successive lists in which the second ("whole") list contained all the words from the first ("part") list plus an equal number of new words. Subjects in one experimental group were informed of the nature of the relationship between lists, while subjects in the other experimental group were not so informed.

**Subjects.** Subjects were 30 male and 15 female Stanford undergraduates, serving to fulfill a requirement for an introductory psychology course. They were tested in groups of from two to six members, with assignment to conditions dependent upon the session which they selected; within the two experimental conditions, assignment to one or the other depended on the order in which subjects entered the testing room.

**Materials.** A pool of 36 words was selected from the 42-10-032 to 49-11-033 frequency range of Kučera and Francis (1967). Words were selected randomly from this range except for the stipulations that the words be two-syllable, non-homonymous, and unique in the letters occupying the initial two positions of the word. The word pool consisted of 22 nouns (of which 9 were concrete and 13 abstract), 8 adjectives, and 6 verbs. Words were each assigned randomly to one of three lists, except for the stipulation that each list consist of 12
words approximately counterbalanced for parts of speech.

The three lists, A, B, and C, comprised the three possible first lists which a subject might receive. Second lists were formed by combining, to get A + B, A + C, and B + C. Experimental subjects received first list–second list combinations such as A, A + B, while control subjects received combinations such as C, A + B. Lists A, B, and C were completely counterbalanced among conditions.

Procedure. As subjects entered the testing room, they were seated alternately at two tables; when all subjects had arrived, they were handed test booklets. In the experimental conditions, subjects at one table formed the informed group, while subjects at the other table comprised the uninformed group.

The experimental session of approximately fifty minutes was divided into three parts. In the first, subjects received four alternating presentation and recall trials on a 12-word list, with instructions to recall in any order as many of the words as they could during the 48 seconds allotted after each presentation. Words were presented orally at a 2-second rate, and were randomly reordered for each trial.

Next, subjects received eight alternating presentation and recall trials on a 24-word list. Presentation rate was again 2 seconds per word, and 180 seconds were allowed for recall plus the other tasks to be described. Following each second-list presentation, the subjects were asked to try to recall in any order all the items seen in the entire experiment—that is, all second-list items and all first-list items in any order. As each item was written in recall, subjects were to record beside it the list in which that word had occurred. If they believed that a word they were recalling had appeared in both lists (i.e., had two “tokens”), they were instructed to write the word twice, indicating a list membership for each token recalled. Moreover, subjects were instructed that during recall they were to assign confidence ratings on a scale from 1 to 6 to each of the list identifications of the words they recalled. Each point on the confidence scale was labeled, with 1 indicating that the subject was “not at all confident” of his list identification and 6 indicating that he was “absolutely confident” of his list identification. Instructions were worded so as to encourage guessing in recall, noting in particular that subjects should indicate their uncertainty about list origin by assigning low confidence ratings, not by withholding words of uncertain list origin. This procedure, of second-list presentation followed by all-item recall plus assignment of a list-identification to each recalled item and a confidence rating to each list identification, occurred for eight presentation-test cycles.

After this second part of the session, recall test booklets were collected, and a second test booklet was handed out. This booklet contained instructions and test items for the final list-identification task. This task required list identifications and confidence ratings for all the old items and for 36 new ones. The new items had been drawn from the same pool in Kučera and Francis (1967) and according to the same specifications as had the old items, but had not been previously presented to the subjects. Subjects were instructed to write 1 if the word had appeared on the first list only, 2 if it had appeared on the second list only, 1 and 2 if it had appeared on both lists, and 0 if it had appeared on neither list. Subjects were again asked to assign confidence ratings to each such list identification, using the same scale as before. Subjects in the experimental groups received an additional questionnaire sheet stapled to their booklets. The sheet contained instructions asking each subject to comment upon any relationship he might have noticed between the items on the first and second lists. Experimental subjects completed this questionnaire as soon as they finished the list-discrimination task. They were permitted to look back at their list-discrimination responses if they so wished. Subjects were allowed as much time as they needed on the list-identification task (and, in the experimental groups, for the questionnaire), and
were dismissed after they had completed the task, handed in the booklet, and received a debriefing sheet.

Results and Discussion

Comparability with previous findings. An initial concern was with whether we had replicated the results typically obtained in previous part–whole transfer experiments. Such experiments have usually shown no significant difference between experimental and control groups in level of first-list (pre-treatment) recall, superior List-2 recall by part-to-whole transfer subjects on the initial trials, and a crossover followed by superior recall by control subjects on the last few trials.

For this comparison, we pooled our experimental groups. The mean proportion of words recalled (pooled over trials) on the first-list task was .73 for control subjects and .69 for experimental subjects, $t(43) = 1.39, p > .05$, thus establishing initial comparability of the two groups. List-2 recall was measured by the proportions of List-2 words recalled and correctly identified on each trial as having been on List 2; these proportions are shown in Figure 1. The figure shows the standard crossover result. By the last trial, control subjects recalled a significantly higher proportion of List-2 words than did experimental subjects (.83 versus .67), $t(43) = 3.85, p < .001$. In general, these results replicate those of past experiments. The amount of initial positive transfer was less than that usually obtained, but the amount of negative transfer evident on the last trial was greater.

Regrouping of experimental subjects. As mentioned previously, experimental subjects were asked at the end of the experiment to comment upon any relationship they might have noticed between items on the first and second lists. Inspection of those comments revealed that some subjects in the uninformed group had recognized the relationship between lists, while not all subjects in the informed group had correctly described the inter-list relationship. While the former outcome was not surprising, the latter one was. Since the question probing for recognition of the relationship between lists had been purposely phrased in a non-specific way, it seemed possible that some of the subjects might not have understood what answer was expected of them. Because of the ambiguity inherent in the questionnaire data, we turned to the list-identification results as a more reliable index of subjects' final knowledge regarding the relationship between lists.

On the final list-identification task, subjects had been asked to identify the list or lists, if any, on which each test word had appeared. For the part–whole groups, of course, no word could appear on List 1 without also appearing on List 2. Therefore, a subject aware of the relationship between lists should never identify a word as having appeared only on List 1. As an operational definition, then, we shall say that an experimental subject who identified one or more words as having appeared only on the first list must have been "unaware" of the exact relationship between

![Fig. 1. Proportions of List-2 words recalled and identified as List 2.](image-url)
lists. By this index, 9 of the 15 subjects in the uninformed group showed themselves to be unaware of the list relationship, whereas 6 of the 15 subjects in the informed group similarly showed themselves to be unaware, presumably because they had forgotten or disbelieved the instructions regarding the relationship between lists. Since we were interested in what the subjects had actually believed rather than in what they had been told about the relationship between lists, we regrouped the experimental subjects on the basis of whether during the recognition task they had shown themselves to be aware or unaware of the list relationship. Had questionnaire replies been used instead, 8 subjects would have been classified differently than was suggested by their identification responses; but 7 of these were subjects who were "aware" according to their list-identifications but failed to mention the part-whole list relationship in the questionnaire.

One undesirable but unavoidable result of the regrouping of the experimental subjects into "aware" and "unaware" groups was the introduction of a slight selection effect into the assignment of subjects to groups. It might be expected that the less able subjects would be less likely to learn the relationship between lists. With "ability" assessed by initial-list learning, this appears to have been the case. While the proportion of first-list words recalled by aware subjects (.71) hardly differed from the proportion recalled by control subjects (.73), the proportion of first-list words recalled by unaware subjects (.69) was slightly lower than that recalled by control subjects, $t(28) = 2.00, p < .06$. Since the mean of both experimental groups differed from that of the control group in the direction which might spuriously produce the appearance of later negative transfer, List-2 recall (described in the next two sections) was assessed by analyses of covariance with over-all first-list recall as a covariate.

Type and token recall. In previous part-to-whole transfer studies, second-list recall has been the criterion for determining whether negative transfer has occurred. But our approach suggests that second-list recall is a composite of several components. We have therefore scored the recall data in several ways, attempting to do justice to its complexity. At a global level, we have distinguished between type and token recall.

Type recall was defined as the total proportion of distinct word-types recalled during the transfer task without regard to the list identification assigned. Experimental subjects had studied 24 different word-types (12 appearing on both lists and 12 appearing on the second list only), whereas control subjects had studied 36 different word-types (12 appearing on the first list and 24 different ones appearing on the second list). Type recall proportions for control subjects can be computed with respect to either all 36 words or just the 24 word-types on List 2. These type recall proportions are plotted in Figure 2. If one assesses performance in terms of proportional recall of all word-types studied, then neither experimental group

![Figure 2](https://example.com/figure2.png)

**Fig. 2.** Proportions of all word-types (List 1 and List 2) recalled irrespective of list identification. Recall of just List-2 types by controls is also shown.
showed negative transfer relative to the control group; in fact, relative to the controls, subjects in the aware group showed significant positive transfer, considered either in terms of the proportion of all types recalled on the final trial, \( t(27) = 2.39, p < .05 \), or in terms of the proportion of all types recalled summed over the eight trials.

The standard patterns emerge, however, when controls are assessed only for recall of their List-2 word types. Compared to the unaware subjects, the controls now show the classic "crossover" effect (though these groups do not differ on the final trial, \( t(27) = 1.68, p > .05 \)). Compared to the aware subjects, the controls begin List 2 recalling fewer word-types, but soon converge to the same recall level as the aware subjects.

Token recall was defined as the total proportion of individual tokens recalled during the transfer task from both lists and without regard to the list identification assigned (except that for overlapping words twice recalled on a given trial, experimental subjects were required to assign different list identifications to each token recalled). For control subjects, token recall is identical to recall of all types, since the number of tokens presented to subjects in this group equalled the number of types presented. The experimental subjects, however, had been exposed to two tokens for every overlapping word, one occurring in the context of List 1, the other in the context of List 2. Like the control subjects, therefore, the experimental subjects had been exposed to 36 individual tokens. Token recall scores for the three groups are plotted in Figure 3. The means for control and aware subjects were close to one another over all trials (with an inconsequential crossover after the second trial). In contrast, token recall by unaware subjects was markedly depressed. Unaware subjects recalled a significantly lower proportion of tokens than did control subjects on the last trial, \( t(27) = 2.30, p < .05 \), and over all trial pooled together. Type and token recall scoring do not take into account the accuracy of the subject's list identifications during recall. This is done by our discriminative recall measure.

**Discriminative recall.** Discriminative recall was defined in terms of the proportion of word-tokens recalled and correctly identified as to list membership. The full set of discriminatively recalled word-tokens was partitioned into three subsets. For experimental subjects, these recalled subsets were (1) unique words presented only on the second list and correctly identified as having appeared on the second list, (2) overlapping words identified as having appeared on the first list, and (3) overlapping words identified as having appeared on the second list. In a "perfect" protocol, each overlapping word-type would be identified as having appeared in both the first and second lists. For control subjects, the second-list words were randomly divided into two arbitrary subsets, one serving as the pseudo-control for the unique words in the list of the experimental subjects, and the other serving as the pseudo-control for the overlapping words.
identified by experimental subjects as having appeared on List 2. (Recall by controls of the words assigned to the two random halves did not differ, $t(14) = .85, p > .05$.) List-1 words formed the third subset of words for control subjects.

Recall of List-2 unique words on the final trial was .78 for aware subjects and .71 for unaware subjects. Recall of the corresponding words for control subjects was .83. Differences in level of recall between the control and the two experimental groups (with covariance adjustments) were not statistically significant, $t(27) < 1$ for each comparison. Differences between control and experimental subjects were even smaller for recall pooled over trials. The exponential learning curve described earlier was fitted to the unique-word recall for each subject. The average learning rate parameter ($\theta$) was .47 for aware subjects, .44 for unaware subjects, and .41 for control subjects. Neither experimental group mean differed significantly from the control group mean, $t(28) < 1$ for each comparison. So, with respect to discriminative recall of unique List-2 words, the three groups do not differ.

A more interesting pattern of results emerged in comparing recall by experimental subjects of overlapping words to recall by control subjects of their corresponding words. Figure 4 compares the aware and control groups for these discriminative recall measures. This figure shows the mean proportions of overlapping words recalled and identified as List 2 by aware subjects, corresponding words recalled and identified as List 2 by control subjects, overlapping words recalled and identified as List 1 by aware subjects, and List-1 words recalled and identified as List 1 by control subjects.

The first result of interest in Figure 4 is the dissociation in level of discriminative recall for aware subjects between overlapping words identified as List-1 and as List-2. Although the two sets of test words were identical, levels of discriminative recall were not. On the first trial, discriminative recall by aware subjects of List-1 words was significantly higher than their recall of overlapping List-2 words, $t(14) = 3.14, p < .01$, presumably because not all subjects had yet become aware of the relation between part and whole lists. By the final trial, however, discriminative recall of the List-2 overlapping words had caught up with and slightly surpassed discriminative recall of the List-1 words, although the two levels of discriminative recall did not differ significantly.

A second result obvious in Figure 4 is that while levels of discriminative recall of List-1 words were identical for control and aware subjects on the first trial, performance of subjects in the two groups diverged after the first trial, with the controls performing below the aware subjects. On the final trial, this difference was significant, $t(27) = 2.14, p < .05$. These results may be summarized in two statements. First, presentation of the overlapping words as List-2 items helped maintain their recall as List-1 items, but only at the original level of performance—aware subjects’ discriminative recall of List-1 items did not improve across
trials. Second, nonpresentation of List-1 words resulted in a typical forgetting curve for recall of these items by control subjects.

The third result of interest in Figure 4 compares the performance of aware subjects on overlapping words identified as List 2 with that of control subjects on their corresponding subset of words. Subjects in the two groups did not differ significantly in their mean level of discriminative recall on the final trial, t(27) = 1.55, p > .05. However, the average slope (θ) of the individual learning curves was .22 for aware subjects and .39 for control subjects, and these means differed significantly, t(28) = 3.02, p < .01. This difference in learning rate may reflect the fact that many subjects classified as “eventually aware” were not aware of the list relationship throughout the trials, but rather became aware of it at various points during the learning trials. Thus, like unaware subjects, the “aware” subjects may show some negative transfer in the recall and identification of overlapping words as List-2, but such negative transfer would presumably be limited to the earlier trials of the transfer task.

The most important results of the study are shown in Figure 5, which depicts discriminative recall of overlapping words by unaware subjects compared to performance of the controls on their corresponding items. In order to facilitate the comparison, control curves are repeated from the preceding figure.

For unaware subjects, discriminative recall of words identified as List 2 began lower on Trial 1 than their recall of words identified as List 1, t(14) = 3.74, p < .01; discriminative recall of the List-2 words also finished lower on Trial 8 than that of the List-1 words, although not significantly so, t(14) = 1.37, p > .05. While recall of first-list tokens on Trial 8 was identical for the aware and unaware subjects, the aware subjects recalled more List-2 token than List-1 tokens, whereas the unaware subjects showed just the opposite pattern, with discriminative recall of List-2 tokens lower than that of List-1 tokens. This pattern of results indicates that the aware and unaware subjects differed only in their discriminative recall of overlapping tokens identified as List 2; recall of overlapping items identified as List 1 was comparable for the two groups.

The unaware subjects, like the aware subjects, showed significantly higher discriminative recall of List-1 tokens than did the control subjects, with t(27) = 2.14, p < .05, for Trial 8. This difference was also significant when List-1 tokens recalled were pooled across trials. The opposite ordering obtained, however, for recall of overlapping List-2 tokens as compared to corresponding items for control subjects. Control performance was considerably higher on the final trial, t(27) = 3.39, p < .01. The subjects in the two groups also differed significantly in the slopes of their individual learning curves, which averaged .39 for control subjects, but only .18 for unaware subjects, t(28) = 3.37, p < .01. Hence, relative to the controls, unaware subjects showed positive transfer for
overlapping words recalled and identified as List 1, but very strong negative transfer for overlapping words recalled and identified as List 2. Clearly, the problem for the unaware subjects was not one of recalling word-types, but rather one of identifying these types as List-2 tokens.

The curves plotted in Figures 4 and 5 show marginal probabilities, that is, that an item on List 1 is recalled and identified as coming from List 1. For overlapping words, one may also examine the joint probabilities that an overlapping word is recalled and jointly identified as coming from both Lists 1 and 2. For aware subjects (cf. Figure 4), this joint probability starts at .39 and rises approximately linearly over trials to .51 on Trial 8; for unaware subjects (cf. Figure 5), this joint probability starts at .22 and increases over trials to .34. Comparison of these joint probabilities with the marginal probabilities (in Figures 4 and 5) reveals that discriminative recall of List-1 and List-2 tokens of a word-type are not statistically independent. As might be expected, the two tokens of a word-type tend to be recalled (or nonrecalled) together. This nonindependence, of course, would be a consequence of any recall strategy by virtue of which word-types were retrieved probabilistically and their successfully-associated list tags (List 1, 2, or both) were "read off."

The discriminative recall measures help us dissect the components contributing to the composite outcome of negative transfer in part-whole recall. The negative transfer is not due to differential recall of unique List-2 words, since all groups were equal in this respect; nor is it due to the general unavailability of the word-types (either overall or for List 2 only), since experimental subjects did not show negative transfer for type recall. Only the unaware subjects show significant negative transfer in discriminative recall and it hinges precisely upon the fact that they do not assign List-2 tags to overlapping words which they recall (they identify them only as "List 1" items). This negative transfer in list-tagging of overlapping words is a very large effect in discriminative recall (Figure 5) as well as in the final list identification test.

These components suffice to account for the composite negative transfer in standard part-whole studies using uninformed (and predominantly unaware) subjects. Since subjects are typically asked to recall only those words which appear on the second list, unaware subjects must learn to discriminate between the two list-tokens of the same word type. This difficult discrimination is not required of aware subjects (who know that all List-1 tokens have List-2 tokens) or of control subjects (who have two independent lists). The locus of the negative transfer is in the learning of List-2 tags for overlapping words for those part-whole subjects who are not aware of the full containment of the first list in the second.

Our analysis hinges upon the type-token distinction, using recall of a word plus its list tag as the primary unit of analysis. Once these distinctions are pressed, the negative transfer effect no longer appears so mysterious. That outcome was dumbfounding so long as we believed that the part-whole transfer group had to learn "only half as many words" as their controls. But while investigators were theorizing in terms of relations among organizations of word types, they were scoring their data in terms of List-2 tokens recalled and instructing their subjects so as to enforce their recalling in terms of List-2 tokens. Perhaps the theoretical lesson we have learned is a strong denial of the premise that "a word is a word is a word" for purposes of experiments on memory for word-lists. Rather, different tokens of a single word type may have independent representations in memory and become related to entirely different associative contexts (see the set-membership relation in HAM, the Anderson & Bower theory, 1973, which is equivalent to our type-token distinction).

Let us now consider the organizational theories of part-whole transfer and how they
apply to our data, since they have been the predominant hypotheses regarding this phenomenon. Organizational hypotheses such as those of Tulving (1966) and Novinski (1972), suggested in the context of experiments measuring second-list recall as customarily defined, seem incapable of accounting for the present results. Because these hypotheses nowhere provide for a type-token distinction, either in conception or in scoring operations, they are unable to explain the obtained dissociations and opposite group trends in our type versus token measures and especially in the discriminative recall of overlapping words. It is also not clear how these theories would predict that the only difficulty for unaware subjects would be localized in the recall of overlapping List-2 tokens. If, in accordance with such hypotheses, the presence (or absence) of first-list organization impedes organization of second-list words, it should do so about equally for both old and new words, since presumably the negative transfer is caused by the merging of new words into old, nonoptimal S-units. If organizational hypotheses were to be modified so as to view the subject as separately organizing the old and new items during List-2 learning, this modification would explain the absence of transfer in learning the unique List-2 words. But such a modification would do so only at the expense of becoming unable to explain why negative transfer occurs for recall of old words, since first- and second-list organizations would then be independent.

The withholding hypothesis of negative transfer (Slamecka et al., 1972) fares no better by these data than do the organizational hypotheses. Although our subjects received instructions designed to encourage guessing, unaware subjects nevertheless exhibited considerable negative transfer relative to the controls in the discriminative recall of overlapping List-2 tokens. Of course, subjects may have failed to follow instructions, withholding words despite our urging them to guess. But this possibility is rendered unlikely by analyses of subjects' confidence ratings for discriminatively recalled words (to be reported later).

A list discrimination hypothesis seems most compatible with our results. This hypothesis, as advanced by Anderson and Bower (1972), suggests that subjects in the unaware group would be able to recall overlapping types with little difficulty; but because they experience negative transfer in learning List-2 tags for these types, they do not recall these word-types as List-2 tokens. Although unaware subjects have difficulty learning List-2 tags, they experience no difficulty in assigning List-1 tags during first-list learning, and the continued presentation and recall of the overlapping types during List 2 trials would enable the unaware subjects to maintain recall of the List-1 tokens at the final level of List-1 recall. Thus, the hypothesis predicts the learning curves for recall of List-1 and List-2 overlapping tokens.

Because aware subjects know the relationship between lists, they have available a rule for assigning List-2 tags. According to the rule, recall of a List-1 token is tantamount to recall of a List-2 token. However, recall of a List-2 overlapping token does not assure its recall as a List-1 token, since a subject may forget that the particular word-type had also appeared as a List-1 token; in this event, he would believe the type to be a unique, List-2 word. Overall, then, the hypothesis predicts overlapping List-2 discriminative recall for aware subjects to be higher than their List-1 discriminative recall, as it was.

The results described above, based upon proportions of words recalled, would seem only indirectly relevant to organizational hypotheses. We therefore investigated the compatibility of organizational hypotheses with our data by analyses of the subjective organization of subjects' recall protocols.

Organizational analyses. According to the organizational hypotheses, perseveration of first-list organization into transfer-task recall results either in negative transfer (Tulving, 1966) or positive transfer (Novinski, 1972).
during second-list learning. Although numerous experiments have been designed to test the effects of perseveration of first-list organization, no experiments have reported organizational analyses which verified the existence of such perseveration. Several experimenters have measured category clustering during second-list recall only, but such measures do not yield information of direct relevance to hypotheses involving subjective organization. Fewer investigators have utilized measures of subjective organization, but not in ways which directly measure the perseveration of first-list organization into transfer-task recall. We have attempted to measure this perseveration directly.

On the transfer task, experimental subjects recalled three subsets of words—overlapping words identified as List 1, overlapping words identified as List 2, and words unique to the second list and so identified. Control subjects may also be conceived to have recalled three sets of words—List-1 words, words pseudo-paired with the overlapping List-2 words, and words pseudo-paired with the unique List-2 words of the experimental groups. In the experimental groups, only the first two subsets of words are relevant to hypotheses regarding the perseveration of first-list organization into second-list learning. In the control group, only the first subset of words is relevant. Specifically, the relevant items for determining organizational perseveration are those types which appeared on the first list. We extracted those types from the transfer-task recall protocols in the order in which they appeared, separating overlapping List-1 and List-2 tokens, and ignoring all other recalled words (which were deemed not relevant to the organizational analysis). Using this procedure, we obtained two sets of words per trial for each experimental subject (overlapping words recalled and identified as List 1 and overlapping words recalled and identified as List 2), and one set of words per trial for each control subject (List-1 words recalled).

Assuming that the best information available regarding final first-list organization was the ordering of words in recall on the final trial of the first list, we computed unidirectional ITR (Bousfield & Bousfield, 1966) comparing the last trial of the first list to each trial of the transfer task for each of the above-mentioned sets of words.

ITR, an acronym for “inter-trial repetitions,” is a measure of the extent to which pairs of words appearing adjacently in recall on a given trial are also recalled adjacently on a specified subsequent trial. In unidirectional ITR, a pair of words is counted only if the words are recalled in the same order on both trials. The observed value of unidirectional ITR is the number of ordered pairs which appears on both trials. The minimum number of such pairs is zero, while the maximum number is one less than the number of recalled words which the protocols for the two trials have in common. Since some such pairs are expected to occur by chance, it is customary to subtract from the observed value of ITR an expected value of ITR. After thus correcting for chance repetitions, the expected value of ITR for randomly generated recall orders is zero.

Using the procedure described above, we obtained for each experimental subject two different ITR measures, one for organization from the last trial of the first list to overlapping words recalled and identified as List 1 (to be called ITR1), and one for the same first-list words to overlapping words recalled and identified as List 2 (to be called ITR2). For control subjects, only ITR1 is defined; ITR2 is undefined because control subjects receive two lists containing no overlapping words. Figure 6 shows a schematic diagram describing the ITR1 and ITR2 measures and another measure, ITR1,2, to be explained later.

The claim of Tulving's organizational theory is that maintenance of first-list organization will impede second-list recall. According to this theory, one should expect both the ITR1 and ITR2 measures to be negatively correlated with discriminative recall of overlapping words, since there is no distinction
THEORIES OF TRANSFER

made between different list tokens of the same word type. Novinski's organization theory, however, supposes that maintenance of first-list organizational facilitates second-list recall. So according to this theory, the ITR measures should be positively correlated with discriminative recall of overlapping words. Our arguments here assume that traditional "second-list recall" measures are actually measures of recall filtered through a list-identification editor which discriminates List-2 from non-List-2 words.

Neither ITR measure showed a clearcut trend across trials, and so trial means were combined to form an overall mean for each condition. The mean level of ITR\(_1\) was .58 for control subjects, .25 for aware subjects, and .40 for unaware subjects. The mean level of ITR\(_2\) was .30 for aware subjects and .23 for unaware ones. The correlation of ITR\(_1\) with discriminative recall of overlapping words identified as List 1 was only -.13, and was virtually the same whether or not control subjects were included in the analysis; the correlation of ITR\(_1\) with discriminative recall of overlapping words identified as List 2 was -.25. The correlation of ITR\(_2\) with discriminative recall of overlapping words identified as List 1 was -.07, and its correlation with overlapping words recalled and identified as List 2 was only -.05.

The first thing to notice about these results is that the mean levels of ITR were all very small, averaging less than one inter-trial repetition unit per trial. There is thus no evidence that persistence of first-list organization was a major factor during transfer-task recall. While the obtained correlations were all negative, as predicted by Tulving's "persistence" theory, none of them approached significance. Therefore, one cannot interpret the correlations as supporting either organizational theory.

A possible weakness in our argument is that the amount of within-list organization developed during List-1 learning was unexpectedly small for all groups. Mean ITR between the last two trials of List-1 learning was only .25 for aware subjects, .20 for unaware subjects, and .47 for controls. An organizational theorist might therefore discount our low ITR\(_1\) and ITR\(_2\) values (and the correlations involving them), since there was hardly any List-1 organization to be carried over into List-2 recall. But the organizational theorist would encounter difficulty in explaining how the observed negative part-whole transfer could have occurred in the near-absence of first-list organization.

Perhaps the failures to obtain significant correlations were due to weaknesses in the organizational measures rather than to weaknesses in the theories they were intended to test. We do not believe this to have been the case; however, if there are those who do, then it is incumbent upon them to propose a better measure, or at least to show in some direct way that organizational factors are responsible for...
the results obtained in standard part–whole experiments.

A novel organizational index, a measure of *intra-trial* repetitions (to be called ITR\textsubscript{1,2}), was computed for the experimental groups between the two arrays of overlapping words recalled on each trial of List-2 (ignoring irrelevant words). This index was intended to measure communality in organization between overlapping List-1 and List-2 words as estimated by the extent to which pairs of List-1 tokens were recalled in the same order as pairs of List-2 overlapping tokens. It should be noted that this index is not a measure of the persistence of first-list organization into transfer-task recall, and that it was possible to compute this index only because subjects had recalled multiple tokens for single word-types. The ITR\textsubscript{1,2} measure is illustrated schematically in Figure 6.

No obvious trends in ITR\textsubscript{1,2} emerged over trials, and so trial means were pooled. The mean level of ITR\textsubscript{1,2} was 4.03 for aware subjects and 3.01 for unaware subjects. These means are considerably higher than those obtained for the ITR\textsubscript{1} and ITR\textsubscript{2} measures, indicating that while subjects retained very little of their final first-list organization on second-list trials, they did tend to recall first- and second-list tokens in stereotyped ways on a given transfer-task trial. The correlation of ITR\textsubscript{1,2} with discriminative recall of overlapping words identified as List 1 was .45, \( p < .05 \); its correlation with discriminative recall of overlapping words identified as List 2 was .84, \( p < .001 \). This latter correlation was so high that we at first thought it might be an obscure artifact of our scoring method. We explored this possibility by computing ITR\textsubscript{1,2} scores for 100 random recall protocols ("statistical subjects") generated by a Monte Carlo procedure. The mean value of ITR\textsubscript{1,2} for these statistical subjects was only 0.02. The correlation for statistical subjects of ITR\textsubscript{1,2} with their discriminative recall of overlapping List-1 words was .03, and with recall of overlapping List-2 words was .07. These Monte Carlo results demonstrate that the correlations obtained for real subjects were not artifactual.

Several speculative interpretations are consistent with the high correlations of ITR\textsubscript{1,2} with List-2 recall. One derives from the Anderson and Bower (1972) theory of retrieval and decision processes in free recall. This theory (see also Anderson, 1972; Anderson & Bower, 1973) supposes that free recall is an outcome of several processes, including (1) subjects’ retrieval of word-types by following tagged associative pathways between word-types in memory, and (2) subjects’ monitoring of each implicit word-type associate for the presence of the list-tag appropriate to the list for which recall is being attempted. The high value of ITR\textsubscript{1,2} may be explained within the framework of this theory: Inter-item associations connect word-types (i.e., meaning-conveying nodes), and as each type-node is activated, the subject “reads off” the list tag or ordered pair of tags attached to that type-node.

Let us illustrate using the letters A, B, C, and D to represent overlapping word-types, X and Y to represent unique List-2 words, and 1 and 2 to represent possible list tags. In this notation, then, a string of inter-item associations developed during List-2 learning from Type A to B to X to C to Y to D might produce a token recall protocol such as A1, A2, B1, B2, X2, C1, Y2, D2. An alternative retrieval strategy would involve two successive passes through the A-to-D associative chain, with the subject reading off different list tags each time—thus, the protocol A1, B1, C1, then A2, B2, X2, Y2, D2. Either of these two hypothetical protocols would yield one observed intra-trial repetition unit. These hypothetical protocols also illustrate two other frequent events: an overlapping word that is as yet identified only as a List-1 word (C1), and another overlapping word which the subject has learned appears on List 2, but has forgotten appeared on List 1 (D2). The important point about these illustrative protocols is that they are exactly of the type to be expected on the basis of the Anderson and Bower theory.
Discriminative recall confidence ratings. Overall confidence ratings were computed for each of the three discriminative recall scores by averaging on each trial the confidence ratings for those words correctly recalled and identified. Confidence ratings for List-1 and for List-2 unique items recalled and correctly identified on the final trial averaged about 5.9 (on the 1 = low to 6 = high scale described previously) in all groups, indicating almost complete confidence in the assignment of list identifications to these two subsets of items.

Final-trial confidence ratings for overlapping (or in the control group, paired) List-2 identifications averaged 5.99 for control subjects and 5.90 for aware subjects, but only 5.33 for unaware subjects. Clearly, unaware subjects were less confident of their list assignments for recalled words than were the other subjects.

We noted previously that although subjects had received low-criterion instructions designed to encourage guessing, unaware subjects nevertheless showed negative transfer with respect to the controls in recalling and identifying overlapping List-2 tokens. These data contradict those of Slamecka et al. (1972); they are incompatible with a withholding hypothesis, unless, of course, unaware subjects failed to follow instructions and still withheld words. But this conjecture seems untenable, because mean confidence ratings for List-2 overlapping tokens were lower for unaware than for control subjects. Apparently, unaware subjects were recalling words of uncertain list origin and assigning them lower confidence ratings, as they had been instructed to do. They were not withholding these words. Despite this, they showed negative transfer in their discriminative recall of overlapping List-2 words. The data therefore fail to support the withholding hypothesis of negative transfer.

Type, token, and discriminative recognition. The final recognition task required list identification of all the old words plus 36 new words. Protocols for the final recognition task were scored in a manner analogous to that for the recall task, and the pattern of results was very similar.

Type recognition was defined as the total proportion of different old word-types recognized as having been old, without regard to the list identification assigned. Type recognition was virtually perfect for subjects in all three groups, with proportions recognized averaging over .99 in each.

Token recognition was defined as the total proportion of individual tokens recognized as having appeared on the training lists. Thirty-six such tokens had appeared on the lists for each group. Mean proportions recognized were .99 for controls, .92 for aware subjects, and .78 for unaware subjects.

Mean token recognition scores must be viewed in conjunction with false alarm scores, which were computed by summing the number of times subjects identified as “old” a distractor which was “new” and had not appeared on either the first or second list. The false alarm rates were .15, .01, and .08 for control, aware, and unaware subjects, respectively. When adjustment was made for false alarms by subtracting the false alarm rate from the hit rate, the mean proportions of tokens recognized became .84 for controls, .91 for aware subjects, and .70 for unaware subjects. Clearly, only unaware subjects showed negative transfer relative to the controls. Thus, while experimental subjects had no difficulty identifying the word-types which had appeared on the training lists, unaware subjects experienced considerable difficulty discriminating which types had occurred more than once.

Discriminative recognition was defined by the proportion of words recognized as old and correctly identified as to list membership. Again, these scores must be viewed in conjunction with false alarm rates. Ninety-two percent of the false alarms by control subjects were new words judged as having appeared on List 1; 76% of the false alarms by unaware subjects were also new words judged as List 1; none of
the three false alarms of aware subjects were incorrect List-1 identifications. The higher false alarm rate for control subjects probably reflects their forgetting of List-1 words due to nonpresentation of these words during second-list trials.

List-1 discriminative recognition proportions were .86 for controls, .84 for aware subjects and .76 for unaware subjects. When adjustment was made for false alarms (new words identified as List 1) by subtracting the false alarm rate from the hit rate, the adjusted mean proportions became .73 for controls, .84 for aware subjects, and .70 for unaware subjects. The adjusted mean for aware subjects was not significantly higher than that for control subjects $t(28) = 1.60, p > .10$.

Effects of false-alarm adjustments in List-2 discriminative recognition scores were trivial, and so uncorrected proportions were analyzed. Mean discriminative recognition proportions for unique words were .99 for controls, .91 for aware subjects, and .84 for unaware subjects. Mean proportions of overlapping (or in the control group, pseudo-paired) words recognized and correctly identified as to List-2 membership were .98 for control subjects, 1.00 for aware subjects, but only .70 for unaware subjects. Clearly, unaware subjects do relatively poorly at identifying overlapping words as having appeared on List 2.

In summary, it appears that for overlapping words identified as List 1, unaware subjects performed at about the same level as control subjects, whereas aware subjects performed slightly better. For overlapping words identified as List 2, unaware subjects were inferior to control subjects, while aware subjects again performed slightly better. These results are similar to those obtained for the discriminative recall task, and provide further support for those findings. Specifically, these recognition results show that difficulties involving list-identification arise for unaware subjects in processing all overlapping List-2 tokens, not just those tokens which unaware subjects happen to recall on a given trial.

**Discriminative recognition confidence ratings.** Confidence ratings for the recognition task showed lower means and higher variances than those for the recall task, probably in part because of the interspersing of new words among the old and in part because those words not recalled tended to be the ones for which subjects had the most difficulty in assigning a list identification. The pattern of results was similar to that for the discriminative recall confidence ratings, and thus does not warrant a separate presentation.

**EXPERIMENT II: WHOLE-PART TRANSFER**

We now proceed to investigate whole-part transfer, considered within the same framework distinguishing type, token, and discriminative recall. The list discrimination theory applies to whole-part as well as to part-whole transfer, and so we were particularly interested in determining whether the negative transfer typically obtained in whole-part experiments could be attributed to list-discrimination difficulties.

**Method**

**Design.** The design of Experiment II was identical to that of Experiment I, except that the second list was half as long as the first, rather than the reverse. The two sets of list items were non-overlapping for control subjects; but for experimental subjects, all the words from the second list had also appeared on the first list. Half of the experimental subjects were informed of this whole-part relationship, and half were not.

**Subjects.** The subjects were 24 male and 21 female Stanford undergraduates, serving to fulfill an introductory psychology course requirement in the academic quarter following that in which Experiment I was conducted. Subjects were tested in groups of from two to six members, and were assigned to conditions as in Experiment I, with 15 subjects in each group. Four subjects who manifestly failed to follow instructions were replaced by four further subjects.
THEORIES OF TRANSFER

Materials. A pool of 48 words was selected from the A and AA categories of Thorndike and Lorge (1944) as reported in Paivio, Yuille, and Madigan (1968). Words were selected at random from the frequency groups, except for the stipulations described in the previous experiment and the new stipulation that only nouns be used. Words were assigned randomly to one of three lists, except for the stipulation that each list consist of 16 words approximately equated for mean imagery value. Lists were completely counterbalanced among conditions.

Procedure. The testing room and initial administrative procedures were identical to those reported for Experiment I. The experimental session of approximately sixty minutes was divided into three parts. In the first, subjects received eight trials on an initial 32-word list, with instructions to recall by writing in any order as many of the words as they could in the 128 seconds allowed after each presentation. Words were presented orally at a 2-second rate, and were randomly reordered for presentation on each trial.

In the second part of the session, subjects received four trials on a 16-word list. The presentation rate was again two seconds per word. Following each presentation trial, 240 seconds were allowed for recall, list identification, and confidence-rating assignments. As in Experiment I, the subjects were asked to recall in any order all words from both Lists 1 and 2 (writing down overlapping words twice), to identify each recalled word as to its list origin, and to assign a confidence rating to each list identification. Instructions were worded so as to encourage guessing.

Following the four presentation-test cycles, the recall test booklets were collected and a recognition booklet was then distributed. The recognition task required list identifications and confidence ratings for all the old items plus 48 new items. The new items were drawn from the same word pool as the old items and met the same specifications, but they had not been presented previously. Instructions were the same as in Experiment I, although experimental subjects were not asked to comment upon any relationship they might have noticed between lists. The recognition task was untimed, and subjects were dismissed after they handed in their recognition test booklets and received a debriefing sheet.

Results and Discussion

First-list recall. First-list recall proportions pooled over trials averaged .68 for experimental-informed subjects, .67 for experimental-uninformed subjects, and .70 for control subjects. Neither experimental group mean differed significantly from the control group mean, t(28) < 1 for both comparisons. One may therefore assume that the groups were of comparable ability for the free recall task.

Pooling of experimental groups. For experimental subjects, all second-list words had previously appeared on the first list. The experimental subject was operationally defined as “unaware” of this relationship between lists if in the final recognition task he identified any word as having appeared only on the second list. Whereas in Experiment I half the experimental subjects had shown themselves to be unaware of the relationship between lists, in this experiment none of the subjects showed themselves to be unaware, that is, no experimental subject identified any word as having appeared only on the second list. Obviously, then, our uninformed group was operationally identical to our informed group. Since our only purpose in forming the two separate groups had been to manipulate awareness of the nature of the relationship between lists, and since subjects’ responses in the recognition task indicated that they had all become aware of this relationship, we shall pool the two experimental groups for the analyses to be reported. (In fact, analyses conducted with the groups separated were no more illuminating than those conducted with the groups combined.)

Why should half the experimental subjects in Experiment I, but none in Experiment II,
have remained unaware of the relationship between word lists? A plausible explanation attributes the differential awareness to the psychological asymmetry of the part–whole and whole–part tasks. Assuming that the eight trials of first-list learning have prepared the subject to recognize an overlapping word as an old type during List-2 presentation, we can conclude that the whole–part arrangement enables the subject to recognize each List-2 item as old. He can thus immediately infer on Trial 1 that List 2 is a subset of List 1. In contrast, while the part–whole subject may recognize overlapping words as old during List-2 presentation, he is in no position to know whether all of the first-list words are included in List 2; most probably, he knows only that some List-1 words are on List 2. But to distinguish the case of partial list-overlap from the case of full inclusion of List 1 in List 2 requires for part–whole subjects a non-trivial feat of remembering and comparing words.

Type and token recall. Type recall was defined as the proportion of all different types recalled during the transfer task from both lists and without regard to the list identification made. The learning curve for experimental subjects was practically indistinguishable from that for control subjects, and the mean proportion of words recalled (pooled over trials) by experimental subjects was .79, identical to that for control subjects. So with respect to type recall of all words, there was neither positive nor negative transfer.

Token recall was defined as the proportion of individual tokens recalled during the transfer task from both lists and without regard to the list identification assigned (except that for overlapping words twice recalled, experimental subjects were required to have different list identifications for each token recalled). By Trial 4, the mean proportion of tokens recalled by experimental subjects was .79, while the mean proportion recalled by control subjects was .85. The difference between means is not significant.

These type and token recall results are similar to those obtained in Experiment I. There was no negative transfer for overall type recall in either experiment. Since all whole–part subjects became aware of the relationship between lists, we would not expect negative transfer on the token recall index, and in fact, none occurred.

Discriminative recall. Discriminative recall was defined in terms of the proportion of words recalled and correctly identified as to list membership. The full set of discriminatively recalled words was partitioned into three subsets for experimental subjects and two subsets for controls. For experimental subjects, these recalled subsets were (1) unique words presented only on the first list and correctly identified as having appeared on the first list, (2) overlapping words identified as having appeared on the first list, and (3) overlapping words identified as having appeared on the second list. For control subjects, the entire set of first-list words served as the pseudo-control for both overlapping and unique words identified by experimental subjects as having appeared on List 1. List-2 words formed the other subset for control subjects.

Mean learning curves for the discriminative recall measures appear in Figure 7. The figure shows mean discriminative recall over trials for words identified by control subjects as List 2, words identified by control subjects as List 1, words identified by experimental subjects as List 2, overlapping words identified by experimental subjects as List 1, and unique first-list words identified by experimental subjects as List 1.

Control subjects showed no forgetting of the nonpresented List-1 words over the four transfer trials. Experimental subjects, on the other hand, actually showed increasing discriminative recall of List-1 unique words over trials, a 14% rise, even though these words were no longer being presented. This latter result is what one would expect according to a list discrimination theory of whole–part transfer. According to such a theory, subjects in the whole–part task are required to learn which
words from the first list were carried over to the second. In complementary fashion, they are required to learn which words from the first list were not carried over to the second list (despite the requirement to continue recalling them as List-1 words). During second-list trials, experimental subjects learn which List-1 words are also on List-2; but that learning simultaneously sharpens their ability to say which of the words they recall are not on List 2. Thus, their discriminative recall improves for List-1 unique words as a consequence of improving discriminative recall of List-2 overlapping words.

By the final trial of the transfer task, the mean proportion of List-1 unique words discriminatively recalled by experimental subjects was .76, while the proportion of List-1 words discriminatively recalled by control subjects was .84. The difference between proportions was not significant, $t(43) < 1$. The proportion of List-1 unique words recalled by experimental subjects, without regard to list identification, was .83, while the similar proportion for control subjects was .85. By the final trial, therefore, experimental subjects had no more difficulty than did control subjects in recalling List-1 unique words, although they may have had slightly more difficulty identifying the list origin of these words.

On the final transfer trial, experimental subjects discriminatively recalled as List 1 an average of .84 of the presented words; this proportion was identical to that for the controls. Clearly, experimental subjects displayed no negative transfer for overlapping words identified as being on List 1.

The major result of this experiment concerns discriminative recall of List-2 words. The mean proportion of such words recalled (averaged over trials) by experimental subjects was .62, while the mean proportion recalled by control subjects was .72. The difference between means was significant, $t(43) = 2.38$, $p < .05$. Considering only the final trial of the transfer task, experimental subjects discriminatively recalled and identified as List 2 77% of the words, whereas control subjects discriminatively recalled 88% of the words. This difference between groups was marginally significant, $t(43) = 1.94$, $p < .07$. However, the mean for the experimental group was clearly inflated by "guessing." Both Tulving's organizational theory and the Anderson-Bower list discrimination theory predict a substantial number of intrusions of unique List-1 words during List-2 recall by experimental subjects. Organizational theory would attribute these intrusions to the carryover of an inappropriate first-list organization containing unique first-list words, while list discrimination theory would attribute the intrusions to subjects' difficulties in discriminating between first-list words which are and are not carried over to the second list. Over the four trials on List 2, experimental subjects accumulated an average of 4.4 intrusions of unique List-1 words per subject, whereas control subjects accumulated only .3 List-1 intrusions on the average. When discriminative recall scores were corrected for guessing by subtracting the number of intru-
sions on each trial from the number of correct identifications, the Trial-4 mean for experimental subjects' discriminative recall of List 2 words decreased to .70, while that for control subjects remained at .88. The difference between these corrected means is now quite significant, \( t(43) = 2.75, p < .01 \).

Discriminative recall results for overlapping words may again be viewed in terms of the joint probabilities that an overlapping word is recalled and jointly identified as coming from both Lists 1 and 2. These joint probabilities ranged from .38 on Trial 1 to .71 on Trial 4 for the experimental subjects. As in Experiment I, discriminative recall of List-1 and List-2 tokens of a word-type are not statistically independent, indicating again that two tokens of a word-type tend to be recalled (or non-recalled) together.

Discriminative recall patterns obtained in Experiment II differed from those obtained in Experiment I in two important ways, both of which would be predicted by a list discrimination theory. First, aware subjects in the whole-part experiment (which in fact included all experimental subjects) showed negative transfer for discriminative recall of overlapping List-2 words, whereas aware subjects in the part-whole experiment did not. The reason for this is clear. In Experiment I, a subject's awareness of the relationship between lists enabled him to record a List-2 overlapping token for every List-1 token recalled. The strategy can not be reversed, however, because in the part-whole task not every List-2 token appears on List 1. On the other hand, in the whole-part experiment, only the reverse strategy can be employed. Awareness of the relationship between lists enables the experimental subject to write down a List-1 overlapping token for every List-2 token recalled. The strategy can not be reversed, however, because in the part-whole task not every List-2 token appears on List 1. On the other hand, in the whole-part experiment, only the reverse strategy can be employed. Awareness of the relationship between lists enables the experimental subject to write down a List-1 overlapping token for every List-2 token recalled, but not vice versa. Thus, awareness would be expected to improve discriminative recall of List-2 overlapping words but not List-1 words in the part-whole experiment, while it would be expected to improve discriminative recall of List-1 overlapping words but not List-2 words in the whole-part experiment. Such were the facts of the case. In the whole-part experiment, therefore, negative transfer would be expected for discriminative recall of overlapping List-2 words, even when subjects are aware of the relationship between lists.

The second way in which the part-whole and whole-part experimental results differed was in the groups' patterns of recall for unique words. In Experiment I, experimental and control subjects showed almost identical patterns of learning for List-2 unique words, as predicted by a list discrimination but not by an organizational theory. In Experiment II, the experimental and control subjects showed quite different patterns of recall. Recall of non-presented List-1 words by control subjects remained constant over all List-2 trials, whereas recall by experimental subjects of List-1 unique words increased over trials. This result can be explained by subjects' learning of unique List-1 tokens as complementary to the List-2 overlapping tokens—that is, experimental subjects have to learn which words are not carried over as well as which ones are. A supplementary factor could be that a formerly unretrieved List-1 unique word might become retrievable because it was associated earlier to an overlapping word-type which becomes retrievable during the course of List-2 learning. The pattern of discriminative recall for unique List-1 words in this experiment, as for unique List-2 words in Experiment I, is interpretable most readily in terms of such an augmented list discrimination theory.

Organizational analyses. Organizational analyses similar to those described in Experiment I were performed upon the recall protocols of subjects in Experiment II. First, we computed ITR\(_1\) and ITR\(_2\). ITR\(_1\) was defined in terms of the number of inter-trial repetitions from overlapping words on the last trial of the first list to overlapping words identified as List 1 on each of the second-list trials; ITR\(_2\) was defined in terms of the number of inter-trial repetitions from the last trial of the first list to overlapping words identified as List 2.
As before, one can compute $ITR_2$ only for experimental subjects, since no first-list words should have been identified as List 2 by control subjects. Recall that Figure 6 was used to clarify the meaning of the $ITR_1$ and $ITR_2$ measures.

Neither $ITR_1$ nor $ITR_2$ showed a clearcut trend across trials, and so trial means were pooled to form an overall mean for each condition. The mean level of $ITR_1$ was .87 for experimental subjects and .68 for controls. The measure correlated .36, $p > .05$, with discriminative recall of overlapping words identified as List 1. The mean level of $ITR_2$ was .89 for experimental subjects, and it correlated only .26, $p > .05$, with discriminative recall of overlapping words identified as List 2. The mean levels of $ITR$ and their correlations with discriminative recall are quite small.

These results should be viewed in terms of the amounts of organization which had developed by the end of the first-list task. The mean levels of $ITR$ between the last two trials of first-list recall were 4.77 for experimental subjects and 4.44 for control subjects. These $ITR$ values for final first-list organization are much higher than those obtained in Experiment I, perhaps because there were twice as many first-list trials, but more likely because the words used were high-frequency, concrete nouns, whereas Experiment I had used low-to-moderate frequency nouns, verbs, and adjectives. Whatever the cause of the difference, it is clear that the obtained low levels of $ITR_1$ and $ITR_2$ and the small correlations between these measures and level of recall cannot be attributed to an absence of any first-list organization to carry over to the second list. Rather, it appears that the existing first-list organization was simply not carried over to influence second-list recall.

In the whole-part paradigm, all the experimental word-types may appear in the recall protocols for both the initial and the transfer list-learning tasks. It was therefore possible to compute two $ITR$ measures from all the word-types on List 1 to all the word-types on List 2. One measure, $ITR_{1m}$, was defined in terms of the number of inter-trial repetitions from all the words recalled on the last trial of the first-list to all words identified as List 1 on each transfer trial. The other measure, $ITR_{2m}$, was defined in terms of the number of inter-trial repetitions from the last trial of the first-list task to all unique tokens identified as List 1 and all overlapping tokens identified as List 2 on each transfer trial. Means and correlations of the $ITR_{1m}$ and $ITR_{2m}$ measures were small, statistically insignificant and similar to those for the $ITR_1$ and $ITR_2$ measures, and hence a separate analysis will not be reported.

The small positive correlations of the $ITR$ measures in this experiment with the discriminative recall measures were in the direction opposite to that predicted by Tulving's organizational persistence theory. However, since negative transfer was obtained for discriminative recall of overlapping words identified as List 2, these non-negative correlations suggest that whatever is causing the negative transfer, it is not the perseveration of inappropriate first-list organization.

The last $ITR$ measure to be computed was the measure of intra-trial repetitions, $ITR_{1,2}$, described in Experiment I. This index was intended to measure within the transfer task the stereotypy of recall for overlapping List-1 and List-2 words. This communality was estimated by the extent to which pairs of List-1 overlapping tokens were recalled in the same order as pairs of List-2 tokens. The mean level of $ITR_{1,2}$ in the experimental group was 7.24. The measure correlated .38, $p < .05$, with discriminative recall of overlapping words identified as List 1, and .61, $p < .001$, with discriminative recall of overlapping words identified as List 2. These high correlations are similar to those obtained in Experiment I. The high values of $ITR_{1,2}$ can be explained as before in terms of a retrieval strategy which, during List-2 learning, utilizes a chain of learned associations among word-types, with the output mechanism "reading off" the list-tags that have been attached to the linked type nodes.
Discriminative recall confidence ratings. Discriminative recall confidence ratings for control subjects averaged 5.9 or above on all measures on all trials, and averaged 6.0 (the scale maximum) on the last trial for any given measure. For experimental subjects, mean confidence ratings increased over trials, starting around 5.3 on the first trial and ending around 5.8 on the last trial. Since negative transfer was obtained even with instructions which encouraged guessing, the lower confidence ratings of experimental than of control subjects indicates that the experimental subjects adopted a lower criterion for discriminative recall than did control subjects. Despite this lower criterion, the experimental subjects still showed negative transfer. There is thus no support for the Slamecka et al. (1972) response-withholding theory of negative transfer.

Type, token, and discriminative recognition. Type recognition scores averaged 1.00 for experimental subjects and .99 for controls. Clearly, subjects had no difficulty recognizing the word-types which they had studied. Token recognition averaged .99 for controls and .89 for experimental subjects. These proportions must again be viewed in light of the group false alarm rates, which were .07 for controls but .00 for experimental subjects. When token recognition means were corrected for guessing by subtracting the false alarm rates from the corresponding hit rates, the control and experimental mean proportions became .92 and .89, which clearly do not differ significantly.

Discriminative recognition of overlapping words identified as List 1 averaged 1.00 for experimental subjects and .95 for controls. Since almost all false alarms by control subjects were to new words identified as List 1, the score of .95 for control subjects is inflated. In any case, it is clear that as in Experiment I, experimental subjects had no difficulty in recognizing those overlapping tokens which had appeared on the first list. However, discriminative recognition of overlapping words identified as List 2 averaged only .90 for experimental subjects compared to .97 for controls. This pattern of discriminative recognition results for overlapping words is identical to that obtained in Experiment I, and is compatible with the list discrimination theory.

One surprising result among the recognition data occurred for discriminative recognition of unique List-1 words. Control subjects correctly identified 95% of such words, while experimental subjects averaged only 77% on this measure. In fact, the discriminative recognition level of experimental subjects was about the same as their discriminative recall level of 76% on Trial 4 of the transfer task. The mean confidence rating assigned to these words was 4.5 among experimental subjects compared to 5.7 among controls. Such results indicate that the greater difficulty for experimental subjects was not in learning which List-1 words were carried over to List 2, but rather in remembering which List-1 words were not carried over. This ordering of difficulty is understandable, since the List-2 overlapping words are presented on every transfer trial, whereas the List-1 unique words are not presented on any transfer trials.

In summary, the results of the discriminative recognition task reinforce again the list discrimination explanation of whole-part transfer, while providing no support for an organizational account.

General Discussion

The results presented and discussed above place the part–whole and whole–part negative transfer effects into a somewhat broader context than that in which they are usually discussed. Specifically, our results illustrate that the appearance of positive or negative transfer in free recall largely depends upon the way in which performance on the transfer task is specified and measured. The usual index, "second list recall," measures only a fraction of the subjects' knowledge about the experimental words and the contexts in which they have appeared.

Overall, the results are most consistent with a list discrimination interpretation of transfer
in the part–whole as well as in the whole–part paradigm. Of course, organizational theory might be modified to incorporate some notion of list tagging. Without prejudging the exact nature of such a modified theory, we would note that results particularly critical for it to explain are the absence of negative transfer in discriminative recall of List-2 unique words in our part–whole experiment, and the improvement in discriminative recall of List-1 unique words by our whole–part subjects. Also, whatever the list-tagging features which might be postulated for such a theory, the outcomes of our organizational analyses were incompatible with those organizational theories which relate whole-list organization to part-list organization. We are not denying the existence or general importance of inter-item associations or organizational processes in free recall learning. Their “existence” is established beyond dispute; in fact, we would cast our allegiance with associative models like FRAN (Anderson, 1972) or HAM (Anderson & Bower, 1973, Chap. 14) which provide concrete specifications of otherwise vague references to organizational factors in free recall. While not denying the simple facts of within-list organization, this paper does question whether those factors have special relevance for understanding the inter-list transfer results, namely, negative transfer in part–whole and whole–part free recall of unrelated words.

The proposed list discrimination model supposes that negative transfer occurs at the level of list-tagging, which is a required stage in most two-process theories of free recall. According to this hypothesis, subjects have difficulty associating List-2 markers (list tags) to items that have previously been associated to List-1 markers. As one consequence of this difficulty, the learning of old words repeated on List 2 is retarded; when subjects are asked to recall List-2 words, old words not yet tagged as appearing on List 2 may be thought of but then edited out by a monitor which checks for the correct list tag before overt recall of a candidate item. Hence, overlapping words may not be recalled as List-2 words, although they may well be recalled as List-1 words.

One may ask whether the list discrimination theory accounts for all the facts known about part–whole and whole–part transfer in free recall. On the face of it, this is doubtful for the simple reason that interlist transfer in free recall is a complexly determined outcome to which the variable of list-identification is but one contributor. For the standard free-recall experimental paradigm, list discrimination provides an obvious explanation, as we have shown in our experimental tests. It also helps to explain several other salient findings in the literature.

As a beginning step, we earlier discussed the important finding by Wood and Clark (1969) and by Novinski (1972) that informing experimental subjects of the relationship between lists in the part–whole task eliminates negative transfer. A list discrimination theory easily accounts for this result. Informed experimental subjects simply use a rule which states that a List-2 tag is to be assigned to any word already having associated to it a List-1 tag. Thus, there is no problem of the monitor editing out from recall overlapping words not tagged as List 2.

One of the more exotic findings in the literature is that of Schwartz and Humphreys (1973), that negative transfer can be obtained even if the first and second lists are identical (provided that subjects are led to believe that the “second list” might be different from the “first list”). Such a finding can be explained by a list discrimination theory: Following incomplete learning of the first list, subjects have difficulty in determining, during “second list” trials, which of the old items are still being presented. They have to commence a separate “List-2 tagging” process, monitoring their recall for such List-2 tags. It is difficult to see how the result could be explained in terms of an organizational theory, since any organization formed for the first list would necessarily be compatible with the second (identical) list.

A recent part–whole experiment by Carey and Okada (1973) found that if experimental
subjects are asked to recall part-list words (without re-presentation of these words) following whole-list trials, they show higher recall of these words than do control subjects for their first list (of unrelated words). This result is incompatible with organizational theory, since during whole-list learning subjects would be presumed to reorganize the S-units they formed during initial part-list learning in an attempt to establish an “optimal” organizational for the whole list. But the new organization formed for the whole list would presumably not be compatible with the subset of words in the part-list, and so experimental subjects should not recall the part-list words better than controls during the final test. The Carey and Okada (1973) result may be seen as essentially identical to our finding of positive transfer for overlapping words discriminatively recalled as List 1 by experimental subjects. This result is compatible with a two-process model of free recall which supposes that associative paths (retrieval routes) between word-types of List 1 are either maintained or increased during List-2 learning (involving these overlapping word types), whereas such paths are forgotten by control subjects who have not been recently studying the initial-list words.

Several studies in the literature presume to show that strong positive transfer is found if the organization of the whole-list is compatible with that of the part-list. Perhaps one of the most striking examples of this phenomenon was in Mandler and Dean’s (1969) Add-1 condition, in which the items presented for study on Trial \(n + 1\) were all those presented (and typically recalled) on Trial \(n\) plus one new word, added at the end of the list. This progressive part-to-whole study produced very high positive transfer. Starting on the initial trials with one, two, or three items, their subjects adopted a strictly serial ordering of items in output, recalling the first \(n\) old items in order before adding the new item as number \(n + 1\) in their recall. While not disputing the fact that the part-list (serial) organization was compatible with that of the whole list, we would also point out that the serial order of output provides for perfect list discrimination of the most-recent versus earlier items, since a subject can readily check whether all prior “part-list” items have been presented in the serial order he has memorized. Moreover, in the Mandler and Dean experiment, subjects might also have showed positive transfer because they were fully informed about the succession of incremental lists, and so had a monitoring rule to recall any retrieved word-type, regardless of which list-tag was associated to it.

Related results have been reported by Birnbaum (1969) and others. Using lists of taxonomic categories, Birnbaum found that part–whole transfer was positive either if the part-list \((a)\) contained all instances of several categories, with these intact categories used along with other categories to compose the later whole list, or \((b)\) contained two or more blocked items from several or all categories which were to appear on the whole list, with the whole list composed by adding a couple more instances to the earlier instances on the part-list. Now, recall of such categorized lists is known to depend upon retrieval of the categories and of the items within categories. Experimental subjects exposed to such part-lists will clearly have an advantage over controls in terms of learning what categories are represented on the list. Moreover, since the number of instances per category in the part-list was small (two to four), Birnbaum’s subjects should have had little difficulty detecting that all part-list instances were carried over into the augmented whole list. In terms of the list discrimination theory, these circumstances made it easy for part-whole subjects to become aware of the inter-list relations. Hence, they should use the rule that “all items with a List-1 tag should also be output as List-2 items.” This rule, combined with heightened category recall, would produce positive transfer for subjects in Birnbaum’s conditions.

Other methods of presentation have been discovered which influence the degree of part–
THEORIES OF TRANSFER

whole transfer. Ornstein (1970), for example, found that positive transfer could be produced by segregating old from new items while presenting the whole list, so that all 12 of the old (part-list) items would be presented as a block before presentation of the 12 new items as a block. Such item segregation may prevent disruption of previous subjective recall units (as in the Mandler and Dean experiment), and this may suffice to explain the positive transfer obtained by Ornstein. It is also plausible to suppose, however, that such blocked presentations help subjects to become more easily aware of the interlist relations (e.g., by counting the number of old items), and may help subjects avoid serious list identification difficulties.

A similar suggestion might be made to explain the positive part–whole transfer observed by Elmes, Roediger, Wilkinson, and Greener (1972). Their study lists were presented either one word at a time (the standard successive method) or all words at once for an equivalent total time. Positive transfer was obtained whenever the whole list was studied by the simultaneous presentation method. Again, we would suggest that subjects receiving simultaneous presentation of the whole-list could more easily recognize old list items and then, perhaps by counting them, infer the inter-list relationship (i.e., that all List-1 items occur in List 2). And we know that such awareness produces positive transfer.

Our analysis of how one or another presentation method affects the learning (or “awareness”) of inter-list relations is plausible but will not become convincing until the relevant facts are collected. The suggestions made above could be easily tested in the Ornstein or Elmes et al. procedures by simply questioning subjects over the first few trials of whole-list learning regarding their knowledge of inter-list relations.

One set of results for which there is no obvious explanation is that reported by Hicks and Young (1972). These authors obtained the standard negative transfer result using common nouns, but obtained positive transfer using adjectives as items. They suggested that these differences might be related to differential imagery values for nouns and adjectives, or to different levels of pre-experimental associations between nouns and between adjectives. However, the exact details are not understood theoretically. The reliability of the outcome needs to be assessed. Our Experiment I used a mixture of nouns, adjectives and verbs, but resulted in strong negative transfer nonetheless. Unfortunately, the number of items in each word-class was too small to permit separate analyses for each part of speech.

Our brief review of results on part–whole transfer in free recall (no comparable literature exists on the whole–part problem) suggests that the list-discrimination hypothesis is a useful adjunct for the attempts of two-process theories (e.g., Anderson & Bower, 1972; Kintsch, 1970) to explain the pattern of negative and positive results obtained in free recall experiments. The organizational persistence hypothesis is clearly not supported by our results, although there is little doubt that in specially contrived circumstances, experimentally induced organizational persistence can produce positive transfer (e.g., Birnbaum, 1970; Bower & Lesgold, 1969; Ornstein, 1970). But it does so in a manner that appears explicable as well by two-process theories of free recall.

Other results on transfer of organizational units from free recall to paired-associate learning and from paired-associate to free-recall learning are of interest in suggesting that inter-item associations (or “organizations”) develop during free recall (see Johnson, 1972; Postman, 1971) and that they can be based upon prior inter-item associations (see Wood, 1969). However, such results have no logical bearing on the central issue of inter-list transfer in free recall. If inter-item associations do play any important role in such transfer, it remains to be demonstrated what this role is, and in what ways (if any) these associations may interact with those from items to list tags.
In closing, it may be noted that negative transfer in free recall has been "reduced to" negative transfer in list-tagging. If list tags are conceived of just like other cognitive elements, then the associating of "Word A—List 1" followed by "Word A—List 2" is like the standard A-B, A-C paradigm of negative transfer in paired-associates learning (except that associations in the former case are assessed by recognition). So this theoretical approach returns us to the basic question of why negative transfer occurs in an A-B, A-C paradigm. That is a very old but poorly understood phenomenon for which there are no well-supported hypotheses. Nevertheless, we feel that theoretical progress is made whenever one phenomenon (negative transfer in part-whole free recall) is shown to be a manifestation of a second phenomenon (viz., negative transfer in associative learning).

REFERENCES


(Received July 5, 1973)