



Structure-Mapping in Metaphor Comprehension

Phillip Wolff,^a Dedre Gentner^b

^a*Department of Psychology, Emory University*

^b*Department of Psychology, Northwestern University*

Received 6 April 2009; received in revised form 12 January 2011; accepted 19 January 2011

Abstract

Metaphor has a double life. It can be described as a directional process in which a stable, familiar base domain provides inferential structure to a less clearly specified target. But metaphor is also described as a process of finding commonalities, an inherently symmetric process. In this second view, both concepts may be altered by the metaphorical comparison. Whereas most theories of metaphor capture one of these aspects, we offer a model based on structure-mapping that captures both sides of metaphor processing. This predicts (a) an initial processing stage of symmetric alignment; and (b) a later directional phase in which inferences are projected to the target. To test these claims, we collected comprehensibility judgments for forward (e.g., “A rumor is a virus”) and reversed (“A virus is a rumor”) metaphors at early and late stages of processing, using a deadline procedure. We found an advantage for the forward direction late in processing, but no directional preference early in processing. Implications for metaphor theory are discussed.

Keywords: Metaphor; Analogy; Structural alignment; Career of metaphor; Embodied metaphor

Metaphor lives a double life. On the one hand, metaphors convey insight from one domain (called the *base* [or *source* or *ground*] of the metaphor) to another (the *target* or *topic*). This view of metaphor is implied by the term *metaphor* itself, which in ancient Greek meant to “carry something across” or “transfer.” This view, the *Directional projection* view, emphasizes that in metaphor, information is projected from a familiar, often concrete base domain to a less familiar or less clear target. For example, in the metaphor “Some suburbs are parasites,” one’s knowledge of the base concept, *parasites*—that they profit from but harm the host—is projected to the target concept, *suburbs*. Or consider a more vivid example, from Cardinal Wolsey’s speech on being stripped of his position (Shakespeare, Henry VIII, Act 3, Scene 2). Here, the known base domain of boys floating on bladders (the 16th-century equivalent of inner tubes) is used to portray the course of ambition from glory to defeat. The point of the metaphor is to reveal the target.

Correspondence should be sent to Phillip Wolff, Department of Psychology, 532 Kilgo Cir., Emory University, Atlanta, GA 30322. E-mail: pwolff@emory.edu

I have ventured,
 Like little wanton boys that swim on bladders,
 This many summers in a sea of glory,
 But far beyond my depth: my high-blown pride
 At length broke under me and now has left me,
 Weary and old with service, to the mercy
 Of a rude stream, that must forever hide me.

Yet on the other hand, an equally persuasive intuition is that metaphor reveals subtle commonalities that alter our understanding of the base as well as the target. These lines from Toby Litt's (1989) poem "For Borges" show how metaphor can illuminate both terms:

The tiger roaring like a fire, the fire
 roaring like a tiger, are metaphors
 by which both fire and tiger are made clear.

This *Emergent commonalities* view of metaphor is reflected in Aristotle's (trans. 1932, pp. III, x, 4ff.) observation that metaphors draw attention to nonobvious commonalities. It is also prominent in theories of language change¹ (e.g., Heine, 1997; Hopper & Traugott, 2003). Indeed, Hock and Joseph (1996) state (p. 228) "The major vehicle through which words acquire new or broader meaning is metaphor." As one example of this kind of metaphor-driven change, Zharikoff and Gentner traced the evolution of *sanctuary* from its Old English meaning of a place of worship (such as a church or temple) to its more abstract meaning of a safe place. Indeed, Heine (1997) suggests that metaphorical extension and semantic abstraction is often the source of the abstract spatial and grammatical terms in a language. Thus, an account of how metaphors are processed—and how they give rise to changes in the meaning of *both* terms—is important not only for our understanding of metaphor itself but also for our understanding of how new abstractions arise.

Theories of metaphor have mostly emphasized one side or the other of this dichotomous nature. Early theories of metaphor tended to focus on the discovery of commonalities, often nonobvious commonalities (Malgady & Johnson, 1976; Miller, 1979). For example, Murphy (1996) suggested that metaphors reflect structural parallels between two domains. Tversky (1977) suggested that metaphors are understood by finding commonalities between the base and target and that this process involves a search for features that optimize the quality of this resemblance. Miller (1979) analyzed metaphor as a kind of comparison statement with parts left out, and speculated that "...metaphor often involves favoring an extended meaning over the core meaning of a word..." (Miller 1979, p. 247). These approaches, based on extracting commonalities, naturally allow alteration in the meanings of both terms of the metaphor. However, it is not clear how they account for the directional inferences that also characterize metaphor.

On the other side, the directional view of metaphor also has strong adherents (Cienki, Cornelissen, & Clarke, 2008; Gibbs, Costa Lima, & Francozo, 2004; Glucksberg & Keysar, 1990; Lakoff & Johnson, 1999; Shen, 1989; Way, 1991). For example, in

Glucksberg and colleagues' attributive category model (Glucksberg, McGlone, & Manfredi, 1997; McGlone & Manfredi, 2001), metaphors are comprehended in terms of abstractions projected from the base to the target. For example, given the metaphor "My surgeon is a butcher," a category is derived from the base term, *butcher* (perhaps "one who cuts flesh crudely") and applied to the target, *surgeon*. In this model, processing is directional throughout. Another variety of directional theory utilized multidimensional space representations and postulated that in analogy (Rumelhart & Abrahamson, 1973a, 1973b) and metaphor (Tourangeau & Sternberg, 1981), the dimensional structure of the base domain is mapped onto that of the target, allowing for further inferences about the target. Ortony's (1979) salience imbalance model attempted to account for directionality within a commonality framework. He proposed that metaphors, like literal similarity statements, are comprehended in terms of shared features. What distinguishes metaphors is strong directionality: The shared features must be of high saliency in the base and low saliency in the target.

Perhaps the best known of the directional accounts is the embodiment account of metaphor championed by Lakoff and his colleagues (Gibbs, 2006; Lakoff & Johnson, 1999; Wilson & Gibbs, 2007). This perspective has manifested itself in two kinds of theoretical accounts. In the first account, the processes involved in metaphor understanding are inspired by bodily experience but can involve enduring conceptual structures, such as image-schemas (Lakoff & Johnson, 1980). For example, the *love as a journey* metaphor includes: "The road was rough and steep but we carried on.... If we pull together we can surmount these hard times. We're having a rocky time and I'm not sure we're going to make it" (Lakoff & Johnson, 1980; Lakoff & Turner, 1989; Turner, 1987). In the second (more recent) account, metaphor understanding operates over embodied simulations² of the terms being compared. In this approach, metaphors are processed via online simulations that draw on sensorimotor encodings stored in modality-specific areas of the brain (e.g., Barsalou, 2005). This second version is intrinsically directional from the start, because the base of the metaphor is a modality-specific sensorimotor experience.

1. Resolving the paradox

Thus, theories of metaphor have split according to whether they attempt to capture the directional projection side of metaphor or the emergent commonalities side. This is not surprising, as at first glance the two phenomena—discovering commonalities and projecting inferences—seem to require very different processes. Whereas finding commonalities is most naturally conceived of as a matching process—and is therefore seen as symmetric—inference projection is clearly asymmetric: The base and target play very different roles. Of course, one solution would be to propose that there are two kinds of metaphor and to characterize them separately. But to take this route would, in our view, miss crucial connections between the two phenomena.

We suggest that commonality discovery and inference projection are part of the same process—one of structural alignment and mapping. This view, which follows from applying

structure-mapping theory to metaphor, reveals linkages between discovering commonalities and projecting inferences. It also makes specific predictions as to the time course of metaphor processing, which we test in four experiments.

2. The structure-mapping process

According to structure-mapping, the processing of metaphor (like the processing of analogy and similarity) includes both highlighting commonalities and projecting inferences. Comprehending a figurative (or literal) comparison entails an initial set of processes that arrive at a structural alignment between the two representations, followed by the directional projection of inferences from the base to the target (Gentner & Bowdle, 2001; Gentner, Bowdle, Wolff, & Boronat, 2001; Gentner & Kurtz, 1995; Gentner & Markman, 1997; Gentner & Wolff, 1997, 2000; Markman, 1997; Markman & Gentner, 1997; Markman & Gentner, 2000; Wolff & Gentner, 2000). For example, given the metaphor “Some suburbs are parasites,” the initial symmetric alignment process yields the common system “existing in dependence on a host.” Then, directional inference processes project further ideas from the base to the target: for example, “harms its host.” Thus, in structure-mapping, initial processing is symmetrical but later processing is directional.

These processing assumptions are embodied in a simulation, the Structure-mapping Engine (SME) (Falkenhainer, Forbus, & Gentner, 1989; Forbus, Gentner, & Law, 1995). SME utilizes a three-stage local-to-global matching process to find the maximal structurally consistent alignment between two representations. The first stage is a parallel local-match stage in which all pairs of identical predicates and their corresponding arguments are placed in correspondence. For example, for the metaphor “Suburbs are parasites,” if the representations of base and target include something like *derives-from(parasite, host, food)* and *derives-from(suburb, city, utilities)*, the two *derives-from* predicates would be matched, leading to the further correspondences *suburb* → *parasite*, *host* → *city*, and *food* → *utilities*. This initial local matching stage typically results in a large number of potential correspondences. In the second phase, structural consistency is enforced; the local matches are coalesced into small, structurally consistent mapping clusters (called *kernels*). In the third stage the kernels are merged into large global interpretations, using a merge algorithm³ (Forbus & Oblinger, 1990) that begins with the maximal kernel, adds the next-largest kernel that is structurally consistent with the first, and continues until no more kernels can be added without compromising structural consistency. SME then produces a structural evaluation of the interpretation(s), using a cascade-like algorithm that favors deep interrelated systems over shallow systems, all else being equal.

Once this alignment is found, processing shifts from a role-neutral alignment process to a directional inference process. Predicates connected to the common structure in the base, but not initially present in the target, are projected as *candidate inferences* in the target. Thus, according to the structure-mapping model, directionality in metaphor comprehension arises *after* the initial stages of symmetric processing and is guided by the alignment.

There is indirect evidence in support of structure-mapping's claim that the initial comprehension processes are symmetric even for highly directional metaphors. In prior research, Wolff and Gentner (2000) selected a set of strongly directional metaphors, using the criterion that participants preferred the forward direction (e.g., "A brain is a warehouse") over the reverse direction (e.g., "A warehouse is a brain"). (The forward direction was preferred 92% of the time.) We then used the metaphor interference technique (Glucksberg, Gildea, & Bookin, 1982; Keysar, 1989) to investigate the early stages of comprehension. In this technique, participants make true-false judgments about three kinds of statements: true class-inclusion statements (e.g., "Some birds are robins"), false class-inclusion statements (e.g., "Some birds are apples"), and metaphorical statements (which of course are also literally false: e.g., "Some jobs are jails"). Glucksberg et al. found that people took longer to reject metaphors than to reject ordinary false statements, indicating that metaphor processing is initiated *before* literal processing has terminated.

Wolff and Gentner (2000) applied this metaphor interference technique to the highly directional metaphors discussed above, with one methodological innovation: We included both forward and reversed versions of each metaphor (across different participants). As in Glucksberg et al.'s original study, we found that metaphors ("Some brains are warehouses") took longer to reject than ordinary false statements. Importantly, however, reversed metaphors took just as long as forward metaphors; both showed the same interference effect on participants' true-false judgments. These findings are consistent with the claim of an early symmetrical stage in comprehension. However, the task of the participants was to make a true-false judgment, leaving open the question of whether such an early symmetric stage is involved in more natural comprehension of metaphors.

In the current research, we used a more straightforward comprehension task to investigate the time course of metaphor processing. Participants read sentences and indicated whether they were comprehensible by pressing one of two keys (comprehensible or not comprehensible). The key materials were forward and reversed versions of the strongly directional metaphors from the Wolff and Gentner (2000) study. We used a mixed-deadline procedure: Participants saw forward and reversed metaphors and rated their comprehensibility at either a short or a long deadline. The idea is that if the initial comprehension process is one of symmetric alignment, then the forward and reversed metaphors should be indistinguishable at early stages (because both involve the same terms, just in different order). However, because later processing stages are held to be directional, the forward versions should be judged as more comprehensible than the reversed versions at later deadlines.

Experiments 1–3 use this mixed-deadline technique to test whether metaphor processing begins symmetrical and then becomes directional. We also tested whether the later stage of processing builds upon the earlier stage, as structure-mapping predicts. In Experiments 4a and 4b, we addressed a potential alternative interpretation for such a pattern by asking whether early symmetric processing could be attributed to preexisting associations in memory.

In Experiment 1, statements were displayed for either 1,200 or 1,800 ms. The earlier deadline of 1,200 ms reflects the RT typically found in studies using the metaphor interference technique (Glucksberg et al., 1982; Wolff & Gentner, 2000). Thus, by roughly 1,200 ms, metaphor processing is clearly underway. The later deadline of 1,800 ms, reflects

the shortest amount of time people tend to spend when asked to fully understand a metaphor (Wolff & Gentner, 2000). After people saw the metaphor for either 1,200 or 1,800 ms, a time-out signal sounded 400 ms after a statement disappeared. Participants were asked to decide whether each statement was comprehensible or incomprehensible; they were told that they would have to do this quickly. Each participant saw half the metaphors in forward and half in reversed direction. In addition to the key forward and reversed metaphors, participants saw literal class-inclusions (e.g., “Some birds are robins”), scrambled statements—both scrambled class-inclusions (e.g., “Some pianos are trees”) and scrambled metaphors (created by swapping terms between different metaphors: e.g., “Some rumors are jails”). These were important to the logic of the study and also served to keep participants from developing a specific metaphoric set. The key question was how the reversed metaphors would compare to forward metaphors in comprehensibility at the two deadlines.

The structure-mapping model makes two predictions. First, like the directional theories, it predicts an effect of direction at later deadlines: Forward metaphors—for example, “Some jobs are jails” and “Some rumors are viruses”—should be rated as more comprehensible than reversed metaphors—for example, “Some jails are jobs.” Second, unlike these theories, it predicts that forward and reversed metaphors should be indistinguishable at early stages of comprehension. The comprehensibility advantage of forward over reversed metaphors should increase with longer deadlines.

However, the second of these key predictions—that forward and reversed metaphors will be alike in their comprehensibility at the early stage—is only interesting if it can be shown that meaningful processing is occurring at this stage. If the lack of difference between forward and reversed metaphors is simply due to noncomprehension, it would prove nothing. Therefore, in addition to the forward and reversed metaphors, we included two other kinds of materials to establish upper and lower comprehensibility baselines: literal class-inclusions (which should be judged comprehensible even at the early deadline) and scrambled statements (which should be incomprehensible throughout). For this method to be valid, we need evidence that meaningful processing is occurring at the early deadline. Thus, at the early deadline, people should find literal class-inclusions—for example, “Some birds are robins”—comprehensible, and they should find scrambled statements—for example, “Some pianos are trees” and “Some rumors are jails”—incomprehensible. Further, people should distinguish metaphors (whether forward or reversed) from scrambled metaphors even at the early stage. If all these are true, then we can assume that comprehension processes are underway even at the early stage. In this case, a failure to distinguish forward from reversed metaphors in the first stage will be informative; it will imply early meaningful processing that is symmetric.

Thus, the logic is as follows: (a) if comprehension processes are underway by the early deadline, then comprehensibility should be high for literal class-inclusions and low for scrambled metaphors; specifically, the comprehensibility of metaphors should be higher than that of scrambled metaphors, at both the early and later deadlines. If this pattern holds, then the stage is set to test the key predictions of structure-mapping; (b) because the initial stage is a symmetric alignment process, forward and reversed metaphors should not differ in comprehensibility at the early stage; and (c) because directional projection occurs at later

stages of processing, the difference in comprehensibility between forward and reversed metaphors should be greater at the later deadline than at the early deadline. If these three results are found, it will support the structure-mapping claim that there is an initial processing stage for metaphors that is role-neutral and that directional processes occur later in processing.

Of course, to test this predicted time course, it is crucial to use metaphors that are clearly directional. As noted above, we used the 32 metaphors from Wolff and Gentner's (2000) Experiment 2. The directionality of these metaphors was established in a ratings task in which 16 Northwestern University undergraduates saw forward and reversed versions of each metaphor simultaneously on a computer screen and chose which version they preferred. Each participant saw all 32 metaphors. Presentation order was randomized across participants and the vertical position of the two sentences on the screen (one above the other) was counterbalanced across four between-subjects groups. The forward versions of the metaphors were chosen 92% of the time (with a range across items from 69% to 100%). The literal statements were also as in Wolff and Gentner's studies.

As a secondary question, we asked how comprehensibility might be influenced by the relational similarity of the metaphors. By *relational similarity* we mean the degree to which the two concepts share relational structure. In a high-similarity⁴ metaphor, such as "Some soldiers are pawns," the target readily shares the base's relational structure (e.g., *they are subordinated to powerful forces*). In a low-similarity metaphor, such as "Some senators are pawns," the target is not normally associated with the idea but can accept it as a metaphoric inference. Prior research has shown that comprehensibility—as measured by metaphorical aptness—correlates positively with relational similarity (Gentner, 1988; Gentner & Clement, 1988; see also Veale, 2003). We would expect, then, that comprehensibility ratings should be higher for high-relational similarity metaphors than low-relational similarity metaphors. To the extent that relational similarity facilitates the initial process of alignment, this similarity effect might be observed at the earliest deadline. However, prior research has shown that the effects of relational similarity tend to appear relatively late in processing (Goldstone, 1994; Love, Rouder, & Wisniewski, 1999); hence, the influence of relational similarity on comprehensibility judgments may not be observed until later in processing.

The relative similarity of the metaphors was established in a ratings task in which 24 Northwestern undergraduates, using a 1 to 7 scale, rated relational similarity, explained as follows: "Things are relationally similar when they participate in the same relations. For example, a cigarette and a time bomb are relationally similar because they both can cause harm after a period of apparent harmlessness..." The high-similarity metaphors had a mean rating of 3.57, and the low-similarity metaphors, 2.34. For details about the characteristics of the metaphors, see Wolff and Gentner (2000).

In Experiment 1, participants made comprehensibility judgments for the metaphors, scrambled and literal sentences under two response deadlines: 1,200 and 1,800 ms. The early deadline was chosen to fit within the 1,100–1,300 ms range found for metaphor interference effects; the later deadline was chosen as a point at which at least some directional processing should have occurred, given that normal comprehension times for metaphors range from 2,000 to 4,000 ms (see Bowdle & Gentner, 2005; Wolff & Gentner, 2000).

To recapitulate, if early processing is symmetric and later processing is directional, we should see (a) an early stage at which forward and reversed metaphors do not differ from each other in comprehensibility but are both significantly more comprehensible than anomalies; (b) followed by a later stage at which forward metaphors are more comprehensible than reversed metaphors (an interaction between direction and deadline). Finally, we expect literal class-inclusions to be highly comprehensible and anomalies to be incomprehensible throughout.

3. Experiment 1

3.1. Method

3.1.1. Participants

The subjects were 32 Northwestern University undergraduates who participated for course credit.

3.1.2. Materials

There were 32 metaphors, constructed from 16 metaphor bases that were each combined with a high- and a low-similarity target. Each of these could appear in forward or reversed order (see Table 1). Four 72-item test lists were constructed, each containing 16 metaphors (four forward high similarity, four forward low similarity, four reversed high similarity, four reversed low similarity), 16 scrambled metaphors, 16 literal class-inclusions, 16 scrambled class-inclusions, and 8 warm-up metaphor (filler) items. The literal class-inclusions are listed in Appendix A. We constructed the scrambled class-inclusion statements by recombining the subject and predicate nouns of the literally true statements. The scrambled metaphors were constructed by recombining the subject and predicate nouns of the metaphors. In both cases, the subject and predicate nouns were re-paired to produce statements that were not readily interpretable. Each of the four 72-item test lists included a unique set of 16 scrambled metaphors that was based on the particular set of 16 metaphors included in that list. Thus, each word in each metaphor appeared in one of the scrambled metaphors with the following constraint: Words appearing as targets in the metaphors appeared as bases in the scrambled metaphor and vice versa for words appearing as bases in the metaphors. The 16 scrambled class-inclusions were derived from the 16 literal class-inclusions in the same way. In addition, a 64-item practice list was constructed containing the same number and kinds of items as in the test: 16 metaphors, 16 scrambled metaphors, 16 literal class-inclusions, and 16 scrambled class-inclusions.

3.1.3. Design

A within-subject design was used: Each participant judged each sentence type at both deadlines within a single session. The design for the metaphors was Direction (forward vs. reversed) \times Similarity level (high vs. low) \times Deadline (1,200 vs. 1,800 ms), making eight kinds of trials. The assignment of particular metaphor bases to these types was

Table 1
Metaphors used in Experiments 1, 2, and 3^a

High Similarity	Low Similarity
Some arguments are wars	Some conversations are wars
Some lawyers are sponges	Some teachers are sponges
Some lies are boomerangs	Some statements are boomerangs
Some hippopotamuses are blimps	Some lions are blimps
Some horoscopes are maps	Some books are maps
Some saunas are ovens	Some rooms are ovens
Some ferries are bridges	Some boats are bridges
Some dragsters are rockets	Some mopeds are rockets
Some exams are filters	Some applications are filters
Some suburbs are parasites	Some towns are parasites
Some giraffes are skyscrapers	Some busboys are skyscrapers
Some auditions are doors	Some plays are doors
Some babies are angels	Some children are angels
Some librarians are mice	Some receptionists are mice
Some stagecoaches are dinosaurs	Some trains are dinosaurs
Some salesmen are bulldozers	Some merchants are bulldozers

Notes. ^aThe metaphors were the same as those used in Experiments 2 and 3 of Wolff and Gentner (2000). The literal class-inclusion statements were a randomly drawn subset of the literal class-inclusion statements used in Experiment 2 of that paper.

counterbalanced across eight between-subject groups, so that no subject saw the same metaphor base more than once.

3.1.4. Procedure

Participants were run on Windows-based computers separated by sound-attenuating carrels. Participants were told that they would see statements like “Some birds are robins,” “Some antiques are fossils,” or “Some zebras are spoons,” and that they should say whether these statements were comprehensible or incomprehensible by pressing the left or right arrow keys, respectively. Participants were told that they would sometimes be shown statements that were comprehensible but also metaphorical such as “Some antiques are fossils.” Participants were instructed to classify such statements as comprehensible. Because the deadline task, by its nature, pushes participants to respond before comprehension is complete, it was expected to be difficult. Therefore, a fairly extensive practice task was given. Participants were given 20 practice trials pressing the left and right arrow keys in response to the words “comprehensible” and “incomprehensible.” They were told to make their responses within 400 ms after the word disappeared. Errors in both choice and timing were indicated. The participants then received 64 practice trials on statements like those in the test phase.

Presentation of the sentences in both the practice and test sessions occurred as follows. First, a line of pound signs appeared. The number of pound signs matched the number of letters in the upcoming sentence. After 300 ms, the line was replaced with a sentence which remained on the screen for either 1,200 or 1,800 ms. Participants were instructed to make a comprehensibility judgment within 400 ms after the sentence disappeared. Responses

occurring either before the metaphor disappeared or after the 400 ms deadline were followed by a message saying that the response was too fast or too slow, respectively. During the practice session, but not the test session, subjects were also informed of incorrect comprehensibility judgments. In particular, participants were informed that their comprehensibility judgment was incorrect if they classified a forward metaphor as “incomprehensible,” or a reversed metaphor as “comprehensible.” Millisecond accuracy was achieved using Pascal procedures provided in Brysbaert, Bovens, and d’Ydewalle (1989).

3.2. Results

We removed from the analyses any comprehensibility judgments that were made either before (1%) or more than 400 ms after (18%) the stimulus item disappeared. The resulting mean proportions and standard errors of items judged comprehensible for literal class-inclusion, forward metaphors, reversed metaphors, and scrambled metaphors at 1,200 and 1,800 ms deadlines are displayed in Fig. 1.

To preview, the results provided partial support for our predictions. As expected, forward metaphors ($M = 0.57, SD = 0.227$) were rated as more comprehensible overall than reversed metaphors ($M = 0.34, SD = 0.211$). More important, however, there was a significant interaction between direction and deadline: as predicted, the advantage of forward over reversed metaphors increased with longer deadlines. However, we did not find support for our stronger prediction, that forward and reversed metaphors would be indistinguishable at the early deadline.

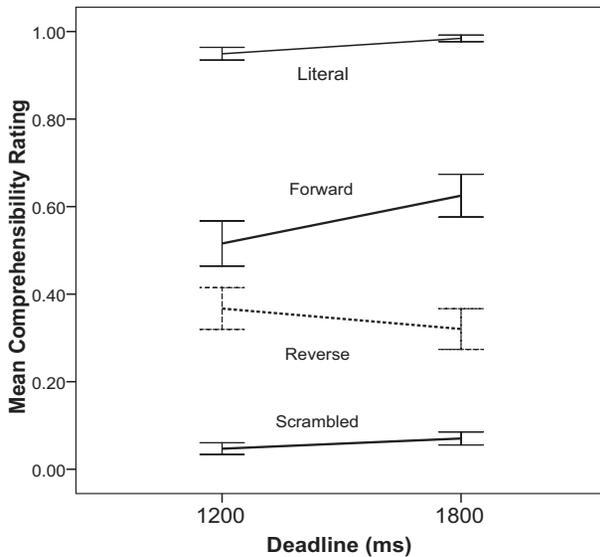


Fig. 1. Results of Experiment 1: Proportion of statements judged comprehensible, with standard errors of the mean, for literal class-inclusions, forward metaphors, reversed metaphors, and scrambled metaphors at 1,200 and 1,800 ms deadlines.

An analyses of variance with Direction, Similarity, and Deadline as factors bore out a main effect of direction, $F_s(1, 31) = 19.01$, $p < .001$, $\eta^2 = .368$, $F_i(1, 15) = 23.58$, $p < .001$, $\eta^2 = .50$, as well as the predicted interaction between direction and deadline, $F_s(1, 31) = 4.2$, $p < .05$, $\eta^2 = .118$, $F_i(1, 15) = 6.0$, $p < .05$, $\eta^2 = .247$. As Fig. 1 shows, the difference in comprehensibility between forward metaphors and reversed metaphors ($M = 0.52$, and $M = 0.37$, respectively, at the early deadline) increased with longer processing time ($M = 0.63$, and $M = 0.32$, for forward and reversed metaphors, respectively, at the later deadline), consistent with the claim of a symmetric-to-directional shift. However, we did not find evidence for early indistinguishability: Forward metaphors were rated more comprehensible than reversed metaphors at both early and late deadlines, $F_s(1, 31) = 5.7$, $p < .05$, $\eta^2 = .155$, $F_i(1, 15) = 5.6$, $p < .05$, $\eta^2 = .157$, for early; $F_s(1, 31) = 19.5$, $\eta^2 = .386$, $p < .001$, $F_i(1, 15) = 37.6$, $p < .001$, $\eta^2 = .654$ for late. Nonetheless, the significant interaction between direction and deadline is encouraging, and it raises the possibility that an earlier deadline might reveal a more symmetric result.

3.2.1. Comprehensibility baselines

As noted above, the results for metaphors are uninterpretable unless it can be demonstrated that meaningful processing is occurring at both the early and late deadlines. Fortunately, the results show a strong separation between the literal class-inclusion statements and the scrambled statements throughout processing. As shown in Fig. 1, literal class-inclusion statements were rated as highly comprehensible at both 1,200 ms ($M = 0.95$) and 1,800 ms ($M = 0.98$). In contrast, the comprehensibility for scrambled metaphors was low for both deadlines ($M = 0.05$ and $M = 0.07$ for early and late, respectively). Comprehensibility for scrambled literal class-inclusions was also low for both deadlines ($M = 0.07$, $SD = 0.10$ and $M = 0.10$, $SD = 0.098$ for early and late, respectively). This strong separation between literal class-inclusions and scrambled statements confirms that comprehension processes were engaged even at the earlier deadline. Importantly, even at the early deadline, both forward metaphors and reversed metaphors were judged more comprehensible than scrambled metaphors: for forward, $F_s(1, 31) = 59.33$, $p < .001$, $\eta^2 = .66$, $F_i(1, 30) = 48.9$, $p < .001$, $\eta^2 = .62$, and reversed metaphors, $F_s(1, 31) = 26.3$, $p < .001$, $\eta^2 = .46$, $F_i(1, 30) = 29.7$, $p < .001$, $\eta^2 = .50$. Note that the scrambled metaphors were based on the same words used in the metaphors, so the difference between the forward and reversed metaphors and the scrambled metaphors cannot be due to the particular words. Rather, the difference must reflect the joint processing of the two terms. Overall, the results indicate that (a) comprehension processes were engaged for *both* forward and reversed metaphors by the 1,200 ms deadline; and (b) the comprehensibility advantage for forward metaphors over reversed metaphors increased with additional processing time.

3.2.2. Further results

Not surprisingly, even the forward metaphors ($M = 0.625$, $SD = 0.277$) were rated as less comprehensible than literal class-inclusions ($M = 0.98$, $SD = 0.043$) even at the late deadline. This pattern, which held across all three comprehension studies, is consistent with prior

results suggesting that the metaphors are less easily comprehended than class-inclusions (McElree & Nordlie, 1999). Finally, as predicted, high-similarity metaphors ($M = 0.50$, $SD = 0.166$) were rated as more comprehensible than low-similarity metaphors ($M = 0.42$, $SD = 0.207$), significant across subjects, $F_s(1, 31) = 4.76$, $p < .05$, $\eta^2 = .133$, although not across items, $F_i(1, 15) = 2.19$, $p = .16$. No other main effects or interactions were significant.

3.2.3. Time-outs

Participants timed out—that is, failed to respond within 400 ms of the stimulus item's disappearance—on 18% of the trials. This level of time-outs is not surprising given that the deadline task, by its nature, pushes participants to respond at a very early stage of comprehension. Importantly, however, the time-out rates were not greater for reversed ($M = 0.24$, $SD = 0.194$) than for forward metaphors ($M = 0.22$, $SD = 0.196$), $F_s(1, 31) = .683$; $F_i(1, 15) = 2.40$, nor did reversed and forward metaphors differ more at the early deadline ($M = 0.31$, $SD = 0.227$ and $M = 0.31$, $SD = 0.291$, respectively) than at the later deadline ($M = 0.17$, $SD = 0.273$ and $M = 0.13$, $SD = 0.191$, respectively), $F_s(1, 31) = 0.497$; $F_i(1, 15) = 0.069$. Thus, the interaction between direction and deadline cannot be attributed to a larger number of time-outs in the reversed metaphor condition.

3.3. Discussion

The results of Experiment 1 show that the comprehensibility advantage for forward over reversed metaphors increases over time, consistent with the claim that metaphor comprehension involves a shift from more symmetric to more directional. However, the results do not provide direct evidence for an initial symmetric stage, because forward metaphors were rated as more comprehensible than reversed metaphors at *both* deadlines. If, as predicted, there is an early stage of symmetric metaphor processing in metaphor comprehension, then there should be a point in time when forward and reversed metaphors are treated equivalently, implying symmetric processing. (Of course, forward and reversed metaphors must also differ from the upper and lower comprehensibility baselines; that is, metaphors must be more comprehensible than scrambled metaphors and less comprehensible than literal statements, etc., indicating that comprehension is underway.)

To test for such a stage, in Experiment 2 we moved the processing window much earlier, setting the earlier deadline at 600 ms. This deadline was chosen because the results of Experiment 1 suggested that the forward and reversed metaphors might converge at roughly this point in time. The later deadline in Experiment 2 was correspondingly shortened to 1,600 ms (from 1,800 ms in Experiment 1). As in Experiment 1, the predictions are as follows. First, there should be an overall effect of direction, with forward metaphors rated more comprehensible than reversed metaphors. Second, there should be an interaction of direction and deadline such that the advantage of forward over reversed metaphors increases over time. The key question is what happens at the early deadline. If we find that at the early deadline, forward and reversed metaphors do not differ from each other (and that both are more comprehensible than scrambled metaphors and less comprehensible than literal

class-inclusions), this will support the structure-mapping claim that the initial processing of metaphors is symmetric.

4. Experiment 2

4.1. Method

4.1.1. Materials and procedure

The materials and procedures were the same as those used in Experiment 1 except that the deadlines here were 600 ms and 1,600 ms instead of 1,200 and 1,800 ms.

4.1.2. Participants

Forty-eight Northwestern University undergraduates participated for course credit.

4.2. Results

Comprehensibility judgments for all stimulus item types occurring either before (2%) or more than 400 ms after (26%) a stimulus item disappeared were removed from the analyses. The resulting mean comprehensibility proportions and standard errors are shown in Fig. 2.

The results replicate and extend the results of Experiment 1. As before, forward metaphors ($M = 0.47$, $SD = 0.264$) were rated as more comprehensible than reversed metaphors overall ($M = 0.36$, $SD = 0.240$), and the difference in comprehensibility between forward and reversed metaphors was greater at the later deadline than at the earlier deadline. Of most importance, forward and reversed metaphors differed from scrambled metaphors and literal class-inclusions at both deadlines, but they differed from each other only at the later deadline. Thus, the results indicate that processing at the earlier deadline was role-neutral, but became role-sensitive by the later deadline, consistent with the predictions of structure-mapping.

The observations were confirmed by analyses of variance with Direction, Similarity, and Deadline as factors. This analysis indicated a main effect of direction, significant across subjects and marginally significant across items, $F_s(1, 47) = 9.4$, $p < .01$, $\eta^2 = .166$, $F_i(1, 15) = 3.7$, $p = .072$, $\eta^2 = .20$. As predicted, there was also a significant Direction \times Deadline interaction, $F_s(1, 47) = 5.2$, $p < .05$, $\eta^2 = .10$, $F_i(1, 15) = 4.7$, $p < .05$, $\eta^2 = .24$. Planned comparisons confirmed the key predictions that forward and reversed metaphors differed at the 1,600 ms deadline ($M = 0.52$ and $M = 0.35$, respectively), $F_s(1, 47) = 16.3$, $p < .001$, $\eta^2 = .23$, $F_i(1, 15) = 14.9$, $p < .001$, $\eta^2 = .50$, but not at the 600 ms deadline ($M = 0.39$ and $M = 0.36$, respectively), $F_s(1, 47) = 0.021$, ns , $F_i(1, 15) = 0.153$, ns .

With respect to the comprehensibility baselines, planned comparisons confirmed that at the 600 ms deadline, both forward metaphors ($M = 0.39$) and reversed metaphors ($M = 0.36$) were more comprehensible than scrambled metaphors ($M = 0.17$, $SD = 0.248$),

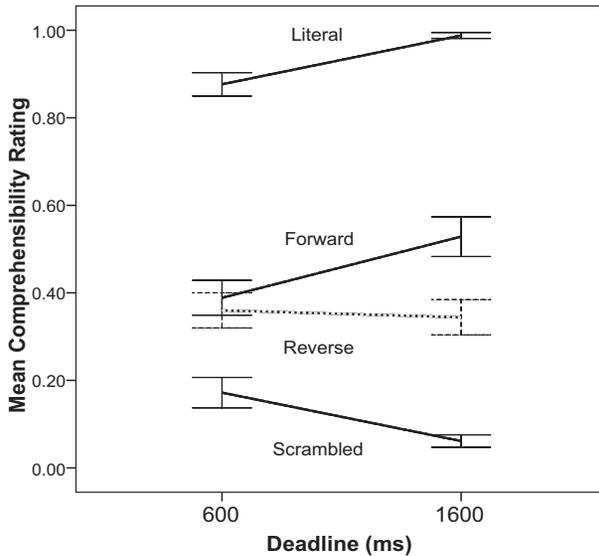


Fig. 2. Results of Experiment 2: Proportion of statements judged comprehensible, with standard errors of the mean, for literal class-inclusions, forward metaphors, reversed metaphors, and scrambled at 600 and 1,600 ms deadlines.

$F_s(1, 47) = 12.7, p < .01, \eta^2 = .30, F_i(1, 30) = 4.3, p < .05, \eta^2 = .16$, for forward metaphors, and $F_s(1, 47) = 9.6, p < .01, \eta^2 = .29, F_i(1, 30) = 7.5, p < .05, \eta^2 = .25$, for reversed. Also as required, forward and reversed metaphors were less comprehensible than literal class-inclusions at the 600 ms deadline ($M = 0.16, SD = 0.214$), $F_s(1, 47) = 30.17, p < .001, \eta^2 = .76, F_i(1, 30) = 48.7, p < .001, \eta^2 = .62$, for forward metaphors, and $F_s(1, 47) = 20.38, p < .01, \eta^2 = .72, F_i(1, 30) = 55.7, p < .01, \eta^2 = .65$, for reversed. The comprehensibility of literal class-inclusion statements was high at both deadlines ($M = 0.88$ and $M = 0.99$), for early and late deadlines, respectively (see Fig. 2). The comprehensibility of scrambled statements was low at both deadlines (for scrambled metaphors, $M = 0.17$ and $M = 0.06$ for early and late deadlines, respectively; for scrambled class-inclusions, $M = 0.14, SD = 0.097$ and $M = 0.08, SD = 0.11$, for early and late deadlines). Thus, we have the required strong separation between literal class-inclusions and scrambled statements at both early and late deadlines, with metaphors (both forward and reversed) intermediate in comprehensibility—exactly what is expected if participants were engaged in meaningful processing at the earliest deadline. The fact that the early processing was symmetric, with no difference between forward and reversed metaphors, is evidence for an initial alignment process. The fact that the later deadline shows an advantage for forward metaphors is consistent with the claim that directional processing follows alignment.

The effect of relational similarity was not significant across either subjects, $F_s(1, 47) = 0.807, p = .373$, or items, $F_i(1, 15) = 0.969, p = .340$. We suspect that the 1,600 ms deadline did not permit an effect of relational similarity on alignment to emerge in this study. No other main effects or interactions were significant.

4.2.1. Time-outs

Time-out rates were no greater for reversed ($M = 0.32$, $SD = 0.194$) than for forward metaphors ($M = 0.29$, $SD = 0.191$), $F_s(1, 47) = 0.65$; $F_i(1, 15) = 0.490$, nor did the time-out rates for reversed and forward metaphors differ more at the early deadline ($M = 0.42$, $SD = 0.298$ and $M = 0.41$, $SD = 0.303$, respectively) than at the later deadline ($M = 0.21$, $SD = 0.213$ and $M = 0.18$, $SD = 0.199$, respectively), $F_s(1, 47) = 0.249$; $F_i(1, 15) = 0.125$. Thus, the interaction between direction and deadline cannot be attributed to a larger number of time-outs in the reversed metaphor condition. Further, the lack of a significant difference between forward and reversed metaphors at the early deadline cannot be explained as due to a lack of power associated with relatively high time-out rates, since forward and reversed metaphors did, in fact, differ from scrambled metaphors (and the means were in any case quite similar: 0.39 and 0.36 for forward and reversed metaphors).

4.3. Discussion

The results of this study extend and strengthen the results of Experiment 1 in supporting the predictions of structure-mapping: Forward and reversed metaphors did not differ in comprehensibility at the early stage, even though the forward metaphors were clearly more comprehensible at the later stage. Both forward and reversed metaphors differed from scrambled metaphors at the earliest deadline. The overall pattern indicates that meaningful processing was underway but was symmetric at the outset. These results provide the first direct evidence for a shift from symmetric to directional processing.

A further prediction concerns the relation between early and late comprehension ratings. According to structure-mapping theory, the later directional processing builds on the earlier structural alignment. For example, inferences are projected when there are additional predicates from the base that are connected to the common structure. One way to examine whether the later stage was based on the results from the earlier stage is to ask whether forward and reversed metaphors differ in the relation between early and late comprehensibility ratings. Specifically, for forward metaphors (such as “Some brains are warehouses”), later processing can build on the earlier alignment. But for reversed metaphors (such as “Some warehouses are brains”), the early alignment does not support later base-to-target inferences. (Indeed, over time such reversed metaphors will often come to seem ill-formed.) Consistent with this reasoning, the results for forward metaphors show a positive correlation between comprehensibility ratings at the 600 ms and 1,600 ms deadlines, $r(15) = .714$, $p = .002$; no such correlation holds for reversed metaphors, $r(15) = .257$, $p = .35$. These correlations suggest that, as predicted by structure-mapping, for the forward, but not reversed metaphors, the process of comprehension was cumulative.

A possible concern is that our conclusion that initial processing is role-neutral is based on a null effect. If this were the only prediction, the case for role-neutral processing would not be strong. However, in addition to this one null prediction, there were 11 other non-null effects for the metaphors, as well as two for the literal statements, all of which were borne out: Forward and reversed metaphors both differed from scrambled metaphors, at both

deadlines; forward and reversed metaphors both differed from literals, at both deadlines; forward and reversed metaphors differed from each other at the later deadline and literals and scrambled literals differed from each other at both deadlines. The occurrence of exactly one predicted null effect in the context of eleven predicted non-null effects is quite striking. Nevertheless, we could be more confident of our conclusion if this overall pattern of results were replicated. In the next experiment we repeated the procedures used in Experiment 2, with one change: We moved the early deadline back from 600 to 500 ms to more fully explore the nature of processing at very early stages in metaphor comprehension. The longer deadline was kept at 1,600 ms.

5. Experiment 3

5.1. Method

5.1.1. Participants

The subjects were 32 Northwestern University undergraduates who participated for course credit.

5.1.2. Materials, design and procedure

These were the same as in Experiments 1 and 2 except for the deadlines, which were set to 500 and 1,600 ms.

5.2. Results

Responses occurring either before (1%) or more than 400 ms after (32%) a stimulus item disappeared were removed from the analyses. The resulting mean comprehensibility proportions and associated standard errors at 500 and 1,600 ms deadlines are shown in Fig. 3.

The pattern of results indicates a shift from symmetric to directional processing, replicating the results of Experiment 2. As before, forward metaphors ($M = 0.46$, $SD = 0.208$) were rated as more comprehensible than reversed metaphors overall ($M = 0.32$, $SD = 0.242$), and the difference in comprehensibility between forward and reversed metaphors was greater at the later deadline than at the earlier deadline. Further, forward and reversed metaphors differed from scrambled metaphors at both deadlines but differed from each other only at the later deadline. Thus, the results indicate that processing at the earlier deadline was role-neutral but became role-sensitive by the later deadline, consistent with the predictions of structure-mapping.

An analysis of variance with Direction, Similarity, and Deadline as factors showed a main effect of direction across both subjects and items, $F_s(1, 31) = 6.48$, $p < .05$, $\eta^2 = .173$, $F_i(1, 15) = 6.68$, $p < .05$, $\eta^2 = .31$. As predicted, there was also a significant Direction \times Deadline interaction, $F_s(1, 31) = 13.23$, $p < .01$, $\eta^2 = .30$, $F_i(1, 15) = 11.76$, $p < .01$, $\eta^2 = .44$. Planned comparisons confirmed the key prediction: Forward and reversed metaphors ($M = 0.55$ and $M = 0.29$, respectively) differed at the 1,600 ms deadline,

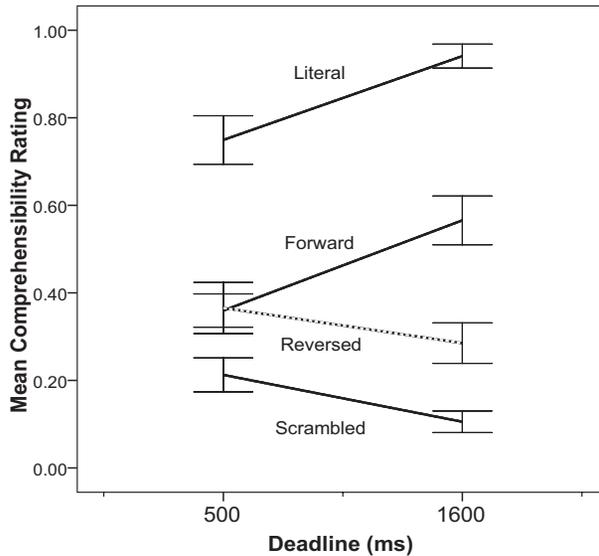


Fig. 3. Results of Experiment 3: Proportion of statements judged comprehensible, with standard errors of the mean, for literal class-inclusions, forward metaphors, reversed metaphors, and scrambled at 500 and 1,600 ms deadlines.

$F_s(1, 31) = 23.76, p < .001, \eta^2 = .43, F_i(1, 15) = 21.9, p < .001, \eta^2 = .59$, but not at the 500 ms deadline ($M = 0.36$ and $M = 0.37$ for forward and reversed metaphors, respectively), $F_s(1, 31) = 0.003, ns, F_i(1, 15) = 0.289, ns$.

The results of the manipulation check were as required. Planned comparisons confirmed that at the 500 ms deadline forward metaphors ($M = 0.36$) were more comprehensible than scrambled metaphors ($M = 0.21, SD = 0.219$) across both subjects and items, $F_s(1, 31) = 7.65, p < .01, \eta^2 = .20, F_i(1, 30) = 4.34, p < .05, \eta^2 = .13$. Reversed metaphors ($M = 0.37$) also differed from scrambled metaphors significantly across subjects, $F_s(1, 31) = 5.41, p < .05, \eta^2 = .15$, and marginally significantly across items, $F_i(1, 30) = 3.48, p = .072, \eta^2 = .1$. In sum, the results bear out the predictions of structure-mapping. As shown in Fig. 3, the comprehensibility of literal class-inclusion statements was high at both 500 ms ($M = 0.75$) and 1,600 ms ($M = 0.94$). Also as expected, comprehensibility for scrambled metaphors was low for both deadlines, ($M = 0.21$ and $M = 0.11$ for early and late, respectively). The comprehensibility ratings for scrambled literal class-inclusions mirrored that of scrambled metaphors for both the early ($M = 0.20, SD = 0.208$) and late deadlines ($M = 0.15, SD = 0.146$). The strong separation between literal class-inclusions and scrambled class-inclusions and between metaphors and scrambled metaphors even at the early deadline confirms that comprehension processes had begun by the time of the early deadline. We suggest that this reflects the initial alignment process.

The effect of similarity was not significant across either subjects or items. However, the interaction between similarity and deadline was significant across subjects, $F_s(1, 31) = 20.25, p < .001, \eta^2 = .40$, though only marginally across items, $F_i(1, 15) = 3.34,$

$p = .088$, $\eta^2 = .18$. Inspection of this interaction indicated that high- and low-similarity metaphors ($M = 0.33$, $SD = 0.222$ and $M = 0.40$, $SD = 0.268$, respectively) did not differ at the earlier deadline but did differ at the later deadline ($M[\text{high}] = .53$, $SD = 0.274$; $M[\text{low}] = .31$, $SD = 0.347$), $F_s(1, 31) = 10.95$, $p < .05$, $\eta^2 = .26$, $F_i(1, 15) = 5.90$, $p = .03$, $\eta^2 = .28$. This delayed effect is consistent with the possibility that relational similarity may have its effects relatively late in processing. No other main effects or interactions were significant.

5.2.1. Time-outs

Time-out rates—that is, failures to respond within 400 ms of the stimulus item's disappearance—were no greater for reversed ($M = 0.34$, $SD = 0.231$) than for forward metaphors ($M = 0.34$, $SD = 0.182$), nor did reversed and forward metaphors differ more at the early deadline ($M = 0.41$, $SD = 0.302$ and $M = 0.45$, $SD = 0.322$, respectively) than at the later deadline ($M = 0.28$, $SD = 0.227$ and $M = 0.24$, $SD = 0.196$, respectively).

5.2.2. Cumulative effects

As in Experiment 2, we found that for forward metaphors there was a positive correlation between comprehensibility ratings at the 500 ms deadline and those at the 1,600 ms deadline, $r(15) = .516$, $p = .041$; no such correlation holds for reversed metaphors, $r(15) = .433$, $p = .094$. These correlations suggest that for forward, but not reversed, metaphors the process of comprehension was cumulative. This is consistent with the claim that, as predicted by the structure-mapping process model, the later directional processes build on the earlier alignment. For the forward metaphors, once the alignment is found, further inferences can readily be made from base to target. But for the reversed metaphors, the best alignment is inconsistent with the given direction of the metaphors—leading participants to decrease their ratings of comprehensibility as time proceeds.

5.3. Discussion

The results of the three studies support the claim that metaphor comprehension entails early alignment and later directional processing. All three studies showed a clear forward advantage at the later deadline. In the two studies with early initial deadlines—Experiments 2 (600 ms) and 3 (500 ms)—there was no difference in comprehensibility at the early deadline. In Experiment 1, with a 1,200 ms early deadline, there was a forward advantage at both deadlines, but in that study (as in the other two) there was a significant interaction of direction and deadline, indicating a greater forward advantage later in processing. In all three studies, forward metaphors were considered more comprehensible than reversed metaphors at the longer deadline, consistent with the many findings showing directionality in untimed metaphor processing. Finally, in all three studies there was a clear separation even at the early deadline between forward and reversed metaphors and both scrambled metaphors and literal statements.

Thus, the evidence appears consistent with the idea that metaphors are processed in a series of stages with a first stage of symmetric alignment, as entailed by structure-mapping

process model. However, before drawing strong conclusions, we must consider an alternative account. Perhaps the early sense of relatedness between the terms in the metaphors stemmed not from an ongoing alignment process but from some other kind of semantic connection. Preexisting semantic associations between the target and base terms could lead to rapid associative priming (Fischler & Goodman, 1978; Shelton & Martin, 1992) that could account for the early sense of active processing. In other words, participants may have experienced an early feeling of semantic resonance between the base and target words, which led them to respond “yes” as to the comprehensibility of the sentence.

The possibility that associative connections could intrude on judgments of metaphoric comprehensibility is brought home by findings of Wisniewski and Bassok (1999), who showed that preexisting semantic connections can influence people’s similarity judgments. For example, people judged pairs with preexisting thematic associations (e.g., milk–cow) to be more similar than equally dissimilar pairs without strong preexisting associations (e.g., milk–horse). More generally, Wisniewski and Bassok argued that the actual processing that people engage in during a task is often a mixture of the processes they are asked to engage in (e.g., similarity judgment) and the processes most strongly invited by the stimuli (e.g., associative processing for pairs with strong semantic connections) (see also Wisniewski, 1997).

What this means for the present work is that before we can conclude that the early mutual activation seen for the metaphoric pairs stems from alignment processes, we must test whether the pairs have semantic connections that could have led to this result. We assessed this possibility in Experiment 4, by using latent semantic analysis (LSA) (Experiment 4a) and with a semantic priming study (Experiment 4b).

6. Experiment 4a

6.1. Testing for preexisting associations using LSA

Frequent co-occurrence of terms in speech and writing is both a symptom and a cause of associations between their semantic content. One way to test for such associations is by using LSA (Landauer & Dumais, 1997), a mathematical method for inducing statistical relationships between words in a large body of text on the basis of their contextual co-occurrence. The important point for our purposes is that co-occurrence measures of word-word association derived from LSA have been found to correlate with word-word associative priming effects (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). These findings bear out LSA’s sensitivity to existing word associations.

Therefore, we used LSA to address the question of whether the early symmetric activation for metaphoric pairs could be explained in terms of associative priming. To do this we submitted the pairs of terms from the metaphors, the scrambled metaphors, the scrambled class-inclusions, and the literal class-inclusions to a LSA. The key question is whether the associative strength (assessed by the cosine of the vectors for the terms) is greater for the pairs used in the metaphors than for the pairs used in the scrambled metaphors. If so, then

the results of our first three studies cannot serve as evidence for early alignment. The alignment account will be strengthened if (a) the metaphor pairs do not differ in associative strength from the scrambled metaphor pairs; and (b) other contrasts show that the measure is appropriately sensitive: specifically, if literal class-inclusion pairs (for which strong prior semantic associations should exist) show more associative strength than metaphor pairs and scrambled metaphor pairs.

The LSA space was accessed from the Latent Semantic Analysis website at the University of Colorado (<http://lsa.colorado.edu/>). We used the 300-dimension LSA space “General Reading Up To 1st Year College,” constructed from texts selected from 37,651 documents, including novels, newspaper articles, and textbooks representative of what the average college freshmen would have read by the end of his/her first year of college. Using the pairwise comparison application at the Latent Semantic Analysis Web site, we obtained the cosine for each pair of terms used in the metaphors, scrambled metaphors, and literal class-inclusions used in Experiments 1–3. We compared each of the terms in both their singular and plural form.

6.2. Results

The results suggest that our key result cannot be explained in terms of associative strength. The average LSA scores for metaphors ($M = 0.058$, $SD = 0.078$) and scrambled metaphors ($M = 0.037$, $SD = 0.042$) did not differ from each other, $t(30) = 0.926$, $p = .364$. Importantly, though, both differed from the LSA scores obtained for literal class-inclusions ($M = 0.461$, $SD = 0.13$). The class-inclusion pairs were rated as far more associated than both the metaphor pairs, $t(30) = 10.647$, $p < .001$, $d = 3.76$, and the scrambled metaphor pairs, $t(30) = 12.43$, $p < .001$, $d = 4.35$. Thus, the LSA measure registered strong semantic associations where they were expected, suggesting that the failure to find a difference between metaphors and scrambled metaphors is not due to LSA’s being insensitive to association strength.

6.3. Discussion

The LSA findings suggest that the terms in the metaphors did not have strong prior semantic associations. These findings run counter to the possibility that the early advantage of forward and reversed metaphors over scrambled metaphors stemmed from preexisting semantic connections, and they support our claim that the early co-activation between the terms in the metaphors occurred via a symmetric alignment process.

However, the LSA measure is based on patterns of co-occurrence. Despite the correlation between co-occurrence and associative strength, they are not the same thing. Thus, the possibility remains that the metaphoric pairs were semantically related in a way that did not manifest itself in co-occurrence. Because this point is crucial to our case, we carried out a more comprehensive test of whether there were preexisting connections between the metaphoric pairs. One standard way to test for such semantic connections is to check for semantic priming effects in lexical decision. It is well established that people are faster and more accurate in a lexical decision to a word that is accompanied or preceded by a semantically related word relative to a semantically unrelated word (e.g., Meyer & Schvaneveldt, 1971;

Neely, Keefe, & Ross, 1989). In one standard version of the lexical decision task, participants are shown two letter strings in sequence and are asked to indicate whether the second letter string is a word (*books maps*) or nonword (*books vixes*). The typical result is that people are faster to make a “word” response if the words are semantically related (*flowers roses*) than if they are not (*giraffes roses*).

In Experiment 4b we used a frequently used task for assessing semantic connections: a lexical decision task. The idea was to test whether the terms in the metaphors would show fast lexical priming, indicative of prior semantic connections. Two letter strings were presented one at a time and people were instructed to indicate whether the second of the two strings was a word. The materials were those used in Experiments 1–3, except that people saw only the two main words in each sentence. These materials were balanced by an equal number of pairs of letter strings in which the second letter string was a nonword.

We expected that lexical decision times to the second term would be faster for class-inclusion pairs (which are semantically related) than for scrambled class-inclusion pairs. The key question is whether lexical decision times to the second term would be faster for metaphor pairs than for scrambled metaphor pairs. If so, this would suggest that the metaphoric pairs had a preexisting semantic connection. If, however, lexical decision times are no faster for metaphor pairs than for scrambled metaphor pairs, this will argue against the preexisting semantic connections explanation and strengthen the alignment account.

7. Experiment 4b

7.1. Method

7.1.1. Participants

The subjects were 24 Emory University undergraduates who participated for course credit.

7.1.2. Materials

The materials were based on the four test lists from Experiments 1–3, except for the filler items, which were not included in this new set of materials. Each sentence from the original set of materials was reduced to a pair of words. For example, “Some arguments are wars” became “arguments” and then “wars.” The word pairs were based on 16 literal class-inclusions, 16 scrambled class-inclusions, 16 metaphors, and 16 scrambled metaphors. These 64 pairs of words constituted the “word” pairs. “Nonword” pairs were constructed, in part, from nonwords from Joordens and Becker (1997) with several minor modifications. In particular, in order to make the “word” and “nonword” items as similar as possible, the nonwords from Joordens and Becker (1997) were pluralized and letters were added to some of the nonwords to equate the lengths of the “word” and “nonword” items: for example, “rensill” became “rensillers.” Additional nonwords were obtained from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002) and were also equated for length. In all 64 nonword pairs, the nonword followed a word. The words in the nonword pairs were

high-frequency nouns not used in any of the other items. All nonwords were checked by searching Merriam Webster's online dictionary; any found to be words were changed by altering a few letters without violating English phonotactics. A practice list was constructed in the same way as the experiment list. The practice list contained 24 items, half of which were "word" pairs and half of which were "nonword" pairs.

7.1.3. Design

A within-subject design was used with each subject receiving each type of word pair, in either a forward or reversed order, within a single session. The order of the word pairs was counterbalanced across lists. Half of the pairs in each pair type (e.g., metaphor, scrambled metaphor, literal class-inclusion, and scrambled class-inclusion) appeared in forward order and half in reversed order.

7.1.4. Procedure

Participants were run on Windows-based computers separated by sound-attenuating carrels. Participants were told that they would be shown pairs of letter strings on the screen one at a time and that their task was to indicate whether the second member of each pair formed a word by pressing the LEFT ARROW key for "yes" or the RIGHT ARROW key for "no." After this instructional phase, participants received 24 practice trials before receiving the 128-item test trials. In both the practice and test phases of the experiment, participants received feedback. For the practice and test trials, participants were instructed to "See how fast you can respond (e.g., under 600 ms) without making mistakes."

Presentation of the letter strings in both the practice and test sessions occurred as follows. First, a blank screen appeared. After 1,000 ms, the first letter string appeared in the middle of the screen. After 700 ms, the word was replaced with the second letter string. The second letter string remained on the screen until a key was hit. Participants then received feedback on whether their response was right or wrong, their response time in milliseconds, and their average percent correct. The feedback screen remained for 1,500 ms until the next trial began with a blank screen. Presentation of the stimulus items and the collection of responses was accomplished using *E-Prime* (version 1.1).

7.2. Results and discussion

Lexical decisions that were incorrect for the word trials (7%) and nonword trials (10%) were removed from the analysis of lexical decision times. The resulting reaction times and standard errors of the mean for literal class-inclusions, scrambled literal class-inclusions, metaphors and scrambled metaphors, with associated error rates, are shown in Table 2.

As expected, lexical decision times were faster for literal class-inclusions ($M = 500$, $SD = 47.37$) than for scrambled literal class-inclusions ($M = 520$, $SD = 55.84$), across both subjects and items, $t_s(23) = 3.86$, $p < .001$, $d = 0.79$, $t_i(15) = 2.84$, $p < .05$, $d = 0.50$. This difference is consistent with prior lexical decision studies and with the assumption that these terms are semantically associated. Turning to the key question, lexical decisions for metaphor pairs ($M = 533$, $SD = 65.79$) were no faster than for scrambled metaphor pairs

Table 2

Lexical decision response times (in milliseconds), standard errors of the mean (in parentheses), and error rates for Experiment 4b

	Reaction Time	Error Rate (%)
Literal class-inclusions	500 (9.67)	4.4
Scrambled literal class-inclusions	520 (11.40)	5.5
Metaphors	533 (13.43)	8.6
Scrambled metaphors	529 (12.32)	8.6
Nonword pairs	584 (16.84)	10.0

($M = 529$, $SD = 60.34$), across both subjects and items, $t_s(23) = .601$, $p = .554$, $t_i(31) = .208$, $p = .836$. Thus, we see no evidence for preexisting semantic connections between the metaphoric terms.

We also analyzed the error rates reported in Table 2. Error rates for literal class-inclusions ($M = 0.044$, $SD = 0.062$) did not differ from those for scrambled literal class-inclusions ($M = 0.055$, $SD = 0.065$) across either subjects or items, $t_s(23) = .659$, $p = .517$, $t_i(15) = .855$, $p = .406$. Similarly, the error rates for metaphors ($M = 0.086$, $SD = 0.102$) did not differ from those for scrambled metaphors ($M = 0.086$, $SD = 0.052$) across either subjects or items, $t_s(23) = 0$, $p = 1$, $t_i(31) = .475$, $p = .638$. In sum, the results provide no evidence for a preexisting association between the terms in the metaphors.

We also checked for directional priming from base to target, as might be expected on the class-inclusion account. No evidence for directional priming was found for metaphors. Lexical decision times when bases (e.g., *virus*) were primed by targets (e.g., *rumors*; $M = 535$, $SD = 61.4$) did not differ from times when targets were primed by bases ($M = 535$, $SD = 75.53$). However, we did find directional priming for the literal class-inclusions. Lexical decision times were faster when categories (e.g., *fruit*) were primed by category members (e.g., *apple*; $M = 486$, $SD = 54.0$) than when category members were primed by categories ($M = 515$, $SD = 51.1$). This effect held across subjects, $t(23) = 3.24$, $p < .01$, $d = 0.66$, but not across items, $t(15) = 1.47$, $p = .164$. The finding of directional priming for literal class-inclusions, but not for metaphors, provides further support for the view that there were preexisting associations in the case of the literal class-inclusions but not in the case of the metaphors.

The lexical decision patterns converge with the findings of the LSA analyses in Experiment 4a. Both findings argue against the possibility that the early symmetric activation of metaphors found in Experiments 1–3 could have resulted from prior semantic connections. If there had been prestored semantic associations, such associations should have resulted in faster lexical decisions for the words drawn from metaphors than from scrambled metaphors. Further, the failure to find a difference cannot be attributed to insensitivity in the lexical decision measure, as differences were found where expected.

Our results in the lexical decision task mirror those of Camac and Glucksberg (1984), who also used lexical decision (but with terms showed simultaneously instead of sequentially) to test for preexisting associative relations in metaphors, scrambled metaphors, literal class-inclusions, and scrambled class-inclusions. As in our study, Camac and Glucksberg

found evidence of priming between the terms in literal class-inclusions, but not between terms in metaphors. In sum, the results of Experiments 4a and 4b support the claim that the effects in Experiments 1–3 reflect an initial process of symmetric alignment.

8. General discussion

We began this paper by noting that there is a dichotomy in theories of metaphor. Some theories focus on the directional projection of inferences—on how the target is changed by the metaphor—and others focus on the emergence of common abstractions that can influence the subsequent representation of the base as well as that of the target. We suggested that a model derived from structure-mapping theory could capture both these phenomena. On this account, metaphor processing begins with an initial symmetric alignment process, which is followed by a later directional stage in which further inferences from the base are projected to the target. Moreover, these processes are intimately connected; the aligned structure provides the basis for the subsequent inferences. The results reported here provide evidence for this view.

Despite the fact that the metaphors in our studies were strongly directional, the directional preference emerged only after an initial symmetric stage of processing. In all three studies of online metaphor processing (Experiments 1–3), we found that the difference in comprehensibility between forward and reversed metaphors increased with processing time. Experiments 2 and 3, with their earlier initial deadlines, provided direct evidence for initial symmetric processing: At the early deadlines, forward and reversed metaphors were rated as equally comprehensible. Finally, in all three studies, even at the earliest deadlines metaphors (in both directions) were rated as more comprehensible than scrambled metaphors (and less comprehensible than literal statements), indicating that meaningful processing had begun. The clear separation between literal statements, metaphors, and anomalies at all time periods shows that the intermediate comprehensibility levels found for metaphors in the early stage were not due to chance. When given scrambled statements, participants unambiguously rejected them at all stages. But both forward and reversed metaphors received an intermediate proportion of “comprehensible” judgments. There was also indirect support for the claim that the later directional stage builds on the earlier symmetric processing, in that there was a positive correlation between the early and late comprehensibility ratings in Experiments 2 and 3 for forward metaphors, but not for reversed metaphors. In Experiment 4b we showed that the evidence for the symmetric processing found in Experiments 2 and 3 cannot be merely due to preexisting semantic connections, since lexical decision times for pairs from metaphors were no faster than for pairs from scrambled metaphors. Likewise, Experiment 4a showed that LSA relatedness scores were no higher for metaphoric pairs than for pairs from scrambled metaphors.

To summarize, (a) at very early time periods (500–600 ms) metaphors were indistinguishable from each other, but clearly distinguishable from both literal statements and anomalies—that is, early processing was symmetric; (b) at later deadlines, forward metaphors were more comprehensible than reversed metaphors—that is, later processing was

directional; and finally, (c) the evidence suggests that the later directional stage built on the earlier symmetric alignment. This pattern is consistent with structure-mapping's prediction of an early symmetric alignment process followed by a stage of directional inference projection.

8.1. Similarity

These studies allow us to ask whether high-relational similarity between the target and base facilitates alignment. Because the high-similarity pairs differed from their low-similarity counterparts chiefly in *relational* similarity,⁵ we expected any such difference to show up relatively late in processing. There is evidence that relational similarity is slower to process than is concrete attribute similarity⁶ (Goldstone, 1994; Love et al., 1999; Lovett, Gentner, Forbus, & Sagi, 2009; V. Sloutsky & A. S. Yarlas, unpublished data). Thus, we expected that any effects of relational similarity would occur at later, rather than earlier, deadlines.

8.2. Implications for theories of metaphor comprehension

Our findings of initial symmetric processing run counter to theories that posit a fixed direction of processing, such as the attributive category model of Glucksberg and colleagues (Glucksberg & Keysar, 1990; Glucksberg et al., 1997) and some versions of the embodiment approach to metaphor. In the attributive category model, the essence of metaphor processing is that the target term is assigned to the abstract category of which the base is a prototypical member. According to this account, the first step in comprehending a metaphor is to find the abstract metaphorical category associated with the base, while simultaneously identifying sets of modifiable dimensions in the target. Thus, both terms are involved in role-specific ways even at the very outset of metaphor processing—contrary to the present evidence that initial processing is role-neutral.

Our findings are also problematic for some versions of the embodied approach to metaphor. Recently a number of researchers have suggested that metaphor understanding can best be explained in terms of embodied cognition (Cienki et al., 2008; Gibbs, 2006; Gibbs et al., 2004; Lakoff & Johnson, 1999; Wilson & Gibbs, 2007). This approach to metaphor has been extremely influential in cognitive linguistics. A guiding principle of the embodied approach is that the sensorimotor system is at the core of cognition (Barsalou, 2008; Wilson, 2002). Thus, cognition is intertwined with perception and action, rather than being centralized and abstract. It is useful to distinguish two views of how this might work, which we will call *moderate embodiment* and *strong embodiment* (for reviews, see Steen, 2008; Wilson, 2002; Wilson & Gibbs, 2007; Weiskopf, 2010). In *moderate embodiment*, sensorimotor representations provide the initial source for metaphors but give rise to abstract conceptual structures. In *strong embodiment*, there are no stable abstract representations. Rather, cognitive processing occurs through embodied simulation; that is, higher level cognition is based on online operations on modality-specific representations in the perceptual-motor systems. Although abstractions may be generated in real-time simulations over stored sensorimotor experiences, they are not retained as enduring representations.

In the strong embodiment account, the process of metaphor understanding involves embodied simulations based on actual sensorimotor systems in the brain. Modality-specific sensorimotor encodings provide the base domain from which metaphors are drawn, and the substrate in which such metaphors are processed. Abstract ideas like pride, goal, and argument would be understood in terms of actions such as grasping, pushing, and chewing (see Wilson & Gibbs, 2007⁷). Thus, when an abstract domain such as time is metaphorically compared to a more concrete experiential domain such as space, (a) the abstract domain derives its structure from the experiential domain; and (b) thinking about the abstract domain is done by invoking sensorimotor experience.

This view of metaphor implies strong directionality. The only possible direction of inference is from modality-specific sensorimotor representations to abstract ideas. It is not clear how this view can accommodate metaphors like “The heart is a pump” and “The liver is a filter,” in which a mechanical contrivance is used as a base domain to structure our own body. Of course, our hearts and livers are invisible to us. But what about metaphors in which mechanical devices are used to structure our own emotions, as in “He blew his top” or “She was really steaming” (instances of the “Anger is a hot fluid under pressure” metaphoric system). Another challenge is that metaphors sometimes involve two different sensorimotor modalities, as in Lakoff and Johnson’s (1999) conceptual metaphor “Seeing is touching,” as in expressions such as “I felt his glance” and “He touched me with his eyes.” The sense of this metaphor is readily understood as likening *an imagined force from the eye to the thing seen to the tactile sense of something that touches our skin*. In a strong embodiment view, this metaphor would appear to be impossible, as the base and target belong to different sensorimotor modalities and must therefore be processed in different cortical regions.

A further theoretical problem for the strong embodiment approach is that it cannot capture the fact that the abstraction that emerges from a metaphor can change the representation of the base concept as well as that of the target concept. Because the sensorimotor substrate that gives rise to metaphors is assumed to be modality-specific, the experiential encodings are not penetrable by the momentary abstractions that may be generated from them. As we review below, research in historical linguistics and in the laboratory suggests, on the contrary, that metaphorical abstraction processes *do* lead to new representations of the base terms. These new representations are not only retained as abstractions, but are in many cases re-used and further abstracted.

In *moderate embodiment* accounts, sensorimotor representations often provide the initial fodder for metaphors but in so doing they give rise to abstract conceptual structures (Boroditsky, 2000; Boroditsky & Ramscar, 2002; Gentner, Imai, & Boroditsky, 2002; Lakoff & Nunez, 2000; McGlone & Harding, 1998). These more abstract conceptual structures could be described as structured representations or as image-schemas. Often the base domain is an experiential gestalt, as in “An argument is a journey” (e.g., “Are you following this discussion?”; “The first steps in the proof...”; “This line of reasoning has reached a dead end.”) (Lakoff & Johnson, 1980). Once an abstraction is developed, it can often be used with a variety of different target concepts: for example, “A marriage is a journey” and “Life is a journey.” The moderate embodied view

receives support from historical linguistics (Heine, 1997; Traugott, 1978) as well as from psycholinguistic studies. For example, the Swahili term *mbele* began as the body-part term “breast”; it was extended to become a more general part term meaning “frontside or front part” and, even more abstractly, a purely locational term meaning “the front” or “in front of” and (still more abstractly) a temporal marker meaning “before” (Heine, 1997). A better known example from historical linguistics is that temporal semantics often arises via metaphorical extensions from space to time (Bierwisch, 1996; Heine, 1997; Traugott, 1978), as in *We are fast approaching the holidays* (the “ego-moving” mapping), and *The holidays are fast approaching* (the “time-moving” mapping). Psychological research suggests that these two mappings are psychologically real, in that switching between them costs processing time (Gentner et al., 2002), and that spatial events prime the corresponding temporal events (Boroditsky, 2000).

8.3. *The career of metaphor*

The results from historical linguistics dovetail with the present results in suggesting that metaphors are processed by aligning two representations and abstracting common structure, and that the resulting abstractions can be retained for further use. A further implication is that new meanings can develop gradually over repeated instances of metaphor use. This is the process proposed in the career of metaphor hypothesis (Bowdle & Gentner, 2005; Gentner & Bowdle, 2001; Gentner & Wolff, 1997, 2000; see also Chiappe & Kennedy, 2001), an extension of structure-mapping theory to metaphorical extension. According to structure-mapping, processing a figurative statement involves a process of alignment that results in a common structure that is typically somewhat more abstract than either term. If this common abstraction is repeatedly invoked, it may become a standard sense of the base term, creating a conventional metaphor. This process of repeated alignment and abstraction is important in the creation of new abstract terms: for example, in relational abstractions such as *sanctuary* or *bridge* (Zharikov & Gentner, 2002), and as in the shifts of meaning documented in historical linguistics (Bybee, 1985; Heine, 1997; Traugott, 1978).

One implication of the career of metaphor framework is that there should be a continuum of conventionality in metaphor. Highly conventional metaphors are those whose bases already possess a salient conventional metaphoric meaning: for example, *goldmine*, used to mean “a source of something valuable.” There is considerable evidence for this continuum in figurative language. For example, conventional metaphors are comprehended faster than novel metaphors (Blank, 1988; Gentner & Wolff, 1997), consistent with the claim that conventional metaphor bases already have an associated abstraction, whereas the metaphoric abstraction must be derived anew for a novel figurative. Further, studies by Giora (1997, 1999, 2007) indicate that for conventional bases, the abstract meaning is often the default sense, accessed early in processing regardless of context. Finally, Bowdle and Gentner (2005) demonstrated that conventionalization can occur *in vitro*, if the same base term is repeatedly used with different targets.

The progressive abstraction sequence proposed in the career of metaphor receives some support from studies tracing the neural activation of sensorimotor metaphors. For example, Chatterjee and colleagues find, using fMRI studies, that metaphorical sentences involving action verbs are processed in areas adjacent, but anterior to, the left occipito-temporal areas activated by literal sentences using the same action verbs (Chen, Widick, & Chatterjee, 2008; Wu, Waller, & Chatterjee, 2007). They note that this suggests a route by which initially sensorimotor representations can become abstracted and stored. Desai, Binder, Conant, Mano, and Seidenberg (2011) used fMRI to compare neural processing of literal versus metaphoric sensorimotor sentences varying in familiarity. They found that metaphoric (but not literal) sentences activated areas involved in processing abstract sentences (see Giora, 2007). Further, they found lower activation of primary motor and motion perception areas for familiar than for unfamiliar examples, leading them to suggest a gradual abstraction process. These findings are consistent with the predictions of the career of metaphor, as well as with the moderate embodiment position: Sensorimotor representations can give rise to metaphors and, with sufficient conventionalization, to metaphorical abstractions.

8.4. Summary

Metaphors lead a double life. On the one hand, metaphor is a way of discovering emergent commonalities that may alter the representation of both terms. On the other hand, metaphors are strikingly asymmetric; often their communicative function is that the base concept provides a way of viewing the target concept. This has led some theorists to conceive of metaphor as a purely directional process, in which the base provides a firm structure that can be imported to the target.

But it is important to distinguish the communicative function of metaphors from the process by which they are comprehended. The present results show that the process of comprehending a metaphor—even a highly directional metaphor—involves an initial symmetric stage. We suggest that this initial stage is an alignment process in which a common structure is found. This structure—which by its nature will be somewhat more abstract than either the base or the target concept—can act to subtly alter the representation of the base as well as of the target. If it is repeatedly invoked, it can become a secondary word meaning or even supplant the original meaning of the base term. Thus, to return to Toby Litt’s metaphor, when you read “the tiger roaring like a fire,” you may (as the metaphor requests) imbue the tiger with characteristics of a fire. But the metaphor will color your sense of fire as well—you will probably imagine a rather ferocious fire. Although the finding of an early symmetric stage in the processing of metaphor may initially seem counterintuitive, we suggest that the early alignment sets the stage for directional inferences that are appropriate to the particular pairing of base and target. More important for the broad scheme of things, the alignment process explains how metaphors can serve to create enduring abstractions of the base concept. This gradual metaphoric abstraction is crucial to explaining change of meaning in language evolution, as well as in history of science and in individual learning and development.

Notes

1. Grammaticalization is a process whereby terms that once served as content words become used as grammatical terms; that is, they lose their concrete lexical meaning and participate in (typically obligatory) grammatical rules (e.g., Heine, 1997).
2. This account is part of a larger approach to mental processing that rejects many of the assumptions of traditional symbol-processing accounts of cognition (e.g., Fodor & Pylyshyn, 1988; Minsky, 1990; Norman, Rumelhart, the LNR Research Group, 1975; Pinker, 1997), especially the assumption that perceptual input is translated into amodal symbolic representations that can be operated on by the same processes that operate over abstract content. In embodiment theory, perceptual and motor systems are seen as the foci where cognition primarily occurs (Barsalou, 2008). Rather than being at the periphery, the sensorimotor system is viewed as being at the core of cognition.
3. This merge algorithm, called the *greedy merge* algorithm, operates in linear time. Although the interpretations it finds cannot be guaranteed to be maximal, the algorithm does very well. Forbus and Oblinger (1990) tested the greedy algorithm on a large set of analogies; on 52 of 56 pairs, its top interpretation was identical to the best interpretation found in an exhaustive merge.
4. Henceforth we abbreviate *high-relational similarity* as *high similarity*, and likewise for *low-relational-similarity*.
5. For example, the difference between the high-similarity metaphor “Some auditions are doors” and its low-similarity counterpart “Some plays are doors” is that “audition” and “door” share a salient relation—*affording access to something*—that is lacking in the corresponding low-similarity target, “play.”
6. The results offer some encouragement for this speculation. In Experiment 3, with its short initial deadline (500 ms), there was an interaction between similarity and deadline, with similarity effects emerging at the later (1,600 ms) deadline. High-similarity metaphors were rated as more comprehensible than low-similarity metaphors at the later but not the earlier deadline. The results in Experiment 2 (with its 600 ms early deadline) showed a similar but nonsignificant pattern. High- and low-similarity metaphors were both at 37.5% comprehensibility at the early deadline, but differed nonsignificantly at the later deadline (high: 53%; low: 40.57%). In Experiment 1, with its later deadlines, high-similarity metaphors were rated as more comprehensible than low-similarity metaphors (significant across subjects) at both early (1,200 ms) and late (1,800 ms) deadlines.
7. It should be noted that Wilson and Gibbs (2007) do not claim that all metaphors (e.g., *Lawyers are sharks*) are necessarily understood in terms of bodily activities and sensations; as a consequence, their position is not that metaphor understanding must involve embodied representations, just that it often does. Thus, this view would not be classified as a strong embodiment view.

Acknowledgments

This research was supported by Office of Naval Research grant N00014-92-J-1098 awarded to the second author. We are grateful to Jeff Loewenstein for suggesting the use of the deadline procedure, to the Similarity and Analogy group at Northwestern University for many discussions of these issues, and to Kathleen Braun for help with the research and analyses.

References

- Aristotle (1932). Rhetoric. In L. Cooper (Trans.), *The rhetoric of Aristotle*. New York: Appleton-CenturyCrofts.
- Barsalou, L. W. (2005). Abstraction as dynamic interpretation in perceptual symbol systems. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories* (pp. 389–431). Carnegie Symposium Series. Mahwah, NJ: Erlbaum.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Bierwisch, M. (1996). How much space gets into language? In P. Bloom & M. A. Peterson (Eds.), *Language and space. Language, speech, and communication* (pp. 3–35). Berlin: Walter De Gruyter & Co.
- Blank, G. D. (1988). Metaphors in the lexicon. *Metaphor and symbolic activity*, 3, 21–36.
- Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75(1), 1–27.
- Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13, 185–188.
- Bowlde, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112, 193–216.
- Brysbaert, M., Bovens, N., & d'Ydewalle, G. (1989). Turbo Pascal timing routines for the IBM microcomputer family. *Behavior Research Methods, Instruments, & Computers*, 21, 73–83.
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Camac, M. K., & Glucksberg, S. (1984). Metaphors do not use associations between concepts, they are used to create them. *Journal of Psycholinguistic Research*, 13, 443–455.
- Chen, E., Widick, P., & Chatterjee, A. (2008). Functional–anatomical organization of predicate metaphor processing. *Brain and Language*, 107, 194–202.
- Chiappe, D. L., & Kennedy, J. M. (2001). Literal bases for metaphor and simile. *Metaphor and Symbol*, 16, 249–276.
- Cienki, A., Cornelissen, J. P., & Clarke, J. (2008, November). The role of human scale, embodied metaphors/blends in speech and gestures of entrepreneurs. Paper presented at the 9th Conference on Conceptual Structure, Discourse, and Language, Cleveland.
- Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., & Seidenberg, M. S. (2011). The neural career of sensorimotor metaphors. *Journal of Cognitive Neuroscience*, 23(9), 2376–2386.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: An algorithm and examples. *Artificial Intelligence*, 41, 1–63.
- Fischler, I., & Goodman, G. O. (1978). Latency of associative activation in memory. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 455–470.
- Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19, 141–205.
- Forbus, K. D., & Oblinger, D. (1990). Making SME greedy and pragmatic. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 61–68). Hillsdale, NJ: Erlbaum.
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 59, 47–59.

- Gentner, D., & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor and Symbol, 16*(34), 223–247.
- Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 199–253). Cambridge, MA: MIT Press.
- Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In G. H. Bower (Ed.), *The psychology of learning and motivation, advances in research and theory* (Vol. 22, pp. 307–358). New York: Academic Press.
- Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space → time metaphors. *Language and Cognitive Processes, 17*, 537–565.
- Gentner, D., & Kurtz, K. (2005). *Relational categories*. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), *Categorization inside and outside the lab* (pp. 151–175). Washington, DC: APA.
- Gentner, D., & Markman, A. B. (1997). Structure-mapping in analogy and similarity. *American Psychologist, 52*, 45–56.
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language, 37*, 331–355.
- Gentner, D., & Wolff, P. (2000). Metaphor and knowledge change. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 295–342). Cambridge, MA: MIT Press.
- Gibbs, R. (2006). *Embodiment and cognitive science*. New York: Cambridge University Press.
- Gibbs, R. W., Costa Lima, P. L., & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics, 36*, 1189–1210.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8*, 183–20.
- Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31*, 919–929.
- Giora, R. (2007). Is metaphor special? *Brain and Language, 100*, 111–114.
- Glucksberg, S., Gildea, P., & Bookin, H. (1982). On understanding non-literal speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior, 1*, 85–96.
- Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*, 3–18.
- Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language, 36*, 50–67.
- Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 3–28.
- Heine, B. (1997). *Cognitive foundations of grammar*. New York: Oxford University Press.
- Hock, H. H., & Joseph, B. D. (1996). *Language history, language change and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*, 2nd ed. Cambridge, England: Cambridge University Press.
- Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1083–1105.
- Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language, 28*, 275–385.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.
- Lakoff, G., & Nunez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.

- Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago: University of Chicago Press.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 259–284.
- Litt, T. (1989, May). For Borges. *Oxford Magazine*.
- Love, B. C., Rouder, J. N., & Wisniewski, E. J. (1999). A structural account of global and local processing. *Cognitive Psychology*, *38*, 291–316.
- Lovett, A., Gentner, D., Forbus, K., & Sagi, E. (2009). Using analogical mapping to simulate time-course phenomena in perceptual similarity. *Cognitive Systems Research*, *10*(3), 216–228. Special Issue on Analogies—Integrating Cognitive Abilities.
- Malgady, R. G., & Johnson, M. G. (1976). Modifiers in metaphors: Effects of constituent phrase similarity on the interpretation of figurative sentences. *Journal of Psycholinguistic Research*, *5*, 43–52.
- Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, *21*, 373–418.
- Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory storage. *Psychological Science*, *8*(5), 363–367.
- Markman, A. B., & Gentner, D. (2000). Structure mapping in the comparison process. *American Journal of Psychology*, *113*(4), 501–538.
- McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, *6*, 486–494.
- McGlone, M. S., & Harding, J. L. (1998). Back (or forward?) to the future: The role of perspective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1211–1223.
- McGlone, M. S., & Manfredi, D. A. (2001). Topic–vehicle interaction in metaphor comprehension. *Memory & Cognition*, *29*, 1209–1219.
- Meyer, D., & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.
- Miller, G. A. (1979). Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (1st ed., pp. 202–250). Cambridge, England: Cambridge University Press.
- Minsky, M. (1990). Logical versus analogical, or symbolic versus connectionist, or neat versus scruffy. In P. H. Winston & S. A. Shellard (Eds.), *Artificial intelligence at MIT, expanding frontiers vol. 1* (pp. 218–243). New York: MIT Press.
- Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*, 173–204.
- Neely, J. H., Keefe, D. E., & Ross, K. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 1003–1019.
- Norman, D. A., Rumelhart, D. E., & the LNR Research Group (1975). *Explorations in cognition*. San Francisco: Freeman.
- Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, *86*, 161–180.
- Pinker, S. (1997). *How the mind works*. New York: W. W. Norton & Company, Inc.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.
- Rumelhart, D. E., & Abrahamson, A. A. (1973a). A model for analogical reasoning. *Cognitive Psychology*, *5*, 1–28.
- Rumelhart, D. E., & Abrahamson, A. A. (1973b). A model for analogical reasoning. *Cognitive Psychology*, *5*(1), 1–28.
- Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(6), 1191–1210.
- Shen, Y. (1989). Symmetric and asymmetric comparisons. *Poetics*, *18*, 517–536.
- Steen, G. (2008). The paradox of metaphor: Why we need a three-dimensional model of metaphor. *Metaphor and Symbol*, *23*(4), 213–241.

- Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive Psychology*, 13, 27–55.
- Traugott, E. C. (1978). On the expression of spatio-temporal relations in language. In J. H. Greenberg (Ed.), *Universals of human language: Vol 3 word structure* (pp. 369–400). Stanford, CA: Stanford University Press.
- Turner, M. (1987). *Death is the mother of beauty: Mind, metaphor, and criticism*. Chicago: University of Chicago Press.
- Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4), 327–352.
- Veale, T. (2003). Systematicity and the lexicon in creative metaphor. In J. Barnden, S. Glasbey, M. Lee & A. Wallington (Eds), *Proceedings of the ACL Workshop on Figurative Language and the Lexicon* (Vol 14, pp. 28–35). Morristown, NJ: Association for Computational Linguistics.
- Way, E. C. (1991). *Knowledge representation and metaphor*. Dordrecht, The Netherlands: Kluwer Academic.
- Weiskopf, D. A. (in press). Embodied cognition and linguistic comprehension. *Studies in History and Philosophy of Science*, 41(3), 294–304.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9, 625–636.
- Wilson, N. L., & Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science*, 31, 721–731.
- Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, 4(2), 167–183.
- Wisniewski, E. J., & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology*, 39, 208–238.
- Wolff, P., & Gentner, D. (2000). Evidence for role-neutral initial processing of metaphors. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 529–541.
- Wu, D., Waller, S., & Chatterjee, A. (2007). The functional neuroanatomy of thematic role and locative relational knowledge. *Journal of Cognitive Neuroscience*, 19, 1542–1555.
- Zharikov, S., & Gentner, D. (2002). Why do metaphors seem deeper than similes? In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 976–981). Fairfax, VA: George Mason University.

Appendix

Literal class-inclusions used in Experiments 1–3

Some soldiers are lieutenants
 Some utensils are forks
 Some fruits are apples
 Some weapons are knives
 Some crimes are murders
 Some tools are hammers
 Some instruments are pianos
 Some birds are robins
 Some vehicles are cars
 Some toys are trucks
 Some dances are waltzes
 Some vegetables are carrots
 Some insects are flies
 Some flowers are roses
 Some trees are maples
 Some ships are destroyers