JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 21, 85–98 (1982)

On Understanding Nonliteral Speech: Can People Ignore Metaphors?

SAM GLUCKSBERG, PATRICIA GILDEA, AND HOWARD B. BOOKIN

Princeton University

How do people understand expressions such as *Some jobs are jails?* A predominant view contends that we must first derive the literal meaning of such statements. Second, we test this meaning against the context. Third, if the literal meaning fails to make sense, only then do we seek an alternative, nonliteral meaning. This serial, three-stage model implies that people can and do ignore the nonliteral meanings of sentences whenever the literal meanings are plausible in context. To test this view we asked subjects to make rapid decisions about the literal truth of sentences such as *Some jobs are jails*. Subjects correctly judged that such sentences are literally false, but the availability of a "true" metaphorical interpretation—for example, that some people are trapped in their occupations—interfered. When metaphoric interpretations of literally false sentences were available, subjects took significantly longer to decide that such sentences were false. This suggests that people do not have the option to ignore the nonliteral meanings of sentences. Instead, people seem to process both the nonliteral and literal meanings of sentences in the same ways, and at the same time.

Some jobs are jails. Is this statement true or false? If we interpret it to be a literal class-inclusion statement, then it is patently false. *Jobs* are not members of the category *buildings that are used to house prisoners*. However, if we interpret it to mean that some people feel trapped in their occupations, then there is some truth to it. How are such expressions normally understood?

A predominant view contends that we must first derive the literal meaning of such statements. Second, we test this meaning against the context of the utterance. Third, if the literal meaning fails to make sense in that context, only then do we seek an alternative, nonliteral meaning. This sequential, three-stage model has been proposed in linguistic (cf. Lyons, 1977) and philosophical (cf. Searle, 1979) as well as psychological (cf. Clark & Lucy, 1975) accounts of how nonliteral expressions are understood. Grice (1975) suggests that people implicitly agree to cooperate when talking with one another. One important component of this cooperative principle is an implicit assumption that statements are both true and informative. When this implicit agreement appears to be violated—as when someone utters an obviously false statement such as "Sam is a pig"—then this apparent violation serves as a cue to reinterpret the statement nonliterally. In this vein, Searle argues: "Where the utterance is defective if taken literally, look for an utterance meaning that differs from sentence meaning" (Searle, 1979, p. 114).

One implication of this view is that comprehending the nonliteral meanings of utterances requires more time and effort than comprehending literal meanings. This need not always be so. Swinney and Cutler (1979) report that people are able to understand familiar idioms, such as *kick the bucket,* as quickly as comparable phrases

We thank Michael McCloskey for the programming system that we used to conduct the experiments, Julia Jorgensen for her help in gathering the rating data for Experiments II and III, and Cynthia Pittman for collecting much of the data for Experiment III. Among the colleagues who read drafts of this paper and provided invaluable criticism, we especially wish to thank Larry Barsalou, Nancy Cantor, Joseph Danks, Arnold Glass, Michael McCloskey, Susan Sugarman, and Andrew Ortony. Reprints and copies of our stimulus materials may be obtained by writing to S. Glucksberg, Department of Psychology, Princeton University, Princeton, N.J. 08544. The research was supported by NSF Grant BNS 78-27943 to Princeton University, S. Glucksberg, principal investigator.

0022-5371/82/010085-14$02.00/0

that are used nonidiomatically. Swinney and Cutler suggest that such familiar idiomatic expressions are dealt with as if they were single lexical items. As such, they pose no special problems for either speakers or listeners. Clark (1979) analyzed merchants' responses to indirect requests such as *can you tell me what time you close?* Whether or not the literal meanings of such requests were responded to at all depended on various characteristics of those requests. In some cases an indirect request can be treated as an idiom, with its literal meaning playing little if any communicative function. In those cases, the literal or direct meanings need not be prior to nonliteral or intended meanings. Ortony, Schallert, Reynolds and Antos (1978) compared the time required to understand how a target sentence fits with (a) a minimal prior context, or (b) a fully adequate prior context. When prior context was minimal, both literal and figurative uses took a long time to understand (4 and 5 seconds, respectively). When the prior context was adequate and appropriate to the target sentence, metaphorical usages took no longer to understand than did literal. Ortony et al. also found, as did Swinney and Cutler, that idioms were as easy to understand as were comparable literal phrases. The consensus seems to be that nonliteral expressions are not necessarily more difficult or time-consuming to understand than are literal expressions. Both the familiarity of the expressions and their contextual supports play important roles (Gibbs, 1979).

Our research is addressed to an issue somewhat different from the relative comprehensibility of literal and nonliteral expressions. The serial, three-stage comprehension model implies a principled distinction between the literal and nonliteral meanings of utterances. Literal meanings are derived automatically, but nonliteral meanings are derived only optionally. The term *automatic* here does not mean that the understanding is not effortful. Instead, it means that a fluent speaker of a language has no voluntary control over whether or not an utterance such as "shut up and drink your beer" will be understood. A fluent speaker does, of course, have control over any actions that might be taken after such a sentence has been understood. Miller and Johnson-Laird (1976) suggest that understanding and acting are analogous to compiling a program and executing it, respectively. Understanding and acting "can be separated by the earliest possible 'point of refusal.' Neither man nor machine can refuse to obey a command until the command has been understood. Presumably, compilation [understanding] occurs automatically without conscious control by the listener; he cannot refuse to understand . . . loss of control over one's compiler may correspond to knowing a language fluently" (Miller & Johnson-Laird, 1976, p. 166).

The serial model of metaphor comprehension asserts that metaphor meanings, unlike literal meanings, are understood only optionally—whenever, and only whenever, an automatically derived literal meaning fails to make sense in context. This view seems wrong. Intuitively, it seems quite difficult, if not impossible, to inhibit our understanding of simple and transparent statements such as *Some salesmen are bulldozers* or *Some hearts are closets,* even though such statements are literally false. Furthermore, it seems to matter not at all whether such statements are plausible in context or not. Both the literal and nonliteral meanings seem to be apprehended without conscious effort or explicit inference.

We designed an analog of the Stroop (1935) color–word interference technique to see if people do ignore the meanings of metaphors, even when the literal meanings of those metaphors are plausible in context. The logic of the technique is straightforward. People are asked to decide, as quickly and accurately as possible, whether each sentence that appears on a CRT display is literally true or false. If people can ignore metaphorical meanings, then literally false class-inclusion sentences such as *Some jobs are jails* should pose no particular difficulties. People should have no problem in deciding that such sentences are false. On the other hand, if metaphorical meanings leap out—i.e., cannot be inhibited or ignored—then it should take longer to judge that such sentences are false. The "true" metaphorical interpretation, if it is made at all, should conflict with the "false" literal interpretation and so should slow up response latencies.

Of course, it could be argued that Searle's (1979) argument can still apply. Searle proposed that whenever the literal meaning of a sentence fails to make sense, a nonliteral meaning is sought. If literally false sentences never make sense—are "defective" in Searle's terminology—then literal falsehood will always trigger a search for nonliteral meaning. This argument encounters difficulties both in the context of everyday discourse and in our experimental situation. It is by no means clear that literal falsehood in everyday discourse acts as an automatic or unconditional trigger for determining nonliteral meanings. People have been known to lie, and people are often mistaken. Furthermore, literal falsehood is not a necessary condition for a statement to have metaphorical content. *No man is an island* is literally true, but it has metaphorical truth as well. In general, literal falsehood will not trigger a search for an alternative, nonliteral meaning unless a listener has reason to suppose that a speaker (a) is truthful; (b) is not mistaken; and (c) intends to be informative (Grice, 1975). The subjects in our experiments have no reason to suppose that the stimulus sentences that appear before them are intended to be informative in any way. Furthermore, half of the sentences in each experiment are false, but only a small proportion of these false sentences are metaphors. Therefore, it seems quite unlikely that subjects will always try to find a nonliteral meaning whenever they encounter a false sentence. If they do, then they should be able to do so quite easily. People can always generate a nonliteral interpretation for any sentence, even ones that randomly pair subject and predicate nouns (Pollio and Burns, 1977). Therefore, to maintain Searle's position despite a demonstration of a metaphor interference effect, one would have to argue that (a) all literally false sentences trigger a search for nonliteral meanings; and (b) it is only with metaphors that this search is completed in time to interfere with a literal-false decision. Even on this view, we would still conclude that people cannot inhibit their understanding of metaphors, even when that meaning is irrelevant to the task at hand.

Experiment I provides a demonstration of just such a metaphor interference effect for sentences of the form *Some X are Y*. Experiment II shows that the metaphor interference effect is attributable to the availability of a metaphorical interpretation per se, and not merely to particular semantic or associative relations between the *X* and *Y* terms of the metaphor sentences. Experiment III extends this finding to sentences of the form *All X are Y*, indicating that the effect applies to universal as well as to existential quantifiers. These findings lead us to argue that people use the same comprehension strategies and mechanisms for understanding the literal and nonliteral meanings of sentences, with neither category of meaning having unconditional priority over the other. A model of metaphor comprehension that is based on ordinary language processing mechanisms is then outlined.

Experiment I: People Cannot Ignore Metaphors

We used standard sentence-verification procedures and materials (cf., McCloskey & Glucksberg, 1979) to see if the availability of metaphorical meanings interfered with literal false decisions. College students responded "true" or "false" to five types of sentences. (a) True high-typical sentences, e.g., *Some birds are robins*. (b)

True low-typical sentences, e.g., *Some birds are penguins*. (c) Standard False sentences, e.g., *Some birds are apples*. (d) Metaphors, e.g., *Some jobs are jails, Some flutes are birds*. These sentences were literally false category-membership statements, but they were readily interpretable if taken nonliterally. (e) Scrambled metaphors, e.g., *Some jobs are birds, Some flutes are jails*. These sentences were also literally false, but were not readily interpretable.

If subjects ignore the nonliteral meanings of the Metaphors, then the metaphor sentences should take no longer to reject than the Scrambled Metaphors. This is because in neither case would a "true" nonliteral meaning conflict with the "false" literal meaning. On the other hand, if subjects automatically register any nonliteral meanings that are available, then the Metaphor sentences should take longer to judge as false than their scrambled counterparts. This is because there would now be a conflict between the "true" nonliteral meanings and the "false" literal ones.

Method

Subjects. Twenty undergraduate students at Princeton University served as paid volunteers. The data of three additional subjects were discarded because of excessive error rates (over 5%) and two subjects' data were lost because of equipment malfunction.

Materials. The high- and low-typical True sentences were constructed by drawing high- and low-frequency items from each of 15 categories of the Battig and Montague (1969) category-exemplar production norms. The Standard False items were constructed by scrambling the subject and predicate nouns of the True sentences. The Metaphor items were constructed to be literally false, but readily interpretable even though, in our judgment, they were relatively novel expressions. The Scrambled Metaphor items were constructed by repairing the subject and predicate nouns of the Metaphors to produce sentences that were not readily interpretable. The experimenters' intuitions were the bases for these choices.

A 120-item practice list was formed, consisting of 60 True (30 high-, 30 low-typical) items and 60 false items. The false items consisted of 30 Standard False, 15 Metaphors, and 15 Scrambled Metaphors. A 160-item test list had the same proportions of these five item types (See Table 1 for examples). Both the practice and test lists were block randomized and each of the 10 test-list blocks contained the same number of each sentence type. Five different orders of the test list were used, and each test list began with 12 warm-up (filler) items, 6 True, and 6 Standard False.

Design and procedure. A within-subjects design was used with each subject verifying each sentence type within a single session. The subjects, tested individually, were seated in a sound-attenuated booth in front of a CRT display. One response key was available to the left-hand index finger. This key simultaneously initiated a trial and started a millisecond timer. Two other response keys were operated with the index and middle fingers of the right hand. For half of the subjects, the index finger was designated "true," and the middle finger "false"; this was reversed for the other subjects.

The subjects were told that they would see sentences of the form *Some X are Y*. They were to decide, as quickly and accurately as possible, whether each sentence was true or false and to indicate that decision by pressing the appropriate response key. The subjects were then given 20 prepractice trials to respond "true" and "false," respectively, to the words "true" and "false" as they appeared on the screen. This ensured familiarity with the response-key assignment. The subjects then received 120 practice trials and any errors that they made were pointed out. This ensured that subjects understood that the task required decisions about literal (i.e., categorical) truth. The test list was presented immediately afterward.

At the beginning of each trial, subjects viewed a fixation point (an asterisk) at the screen location where a test sentence would appear. When the left-hand start button was pressed the asterisk was replaced by a sentence, and a timer would start. This timer was stopped when the subject pressed a decision-response button. Stimulus display and response timing were under computer control (Commodore PET model 2001, with 1-msec response timing accuracy). At the end of the test session subjects were asked to write as many of the test sentences as they could recall, and were given 10 minutes to do so.

Results and Discussion

We found a substantial metaphor interference effect. The mean RT's were calculated from correct responses to the test sentences, and RT's exceeding 4000 milliseconds were treated as errors. These data are shown in Table 1. Analyses of variance with both items and subjects as random factors (Clark, 1973) were applied to these data. A preliminary analysis revealed that the mean RT for high-typical True sentences was faster than that for low-typical True sentences. This comparison serves as a manipulation check to ensure that subjects in this task are processing sentences as they normally do in sentence-verification experiments, and this result is typical. (See, e.g., McCloskey and Glucksberg, 1979; Smith, Shoben, & Rips, 1974). Since this comparison is of no further interest, the True-sentence data were pooled, yielding four sentence types for more detailed analysis (See Table 1).

The mean RT for these four sentence types differed reliably, *min* $F'(3,202) = 4.51$, $p < .01$. A Duncan multiple range test (alpha level = .05) indicated that (a) mean RT for True sentences was reliably faster than for any other sentence type; (b) Standard False and Scrambled Metaphors were not reliably different; (c) Metaphors were reliably slower than Standard False and Scrambled Metaphors.

The first finding, that True sentences are verified faster than False, is a typical finding in the literature (cf., McCloskey & Glucksberg, 1979; Smith et al., 1974), and need not concern us further. The second finding, that Standard False and Scrambled Metaphors did not differ in mean RT, suggests that the lexical items that we used to construct the various sentence types are relatively homogeneous. At the very least, the words that we used to construct the Metaphor sentences are no more difficult to

TABLE 1
MEAN REACTION TIME (MSEC) TO MAKE LITERAL TRUE–FALSE DECISIONS AS A FUNCTION OF SENTENCE TYPE, EXPERIMENT I

	Reaction time			
	Mean	SD		
Sentence type, with examples		Subjects	Items	Percent errors
True ($n = 80$) Some fish are trout Some birds are eagles	1114	153	114	3.0
False ($n = 40$) Some fish are eagles Some birds are trout	1185	194	95	5.3
Scrambled metaphors ($n = 20$) Some jobs are snakes Some roads are jails	1162	198	117	0.75
Metaphors ($n = 20$) Some jobs are jails Some roads are snakes	1239	198	178	3.3

process than the words used in the Standard False sentences. The third finding is the critical one for our argument. The Metaphor sentences took significantly longer to judge as false than their scrambled counterparts. This suggests that our subjects were, at some level, apprehending the nonliteral meanings of the metaphor sentences. When these nonliteral meanings were apprehended, they produced a conflict in truth value: the literal meanings were false, the nonliteral true. This conflict, in turn, delayed or slowed down the final response.

The recall data are consistent with this interpretation. The best recall was for True sentences (24.9%). This is not surprising. These sentences are meaningful and they were categorized, and the recall protocols reflected their category structure. Recall performance for Standard False and Scrambled Metaphors was quite poor (1.1 and 1.8%, respectively). Recall for the Metaphor sentences was considerably higher (12.3%), even though these were also judged "false" during the sentence-verification phase of the experiment. This is consistent with the notion that the Metaphor sentences were processed more deeply than the other false-sentence types (see Craik & Lockhart, 1972, on the relationship between depth of processing and memory performance), as well as with the notion that these sentences were more meaningful than the other false-sentence types (see Miller & Selfridge, 1950).

There is, however, an alternative interpretation of both the verification latency and sentence recall data. We know that subjects take longer to reject false sentences with related subject and predicate nouns than with unrelated nouns. For example, it takes longer to judge that *vegetables are apples* is false than that *vegetables are hammers* is false (McCloskey & Glucksberg, 1979). The subject and predicate nouns of the Metaphor sentences are, by definition, related to each other. Were it to be otherwise, the Metaphors would not be readily interpretable. The critical question is: Are these subject and predicate nouns related in isolation, or does it require a sentence interpretation to make them so?

Experiment II addresses this question by keeping the relevant relations between subject and predicate nouns constant while independently varying metaphor quality.

Experiment II: Poor Metaphors Can Be Ignored

The quality of a given metaphor can be affected by the quantifier that is used. For example, the statement *Some surgeons are butchers* is intuitively more plausible than *All surgeons are butchers*. If the metaphor interference effect of Experiment I is attributable entirely to semantic or associative relationships between subject and predicate nouns, irrespective of sentence context, then the quantifier should not influence that metaphor interference effect. The differences in RT between Metaphor and Scrambed Metaphor sentences should be equal for *All* and *Some* statements.

Alternatively, if the *Some* version of a statement produces a better and more readily interpretable metaphor than the corresponding *All* version, then the differences in RT between Metaphors and Scrambled Metaphors will not be equal for *All* and *Some* statements. Instead, that difference will be greater for *Some* than for *All* statements. We should, therefore, expect a robust metaphor interference effect with *Some*, and either a reduced or no interference effect with *All*.

Method

Subjects. Sixty undergraduate students at Princeton University provided metaphor ratings. An independent group of 25 students participated in a sentence-verification task modeled after that of Experiment I. Four additional subjects were excluded because of high error rates (over 5%), and 1 subject because of equipment failure.

Materials. We first constructed a set of 141 candidate metaphor sentences, similar to those of Experiment I. We then constructed three different questionnaires so that each item appeared equally often in each of three quantifier contexts: *All X are Y; Some X are Y; An X is a Y;* for example, *All surgeons are butchers, Some surgeons are butchers, A surgeon is a butcher.* Only one version of any sentence appeared in any given questionnaire. Three groups of 20 subjects were asked to rate each sentence on a 7-point scale of how good a metaphor each one was, and how familiar each sentence was.

After we had obtained the ratings we selected the 24 items that had the largest difference in metaphor-goodness rating between their *Some* and *All* versions. The *Some* versions were in all cases rated as better metaphors than the corresponding *All* versions. For example, *Some marriages are iceboxes* had a mean rating of 4.95, while its counterpart, *All marriages are iceboxes,* had a mean rating of only 2.15. The mean metaphor-goodness ratings for the *Some* items was 5.33, range 4.50–6.45. The mean rating for the *All* items was 3.65, range 2.15–4.50. The familiarity ratings were not used for item selection. These, however, also differed as a function of quantifier: 4.84 and 3.80 for the *Some* and *All* versions, respectively. The ratings for the *A* (indefinite article) versions of the metaphor statements were collected for other purposes and were not used here.

The subjects and predicates of these 24 Metaphor sentences were re-paired to generate a list of 24 Scrambled Metaphors. The literal True and Standard False statements of Experiment I were also used here, but in both *Some* and *All* versions. There were also two types of Standard False *All* sentences: those with unrelated subject and predicate nouns (e.g., *All vegetables are hammers*), and those with related subject and predicate nouns (e.g., *All vegetables are potatoes*). This last sentence type was included to ensure that subjects would attend to the quantifiers. If they did not, then they would make errors on the high-related false sentences because these sentences are true in their *Some* versions.

A practice list of 96 sentences was constructed to have the same proportions of sentence types as a 192-item test list. Both practice and test lists had 12 filler sentences as warm-up items.

Procedure. We followed the procedures of Experiment I. Test lists were assigned to subjects in the following way. All subjects received the same set of True and Standard False sentences. Each subject had half the Metaphors with *Some,* and half with *All.* Each Metaphor appeared equally often in each version, and each subject had a different assortment of the *Some* and *All* versions. The Scrambled Metaphors were assigned to lists in the same fashion. This procedure confounds subjects and items and eliminates the need for separate analyses for subjects and for items. A different block-randomized order was used for each subject's test list.

Results and Discussion

The mean RT's were calculated for correct responses only, and RT's exceeding 4000 milliseconds were automatically counted as errors. As in Experiment I, a preliminary analysis revealed the expected typicality effect—True high-typical items (1419 msec) were verified faster than True low-typical items (1580 msec), $p < .01$, Duncan multiple range test. Since this effect is of no further interest here, the True-sentence data were pooled for further analyses.

Mean RT's as a function of sentence type and quantifier are presented in Figure 1. Analysis of variance of these data yielded a significant effect of sentence type, $F(3,72) = 5.30$, $p < .01$; no reliable main effect of quantifier, $F(1,24) = 1.26$, $p > .05$; and a significant sentence-type $\times$ quantifier interaction, $F(3,72) = 17.05$, $p < .001$. The error rates were comparable to those of Experiment I, ranging from less than 1 (*False–Some* sentences) to 6% (*False–All* sentences).

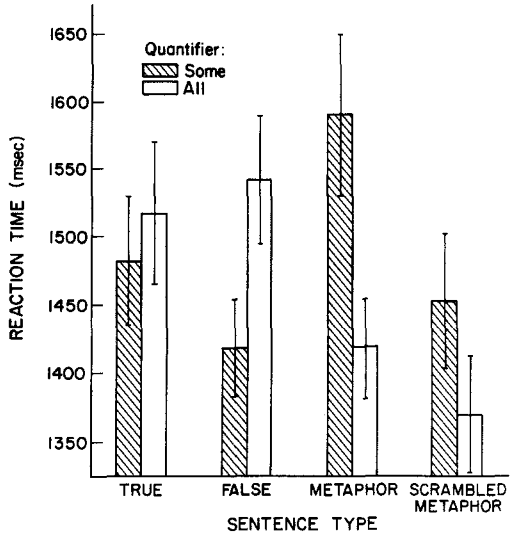

FIG. 1. Mean reaction time (and standard errors of the means) to make literal true–false decisions as a function of sentence type, Experiment II.

We obtained a metaphor interference effect with the better metaphors (those quantified with *Some*), but not with the poorer ones (those quantified with *All*). The *Some* versions of the Metaphor sentences were reliably slower than the Scrambled *Some* Metaphors, $p < .01$ (Duncan multiple range test in this and in all subsequent pairwise comparisons). This replicates the analogous finding of Experiment I. The *All* versions of these very same sentences were not slower than either of the Scrambled Metaphor types, $p > .05$. Finally, the *Some* Metaphors were significantly slower than the *All* Metaphors and slower than the Scrambled *All* Metaphors. Apparently, the quantifiers *Some* and *All*, which had made a difference in the metaphor ratings, also made a difference in how quickly the metaphors could be judged literally false.

These results cannot be attributed simply to the quantifiers themselves. With the True sentences, *Some* and *All* versions were not reliably different. With Standard False sentences, the *All* versions were reliably slower than *Some* ($p < .01$), an effect that is opposite to the effect for Metaphors. The slower RT for the *All* versions of the Standard False sentences is attributable to the relative difficulty of such high-related False sentences as *All vegetables are potatoes* as compared to *All vegetables are hammers*. Finally, the *Some* and *All* versions of the Scrambled Metaphors did not differ reliably from each other.

Because the subject- and predicate-noun pairs of the *Some* and *All* versions of the metaphors were identical, the metaphor interference effect of the *Some* versions cannot be attributed solely to the semantic relatedness of those noun pairs, or to the associative relations between those pairs either. The fact that we obtained a difference in RT between *Some* Metaphor sentences and their scrambled counterparts (a metaphor interference effect), but no such difference between the *All* versions and their scrambled counterparts, must be attributable to some sentence characteristic that varies with the quantifier. The items were, of course, chosen because they differed in metaphor ratings. They also differed in ratings of familiarity, with the *All* versions rated as somewhat less familiar than the *Some* versions. It is rather likely that both metaphor goodness and familiarity can contribute to the relative transparency or comprehensibility of a metaphor statement. The more transparent its figurative meaning, the more likely will that meaning "leap out," and hence the more difficult will it be to judge literally false.

Still another possibility is that *Some* metaphors cannot be ignored, but *All* can be. It may be the case that only existentially quantified statements are processed automatically for nonliteral meanings, while universally quantified statements are not. It is somewhat difficult to imagine why this might be so, even in the context of a sentence-verification experiment. Still, we decided to see if we could obtain evidence for nonoptional processing of the figurative meanings of universally quantified statements. Experiment III was designed to demonstrate a metaphor interference effect with *All* versions of metaphor sentences that (a) had about the same mean familiarity ratings as the *All* metaphors of Experiment II; but (b) had higher metaphor goodness ratings than the *All* metaphors of Experiment II. Accordingly, we selected metaphor items that had just as high metaphor goodness ratings in their universal (i.e., *All*) form as in their existential (i.e., *Some*) form. If we now obtain differences in RT between metaphors and scrambled metaphors regardless of quantifier, then rated familiarity per se is not a sufficient condition for the metaphor interference effect. It may still, of course, be a necessary condition.

Experiment III: Neither All nor Some Metaphors Can Be Ignored

Method

Subjects. Twenty-four college students served as paid volunteers. All were native English speakers and none had had prior experience in sentence-verification or language-rating experiments. Three additional subjects were excluded from data analyses because of excessive error rates (over 5%), and 1 subject's data were lost because of equipment failure.

Materials and procedure. The Metaphors and Scrambled Metaphors of Experiment II were replaced by 24 other Metaphors on the basis of their metaphor and familiarity ratings. These 24 Metaphors each had a mean rating of 4.00 or better on the 7-point rating scale, and their *All* versions (4.62) did not differ reliably from their *Some* versions (4.78). For example, *Some rumors are diseases* and *All rumors are diseases* had the same metaphor rating, 5.50. The mean familiarity ratings for the *Some* and *All* versions of these Metaphor sentences were 3.91 and 3.69, respectively. Recall that the mean familiarity ratings for the *All* Metaphors of Experiment II was 3.80. If we obtain a metaphor interference effect with the *All* Metaphors of this experiment, then familiarity rating cannot be a sufficient condition for that effect.

In every other respect the materials, design, and procedures of this experiment were identical to those of Experiment II.

Results and Discussion

Mean RT's were calculated from correct test responses only, and responses exceeding 4000 milliseconds were automatically treated as errors. Again, mean RT for high-typical True sentences was significantly faster than for low-typical True sentences, and the True-sentence data were pooled for further analyses.

Mean RT as a function of sentence type and quantifier are presented in Table 2. Analysis of variance yielded a significant effect of sentence type, $F(3,69) = 10.73, p < .001$. Neither the main effect of quantifier nor the interaction of quantifier by sentence type was reliable. As the data in Table 2 clearly show, the effect of quantifier on the Metaphor and Scrambled Metaphor reaction times is negligible. Duncan multiple range tests show that the RT for Metaphors is significantly greater than the RT for Scrambled Metaphors, $p < .01$. The quantifier in this set of Metaphor sentences had no effect on their metaphor ratings, and also had no effect on their relative verification latencies. While the evidence on the

TABLE 2
MEAN REACTION TIME (MSEC) TO MAKE LITERAL TRUE–FALSE DECISIONS AS A FUNCTION OF SENTENCE TYPE AND QUANTIFIER, EXPERIMENT III

	Quantifier			
	Some		All	
Sentence type	Mean	SD	Mean	SD
True	1364 (4.0)	307	1381 (3.0)	317
Standard false	1305 (2.0)	255	1374 (8.0)	246
Scrambled metaphors	1241 (0.3)	279	1236 (1.0)	277
Metaphors	1303 (0.3)	322	1314 (1.0)	353

Note. Percent errors are shown in parentheses.

role of familiarity cannot be conclusive because we have yet to perform a direct test of that role, the data do suggest that the interference we have observed in all three experiments is attributable to the nonliteral meanings of the metaphor sentences. In particular, the *All* metaphor items in Experiment III had a marginally lower mean familiarity rating than those of Experiment II, yet we obtained a metaphor interference effect with the former and not with the latter. This finding argues that familiarity, while perhaps a necessary condition, is not a sufficient one for that effect.

General Discussion

We have shown that sentences of the form *Some/All X are Y* take longer to judge as literally false when they have readily interpretable nonliteral meanings than when they do not. This result suggests that people can no more easily refuse to understand statements such as *Sam is a pig* than statements such as *Tomatoes are red.* In this sense, nonliteral comprehension is automatic—even in the relatively impoverished context of a laboratory experiment. People do not seem to have the option to ignore simple metaphors.

What is the nature of this nonliteral comprehension? One possibility is that we have a special, nonliteral comprehension device that is triggered automatically whenever a nonliteral meaning is available in an utterance. Not only is this view unparsimonious, it also has some serious difficulties. To begin with, nonliteral meanings are, in principle, always available, and so availability per se cannot serve as a reliable trigger for such a nonliteral comprehension device. What can serve this purpose? If the failure of a literal meaning to make sense is not a necessary condition for triggering metaphor comprehension, then we must assume that normal comprehension strategies and mechanisms must be capable of dealing with at least the initial stages of nonliteral comprehension. Otherwise, there would be no way to call up those special subroutines that might be needed for nonliteral comprehension.

An alternative and more parsimonious hypothesis can be entertained. People may use the same set of strategies and comprehension mechanisms for interpreting both the literal and nonliteral meanings of sentences, and for making decisions about the truth values of such sentences. Miller (1979), among others, argues that metaphors are implicit forms of comparison statements. A sentence such as *Some salesmen are bulldozers* can be paraphrased as *Some salesmen are like bulldozers.* In either the implicit or explicit form, the basis or ground for the similarity assertion must be discovered or apprehended. Either form of the statement is said to be understood when the relevant properties of *bulldozer* are appropriately applied to *salesmen.* This set of properties constitutes the ground of the metaphor (Richards, 1936), and often involves a salient, perhaps stereotypical property of the sentence predicate. This property then becomes activated in the sentence subject. Being inexorably and insensitively pushy may well be a stereotypical property of bulldozers, and an easily activated and relevant property of at least a subset of salesmen.

This view of metaphor processing implies that people assess the degree to which certain properties of a predicate noun are shared by the subject noun, or at least are applicable to that subject noun, and that people do so whenever statements of the form *X is Y* are encountered. In addition, such a comparison process may well be used to make verification decisions. Both McCloskey and Glucksberg (1979) and Smith et al. (1974) have argued that people assess the truth of category-membership statements by assessing the extent to which properties are shared between subject- and predicate-noun concepts. If this is so, then it should make no difference if a comparison statement is implicit, as in *X is Y,* or explicit, as in *X is like Y.* In either case, people will evaluate the similarity between subject- and predicate-noun concepts, and presumably will also apprehend the basis for that similarity.

These arguments suggest just what it might mean for metaphor comprehension to be automatic (Barsalou, note 1). When people try to understand statements of the form *X is Y,* they normally follow Grice's (1975) cooperative principle and assume that the statement is informative. People also assume that the speaker (or writer) is following the conventions for marking given and new information (Clark & Haviland, 1977). If this is true, then people will look for new information in *Y* that would be informative about *X*. One way to satisfy these conditions is (a) for the information in *Y* to be salient and normally activated in the representation of *Y* whenever *Y* occurs; and (b) for that information to be unactivated in the representation of *X* when *X* occurs in isolation. When *X* occurs as the subject of a sentence with *Y* as the sentence predicate, the salient property or properties of *Y* become automatically activated in the representation of *X*. This raises two important questions. Why is this activation automatic, and just what properties of *Y* behave in this way?

Consider first the automaticity question. Ortony (1979) suggests that implicit comparison statements may be a simple and common form of predication. Such predication occurs continually in everyday discourse, and it involves the activation of previously unactivated properties in the representation of a statement's topic concept. Because this is such a well-practiced set of events, it has become automatic. If it has become automatic, then we would expect that it would be difficult, if not impossible, to control voluntarily (cf. Posner & Snyder, 1975; Shiffrin & Schneider, 1977). In contrast, verification, particularly of isolated statements that express analytic truths and falsehoods, occurs quite rarely outside the psychology laboratory. It is therefore not surprising that ordinary predication is not inhibited even when it may interfere with verification performance. Viewed in this way, our results are not particularly novel or noteworthy. What is novel and potentially noteworthy is that the automaticity of language comprehension should be extended to include both literal and figurative predication.

Given that predication will occur automatically when people read or hear simple statements of the form *X is Y,* what determines just what properties of *Y* are predicated of *X?* Consider the statement *Some surgeons are butchers*. The subject of the sentence, "surgeons," is the topic of the metaphor. The sentence predicate, "butchers," is the vehicle. The relevant relationship between topic and vehicle, the ground of the metaphor (Richards, 1936), consists of those properties of the vehicle that are predicated of the topic. What are those properties? For the statement *Some surgeons are butchers,* most adult English speakers would agree that it is a negative comment about some surgeons. Indeed, to say that anyone is a butcher is to say something negative, unless one intends to identify a person's actual occupation.

It may be that *butcher* is a stock vehicle that is used and understood in cliché-like fashion. The statement *X is a butcher* can always be taken to mean that *X* is negatively evaluated, and that *X* is grossly and characteristically incompetent as well. The particular way that *X*'s incompetence is instantiated will depend on who or what *X* might be. If *X* is a surgeon, then the incompetence takes the form of botched operations, with bleeding, disfigurement, and death among the likely consequences. If a pianist is a butcher, then the incompetence is not merely the forgetting of certain parts of piano pieces or a lack in the repertoire, but rather that the music is plowed through insensitively, too loudly, without any hint of subtlety or beauty. If a cabinet maker is a butcher, then he does not merely make mistakes, but really messes up the wood. In this sense, the topic and vehicle of a metaphor interact (Black, 1962; 1979). The topic selects from among a small set of alternatives those properties of a vehicle that are appropriate. In the case of the vehicle "butcher," these properties will include: (a) negative evaluation; (b) if competent, then murderous. This would be a fairly straightforward extension of the meat-cutting property of a literal butcher from edible livestock to animals in general including, in the extreme, human beings. (c) If incompetent, then grossly and vulgarly so. "Butchers" are not merely incompetent, they really make a mess of things. (d) The instantiation or realization of incompetence will be determined by the character of the activities that are normally associated with the topic concept. In this way, the general, stereotypical properties of the vehicle are made specifically relevant to the topic. This way of specifying particular subsets of more general characteristics is not limited to figurative predication. Ordinary adjectival predication operates in precisely the same way (Vendler, 1963). To say that *X is good* implies a positive evaluation of *X*. The particular characteristics of "goodness" will vary as a function of what *X* is. A good poem is good in a way different from a good symphony, and both are different in their goodness from such things as good whippings, good men, or good automobiles. Stock vehicles, such as "butcher" could operate in precisely this way: providing a set of general characteristics which are then specified by the topic of a metaphor.

If "butcher" is a stock vehicle, then perhaps other concepts can play this role as well. If we consider just those metaphors that we have used in these experiments, it soon becomes apparent that many, if not all, of the metaphor predicates can be readily characterized. For example, if anything is an icebox, then it is negatively evaluated and it is cold in some sense. If anything is a closet, then it is private and secretive in some sense. Assume that people in any given language community have a relatively small stock of such conventional vehicles (see Lakoff & Johnson, 1980). Each vehicle may be used to characterize any one of a large and open class of potential metaphor topics. The ground of the metaphor can sometimes be determined entirely by the vehicle. More often, the particular ground will be specified from a relatively small set of alternatives by the topic concept. The vehicle provides a general ground, the topic permits an instantiation of that general ground.

This notion can be described as the used-car-lot theory of figurative predication. A relatively small set of stock vehicles is available. Whenever they are used in noncategorial statements, they provide stereotypical predications of their metaphor topics. This view is consistent with at least one important property of nonliteral comparison statements, their lack of symmetry (Ortony, 1979). When the subject and predicate nouns of a metaphor are interchanged, the statement either becomes nonsensical, or the ground of a metaphor is changed, often quite drastically. Reversing a metaphor will yield nonsense (a) if the original topic is not a stock vehicle (e.g., *Sam is a pig* vs *A pig is Sam*); or (b) if the stereotypical vehicle properties of the original topic cannot be sensibly predicated of the new topic, (e.g., *Some marriages are iceboxes* vs *Some iceboxes are marriages*). A metaphor can be reversed to produce a different and intelligible statement only when the new vehicle's stereotypical properties can be appropriately predicated of the new topic. The statement *Some surgeons are butchers* expresses a negative comment about surgeons. The statement *Some butchers are surgeons* expresses a positive comment about butchers. This asymmetry seems to be characteristic of comparison statements (Tversky, 1977) and of nonliteral comparison statements in particular.

This view of how any number of novel metaphorical expressions may be constructed from a reasonably manageable stock of stereotypical vehicles leaves open the important question of how speakers may generate and how listeners may comprehend metaphors that do not employ conventional vehicles. Perhaps Searle's (1979) view of how metaphors are understood is applicable to that subset of figurative expressions that do not use stock vehicles, but instead use a brand new vehicle for the first time. In such cases, we would not expect automatic and facile comprehension. Instead, we would expect some initial puzzlement, and we would not be surprised to find differences of opinion as to how such novel expressions should be interpreted. Indeed, literary criticism provides no end of such examples for literal as well as for poetic or figurative texts. If literature is too esoteric an example, recall instead the kind of literal instructions that all too often accompany children's toys, stereo equipment, or other things that need to be assembled at home. If initial puzzlement and occasional errors occur with literally intended language, we should not be very surprised when puzzlement and error occur with figurative language as well. Conventionality of content and conventionality of expression may well be the most important determiners of the ease and accuracy of language comprehension. Whether an expression is intended literally or not may have little or no effect upon either the ease with which that expression is understood, or upon the selection of comprehension strategies that may be used to accomplish that understanding.

References

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monographs*, 1969, **80** (3, Pt. 2).

Black, M. Metaphor. In M. Black (Ed.), *Models and metaphors*. Ithaca, N.Y. Cornell Univ. Press, 1962.

Black, M. More about metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge: Cambridge Univ. Press, 1979.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Clark, H. H. Responding to indirect speech acts. *Cognitive Psychology*, 1979, **11**, 430–477.

Clark, H. H. & Haviland, S. E. Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension*. Norwood, N.J.: Ablex, 1977.

Clark, H. H., & Lucy, P. Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 56–72.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Gibbs, R. W. Contextual effects in understanding indirect requests. *Discourse Processes*, 1979, **2**, 1–10.

Grice, H. P. Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts*. Vol. 3. New York: Academic Press, 1975.

Lakoff, G., & Johnson, M. *Metaphors we live by*. Chicago: Univ. of Chicago Press, 1980.

Lyons, J. *Semantics*. Vol. 2. Cambridge: Cambridge Univ. Press, 1977.

McCloskey, M. & Glucksberg, S. Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, 1979, **11**, 1–37.

Miller, G. A. Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge: Cambridge Univ. Press, 1979.

Miller, G. A., & Johnson-Laird, P. N. *Language and perception*. Cambridge, Mass.: Belknap/Harvard Univ Press, 1976.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *American Journal of Psychology*, 1950, **63**, 176–185.

Ortony, A. Beyond literal similarity. *Psychological Review*, 1979, **86**, 161–180.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17**, 465–477.

Pollio, H. R., & Burns, B. C. The anomaly of anomaly. *Journal of Psycholinguistic Research*, 1977, **6**, 247–260.

Posner, M. I., & Snyder, R. L. Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N. J.: Erlbaum, 1975.

Richards, I. A. Metaphor. In I. A. Richards, *The philosophy of rhetoric*. London: Oxford Univ. Press, 1936.

Searle, J. R. Metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge: Cambridge Univ. Press, 1979.

SHIFFRIN, R. M., & SCHNEIDER, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 1977, **84**, 127–190.

SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, **81**, 214–241.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

SWINNEY, D. A., & CUTLER, A. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 1979, **18**, 523–534.

TVERSKY, A. Features of similarity. *Psychological Review*, 1977, **84**, 327–352.

VENDLER, Z. The grammar of goodness. *Philosophical Review*, 1963, **72**, 446–465.

REFERENCE NOTE

1. BARSALOU, L. Personal communication.

(Received March 6, 1981)