Memory & Cognition
1978, Vol. 6 (4) 354-363

Acquisition-test interactions between different dimensions of encoding

C. DONALD MORRIS
Vanderbilt University, Nashville, Tennessee 37240

The effects of changes in context on recognition memory were investigated in two experiments. In Experiment 1 target items were embedded in congruous or incongruous sentence frames and then tested in the original, new congruous, or new incongruous contexts. Experiment 2 included a third encoding condition; targets were also embedded in abstract definitional sentences. Both experiments revealed a reliable crossover interaction between the initial encoding condition and the mode of the changed context. Congruous items were recognized better in new congruous contexts, whereas incongruous items were recognized better in new incongruous contexts. Recognition of definitional items was equally impaired by both context changes. The results suggest the importance of the symmetry between acquisition and test encoding dimensions. These experiments also point out the need to distinguish between different ways in which the concept "integration" is used in current research.

A recent study by Baker and Santa (1977) examined a particular aspect of the relationship between initial learning and subsequent retrieval conditions, namely, the interaction between the degree of congruity or integration of initial encoding and subsequent changes of context at time of test. Subjects were presented target items embedded in congruous, incongruous, or anomalous sentence contexts and were then tested for recognition of these items in either the original context or a changed sentence context of the same semantic type. Results indicated that a change in context at time of test was most detrimental for targets embedded in congruous sentence frames. To account for these results, Baker and Santa (1977) suggest that "the better the original encoding, the harder it is to break that context and recognize the material in a new way" (p. 151).

It is instructive to examine Baker and Santa's claim in more detail. They suggest that better learning entails a greater degree of integration of the initial representation, hence making it more difficult to access a target in a changed context. Intuitively, however, there would appear to be many instances in which effective learning requires that one be capable of flexibly yet appropriately transferring learned material to new situations or contexts (see Bransford & Franks, 1976; Bransford & Nitsch, 1978). In order to clarify this issue, it is important to realize that the term "integration" is often used in a number of different senses. At times the term refers to the degree to which an episodic event is consistent with one's semantic knowledge of the world. For example, Santa (Note 1) has stated that this is the intended interpretation of the term integration in the Baker and Santa (1977) study. At other times, however, the term integration refers to the extent to which a target item and its immediate context may be unified. The research conducted by Rosenberg (Rosenberg, 1968, 1969; Rosenberg & Jarvella, 1970), for example, appears more consistent with the latter interpretation.

It should be pointed out that there is often no clear-cut distinction between the two senses of integration. Thus it is difficult to determine whether specific results may be attributable to the consistency of stimulus materials with prior knowledge, to trace unification, or to some combination of the two. One of the aims of the present experiments was to replicate and extend Baker and Santa's work while distinguishing between the two senses of integration. To this end, data were analyzed so that it would be possible to examine the interdependence of sentence components. At the same time, experimental materials were designed so as to vary the degree to which sentences were consistent with one's background knowledge of the world.

In addition to distinguishing between the different ways in which integration is used, the present experiments also utilized a more comprehensive design. Baker and Santa's (1977) design manipulated encoding conditions during acquisition and then tested for target recognition in old contexts and in changed sentence contexts of the same semantic type. A growing number of theorists have argued, however, that memory can be most fruitfully examined by orthogonally varying acquisition and test conditions (e.g., Bransford, Franks, Morris, & Stein, in press; Tulving, in press). Indeed, there is a substantial body of

This article is based on a doctoral dissertation submitted to Vanderbilt University. The author gratefully acknowledges Jeffrey J. Franks and John D. Bransford for their many important and helpful comments. This research was supported in part by National Science Foundation Grant BNS77-077248, awarded to J. D. Bransford and J. J. Franks.

evidence indicating that there are strong interactions between encoding and retrieval conditions (Fisher & Craik, 1977; Morris, Bransford, & Franks, 1977; Stein, in press). Therefore, the present design included a third testing condition as well: Items were also tested in changed sentence contexts of the different semantic type (e.g., congruous acquisition/incongruous test, incongruous acquisition/congruous test).

EXPERIMENT 1

Experiment 1 was designed as a basic replication and extension of the Baker and Santa (1977) study. Subjects were presented items in either congruous or incongruous contexts. Varying the nature of the sentence context was expected to affect the mode of semantic processing carried out on targets. Furthermore, the congruous and incongruous modes of encoding could be viewed as forming some sort of background context for the targets. On a subsequent recognition test, the targets appeared in the original encoding context, in a new sentence context of the same semantic type, or in a new sentence context of the different semantic type. Thus, changed sentence contexts could vary from the original contexts in their specific content and in terms of the general mode or dimension they represented.

The present experiments utilized congruous and incongruous encoding conditions as in the Baker and Santa (1977) study. It should be noted, however, that Baker and Santa's incongruous condition is unlike the incongruous conditions of other recent experiments. For instance, compare Baker and Santa's incongruous sentence, "The TRUCK was parked on top of the high school" to Schulman's (1974) queries (e.g., Is VELVET brave?"), or to Craik and Tulving's (1975) sentences (e.g., "The boy met a _____ on the street: SPEECH"). Baker and Santa's incongruous sentences are clearly not incongruous in the same way as these other sentences, suggesting that the term incongruity also means different things to different people. This suggestion is supported by contrasting one result from Baker and Santa's study with the results of some previous studies of incongruity. Specifically, Baker and Santa reported that recognition of incongruous items in old contexts was superior to recognition of congruous items in old contexts (.79 vs. .70). This finding is at odds with data reported by others (e.g., Arnold, Bower, & Bobrow, 1972; Craik & Tulving, 1975; Schulman, 1974). While recognizing the inconsistencies in usage of the term incongruous, the present experiments employed the term incongruous as it was used in the Baker and Santa (1977) study.

It was hypothesized that decrements in recognition performance would be greater for congruously encoded words than for incongruously encoded words. This hypothesis is congruent with Baker and Santa's data and, if supported, would replicate their results.

It was further hypothesized that context-related decrements within an encoding condition would be smaller when the mode of the changed context was the same as the mode of the original encoding (e.g., congruous-congruous, incongruous-incongruous) than when the mode of the changed context was different from the original mode of encoding (e.g., congruous-incongruous, incongruous-congruous). Support for this hypothesis would also suggest that the abstract semantic dimension along which items are processed may form part of the encoding context of the items. Text contexts that reinstate the domain of the original encoding should provide superior performance to test contexts that fail to reinstate the original encoding domain.

Method

Materials. Forty-eight concrete nouns were embedded within sentence contexts and presented to subjects in list form. The sentences were typed with the target nouns appearing in capital letters. Half of these words were embedded within congruous sentence frames (e.g., "The PICKLE was served with the slaw") and half were embedded within incongruous frames (e.g., "The PICKLE jammed the saxophone"). Congruous contexts described regular, normatively typical situations in which the target noun could occur; incongruous contexts described unique, somewhat unusual situations that could possibly, although not likely, occur.

Two study lists were used in order to counterbalance target items with sentence context: A target that appeared in a congruous context in one list appeared in an incongruous context in the other list. The part of speech of a target (subject noun, direct object, etc.) and its position within the sentence were constant across both lists. A different random order of target presentation was used for each list.

Thirty-six of the 48 nouns presented during acquisition were designated as targets on a subsequent recognition test. The 12 remaining items, which were the same for each list, were designated as filler items. In addition, 24 new concrete nouns were chosen as foil items. Thus, a total of 72 critical nouns was used in this experiment. None of these nouns appeared elsewhere on the study list.

The recognition test consisted of 60 of the 72 nouns, embedded in old, new congruous (e.g., "The PICKLE was on top of the sandwich"), or new incongruous (e.g., "The PICKLE was cut by the chain saw") contexts. Twelve of the 36 targets (six from each of the two acquisition encoding conditions) appeared in each of the three test contexts. One-hundred-and-forty-four sentences (36 nouns by 4 conditions) were used in conjunction with the 36 target items. Twelve of the foil items appeared in completely new sentence contexts; the remaining 12 foils were placed in the sentence frames affiliated with the 12 acquisition filler items. In both of these cases, half the sentence contexts were congruous and half were incongruous. Twelve sentences were constructed for foils in completely new contexts and 24 sentences were used for foils in filler frames (12 items by 2 conditions). Overall, a total of 180 sentences was used in this experiment.

Six different forms of the recognition test were given, three for each of the two study lists. These forms corresponded to the six possible combinations of acquisition encoding condition and test context. Thus, each target was encoded in a congruous or an incongruous context and tested in its old, a new congruous, or a new incongruous context. Foil items that appeared in completely new contexts were constant across all six test forms. Foils that appeared in old filler frames appeared in one context for half the test forms and in the other

context for the remaining test forms. Each of the six test forms utilized a different random order of target presentation. The part of speech of a target and its position within a sentence were held constant for all test forms. Different sentence contexts for an item always biased the same sense of a word in order to control for homographic or polysemic factors.

Design. A 2 by 3 by 2 by 6 mixed factorial design was utilized. Three factors were manipulated within subjects: initial encoding condition (congruous or incongruous), sentence context at time of test (old, new congruous, or new incongruous), and type of embedded item (target or distractor). The fourth factor (lists) was varied between subjects. Thirty-six students from an introductory psychology class served as subjects in this experiment. Eighteen subjects received each of the two acquisition lists; six subjects were tested on each of the six forms of the recognition test.

Procedure. Subjects were given examples of congruous and incongruous sentence contexts and instructed that they were to classify the acquisition sentences according to these semantic contexts. Subjects were paced by the experimenter and asked to read the acquisition sentence carefully, to pay particular attention to the capitalized target, and to perform the context classification task. A 4-sec interval was allowed for each sentence. Subjects were informed that they would later be tested for memory of both sentence contexts and capitalized words.

The recognition test involved three separate tasks. First, subjects were asked to classify each test sentence as either congruous or incongruous. This task was included to induce semantic processing of the test sentence frames and targets in congruous and incongruous modes comparable to the two types of acquisition processing modes. Second, subjects were asked to decide whether a given sentence frame was old or new, regardless of whether the capitalized word was the same or different. The final task was to judge whether a given capitalized word had been seen before, regardless of whether the sentence context was old or new. Subjects were given as much time as necessary to complete the tasks.

Results

Analyses were performed on: (1) target recognition scores, (2) sentence-frame recognition scores, and (3) the conditional recognition of targets in old contexts given recognition of the corresponding sentence frame. Each of these analyses was based on recognition scores as they fit into experimentally defined categories (i.e., encoding condition by sentence context).

The analysis for targets were based on scores obtained in the following manner: Hit rates were calculated for each of the six experimental cells (2 encoding conditions by 3 sentence contexts). Two different false positive rates (see Table 1) were subtracted from these hit rates to serve as a correction for guessing. One rate was based on false positive responses to new target items appearing in old sentence frames. This correction was used to account for any tendencies to respond positively to target items merely because an old sentence frame was present. The second false positive rate was based on false positive responses to new target items appearing in new sentence frames. This correction was used to account for any differential response-bias effects related to the congruous and incongruous dimensions of encoding.

It is important to note that the former correction procedure is based on the assumption that recognition of target items and sentence frames are independent. This assumption is highly questionable, particularly for studies where the degree of integration between target and sentence frame is an experimental issue. For these reasons, the present analyses were based on target hit rates corrected by false positive rates to new items in new frames. It should be noted that the same pattern of *difference* scores was obtained from both of the correction procedures; thus, interpretation of the results cannot be equivocated on these grounds. The mean recognition scores for target items and both false positive rates are presented in Table 1.

A 2 by 3 by 6 analysis of variance (encoding condition by sentence context by lists) was performed on the corrected target recognition scores. There was a significant main effect for sentence context at time of test [$F(2,60) = 38.26$, $p < .001$], indicating that recognition performance was better when items were tested in their old context. No other main effects were significant. There was a significant interaction between encoding condition and sentence context [$F(2,60) = 5.38$, $p < .007$]. The nature of this interaction is described more fully below. There was a significant Lists by Sentence Context interaction [$F(10,60) = 3.01$, $p < .004$], and the three-way interaction approached, but failed to reach, conventional levels of significance.

Dunn's planned comparisons were performed on means associated with the Encoding Condition by Sentence Context interaction. These comparisons revealed that congruously encoded items were better recognized in their old contexts than in either new congruous contexts or new incongruous contexts (both $ps < .01$). Likewise, incongruously encoded items were

Table 1
Target Recognition and Probabilities of Hits and False Alarms (Experiment 1)

	Encoding Condition							
	Congruous				Incongruous			
Text Context	p(R)*	p(H)	p(FA-N)	p(FA-O)	p(R)	p(H)	p(FA-N)	p(FA-O)
Old	.775	.932	.157	.191	.747	.928	.181	.120
New Congruous	.594	.751	.157	.191	.596	.777	.181	.120
New Incongruous	.532	.689	.157	.191	.641	.822	.181	.120

*Note–p(R) = p(recognition; p(H) = p(hits); p(FA-N) = p(false alarms to new targets in new frames); p(FA-O) = p(false alarms to new targets in old frames. *Correction for guessing is p(H) – p(FA-N).*

better recognized in their old contexts than in either new congruous contexts ($p < .01$) or new incongruous contexts ($p < .05$). There were no significant differences between congruous and incongruous acquisition for words tested in old contexts or in new congruous contests. Incongruously encoded words were significantly better recognized than congruously encoded words when tested in new incongruous contexts ($p < .01$).

An additional specific comparison was made in order to determine the presence of a "transfer-appropriate processing" effect. Such a comparison tests the degree to which recognition scores for congruously encoded and incongruously encoded words differentially vary as a function of the mode of the changed testing context. For example, the mean corrected recognition score for congruous items appearing in new congruous contexts is .594, compared to .532 for items appearing in new incongruous contexts. The corresponding scores for incongruous tems are .596 and .641, respectively. A t test of the difference between differences suggests evidence for a transfer-appropriate processing effect [$t(60) = 3.45$, $p < .01$]. Recognition performance improves when the abstract encoding dimensions underlying acquisition and test are symmetrical.

Baker and Santa (1977) report that congruous items suffered the most from changes in context. The present experiment, however, utilized two different kinds of changed contexts–one in which the dimension of the changed context was the same as the initial encoding dimension (hereafter, the transfer-appropriate context) and one in which the two dimensions are different (the transfer-inappropriate context). A final set of comparisons tested the extent to which congruously and incongruously encoded items were differentially affected by these changes of context. T tests of the difference between differences were performed for changes from the old context to (1) the transfer-appropriate context and to (2) the transfer-inappropriate context. The first of these tests revealed that the difference score for the congruous condition (.775 – .594) was significantly larger than the difference score for the incongruous condition (.747 – .641) [$t(60) = 2.43$, $p < .02$], thus replicating Baker and Santa's findings. The second test revealed a similar pattern. The difference score for the congruous condition (.755 – .532) was significantly larger than the difference score for the incongruous condition (.747 – .596) [$t(60) = 2.98$, $p < .01$]. These two comparisons suggest that incongruously encoded items are less susceptible to context-related decrements whether the mode of the changed context is the same as or different from the original encoding mode.

One possible reason for the above effect is that changes in meaning from the original incongruous contexts to new incongruous contexts were simply not as great as the changes from the original congruous contexts to new congruous contexts. There are no established criteria for comparing such changes; nevertheless, there is a need to examine this issue. (Baker and Santa failed to address this concern explicitly.) Four judges were given the set of four sentences (two congruous and two incongruous) affiliated with each target and asked to judge by whatever criteria they wished whether the two congruous sentences were more or less similar than the two incongruous sentences. Seventy-four percent of the pairs of congruous sentences were rated as being more similar in meaning, whereas only 11% of the pairs of incongruous sentences were rated as being more similar. The remaining 15 percentage points reflect instances where congruous and incongruous pairs were judged to be equally similar. Thus, the larger context-related decrements associated with the congruous encodings may not be attributed to a greater dissimilarity of acquisition and test contexts for congruous items.

In addition to the target analyses, sentence-frame scores were also examined. A 2 by 2 by 6 (encoding condition by embedded item type by lists) analysis of variance was performed on corrected sentence-frame recognition scores. As was the case for target recognition scores, false positive corrections were based on responses to new items in new frames. A summary of the mean corrected sentence-frame recognition scores is presented in Table 2.

The analysis of variance revealed a significant main effect for embedded item type [$F(1,30) = 30.80$, $p < .001$], indicating that sentence frames were better recognized when they contained the original target item than when they contained a foil item. There were no other significant main effects. The Embedded Item Type

Table 2
Sentence-Frame Recognition and Probabilities of Hits and False Alarms (Experiment 1)

Embedded Item Type	Encoding Condition							
	Congruous				Incongruous			
	p(R)*	p(H)	p(FA-N)	p(FA-O)	p(R)	p(H)	p(FA-N)	p(FA-O)
Target	.843	.811	.032	.074	.869	.801	.068	.068
Distractor	.605	.573	.032	.074	.593	.525	.068	.068

*Note–p(R) = p(recognition); p(H) = p(hits); p(FA-N) = p(false alarms to new frames with new targets); p(FA-O) = p(false alarms to new frames with old targets). *Correction for guessing is p(H) – p(FA-N).*

by Lists interaction was significant [$F(5,30) = 2.98$, $p < .03$].

Baker and Santa (1977) claimed that congruously encoded items are better integrated than incongruously encoded items. Assume that congruous encodings do provide a greater degree of integration in the sense of consistency with preexperimental knowledge. Granted that this is so, then Baker and Santa's results and the present results indicate that the greater the degree of integration, the harder it is to recognize items when they appear in new contexts. But it may also be the case that congruous encodings provide traces that are better integrated in the sense of having a greater degree of unification between the target and its sentence context (see, for example, Rosenberg, 1969). This possibility was not explicitly examined in the Baker and Santa study. Investigation of this possibility will permit a clearer understanding of how integration affects recognition performance.

In order to determine the extent of unification between targets and their contexts, a conditional analysis was performed on recognition scores of targets in old contexts. The scores used in this conditional analysis are presented in Table 3. The top row of scores presents hit rates for targets given that subjects correctly recognized the corresponding sentence frame. The bottom row, on the other hand, presents the raw target hit rates as presented in Table 1; these scores were obtained without regard to recognition of the corresponding sentence frame. The scores in the middle row represent target hit rates given that subjects did not correctly recognize the corresponding sentence frame; these scores were not included in the actual analysis. Note that the conditional scores in the top row were not obtained by multiplying the raw target hit rate by the raw sentence-frame recognition rate. Rather, these conditional scores represent the proportion of targets recognized given that the corresponding sentence frame had also been recognized. Note further that the contingency analysis was performed on uncorrected hit rates, since the false positive corrections for target recognition were the same for the conditional and nonconditional scores.

A 2 by 2 by 6 (encoding condition by conditional vs. nonconditional by lists) analysis of variance revealed a significant main effect for the conditional vs. nonconditional factor [$F(1,30) = 25.01$, $p < .001$] and a significant interaction between this factor and encoding condition [$F(1,30) = 6.09$, $p < .02$]. The first finding indicates that the nonconditional recognition scores were significantly higher than the conditional scores; thus, subjects could to a certain extent recognize target items without recognizing the corresponding sentence context. The interaction effect is due to the fact that the conditional score for congruous items (.852) was substantially lower than the conditional score for incongruous items (.901) and the nonconditional scores (.932 for congruous items, .928 for incongruous items). These scores indicate that there were substantially fewer cases in which an incongruous target was recognized but its sentence context was not than was the case for congruous encodings. These results suggest that incongruously encoded items are more highly integrated with their encoding contexts than are congruously encoded items, where "integrated" refers to the interdependence of target recognition and recognition of the corresponding sentence frame.

Discussion

The results of Experiment 1 are in general agreement with the experimental hypotheses, thus serving to replicate and extend the results of Baker and Santa (1977). Incongruously encoded items were less susceptible to changes in context than were congruously encoded items. This effect was found regardless of whether the mode of the changed context was the same as or different from the initial encoding mode.

The present results also indicate the presence of a "transfer-appropriate processing" effect (see Morris et al., 1977). Incongruously encoded targets were better recognized in new incongruous contexts than in new congruous contexts; for congruous items the converse was true. Baker and Santa's (1977) data point to the importance of the semantic content of the sentence frames which comprises an item's encoding context. The present data point to the importance of the dimension of encoding which underlies the content as well. Thus, orthogonally varying acquisition

Table 3
Probabilities of Hits for Targets (Experiments 1 and 2)

	Encoding Condition				
	Experiment 1		Experiment 2		
Target Contingency	C	I	D	C	I
Conditional Target Hit Rate, given recognition of the corresponding sentence frame	.852	.901	.767	.647	.781
Conditional Target Hit Rate, given that the corresponding sentence frame had not been recognized	.080	.027	.133	.188	.072
Nonconditional Target Hit Rate	.932	.928	.900	.835	.853

Note–C = congruous; I = incongruous; D = definitional.

and test contexts allows one to determine more exactly the relational nature of memory.

The present study may help clarify the sense in which the term integration is important for the present paradigm. For example, the present results indicate that items that are less consistent with one's general knowledge of the world (i.e., incongruous) are more easily transferred to new contexts. The degree to which items are unified with their immediate context does not appear to be the critical factor accounting for Baker and Santa's (1977) data or the present data. It is important to note, however, that this factor may play an important role in other experimental situations (e.g., Rosenberg & Jarvella, 1970).

Baker and Santa's (1977) study and the present study have assumed that congruous and incongruous encodings permit varying degrees of differentiation of experimental materials in terms of one's general knowledge of the world (as opposed to differentiation of stimulus materials within the context of a given experiment; see Moscovitch & Craik, 1976; Stein, 1977). More specifically, it has been assumed that incongruous encodings provide representations that are more distinctive from background knowledge than do congruous encodings. These incongruous encodings may thus be viewed as "episodes" that are somehow noted or referenced as being relatively distinct from background experiences. They may therefore be more easily recognized in new contexts.

It is interesting to consider what would happen if a third dimension of encoding was also used. For example, an abstract definitional mode of encoding would represent a case in which sentences were relatively poorly differentiated from background knowledge compared to congruous and especially incongruous encodings. This is because definitions are by their nature very general and capable of incorporating a wide range of possible instantiations. Thus, given that abstract definitional encodings are relatively poorly differentiated from background knowledge, one would expect even greater decrements in recognition performance as a function of context change than were found for congruous encodings.

A problem arises in examining this possibility. The methodology of transfer-appropriate processing requires that initial encoding conditions and subsequent testing conditions be orthogonally crossed. It is difficult, if not impossible, to include a testing condition involving new definitional contexts. Alternate definitional contexts for a target generally result in a mere paraphrase of the initial definition or they suggest a different meaning of the target. In spite of this limitation, it is still possible to investigate the effects of changed contexts on items encoded on an abstract definitional dimension.

Note that in Experiment 1 the recognition decrement for congruous items from old to new incongruous contexts was greater than the decrement for incongruous items from old to new congruous contexts. These context changes have been referred to previously as transfer-inappropriate changes. It should therefore be possible to determine how transfer-inappropriate changes of context affect definitionally encoded items.

Two pairs of experimental hypotheses were made: (1) The recognition decrement from old to new incongruous contexts should be larger for definitional items than for congruous items. Likewise, the decrement from old to new congruous contexts should be larger for definitional items than for incongruous items. (2) The recognition decrement from old to new incongruous contexts should be larger for definitional items than for incongruous items. Likewise, the decrement from old to new congruous contexts should be larger for definitional than for congruous items. Note that the first pair of hypotheses is based on transfer-inappropriate changes of context for congruous and incongruous items, whereas the second pair is based upon transfer-appropriate changes. Both pairs of hypotheses are based on the assumption that definitional items are less differentiated from background knowledge and are therefore not as easily transferred to new contexts.

EXPERIMENT 2

Method

Materials. Sixty concrete nouns were embedded within definitional, congruous, or incongruous sentence frames. In many cases the congruous and incongruous frames from Experiment 1 were used. It was, however, necessary to modify some of the congruous sentences that tended to be somewhat abstract and similar to the definitional sentences. These sentences were altered to provide more concrete, particular contexts. For example, the sentence, "The food was cooked on top of the STOVE," was changed to "The bacon was fried on top of the STOVE." Definitional sentences were based on primary dictionary definitions but were modified in order to make them more "readable" and to equate for sentence-frame length (e.g., "A TRUCK is a vehicle for transporting heavy articles").

The recognition test consisted of 75 nouns, embedded in old, new congruous, or new incongruous contexts. Forty-five of these nouns were designated as test targets and had appeared in the study lists. The remaining 15 acquisition items served as filler items. Thirty additional items served as recognition foil items. The acquisition and test lists were constructed and counterbalanced in a manner similar to those of Experiment 1.

Design. A 3 by 3 by 3 by 6 mixed factorial design was utilized. The design was similar to that of Experiment 1 except for the addition of a third acquisition encoding condition (the definitional condition). Forty-five volunteer subjects from the university community served as subjects in this experiment.

Procedure. The procedure was identical to that of Experiment 1 except that test sentences could be classified as definitional, congruous, or incongruous.

Results

The same basic analyses were performed as in Experiment 1. These analyses were based on scores corrected by false positive responses to new items in

Table 4
Target Recognition and Probabilities of Hits and False Alarms (Experiment 2)

	Encoding Condition											
	Definitional				Congruous				Incongruous			
Text Context	p(R)*	p(H)	p(FA-N)	p(FA-O)	p(R)	p(H)	p(FA-N)	p(FA-O)	p(R)	p(H)	p(FA-N)	p(FA-O)
Old	.739	.906	.167	.200	.679	.829	.150	.222	.637	.854	.217	.206
New Congruous	.538	.705	.167	.200	.560	.710	.150	.222	.456	.673	.217	.206
New Incongruous	.528	.695	.167	.200	.467	.617	.150	.222	.549	.766	.217	.206

*Note–p(R) = p(recognition); p(H) = p(hits); p(FA-N) = p(false alarms to new targets in new frames); p(FA-O) = p(false alarms to new targets in old frames). *Correction for guessing is p(H) – p(FA-N).*

new frames. The mean corrected recognition scores for target items are presented in Table 4.

A 3 by 3 by 9 (encoding condition by sentence context by lists) analysis of variance was performed on the corrected target recognition scores. There was a significant main effect for sentence context at time of test [$F(4,108) = 28.25$, $p < .001$], indicating that recognition performance was better when items were tested in their old context. No other main effects were significant. There was a significant interaction between encoding condition and sentence context [$F(32,108) = 3.01$, $p < .02$]. The nature of this interaction is described more fully below. Finally, the three-way interaction was significant [$F(32,108) = 1.71$, $p < .02$].

One of the aims of the present experiment was to replicate Experiment 1. To this end, Dunn's planned comparisons were performed on the means of the congruous and incongruous encoding conditions. These comparisons revealed that congruous items were significantly better recognized in old contexts than in either new congruous contexts ($p < .05$) or in new incongruous contexts ($p < .01$). Incongruous items were significantly better recognized in old contexts than in new congruous contexts ($p < .01$). These findings replicate the results of Experiment 1. It is important to note, however, that the recognition score for incongruous items in new incongruous contexts failed to differ significantly from either the score for incongruous items in old contexts or the score for congruous items in new incongruous contexts. Each of these differences was significant in Experiment 1, although only the difference for incongruous items in old vs. new incongruous contexts is of theoretical relevance.

An additional specific comparison tested for the presence of a transfer-appropriate processing effect for the congruous and incongruous conditions. A t test of the difference between differences [(.560 – .467) – (.456 – .549)] indicates that such an effect was indeed present [$t(108) = 4.48$, $p < .001$]. Congruous items were better recognized in new congruous contexts than in new incongruous contexts, while incongruous items were better recognized in new incongruous contexts than in new congruous contexts. This finding replicates Experiment 1 and suggests that the congruous and incongruous modes activated during acquisition interact substantially with the abstract dimensions that underlie the changed test contexts.

A further set of comparisons tested the extent to which congruous and incongruous encodings were differentially affected by changes of context at time of test. T tests of the difference between differences were performed for changes from the old context to (1) the transfer-appropriate context and to (2) the transfer-inappropriate context. The first test compared the difference between congruous items tested in old and new congruous contexts (.679 and .560, respectively) with the difference between incongruous items tested in old and new incongruous contexts (.637 and .549, respectively). The second test compared the difference between congruous items tested in old and new incongruous contexts (.679 and .467, respectively) with the difference between incongruous items tested in old and new congruous contexts (.637 and .456, respectively). Neither of these comparisons revealed any significant differences ($t < 1.00$ for both tests). Although these findings are in the direction predicted by the results of Experiment 1, it is important to note that the present findings fail to replicate the results of Experiment 1.

Two final sets of comparisons were performed in order to determine whether the context-related decrements for definitional encodings would be larger than the decrements for congruous and incongruous encodings, since they were less differentiated from background knowledge. The first set of comparisons involved transfer-inappropriate changes of context for the congruous and incongruous items. The decrement for definitional items from old to new incongruous contexts (.739 – .528) was compared to the decrement for congruous items from old to new incongruous contexts (.679 – .467). Also, the decrement for definitional items from old to new congruous contexts (.739 – .538) was compared to the decrement for incongruous items from old to new congruous contexts (.637 – .456). Neither of these comparisons was significant ($t < 1.00$ for both comparisons). These results, along with the results reported in the above paragraph, fail to support the

hypothesis that encodings that are less differentiable from background knowledge should show larger context-related decrements than better differentiated encodings.

The second set of comparisons involved transfer-appropriate changes of context for the congruous and incongruous items. The decrement for definitional items from old to new congruous contexts was compared to the decrement for congruous items from old to new congruous contexts (.679 – .560). This comparison was significant [$t(108) = 1.98$, $p < .05$]. The decrement for definitional items from old to new incongruous contexts was compared to the decrement for incongruous items from old to new incongruous contexts (.637 – .549). This comparison was also significant [$t(108) = 2.96$, $p < .01$]. The latter two comparisons, interpreted in conjunction with the comparisons discussed in the previous paragraph, suggest that recognition decrements due to changes in context are not greatly affected by the dimensions of encoding per se. The transfer appropriateness of the test encoding dimension relative to the original acquisition dimension of encoding, on the other hand, has substantial effects on recognition performance.

A 3 by 2 by 9 analysis of variance (encoding condition by embedded item type by lists) was performed on the corrected sentence-frame recognition scores. As in Experiment 1, false positive corrections were based on responses to new items in new frames. The mean corrected sentence-frame recognition scores are presented in Table 5.

There was a significant main effect for embedded item type [$F(1,27) = 4.36$, $p < .04$], indicating that sentence frames were better recognized when they contained the original target than when they contained a foil. No other main effects were significant. There was a significant Lists by Embedded Item Type interaction [$F(8,27) = 2.68$, $p < .03$] and a significant three-way interaction [$F(16,54) = 1.87$, $p < .05$].

Finally, a conditional analysis was performed on recognition of targets in old contexts, given recognition of the corresponding sentence frame. The conditional and nonconditional hit rates for Experiment 2 are presented in Table 3. A 2 by 3 by 9 analysis of variance (conditional vs. nonconditional by encoding condition by lists) revealed a significant main effect for the conditional vs. nonconditional factor [$F(1,27) = 39.01$, $p < .001$] and a significant interaction between this factor and encoding condition [$F(2,54) = 3.65$, $p < .04$]. The first finding indicates that the nonconditional recognition scores were significantly higher than the conditional scores. Thus, to a certain extent, subjects were able to recognize targets without recognizing the corresponding sentence context. The interaction effect is attributable to the fact that the difference between conditional and nonconditional scores is larger for the congruous condition (.646 and .835, respectively) than for the incongruous condition (.781 and .853, respectively) [$t(54) = 3.85$, $p < .001$]. Likewise, the difference for the definitional condition (.767 and .900, respectively) is larger than the difference for the incongruous condition [$t(54) = 2.01$, $p < .05$]. The differences for the congruous and definitional conditions did not differ significantly. It appears from these analyses that incongruous items are somewhat better unified with their encoding contexts than congruous items, as in Experiment 1. Definitional items appear to be intermediate with respect to degree of integration. These results, together with the target recognition results, suggest that the term integration may refer to two somewhat distinct factors and that it is important to specify which of these factors might be responsible for the observed results.

Discussion

Experiment 2 was designed with two purposes in mind. One purpose was to replicate the results of Experiment 1 using the same congruous and incongruous encoding conditions. The second purpose was to investigate the effects of changes in context on abstract definitionally encoded items.

The results of Experiment 2 replicate the results of Experiment 1 in a number of important respects. Most importantly, a transfer-appropriate processing effect was again demonstrated for the incongruous and congruous encoding conditions. Congruous items were better recognized in new congruous contexts than in new incongruous contexts, whereas incongruous items were better recognized in new incongruous contexts than in new congruous contexts. This effect is consistent with the encoding specificity principle (e.g., Tulving, in press). It is necessary to note, however, that it is important to consider the abstract dimension

Table 5
Sentence-Frame Recognition and Probabilities of Hits and False Alarms (Experiment 2)

Embedded Item Type	Encoding Condition											
	Definitional				Congruous				Incongruous			
	p(R)*	p(H)	p(FA-N)	p(FA-O)	p(R)	p(H)	p(FA-N)	p(FA-O)	p(R)	p(H)	p(FA-N)	p(FA-O)
Target	.694	.766	.072	.053	.653	.708	.055	.075	.781	.848	.067	.082
Distractor	.639	.711	.072	.053	.589	.644	.055	.075	.078	.745	.067	.082

*Note–p(R) = p(recognition); p(H) = p(hits); p(FA-N) = p(false alarms to new frames with new targets); p(FA-O) = p(false alarms to new frames with old targets). *Correction for guessing is p(H) – p(FA-N).*

underlying the encoding context as well as the actual semantic content of the encoding context.

The data from Experiment 2 also indicate that incongruous encodings were more highly unified than were congruous encodings, as in Experiment 1, whereas definitional encodings were intermediate with respect to degree of unification. These findings, together with the context-related decrements in recognition performance, emphasize the need to clarify what is meant by the term integration.

The results of Experiment 2 fail to replicate Experiment 1 in one important respect. Incongruously encoded words were not significantly less susceptible to changes in context than congruously encoded words. This finding also runs counter to Baker and Santa's (1977) findings. It is important to note, however, that the present pattern of data is similar to that of Experiment 1 and to that reported by Baker and Santa. This failure to replicate suggests that the incongruous dimension of encoding per se may not necessarily result in traces that are relatively easily transferred to new contexts.

One possible reason for the failure to replicate either Experiment 1 or Baker and Santa (1977) is based on the different overall context provided by the encoding conditions used in the different experiments. Baker and Santa presented congruous, incongruous, and anomalous encoding conditions, whereas Experiment 1 utilized congruous and incongruous conditions. In both of these experiments, the overall context of the study was shifted away from the congruous mode of encoding, relative to Experiment 2. In Experiment 2 the majority of sentences reflected normative knowledge about the world (i.e., both the congruous and definitional conditions provided normatively typical contexts for the target items); thus, subjects may have had some sort of set toward normativity during the recognition test.

The second purpose of Experiment 2 was to investigate the extent to which changes in context affected definitionally encoded items. The difference scores for these items in old vs. new contexts indicate that changes in context debilitate memory performance, but that the extent of debilitation is roughly equivalent to that found for congruous and incongruous items when tested in transfer-inappropriate contexts. At the same time, the difference scores for definitional items are significantly larger than the decrements found for congruous and incongruous items tested in transfer-appropriate contexts. Thus, these results can be interpreted as indicating that the crucial factor determining memory performance in the present experiment is whether or not the encoding dimension specified by the testing context is appropriate to the original encoding dimension.

It is important to consider one final point. The conclusions regarding context-related decrement scores for definitional items may be equivocated somewhat by the higher initial levels of recognition (i.e., in old contexts) for these items. Thus, the difference scores for the definitional items may not be strictly comparable to the difference scores for congruous and incongruous items.

CONCLUSIONS

Overall, the results of Experiments 1 and 2 raise a number of points about studies investigating integration and transfer of learned material to new contexts. Integration may refer to the degree to which items are consistent with prior knowledge or to the degree of unification between a target and its immediate context. The results of Baker and Santa's (1977) study and the present studies support the conclusion that more highly integrated encodings are more difficult to utilize in new situations, *if* one interprets integration to mean consistency with prior knowledge. A quite different conclusion would be reached, however, if one chose to interpret integration in the second sense noted above: More highly integrated encodings are not necessarily more difficult to transfer to new situations.

Second, the present studies serve to clarify further the relation between initial learning and subsequent retrieval conditions. A number of theorists (e.g., Bransford et al., in press; Tulving, in press) have argued that claims about the goodness of acquisition encoding activities must take into account the nature of the testing situation. In the present experiments, there was a strong crossover interaction between the congruous and incongruous encoding dimensions, providing evidence for the above view. Note that this crossover interaction depends not only on the content of the semantic contexts, but on the abstract dimensions of encoding underlying the contexts as well.

Given the effects of abstract encoding dimensions on recognition performance, it appears reasonable to speculate that the overall list contexts in which the dimensions are embedded may also exert some influence on these kinds of memory phenomena. For example, the different overall contexts provided by Experiments 1 and 2 may account for the presence or absence, respectively, of the smaller context-related decrements associated with the incongruous encoding condition. The extent to which such overall contexts may affect memory performance must await further research. If these effects are present, they would provide important constraints about what kinds of conclusions may be drawn from this and similar types of research.

REFERENCE NOTE

1. Santa, J. Personal communication, 1977.

REFERENCES

ARNOLD, P. G., BOWER, G. H., & BOBROW, S. A. Mediated semantic compatibility in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 239-242.

BAKER, L., & SANTA, J. Semantic integration and context. *Memory & Cognition*, 1977, **5**, 151-154.

BRANSFORD, J. D., & FRANKS, J. J. Toward a framework for understanding learning. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1976.

BRANSFORD, J. D., FRANKS, J. J., MORRIS, C. D., & STEIN, B. S. An analysis of memory theories from the perspective of problems in learning. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory*. Hillsdale, N.J: Lawrence Erlbaum, in press.

BRANSFORD, J. D., & NITSCH, K. E. Coming to understand things we could not previously understand. In J. F. Kavanagh & W. P. Strange (Eds.), *Speech and language in the laboratory, school, and clinic*. Cambridge, Mass: MIT Press, 1978.

CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, **104**, 268-294.

FISHER, R. P., & CRAIK, F. I. M. Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 701-711.

MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 519-533.

MOSCOVITCH, M., & CRAIK, F. I. M. Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 447-458.

ROSENBERG, S. Association and phrase structure in sentence recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1077-1081.

ROSENBERG, S. The recall of verbal material accompanying semantically well-integrated and semantically poorly-integrated sentences. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 732-736.

ROSENBERG, S., & JARVELLA, R. J. Semantic integration and sentence perception. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 548-553.

SCHULMAN, A. I. Memory for words recently classified. *Memory & Cognition*, 1974, **2**, 47-52.

STEIN, B. S. The effects of cue-target uniqueness on cued recall performance. *Memory & Cognition*, 1977, **5**, 319-322.

STEIN, B. S. Depth of processing reexamined: The effects of the precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior*, in press.

TULVING, E. Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory*. Hillsdale, N.J: Lawrence Erlbaum, in press.

(Received for publication October 4, 1977;
accepted March 13, 1978.)