A study was designed to test the hypothesis that memory for metaphor was primarily a function of the structure of the metaphor itself. Eighty undergraduate students rated the quality of subsets of 80 metaphors and later freely recalled them, while another 40 students simply read metaphors in their extended contexts and later received a surprise cued recall test. Student ratings of metaphoric quality had only a slight relationship to the frequency with which they recalled those metaphors. There was no relationship between the 80 student ratings of quality and the cued recall frequencies from the other 40 students who merely read the same metaphors. Ratings of conceptual similarity were only minimally related to free recall from students who rated the quality of metaphors in extended context. The metaphors with presumably the poorest general comprehension were not consistently the most poorly recalled, nor were the best understood metaphors the most frequently recalled ones. The tenors of the metaphors were recalled significantly more often than vehicles in the cued recall measures. These results indicate that memory for metaphor is related to its structure but not to its quality, conceptual similarity, or comprehensibility. That is, metaphoric terms (vehicles) are more effective in reminding people of things that are likened to them (tenors) than vice versa, suggesting that vehicles provide the schema to which tenors are assimilated. (RL)
When we say that a literary work is memorable we usually mean that it is good, but what does memory for literature really reflect? So far, only small portions of literature such as metaphors and proverbs have received attention. Such research shows that figurative language is not the aberrant, esoteric phenomenon it was previously thought to be. Pollio & Burns (1977) found that anomalous sentences were learned as easily as natural ones provided subjects interpreted them first. Harris found that metaphors were recalled or recognized as well as nonmetaphors (Harris, 1979) or even more easily than nonmetaphors (Harris, in press). Verbrugge & McCarrell (1973, 1977) found that the implicit basis of a metaphoric comparison was almost as effective a recall prompt as the explicit terms of that comparison, indicating that that implicit basis had been inferred by the subject at the time of reading. Recall of proverbs, another form of figurative language, is somewhat enhanced by enhanced comprehension (Honeck, 1973; Honeck, Reichman, & Hoffman, 1975). Many students of metaphor (e.g., Malgady & Johnson, 1976; Ortony, 1979; Tourangeau & Sternberg, 1978) have discussed the relationship between metaphoric quality and similarity. But no one has looked at the relationship between metaphoric quality and memory. Contrary to the popular belief mentioned at the outset of this paper, my principle hypothesis is that memory for metaphor is less a function
of aesthetic worth and related attributes such as similarity and comprehension than it is of the structure of metaphor itself.

Most students of metaphor have used single sentences as stimuli (e.g., Koen, 1965; Kozlowski, 1975; Malgady & Johnson, 1976; Johnson & Malgady, Note 1 & in press; Ortony, Reynolds, & Arter, 1978; Tourangeau & Sternberg, #13, 1978; Verbrugge, 1975a, b,; 1977b, in press b). One exception to this is Harris (in press), who found different results studying metaphors in the context of play synopses than he did studying metaphors in isolated sentences (Harris, 1979). Thus, in pursuit of ecological validity (Neisser, 1976), I selected 80 metaphors in the context of passages of modern prose fiction. (When contextual paragraphs have been provided (Ortony, Schallert, Reynolds, & Antos, 1978; Harris, in press), the researchers have contrasted metaphorical with literal interpretations rather than contrasting the relative quality of different metaphors.)

My first question is whether memory directly reflects the quality of a metaphor. That is, do we remember a metaphor to the extent that it is good? Eighty undergraduates rated the quality of subsets of the 80 metaphors in their extended contexts and later freely recalled them. Two English graduate students also rated the quality of all the metaphors. Another 40 undergraduates simply read metaphors in their extended contexts and later received a surprise cued recall test. All 3 measures of metaphoric quality were compared to both the free recall and cued recall frequencies, for one within-subjects and five between-subjects tests of my hypothesis.
Undergraduates' ratings of metaphoric quality in extended context have only a slight relationship to the frequency with which they recall those metaphors ($r(80) = .26, p < .01$), accounting for only 7% of the variance in those recall scores. There is no relationship between these undergraduates' ratings of quality and the cued recall frequencies from other undergraduates who merely read the same metaphors. Nor is there any relationship between the two English graduate students' estimations of quality and either recall measure. In short, we do not remember a metaphor to the extent that it is good. Or, in other words, memory is a poor index of quality.

My second question is whether memory for metaphors in extended context is instead a function of the conceptual similarity of its constituents. As I have noted, there has been much discussion of the relationship between similarity and the quality of metaphor, and we have seen that memory does not reflect quality. But perhaps memory for metaphor more reflects the static conceptual structure underlying similarity ratings than it does the quality of a metaphor in a specific, extended context, inasmuch as the quality of a metaphor in context is not explained by the conceptual resemblance of its tenor and vehicle (McCabe, 1980). To test this notion, 20 undergraduates rated the similarity of the metaphoric constituents in the context of simple identity statements (e.g., "A white frame house is a respectable middle-aged woman."), which were extracted from the extended contexts. These ratings were then compared (between subjects) with the
aforementioned free recall and cued recall frequencies for two tests of this question. (I collected ratings of similarity in minimal context because I distinguish conceptual resemblance in minimal, general context from contextual resemblance in extended, specific context. The latter is less theoretically interesting in its relationship to metaphor since contextual resemblance could be either the cause or the result of the contextual quality of a metaphor—a chicken-or-egg dilemma; see McCabe, 1980, for further explanation.)

Ratings of conceptual similarity are only minimally related to free recall from undergraduates who have rated the quality of metaphors in extended context ($r(80) = .18, p < .05$), accounting for only 3% of the variance in the latter measure. Such similarity ratings are unrelated to the cued recall by undergraduates who merely have read metaphors in their extended contexts. In short, we do not remember metaphors because the concepts they compare resemble each other.

A third question concerns whether memory for metaphors is enhanced by enhanced comprehension of them. I could test for this insofar as valid interpretation presupposes accurate comprehension. (Insofar as my experiment is concerned, valid interpretation requires more than accurate comprehension; it requires accurate written expression of that comprehension.) To test this notion that we remember a metaphor to the extent that we comprehend it, each of the 80 metaphors in their extended contexts was interpreted by 8 undergraduates and 2 English graduate students.
Each interpretation, in turn, was scored for validity by one or two psychology graduate students. They agreed on the category of validity of 70.5% of the 295 interpretations they both scored. I looked to see whether metaphors that occasioned the most invalidity (presumably the most poorly comprehended metaphors) also were the most poorly recalled metaphors. This was tested between-subjects, since interpretation itself took too long to ask that the same subjects recall the metaphors they interpreted.

The metaphors occasioning the most invalidity and thus presumably the poorest general comprehension are not consistently the most poorly recalled—nor the three cued recall frequencies I collected. Nor are the most frequently validly interpreted and hence best understood metaphors the most frequently recalled ones. In short, we do not remember metaphors to the extent that we comprehend them.

We have seen then that memory for metaphor does not reflect the aesthetic worth of that metaphor, nor the similarity of its constituents (our static conceptual structure), nor the metaphor's comprehensibility—these three things being the aspects of metaphor most often discussed by students of metaphor from Aristotle on. But what about the structure of a metaphor? Might not memory be related to the structure of a metaphor rather than these attributes? The structure of a metaphor consists of a tenor (the literal subject) and a vehicle (the metaphoric term).

Among other students of metaphor, Malgady & Johnson (1976) argue that the quality of a metaphor is a function of the
overlapping static conceptual resemblance of tenor and vehicle, implying that tenor and vehicle should be equally effective in prompting each other. In contrast to this view, Black (1962) argues that tenor and vehicle interact in the process of metaphor, with the vehicle acting as a "screen" or "filter" through which the tenor is viewed. This latter theory of metaphor implies that the vehicle would prompt recall of the tenor more often than vice versa. I was able to compare the two types of prompts in my three cued recall measures, since each of my 80 metaphors was prompted half the time by its tenor, half by its vehicle, and each subject received half tenor, half vehicle prompts.

In free recall, tenors and vehicles are usually recalled as a pair if they are recalled at all (the frequency of free recall of tenor and vehicle as a pair accounts for 94% of the variance of the summed frequency of recall of tenors plus vehicles; the free recall of tenors is correlated .95 with the free recall of vehicles). This means that people resolve comparisons into conceptual wholes, presumably by means of the extended contexts provided to them in this experiment and in natural settings. This also means that tenors and vehicles are recalled equally often in free recall, or, in other words, that there is no difference in the relative inherent memorability of the items per se. This is not surprising given that in different metaphors a number of items were used alternately as tenors and vehicles (e.g., stars, eyes).
But in all three cued recall measures, tenors are recalled significantly more often than vehicles ($t(79) = 2.27, p < .03$; $t(79) = 3.07, p < .003$; $t(79) = 5.55, p < .001$). This means that vehicles are superior to tenors as cues, since vehicles cued tenors and vice versa. Verbrugge & McCarrell's (1973, 1977) data also showed vehicles to be superior to tenors as prompts (13.1 versus 12.1 successful promptings for vehicles and tenors, respectively), but their small number of metaphoric sentences (14) and the consequent near perfect recall obscured the significance, the existence of this superiority; the authors did not remark it. Thus the vehicle specifies the properties by which the tenor is construed and later recalled more than the tenor specifies the properties by which the vehicle is recalled. This supports the interaction view of metaphor, since the normal, decontextualized tenor term seems different enough from the tenor as tenor in a metaphor that when that decontextualized term is presented as a recall cue, it depresses recall of its partner in metaphor. In other words, the tenor's normally salient properties are frequently sufficiently unrelated to the metaphor so as to obstruct recall of the vehicle. Conversely, the vehicle's normally salient properties are frequently sufficiently pertinent to the metaphor so as to enhance recall of the tenor. The tenor and vehicle are not left unchanged by their participation in the metaphor, as the interaction view argues.

(Note that no scatter plots of any correlations above showed deviation from homoscedasticity or rectilinearity. Also note that all lists are randomized.)
Examples serve to bring these points home. Compare metaphor #25 with metaphor #95 (see Appendix). Metaphor #25 was rated bad (2.8, 5, 2 out of 6 by undergraduates and two English graduate students, respectively), while metaphor #95 was rated good (4.8, 6, 6 out of 6 = high). The tenor and vehicle of metaphor #25 were rated dissimilar (2.6 out of 6 = high by undergraduates), while the tenor and vehicle of #95 were rated similar (4.4 out of 6). Metaphor #25 was only misinterpreted by 2/10 people, while metaphor #95 was misinterpreted by most people (7/10). Yet despite these differences on all three attributes, both metaphors were freely recalled as a pair of concepts by 14/20 people. And despite its superiority in quality and similarity of concepts, metaphor #95 was cued less frequently than metaphor #25 (30/60 versus 43/60, respectively, summing over all three cued recall measures). Summing together the three recall measures, the tenor of both metaphors was recalled more frequently than the vehicle (the tenor was recalled 23/30 times compared to the vehicle's 20/30 times for metaphor #25; the tenor was recalled 17/30 times compared to the vehicle's 13/30 times for metaphor #95).

Metaphor is an important ingredient of everyday speech as well as literature. This study found that memory for metaphor is related to its structure but not to its quality, conceptual similarity, or comprehensibility. That is, metaphoric terms (vehicles) are more effective in reminding people of things that are likened to them (tenors) than vice versa, which suggests that the vehicle provides the schema to which the tenor is assimilated.
Popular belief holds that literature that is good is memorable, and vice versa. Contrary to this, memory for metaphor is not greatly reflective of the quality of a metaphor. A number of recent researchers (e.g., Fredericksen, 1975; Kintsch, 1974) have used memory for certain propositions in a text as a corroborative measure of the relative importance of those propositions within that text. I hope the present findings forestall the use of memory as an index of the relative aesthetic worth of metaphors as well as the literature that is comprised of them.
FOOTNOTES

REFERENCE


Frederiksen, C.H. Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology, 1975, 7, 371-458. (b)


Harris, R.J. Memory for literary metaphors. Bulletin of the Psychonomic Society, in press.


Verbrugge, R.R. Perceiving invariants at the invitation of metaphor. Paper presented at the meeting of the American Psychological Association, Chicago, August 30, 1975. (b)


3. Metaphor #95 was composed by Hawthorne, N. *The scarlet letter*. 