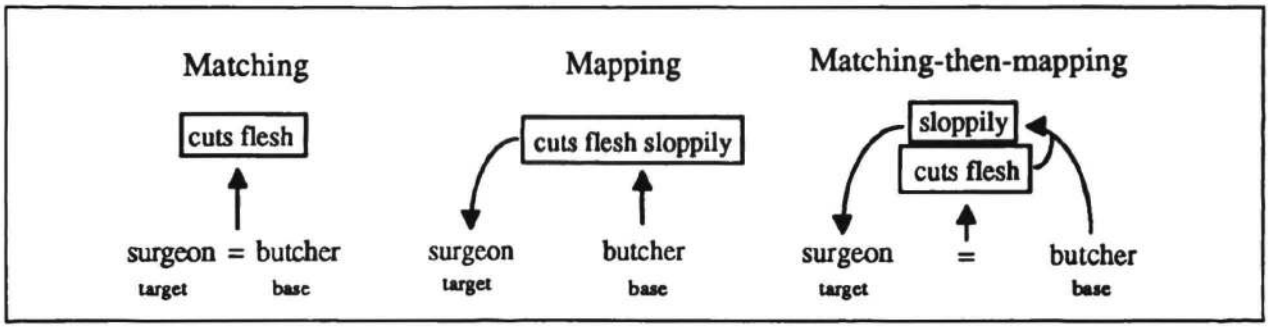

Figure 1. Models of metaphor comprehension

Structure-mapping Engine (SME), processing begins by matching the representations of the base and target. Once a global alignment is discovered, predicates may be carried over from the base to the target using the systematicity criterion: given a common system of interconnected predicates, predicates that belong to that system in the base but are not yet present in the target are mapped to the target as *candidate inferences* (Falkenhainer, Forbus, & Gentner, 1989). Returning to the metaphor *A surgeon is a butcher*, a Match-then-map process asserts that people first find the commonality that both surgeons and butchers cut flesh; then further properties belonging to this common 'cutting-flesh' system are mapped from butchers to surgeons: e.g., that butchers cut sloppily.

Since the research here primarily addresses the initial stages of processing, we will lump match-then-map models with the simple matching models and call them collectively *match-first models*; these will then be contrasted with *map-first* models.

Strengths of the Processing Accounts

Match-first models capture the intuition that metaphor shares with similarity a focus on commonalities (e.g., Tversky, 1977). A further advantage of the match-first models is their ability to deal with the problem of property selection when the same base is compared with different targets. For example, consider the metaphors *The surgeon is a butcher* and *The general is a butcher*. Though they have the same base, they convey quite different meanings: the first suggests a clumsy surgeon, the second a ruthless and efficient general. This property selection problem is not as easily handled by map-first models, which must account for how *butcher* gives rise to two different abstractions in these two contexts (Map-first models can be augmented with the assumption that people try abstractions sequentially, until one fits the target, though this explanation seems cumbersome at best.) Finally, the match-first view can predict further inference as part of a secondary mapping stage.

The map-first perspective, as exemplified by Glucksberg and Keysar's (1992) Category-mapping theory, has its appealing aspects as well. First, it captures the intuition that there should be an intimate relation between metaphor and categorization. Second, the map-first view explains why metaphors are often directional. Just as the class-inclusion statement "A surgeon is a doctor." cannot be reversed to make "A doctor is a surgeon.", neither can the metaphorical statement "A vacation is a doctor." be reversed to make "A doctor is a vacation." The map-first view also offers an intuitive explanation for our ability to understand metaphors that convey new inferences about the target. On hearing "The waiter is a skyscraper" we understand that the waiter is tall even through "tallness" is not necessarily a feature present in our prior representation of waiters. This importing of new features in to the target is a problem for simple matching models.

Testing the Models

The match-first and map-first models make different predictions for the time course of processing of metaphors. According to the Category-mapping model (a map-first model), processing begins with the base term, from which a category must be derived to apply to the target term. In contrast, according to match-first models, processing begins with a comparison of the two terms. This suggests a way to test these theories. If processing begins with the base, as is implied by a category-mapping process, then metaphoric processing should be facilitated if people see the base in advance of the metaphor. More specifically, there should be greater facilitation when the base is given in advance than when the target is given in advance. (Some facilitation is expected in either case, under any model, simply by virtue of permitting a head start in encoding.)

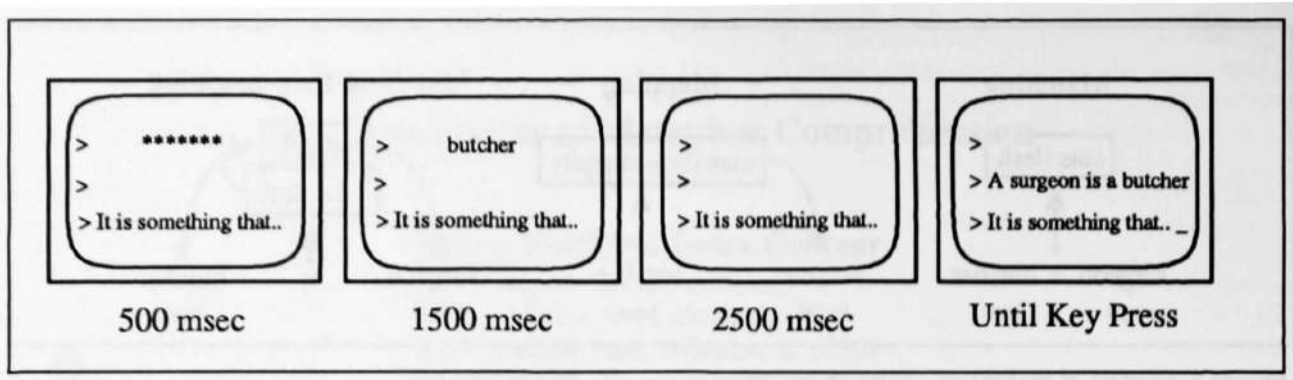

Figure 2: General order and timing of presentations in Experiments 1, 2 and 3.

In contrast, if metaphor interpretation begins with a matching process, there should be no differential advantage for seeing the base vs. the target prior to seeing the metaphor.

Based on this logic, we carried out three experiments to compare the adequacy of these models as psychological accounts of metaphor interpretation. In all three experiments, we recorded subjects' time to interpret metaphors. The metaphors could be primed by an immediately prior presentation of the base term, the target term, both terms, or nothing (no prime). The key question is whether there is more facilitation for base primes than for target primes.

Experiment 1

Method

Subjects. The subjects were 24 Northwestern University students.

Materials. Twenty-four metaphors were drawn from the literature. Prior to the study, in order to establish the preferred direction of the metaphors, we asked 40 subjects to order each pair of terms to make their preferred metaphor. In this way we were able to independently determine the most natural base and target.

Procedure. Subjects were run on individual computers in groups of 2 to 5 subjects at a time. Subjects were shown a series of metaphors, randomly ordered, and instructed to begin typing in an interpretation to each metaphor as soon as they had it well formulated. (These instructions were given to forestall subjects' adopting a strategy of starting to type immediately and then pausing to develop an interpretation.) Prior to seeing each metaphor, subjects were shown either the target, the base, or a blank line. Figure 2 shows the order in which information appeared. Subjects were instructed to use the preceding words to get a head start on their interpretation. Interpretation-time was recorded from the moment the metaphor appeared on the screen until the subject pressed the first key for his or her interpretation. The time to type each interpretation was also recorded.

Design. All prime types (i.e. TARGETs, BASEs, or BLANKs) were presented to each subject and counterbalanced across all metaphors.

Results and Discussion

The mean interpretation times are shown in Table 1. (All analyses throughout this paper are based on subjects' mean responses.) Contrary to the predictions of the Category-mapping model, there was no advantage for BASE primes over TARGET primes; indeed, there was a nonsignificant difference in the reverse direction. A one-way repeated-measures analysis of variance showed a significant effect of Prime type, F (2,46) = 7.72, $p < 0.001$. Pairwise Bonferroni tests indicated that at the 0.05 level, interpretation times were faster for BASEs than for BLANKs, $t(23) = 3.25$, and also for TARGETs than for BLANKs, $t(23)=3.16$. Interpretation times for TARGETs and BASEs, however, did not differ significantly, $t(23)=0.63$.

These results provide no evidence for the category-mapping prediction that processing begins with the base. The results are consistent with the match-first models (although only by default). However, one concern in interpreting these results is that the method may not have been fair to the Category-mapping model, since subjects were not told whether the prime was a base or a target. One might argue that the assumptions of the Category-mapping model are that subjects treat the base term differently from the target term. On this reasoning, it is to be expected that subjects would be unable to begin

TABLE 1

Mean Times to Begin an Interpretation from Experiment 1

Prime Type	BASE	TARGET	BLANK
Means	4951	4828	5648
Standard Error	996	1090	1475

TABLE 2

Mean Times to Begin an Interpretation from Experiment 2

Prime Type	BOTH	BASE	TARGET	BLANK
Means	3048	4086	4420	4829
Standard Error	1576	1474	1658	1436

interpretation until they know that they have the base term. To address this possibility, in Experiment 2 the prime's role (i.e. whether it was the target or base) was made explicitly clear. This was done to encourage the fullest possible use of the primes. One other change was to add a fourth condition in which both primes were shown together. This amounts to simply showing the whole metaphor from the start. This condition tested the prediction of the match-first views that having both terms at the outset should be faster than having only one (since interpretation can begin only when both terms are present). The Category-mapping model makes no strong predictions as to whether seeing both primes should lead to faster interpretation times than seeing the base. Thus, finding an advantage for both terms over one would not distinguish match-first from map-first positions, but *failure* to find such an advantage would count against match-first models.

Experiment 2

Method

Subjects. The subjects were 40 Northwestern University students.

Materials. The metaphors were the same as those used in Experiment 1.

Procedure. The procedure was the same as that used in Experiment 1 with two exceptions. First, the roles played by the primes were made explicit by putting them into sentence frames. So, if *butcher* was used as a base prime, subjects saw *A something is a butcher.* Similarly, if *surgeon* was used as a target prime, subjects saw *A surgeon is a something.* In the BLANK condition, subjects saw *A something is a something,* and in the newly added BOTH condition, subjects saw the complete metaphor: e.g. *A surgeon is a butcher.*

Design. As in Experiment 1, all prime types (i.e. TARGET, BASE, BLANK, BOTH) were presented to each subject and counterbalanced across all metaphors.

Results and Discussion

The mean interpretation times are shown in Table 2. Again contrary to the predictions of the Category-mapping model, there was no significant advantage for BASE primes over TARGET primes, although this time the difference was in the predicted direction. A one-way repeated measures analysis of variance indicated an overall significant effect of Prime type, $F(3,114) = 22.24$, $p < 0.001$. Pairwise contrasts using the Bonferroni t statistic at the .05 level indicated that interpretation times for BASEs (i.e., for metaphors given BASE primes) were faster than for BLANKs, $t(40) = 3.14$. In addition, BOTHs were faster than BASEs, $t(39) = 5.05$; and BOTHs were faster than TARGETs, $t(39) = 5.945$. No significant difference was found between TARGETs and BLANKs, $t(40) = 1.65$. Coming to the key result, no significant difference was found between BASEs and TARGETs, $t(40) = 1.79$.

Thus, as in Experiment 1, no support was found for the Category-mapping model. The match-first account fared better: as predicted, seeing BOTH primes resulted in faster interpretation times than seeing either BASEs or TARGETs alone. While this is not inconsistent with category mapping, it is a central prediction of the matching accounts.

However, although the Category-mapping model has received no strong support, there are some patterns that deserve consideration. First, although there is no significant BASE advantage over TARGETs, the direction of the means is consistent with the predictions of Category mapping. In addition, BASEs, but not TARGETs, were found to show an advantage over BLANKs. To examine the data more closely, we plotted cumulative curves of reaction times, as shown in

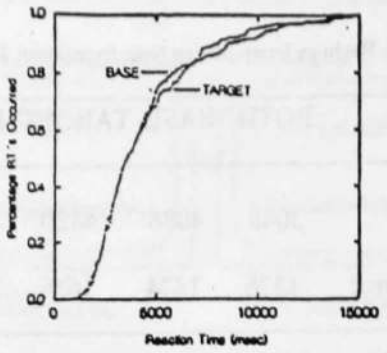

Figure 3: Exp. 2 Cumulative Percentage of Interpretation Times for Targets and Bases.

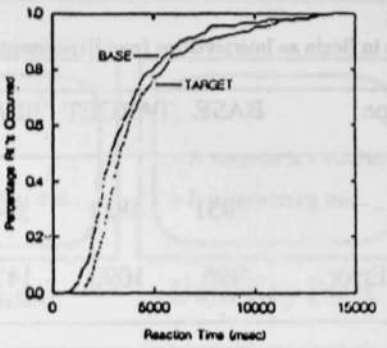

Figure 4: Exp. 4 Cumulative Percentage of Interpretation Times for Targets and Bases.

Figure 3. That is, we plotted the cumulative percentage of responses that had occurred by each duration from the beginning (i.e., the point when the metaphor appeared). If BASEs have an early advantage over TARGETs, we should see an initial difference between the cumulative curves (that is, there should be more BASE responses than TARGET responses at short durations). In fact, the two curves lie virtually on top of one another in the initial part of the distribution. The (nonsignificant) advantage of BASEs over TARGETs appears only at longer durations [2].

In summary, even when a prime's role was explicitly marked, no evidence was found for a category mapping. So far, the results are consistent with a match-first processing account, and inconsistent with a category-mapping account. In Experiment 3 we considered another factor that might affect whether people use a category-mapping model, namely, the conventionality of the metaphors. In the first two experiments the metaphors used were for most people relatively novel. Possibly this degree of novelty contributed to the apparent superiority of the match-first model over the map-first model in accounting for the patterns. Perhaps the category-mapping process is most probable when people are given metaphors whose bases have pre-stored 'stock' metaphorical interpretations. This would be reasonable, because with stock metaphors people could draw on relatively unambiguous existing abstractions. In Experiment 3 we tested this possibility. We applied the same basic priming procedure as in Experiment 2 to metaphors with conventional bases: e.g., *That waiter is a skyscraper* and *That plane is a dinosaur.*

[2] Conceivably, this pattern of BASE advantage that appears only for interpretation times of longer duration could suggest a Match-then-map model in which priming of the base term facilitates a later mapping and inference process.

Experiment 3

Method

Subjects. The subjects were 40 Northwestern University students.

Materials. Twenty metaphors were constructed with bases that possessed a stock metaphorical meaning.

Procedure. The procedure was the same as that used in Experiment 2.

Design. As in Experiment 3, all prime types (i.e. TARGETs, BASEs, BLANKs, and BOTHs) were presented to each subject and counterbalanced across all metaphors.

Results and Discussion

In Experiment 3, there was a major shift in the pattern of results: BASE primes led to significantly faster responding than TARGET primes, as predicted by the category-mapping model. A one-way repeated measures analysis of variance indicated a significant effect of Prime type, F (3,117) = 23.4, $p < 0.001$. Pairwise contrasts at the .01 level using the Bonferroni t statistic indicated that BASEs (M=3577) were faster than TARGETs (M=4218), t(39) = 3.08. In addition, BOTHs (M= 2770) were faster than BASEs, t(38) = 3.74; BOTHs were faster than TARGETs, t(38) = 6.620; and BASEs were faster than BLANKs (M= 4632), t(39) = 4.05. No significant difference was found between TARGETs and BLANKs, t(39) = 1.62.

This effect is consistent with the predictions of the Category-mapping model. This conclusion is supported by an inspection of the TARGET and BASE cumulative graphs in Figure 4, which show the BASE distribution clearly precedes the TARGET distribution. That is, in Experiment 3, subjects who saw BASE primes had a higher percentage of early responses than subjects who

saw TARGET primes. This suggests that category-first models may apply well to metaphors whose bases have stock metaphorical meanings, but not to novel metaphors.

As in Experiment 2, BOTHs were faster than BASEs and TARGETs, as is compatible with both the Category-mapping and the Match-first. Also as in Experiment 2, BASEs were faster than BLANKs, while TARGETs were not.

Summary

The results of these three experiments can be summarized by the following conclusions: 1) Interpretation of novel metaphors begins with matching, and 2) Interpretation of stock metaphors may begin with mapping. The first conclusion was supported by the failure to find an interpretation advantage for metaphors preceded by bases in the first two experiments (which used relatively novel metaphors). The second conclusion was supported by the results of Experiment 3, in which a clear advantage was observed for metaphors following base primes when the metaphoric bases possessed stock metaphoric meanings.

This suggests implications for psychological theories of metaphor as follows. Match-first theories, such as those of Ortony and Falkenhainer, Forbus & Gentner, best reflect how people understand metaphors that are relatively novel. But match-first theories, notably Glucksberg and Keysar's recent proposal, best represent how people understand metaphors with relatively conventional 'stock' base meanings.

The findings of these experiments connect well with recent work on conventional metaphors in natural language understanding. Both Martin (1991) and Gibbs (in press) have noted that any proper treatment of metaphor must not ignore the influence of preexisting structures in long-term memory. The findings of these experiments affirm the importance of these preexisting structures by demonstrating how they might have an effect on processing.

These findings also highlight a more general consideration. Most theorists would agree that the target and base of a metaphor either belong to some common category or share some common attributes. However, a question remains as to whether it is the common category that determines which features two things share, or whether it is shared features that determine a common category. We suggest that in the case of relatively novel pairings, a common category is derived from a Matching/Mapping process. That is, a process of matching (and subsequent mapping) leads to a common system which can serve as a common category. In time, this category may come to have conventional status, in which case the time course of subsequent metaphors is changed.

References

Burstein, M. (1983). Concept formation by incremental analogical reasoning and debugging. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine Learning* (Vol. 2, pp.351-369). Los Altos, CA: Morgan Kaufman.

Carbonell, J.G. (1983). Derivational Analogy: A theory of reconstructive problem solving and expertise acquisition. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine Learning* (Vol. 2, 371-392). Los Altos, CA: Morgan Kaufman.

Falkenhainer, B., Forbus, K. D., and Gentner, D. (1989/90). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.

Gibbs, R. W. (in press). Categorization and metaphor understanding. *Psychological Review*.

Gibbs, R. W. (1990). Psycholinguistics studies on the conceptual biases of idiomaticity, *Cognitive Linguistics*, 1, 207-246.

Gentner, D., Falkenhainer, and Skorstad. (1988). *Viewing metaphor as analogy*. In D.H. Herman (ed.), Analogical Reasoning (pp. 171-177), New York: Kluwer Academic Publishers.

Glucksberg, S., Keysar, B. (1990). Understanding metaphorical comparisons: Beyond Similarity. *Psychological Review*, 97, 3-18.

Greiner, R. (1988). Abstraction-based analogical inference. In D.H. Herman (ed.), *Analogical Reasoning* (pp. 147-170), New York: Kluwer Academic Publishers.

Hofstadter, D. R., Mitchell, M., & French, R. M. (1987). *Fluid concepts and creative analogies: A theory and its computer implementation* (Tech. Rep. No. 87-1). Ann Arbor, MI: University of Michigan, Fluid Analogies Research Group.

Kedar-Cabelli, S. (1985). Purpose-directed analogy. *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*. (pp. 150-159), Irvine, California.

Martin, J. H. (1991). *Computational theory of metaphor*. San Diego: Academic Press.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, 66, 161-180.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

Winston, P.H. (1980). Learning and reasoning by analogy. *Communications of the ACM*, 23, 689-703.