

The effects of complexity on confidence ratings in linguistic integration

CARLTON T. JAMES, MICHAEL L. HILLINGER, and BRIAN J. MURPHY
Rutgers University, New Brunswick, New Jersey 08903

Bransford and Franks (1971) introduced a paradigm for the study of linguistic integration. Their primary measure was subjects' rated confidence in their responses. The present paper considers the appropriateness of this dependent variable, and concludes that the relation between sentence complexity and rated confidence is too small and unstable to justify the emphasis which confidence ratings have received. Further, certain differences between concrete and abstract sentences were obscured by a confidence analysis. The proportion of old responses is championed as a more appropriate dependent variable. Finally, it is shown that increasing the amount of specific memory necessarily decreases the size of the complexity effect.

There is growing literature concerned with the issue of linguistic integration, primarily stimulated by the work of Bransford and Franks (1971, 1973; Franks & Bransford, 1972, 1974). The paradigm introduced by these authors involves the presentation of sentences containing different combinations of simple ideas derived from a complex idea. The following four simple ideas, *The ants were in the kitchen*, *The ants ate the jelly*, *The jelly was sweet*, and *The jelly was on the table*, can be presented individually or in combinations of two (*The sweet jelly was on the table*), three (*The ants in the kitchen ate the jelly which was on the table*), or four ideas (*The ants in the kitchen ate the sweet jelly which was on the table*). In acquisition, sentences derived from each of several fours are presented, usually under a procedure not encouraging memorization of form. During recognition, both old and new sentences are tested. (New sentences are permissible combinations of ideas derivable from the four which were not presented during acquisition.) Subjects are required to label each recognition sentence as "old" or "new" and to assign a confidence rating to their categorizations.

Bransford and Franks (1971) reported two facts of primary interest: Confidence ratings (CRs) increased sharply with complexity of the recognition sentence, and the functions for olds and news were quite similar; ratings for new ones were lower than for old ones, but otherwise it appeared that subjects could not distinguish olds from news. Bransford and Franks concluded that the complex idea was abstracted from the exemplars, no traces of which were retained. Hence, olds and news were indistinguishable, and the closer

the match between a recognition sentence and the (abstracted) four stored in memory, the more confident the subject would be that the sentence was old.

The present paper focuses on some issues attendant to the use of CRs as a dependent measure. Bransford and Franks (1971), ignoring accuracy, simply computed the arithmetic mean of all CRs from all subjects for all sentences in each category (such as old twos). This decision reduces the amount of information to be communicated, but the economy is purchased at the risk of misrepresenting underlying processes.

An illustration of the potential for misunderstanding is provided by Singer and Rosenberg (1973). Using an alternative data analysis, they claimed that confidence did not increase with complexity, that subjects were moderately confident of all their responses. On the other hand, the probability of an old response did vary with complexity. If these results are accurate, then integration has nothing to do with subjects' confidence in a response, and the usefulness of the confidence measure would be ended.

Unfortunately, Singer and Rosenberg's conclusion was based on a misunderstanding of the predictions which the integration hypothesis would make about absolute CRs. For example, an analysis of variance on the mean absolute CRs for the four levels of complexity was not significant. That analysis was intended to test the prediction that CRs are directly related to complexity for both old and new responses. In fact, the hypothesis predicts an inverse relation for new responses. If a subject is most confident in calling a four old because it most closely matches the integrated memory trace, then the subject should show the least confidence in calling a four new. New responses to simpler sentences should have the higher CRs, being more distant from the stored concept.

The integration hypothesis need not, therefore, predict an effect of complexity on absolute CRs. For fours, high absolute CRs for old responses are averaged

Portions of this research were supported by grants from the Rutgers Research Council and NIMH Grant 19859. The authors gratefully acknowledge the assistance of Jack G. Thompson. Requests for reprints should be sent to Carlton T. James, Department of Psychology, Busch Campus, Rutgers University, New Brunswick, New Jersey 08903.

with low absolute CRs for new responses. For ones, low absolute CRs for old responses are averaged with high absolute CRs for new responses. The result can easily be a function of zero slope. Consequently, the analysis does not eliminate the possibility that complexity has some effects on rated confidence.

In a response to Singer and Rosenberg, Franks and Bransford (1974) objected to the rejection of the confidence measure. (They apparently failed to realize that Singer and Rosenberg's predictions were incorrect.) They reanalyzed the data from their third experiment (Bransford & Franks, 1971), breaking data into four categories depending on the actual familiarity (old or new) of a sentence and its rated familiarity. Confidence ratings did vary significantly with complexity for old but not for new responses. (The authors provide no explanation for the latter failure, although their hypothesis should predict old *and* new CRs.)

Franks and Bransford's (1974) significant results for old responses cannot, however, be taken as establishing even a limited usefulness for CRs, because their analysis almost certainly inflated the range of their dependent variable, an argument developed below. Consequently, the existence of a relation between sentence complexity and CRs remains problematic. The present paper attempts to clarify the issue.

EXPERIMENT 1

Franks and Bransford (1972) concluded that the integration process was unaffected by variations in concreteness. Since that conclusion was based on a comparison of data from separate experiments, subtle differences may have escaped detection. It is worthwhile, in light of Paivio's (1971) intensive investigations of the concreteness dimension, to seek a more definitive comparison through the use of a within-subjects design. The results will reveal that certain detailed differences in the integration of abstract and concrete materials are concealed by the gross confidence ratings analysis.

Method

Design. There were two between-groups variables, form and order, and three within-subjects variables, concreteness (abstract-concrete), familiarity (old-new), and complexity (one to four). The total set of possible sentences was divided into two mutually exclusive subsets, or forms (matched on the within-subjects variables), and half the subjects received each of these subsets during acquisition. Two random recognition-list orders were used for each form of the acquisition list. Twenty-four paid subjects (all introductory psychology students) were randomly assigned to each of the four conditions, and were tested in groups.

Materials. There were six complex ideas, three concrete and three abstract. The nouns in the concrete ideas had a mean concreteness rating (Paivio, Yuille, & Madigan, 1968) of 6.70, as compared to a mean rating of 3.02 for the nouns used in abstract ideas. The sentences were as follows: *The photograph in the magazine showed the dead snake near the fireplace; The fire in the village spread to the high grass behind the factory; The musician from the university left the valuable harp in the church; The pledge taken by the ownership improved the hostile*

attitude held by the majority; The unbeliever at the hearing denied the theoretical evidence for an afterlife; The interest in the promotion caused a serious crisis throughout the franchise. Each complex idea generated 12 exemplars: four ones, four twos, three threes, and one four, the complex idea itself.

Acquisition lists contained 36 sentences, six exemplars (two ones and twos, and either two threes or a three and the four) from each of the six complex ideas. There were two forms of the acquisition list, and none of the exemplars contained in one form appeared in the other. The acquisition list was broken into randomized blocks of six, each block containing one exemplar from each complex idea.

Recognition lists contained all possible exemplars from all complex ideas (72 in all), along with six noncases. There were three types of noncases (an abstract and concrete example of each): Type A—A modifier was added to a previously unmodified noun, and a synonymous modifier was substituted for an already existing one (e.g., *The roaring fire in the village spread to the tall grass behind the factory*). Type B—Ones were combined across all topics to arrive at a new four. Type C—The order of prepositional phrases in the original four was changed. Bransford and Franks used Type B.

The 78 recognition sentences were presented in a randomized block format, as in acquisition, with noncases interspersed randomly. Two orders of each recognition list were used, one being the reverse of the other.

Procedure. An incidental learning task was used in acquisition. The experimenter read a sentence aloud, and then held up a card with four colors; after the subjects read the colors aloud in unison, the experimenter asked an elliptical question about the sentence. After subjects had written their answers, the next sentence was read. This procedure required 10-15 sec per sentence. Instructions mentioned no further test.

In recognition, subjects judged each sentence as old or new and rated their confidence in the response on a scale from -5 to +5. A rating of -1 indicated little confidence in a new response, -5 indicated high confidence; for old responses, a rating of +1 indicated little confidence, +5, high confidence.

Results

Noncases. The overall mean confidence ratings for noncases (collapsing across concreteness) were -1.5, -4.4, and -1.9 for Types A, B, and C, respectively. These results replicate those reported by Bransford and Franks (1971) for Type B, and by Singer and Rosenberg (1973) for Type C. As Type A noncases contained two words not previously heard in the experiment, one might have expected their presence to be highly noticeable. In fact, these changes were not appreciably more detectable than the syntactic changes in Type C.

Overall CR analysis. For each stimulus category, the mean of all CRs (correct and incorrect) was computed for each subject (see Figure 1). Each subject contributed 16 scores to an analysis of variance, which also included between-groups factors of form of acquisition list and order of recognition sentences.

The only factors to achieve reliability at the .01 level were the main effects of complexity and familiarity [$F(3,60) = 145.37$, $MSe = 3.06$, and $F(1,20) = 55.59$, $MSe = 1.77$, respectively]. The main effect of concreteness approached significance [$F(1,20) = 5.39$, $MSe = 5.64$, $.01 < p < .05$]. The complexity effect (CRs increasing monotonically with complexity of test sentence) was reported by Bransford and Franks;

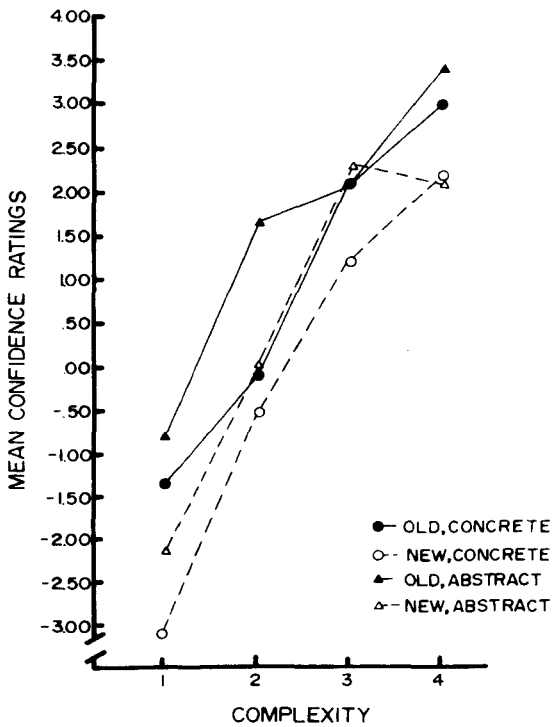


Figure 1. Mean overall CRs for concrete and abstract sentences as a function of complexity.

the familiarity effect (higher CRs for old than for new test sentences) was not, and it is important to examine this factor more closely.

Bransford and Franks (1973) asserted that specific memory, if any, would be restricted to the simpler sentences (i.e., the ones). In the present experiment, the specific memory effect, although largest for ones, did not reliably decrease with increased complexity ($p < .10$). This conclusion is supported by the obtained values of d' , a bias-free measure of recognition accuracy. The average values of d' (see Table 1) do not uniformly decrease across increasing levels of complexity, although there is a tendency for ones to be recognized most accurately. Clearly, Bransford and Franks' original proposal was too strong; some memory traces at all levels survive integration.

Paivio (1971) argued that tests for memory for form (as opposed to memory for meaning) should favor abstract sentences, which are supposedly encoded verbally. However, neither of the above analyses revealed superior memory for abstract sentences.

Conditionalized CRs. The significant complexity effect reported above may be a result of actual differences in confidence, of differences in the probability of calling a sentence old, or both. To disentangle the two factors requires a closer examination of the data than is afforded by the gross average CR.

Following Franks and Bransford (1974), CRs were computed separately for sentences jointly classified

as to actual and rated familiarity. Correct responses may be divided into old/old (read "old called old") and new/new, while old/new and new/old responses are errors. These functions provide a more sensitive test for variations in confidence as a function of complexity. If CRs are based on the degree of deviation from a complex idea, then, for all four response categories, the functions relating rated confidence to sentence complexity should be similar to the function obtained when they are averaged together (as in Figure 1).

For old responses, confidence should decrease with decreases in complexity, assuming the lowest values for ones, which differ maximally from the stored representation. To the extent that old and new test sentences are discriminable, the old/old and new/old functions will have different intercepts but similar slopes. The same prediction holds for new responses: Subjects should be most confident (-5) that ones are new, since they presumably match least closely the stored representation of the complex idea. Few fours should be called new, and they should receive lower CRs.

In agreement with Franks and Bransford (1974), the slopes were shallow (see Figure 2), and did not increase monotonically in all instances, as predicted by the integration hypothesis. The analyses of variance performed on subjects' mean CRs are, however, of questionable validity.

Figure 2 presents the functions obtained by summing all CRs in each particular category, and then dividing by the total number of cases. The alternative is to compute each subject's mean CR at each point, and then to take the average of the means. In most instances, the choice between the two alternatives is of little consequence. However, the second alternative can produce serious distortions when a category contains few observations. For example, only six old threes were tested; one subject may call all old, while another may call only one old. Both scores are weighted equally if the mean of the means is computed, even though the scores are not equally reliable. Further, the great majority of CRs were in the 3-5 range. The relatively small number of lower ratings have little effect on an overall mean, but can assume greater influence when calculating the mean of the means. If a subject assigns a CR of -1 to an old three, and calls no other sentences of this type new, then that one score counts for 1/24

Table 1
Average d' and β Values for Concrete and Abstract Sentences as a Function of Complexity (Experiment 1)

Imagery Condition	Complexity							
	One		Two		Three		Four	
	d'	β	d'	β	d'	β	d'	β
Concrete	.62	1.69	.13	1.06	.34	.87	.29	.80
Abstract	.39	1.22	.49	.93	-.09	1.30	.46	.69

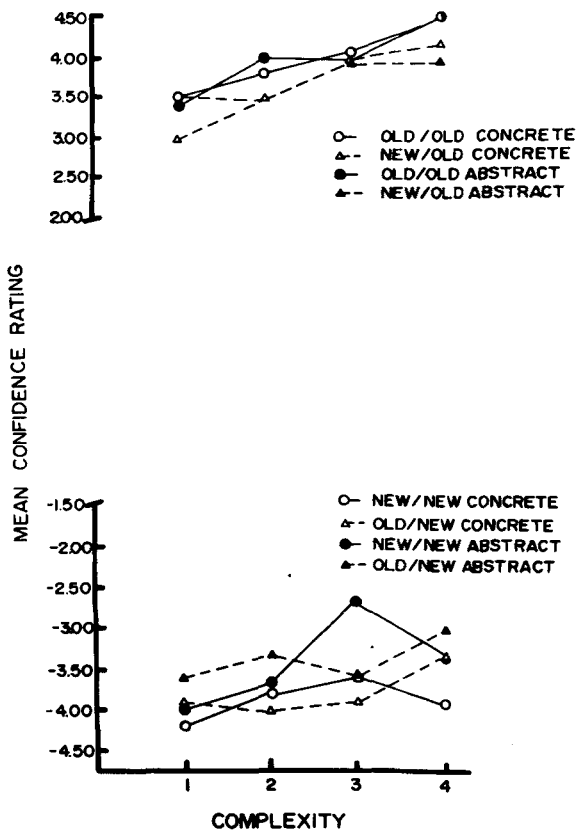


Figure 2. Mean conditionalized CRs for concrete and abstract sentences as a function of complexity.

(in the present experiment) of the final mean. This procedure assigns disproportional influence to deviant ratings and increases the range.

This is illustrated in Table 2, which presents the means calculated both ways, along with the resulting ranges among these means. (Fours were not included for new responses because of the small number of cases involved.) In every comparison, a larger range (as much as four times greater) was obtained by taking the mean of the means. (The same pattern was obtained in Experiment 2.)

The larger range increases the likelihood of a significant F ratio. Indeed, the main effect of complexity was reliable for all four response categories. (Franks and Bransford, 1974, found only old response analyses significant.) The nonsensicalness of such analyses is highlighted by the old/new category for concrete sentences. The means of the means had a range of .86, and the associated analysis seemed to reveal a solid complexity effect [$F(2,40) = 5.30$, $MSe = 1.59$, $p < .01$]. However, the overall means had a range of only .20. It is difficult, in light of these results, to justify an analysis of variance on subjects' mean CRs.

How, then, may one judge whether complexity did influence rated confidence? One reply is that the

answer is clearly before us in Figure 2, and need not be submitted to blind statistical procedures: Integration produced little change in CR, and the changes were not always systematic. Nevertheless, the distribution of CRs among the levels of complexity were analyzed, using complexity (one-three) as the row variable and response (1-5) as the column variable. (Fours were omitted because of their relative scarcity.) The resulting statistic is not distributed as chi square, since the cell entries in the contingency table are not independent. (Each subject contributed scores to many cells.) However, one outcome (acceptance of the null hypothesis) can be interpreted clearly. Since nonindependence inflates chi square, a nonsignificant result strongly indicates that equivalent distributions of CRs were associated with all levels of complexity. Large values cannot be interpreted unambiguously because of the unknown α values.

For concrete sentences, old/old and new/old categories were significant [$\chi^2(8) = 43.87$ and 24.14 , respectively]; the new/new category approached significance [$\chi^2(8) = 17.61$, $.01 < p < .025$], but the old/new category did not. For abstract sentences, only the new/new category achieved the .01 level [$\chi^2(8) = 22.96$], but ratings did not increase monotonically. The remaining analyses were not significant, although the old/old category was borderline [$\chi^2(8) = 18.41$, $.01 < p < .025$].

Table 2
Comparison of Overall Confidence Ratings and the Mean of Individual Subjects' Means (Experiment 1)

Imagery Condition	Ones	Twos	Threes	Fours	Range
Olds Called Old					
Concrete					
Overall	3.40	3.80	4.10	4.50	1.10
Mean of Means	3.14	3.95	4.33	4.50	1.36
Abstract					
Overall	3.40	4.00	4.00	4.50	1.10
Mean of Means	3.01	3.99	4.04	4.45	1.44
Olds Called New					
Concrete					
Overall	-3.90	-4.00	-3.80		.20
Mean of Means	-3.75	-3.83	-2.97		.86
Abstract					
Overall	-3.60	-3.40	-3.60		.20
Mean of Means	-3.48	-2.55	-2.59		.93
News Called Old					
Concrete					
Overall	3.00	3.50	4.00	4.20	1.20
Mean of Means	2.21	3.50	3.97	3.98	1.77
Abstract					
Overall	3.50	3.50	3.90	3.90	.40
Mean of Means	2.71	3.61	3.96	3.23	1.25
News Called New					
Concrete					
Overall	-4.20	-3.80	-3.61		.59
Mean of Means	-4.18	-3.71	-3.18		1.00
Abstract					
Overall	-4.00	-3.70	-2.70		1.30
Mean of Means	-3.91	-3.82	-2.10		1.81

The results provide scant support for the integration hypothesis' prediction of monotonically increasing confidence ratings, but further comment will be deferred until after presentation of the results from the second experiment. Certainly, the limited range of conditionalized CRs illustrates the misleading impression created by reporting gross CRs; the steep slope of the functions in Figure 1 is not attributable to large differences in the actual assignment of CRs.

Response bias. In agreement with Singer and Rosenberg (1973), there was a strong old bias for fours and a strong new bias for ones (see Table 3). It is clear that the biases were primarily responsible for the dramatic slopes reported by Bransford and Franks and replicated here in Figure 1. The mean CR was only slightly more positive for fours than for ones in all eight functions in Figure 2, but the great majority of the four responses were old, and the great majority of the one responses were new. When old/old and old/new functions were averaged to produce the overall old function in Figure 1, the concentration of old responses to complex sentences elevated that area of the function, while the preponderance of new responses to simpler sentences depressed the function in that region. As complexity primarily influenced the probability of an old response, this measure is clearly superior to CRs, which varied little as a function of complexity.

Another argument against the use of the gross CRs is revealed by a closer analysis of the concreteness factor. The overall CRs plotted in Figure 1 suggest that subjects were more confident in their responses to abstract sentences. In fact, just the opposite was true. The functions in Figure 1 are again the result of different old and new response biases. A comparison of the conditionalized CRs in Figure 2 shows no difference in confidence for old sentences as a function of concreteness; however, subjects were more confident in new responses to concrete than to abstract sentences. In addition, there was an overall old bias (.55) for abstract sentences and a new (.53) bias for concrete sentences [$\chi^2(1) = 10.10, p < .01$].

These conclusions are consistent with the results of the signal-detection analysis. The β values (see Table 1) were higher for concrete than for abstract sentences, except for threes, which yielded a negative d' in the abstract condition. Basically, a more conservative response criterion was used for concrete sentences: Subjects were more willing to say "old" for weak

abstract sentences. The criterion decreased as complexity increased for both concrete and abstract sentences (again with the idiosyncratic three result). Subjects were conservative for ones, but became progressively liberal with increases in complexity.

Although CRs were greater for concrete sentences, Figure 1 implies the reverse. The bias differences are responsible. The overall old functions average the old/old and old/new categories. There were more responses in the former category for abstract sentences, although CRs did not differ as a function of concreteness in this category. Likewise, there were more responses in the old/new category for concrete sentences, and the mean CR was more extreme (more negative) for concrete materials. Averaging yields an overall old function for concretes that falls below the abstract function. Similarly, the new function averages the new/old and new/new categories; the latter is weighted more heavily for concrete sentences, so that the overall new function for abstract sentences lies above the function for concrete sentences.

Figure 1 implies that subjects were more confident in their responses to abstract sentences, but a careful inspection of conditionalized CRs and bias data shows just the opposite: Depending upon whether a sentence was called old or new, there was either no difference in CR, or subjects were more confident in their responses to concrete sentences. Reporting only an analysis of averaged CRs conceals one difference between concrete and abstract sentences (different biases for old and new responses) and misleads about another difference (confidence). Whether or not complexity materially influences rated confidence, the potential for distortion cautions against reliance on gross CRs as the sole measure of integration.

EXPERIMENT 2

Bransford and Franks (1973) and Cofer (1973) both briefly mention experiments in which subjects were informed of a subsequent memory test. Cofer merely reported that performance was not substantially altered by the intentional learning instructions, while Bransford and Franks observed that the slope of the function relating CR and complexity was present but reduced. Both experiments tested only new sentences during recognition. These shortcomings will be corrected here, permitting us to point out a necessary relationship between the complexity and specific memory effects. The experiment will also provide another set of conditionalized CRs to aid in the resolution of the complexity-confidence question.

Method

Materials. Complex ideas were those used by Bransford and Franks (1971). As in Experiment 1, half the sentences from each complex idea were used in acquisition, and the entire set was presented in recognition. Across all complex ideas, half of the

Table 3
Proportion of Sentences Called Old (Experiment 1)

Imagery Condition	Complexity			
	One	Two	Three	Four
Concrete	.25	.47	.69	.78
Abstract	.32	.60	.75	.80

sentences of each complexity were presented in acquisition. Two groups of subjects were tested, in order that each sentence would receive both new and old tests. There were no noncases.

Subjects. Sixteen subjects were recruited from introductory psychology courses in return for additional points toward their final grades. Eight were assigned to each of the two groups.

Procedure. In the only procedural change, subjects were told that they would be tested later on the sentences heard in the first part. They were urged to memorize the sentences word for word, because the memory test would be difficult.

Results

Overall CR analysis. Two facts are of importance in the overall analysis (see Figure 3). First, CRs increased with complexity [$F(3,42) = 20.05$, $MSe = 1.96$, $p < .01$], but the effect was not as striking as in Experiment 1. These results agree with those of Bransford and Franks (1973), in showing a reduced complexity effect with intentional instructions.

Second, CRs were higher for old than for new sentences [$F(1,14) = 27.92$, $MSe = 2.08$, $p < .01$], suggesting a greater degree of specific memory. This conclusion is supported by estimates of recognition accuracy (see Table 4). Of course, subjects should remember more under intentional instructions, and it is noteworthy that the improvement in memory was not more pronounced. Clearly, the discriminations required are quite difficult.

Conditionalized CRs. As in Experiment 1, the functions presented in Figure 3 suggest that confidence

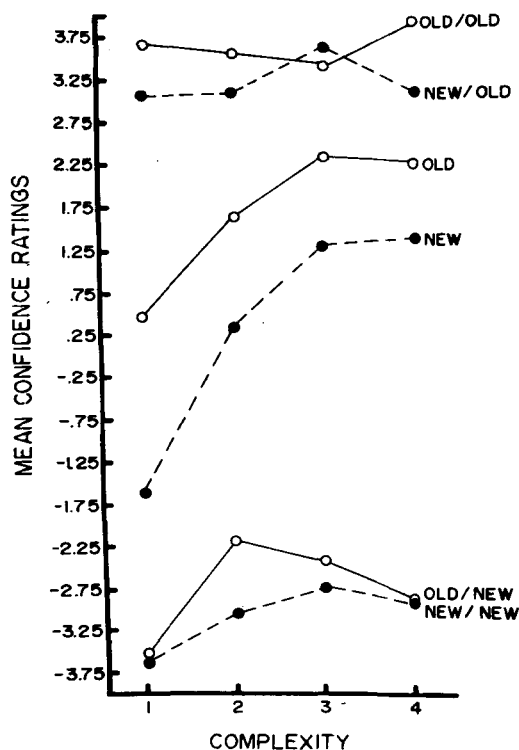


Figure 3. Mean overall and conditionalized CRs as a function of complexity for sentences in Experiment 2.

Table 4
Probability of Hits, False Alarms, and Resulting Accuracy Estimates (Experiment 2)

	Complexity			
	Ones	Twos	Threes	Fours
P (Hit)	.55	.67	.81	.75
P (False Alarm)	.30	.55	.64	.72
Accuracy	.25	.12	.17	.03

is little affected by complexity. Only the old/new category achieved significance [$\chi^2(8) = 29.39$], but the ratings clearly did not increase monotonically. Among the remaining categories, only the new/new analysis approached significance [$\chi^2 = 19.88$, $.01 < p < .025$].

The additional data permit a more thorough evaluation of the relation between CRs and complexity. Two experiments have provided 12 tests of that hypothesis, three tests for each of the four conditionalized response categories. There were only four significant chi squares, and in two of these cases ratings were not monotonically increasing. In two of the three borderline analyses, CRs were not strictly monotonic. These results clearly indicate that any relation that may exist between complexity and CRs is too small and variable to be reliably obtained.

Confidence ratings could maintain limited usefulness if the significant chi squares clustered in certain categories. Indeed, complexity does appear to be related to CRs for correct responses, with the relation being closer for new (three $ps < .05$) than for old (two $ps < .05$) sentences. The generality of even this statement is suspect, considering that Franks and Bransford (1974) did not find significant variation in CRs for the new/new category.

Taking the results on the whole, rated confidence is an unreliable dependent variable whose continued use is unjustified, given the existence of a more appropriate measure.

DISCUSSION

The present results are important for interpreting the claims of Reitman and Bower (1973), who also obtained a reduced complexity effect and a relatively large old-new difference. As they used strings of letters and numbers, Reitman and Bower concluded that previous failures to find specific memory could be traced to the extra confusion introduced by syntactic and semantic information. The present authors agree with their assertion that abstraction and integration need not result in an erasure of all prior specific traces (cf. Experiment 1). However, Reitman and Bower were incorrect to attribute the result wholly to their materials. Their acquisition procedure increased the amount of specific memory by requiring subjects to concentrate on specific order information, and the

intentional learning factor was confounded with any materials-related effects.

Small's (1975) procedure raises similar problems of interpretation. He required subjects to copy input (strings of words), which would apparently encourage a greater amount of intentional memorization than the standard incidental task. Consequently, it is not surprising that he found a specific memory effect. In any case, Small is wrong to overemphasize the co-occurrence of a specific memory effect and a complexity effect. As demonstrated in Experiment 1, that result can also be obtained with sentences. Although the integration hypothesis does strongly interconnect the two effects, it is a matter of degree. That is, in the ideal case, there will be a large complexity effect and absolutely no specific memory. The presence of both results does not imply the absence of integration, but perhaps it does imply that the process is less than complete.

In fact, there is a general relation between the size of the complexity and specific memory effects, as they are inferred from the overall CR function. As specific memory increases, the complexity effect will decrease. To see this, first assume that CRs are constant for all levels of complexity regardless of accuracy of response. As argued earlier, the complexity effect primarily results from a systematic increase in the old-response bias with complexity. The slope of the function will vary between the boundary conditions of total memory and zero memory. With no memory, all responses are determined by the bias. But with total memory, there should be no recourse to guessing: Bias will be eliminated, because old and new sentences are discriminated perfectly. The functions relating the probability of an old response to complexity will have zero slope, and intercepts of 0 and +1 (for new and old sentences, respectively). As specific memory increases from a zero point, more items are correctly discriminated, leaving progressively fewer to be subject to the bias. Clearly, as the separation between old and new functions increases, their slopes must decrease.

The above argument assumes that CR is not related to complexity. The strength of the association has been shown to be minimal, and can be estimated by inspection of the slopes of conditionalized CRs. These estimates do not change the argument, but only (slightly) modify the predicted slopes and intercepts.

Evidence for this general relationship can be inferred from the literature. Reports of reduced complexity (Bransford & Franks, 1973; Reitman & Bower, 1973; Schwartz & Witherspoon, 1974; Small, 1975; the present Experiment 2) occur when the pool of unknown items (to which the bias applies) has been reduced by increasing the pool of known items. That is, reduced complexity effects have been reported whenever evidence for retention of specific traces has been found.

None of the data presented here are inconsistent with the integration hypothesis of Bransford and Franks,

but they do place certain restrictions on their claims. Bransford and Franks' original demonstration of no memory for complex sentences lacks generality, as shown by the specific memory effects obtained in Experiment 1 (and replicated several times in our laboratory). Although the specific memory was not marked, the confusion and interference inherent in the paradigm are so great that memory was poor even when subjects were trying to memorize order information (Experiment 2). Consequently, it is important to demonstrate that, in the face of great interference, some complex memory traces can persist, even though learning is incidental. If a less confusing paradigm were available, specific memory might well be greater. Integration does *not* erase all traces.

The present paper's primary concern was to determine if complexity influences subjects' confidence in their responses. Although there may be some association for correct responses, the effects are small and sufficiently unstable as to counsel against adoption of CRs as a measure of integration; the proportion of old responses is a superior index. Again, this conclusion does not weaken the integration hypothesis, but does more accurately specify the behavioral measure which should be chosen to assess that hypothesis.

REFERENCES

- BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350.
- BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas: A review. *Cognition*, 1973, 1, 211-250.
- COFER, C. N. Constructive processes in memory. *American Scientist*, 1973, 61, 537-543.
- FRANKS, J. J., & BRANSFORD, J. D. The acquisition of abstract ideas. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 311-315.
- FRANKS, J. J., & BRANSFORD, J. D. A brief note on linguistic integration. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 217-219.
- PAIVIO, A. Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches*. New York: Academic Press, 1971.
- PAIVIO, A., YUILLE, J., & MADIGAN, S. Concreteness, imagery and meaningfulness values of 925 nouns. *Journal of Experimental Psychology: Monograph Supplement*, 1968, 76(No. 1, Part 2).
- REITMAN, J. S., & BOWER, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 1973, 4, 194-206.
- SCHWARTZ, S., & WITHERSPOON, K. D. Decision processing in memory: Factors influencing the storage and retrieval of linguistic and form identification. *Bulletin of the Psychonomic Society*, 1974, 4, 127-129.
- SINGER, M., & ROSENBERG, S. T. The role of grammatical relations in the abstraction of linguistic ideas. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 273-284.
- SMALL, D. W. The abstraction of arbitrary categories. *Memory & Cognition*, 1975, 3, 581-585.

(Received for publication December 9, 1976;
accepted December 29, 1976.)