Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 5, 1082–1092

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.5.1082

The Importance of Being Nonalignable: A Critical Test of the Structural Alignment Theory of Similarity

Zachary Estes
University of Georgia

Uri Hasson
Princeton University

The structural alignment theory of similarity distinguishes 2 types of difference that may occur between stimuli: Alignable differences are those related to a commonality, whereas nonalignable differences are not related to a commonality. Alignment theory predicts that alignable differences should be more heavily weighted than nonalignable differences in similarity judgment. Experiments 1 and 2 demonstrate that, contrary to this prediction, nonalignable differences exerted a greater impact than alignable differences in similarity and difference judgments of geometric stimuli. Experiment 3 revealed that the relative weight accorded a given difference was also affected by contextual constraints. Thus, although the experiments supported the validity of the distinction between alignable and nonalignable differences, results were discordant with the specific prediction of structural alignment theory.

Structural alignment theory has provided many important contributions to cognitive science over the last 2 decades. Since its introduction by Gentner (1983), alignment theory has successfully explained a broad range of cognitive phenomena in such domains as analogy (Gentner & Markman, 1997), metaphor (Gentner & Wolff, 1997), concept combination (Wisniewski, 1997), categorization (Lassaline & Murphy, 1998; Markman & Wisniewski, 1997), memory (Markman & Gentner, 1997), choice (Markman & Medin, 1995; Medin, Goldstone, & Markman, 1995), and similarity and difference judgments (Gentner & Markman, 1994; Goldstone, Medin, & Gentner, 1991; Markman, 1996; Markman & Gentner, 1996; Medin, Goldstone, & Gentner, 1990, 1993). Essentially, alignment theory provides a unified account of comparative cognitive processes.

Structural Alignment Theory

Simply stated, structural alignment theory posits that comparison and related cognitive operations are accomplished by putting the structure of one concept into alignment or correspondence with the structure of the other concept to which it is compared. When comparing a rose and a violet, for instance, one aligns the two concepts. This alignment process may yield *commonalities* (e.g., both roses and violets have petals), *alignable differences,* which are related to commonalities (e.g., roses are red, violets are blue), and *nonalignable differences,* which are not related to commonalities (e.g., roses have thorns, violets do not). Put another way, alignable differences occur when the two concepts have different values (e.g., red v. blue) on a common dimension (i.e., color of petals).[1] Nonalignable differences occur when one concept possesses an attribute that has no correspondence in the other concept (e.g., thorns). This distinction between alignable and nonalignable differences is unique to structural alignment theory (Gentner & Markman, 1997); feature-based models of similarity (e.g., Tversky, 1977) distinguish only between commonalities and differences. The further distinction between alignable and nonalignable differences is critical because it provides the basis of alignment theory's predictions about how people evaluate similarity. A central claim of alignment theory is that "alignable differences are more salient than nonalignable differences" (Gentner & Markman, 1997, p. 50). As delineated below, this differential "salience" is manifest as a difference in the relative weighting of alignable and nonalignable differences in various cognitive tasks, such as similarity and difference judgments.

For example, Markman and Gentner (1996) presented a standard stimulus and two alternative stimuli, and participants were instructed to choose the alternative most similar to the standard. To illustrate, one standard stimulus was a drawing of an archer aiming an arrow at a target (see Figure 1A). The alternative stimuli always consisted of an alignable difference stimulus and a nonalignable difference stimulus. In this example, the alignable difference alternative was identical to the standard stimulus except that, in the alternative, the archer was aiming at a bird (see Figure 1C). The nonalignable difference alternative was identical to the standard

Zachary Estes, Department of Psychology, University of Georgia; Uri Hasson, Department of Psychology, Princeton University.

This research was funded in part by a Faculty Research Grant to Zachary Estes from the University of Georgia Research Foundation.

We thank Mark Keane, Art Markman, Brian Ross, and Barbara Spellman for insightful and formative comments on numerous drafts of this article. We are especially grateful to Art Markman and Brian Ross for suggesting Experiment 2. Experiment 1 was presented at the 23rd Annual Conference of the Cognitive Science Society in Edinburgh, Scotland, August, 2001.

Correspondence concerning this article should be addressed to Zachary Estes, Department of Psychology, University of Georgia, Athens, GA 30602. E-mail: estes@uga.edu

[1] Alignable differences may also be conceptualized as different arguments to the same relation, such as *above* (square, circle) and *above* (square, triangle). In this example, the circle and triangle arguments constitute an alignable difference.

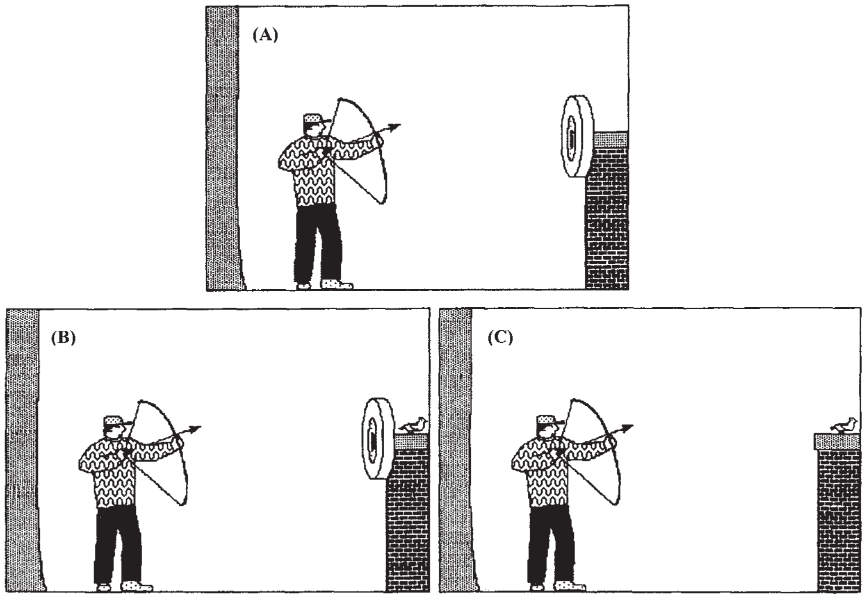

Figure 1. Examples of stimuli from Markman and Gentner (1996). Panel A is the standard, Panel B is the nonalignable difference alternative, and Panel C is the alignable difference alternative. From "Commonalities and Differences in Similarity Comparisons," by A. B. Markman and D. Gentner, 1996, *Memory & Cognition, 24,* p. 242. Copyright 1996 by the Psychonomic Society. Reprinted with permission.

but additionally included a bird located behind the target (see Figure 1B). Alignment theory claims that, by virtue of their relation to a commonality, alignable differences are more salient than nonalignable differences. As a result, the alignable difference should be more heavily weighted than the nonalignable difference, and hence the alignable difference alternative should be judged less similar to the standard than should the nonalignable difference alternative (Gentner & Markman, 1994, 1997; Markman & Gentner, 1996). Markman and Gentner's (1996) results supported this prediction—the nonalignable difference stimuli were selected as more similar to the standard on 88% of the trials. Thus, the distinction between alignable and nonalignable differences accounted for the pattern of similarity judgments.

Markman and Gentner (1996) further reasoned that, if alignable differences are more salient than nonalignable differences, then variation in an alignable difference should have a greater effect on similarity judgments than should variation in a nonalignable difference. Therefore, in their Experiment 2, stimuli consisted of a standard that was compared with four different alternatives. The alternatives varied orthogonally in alignability (i.e., whether the difference was alignable or nonalignable) and variation (i.e., whether the difference was a similar or a different object). For example, one standard depicted a robot fixing a car (see Figure 2A). To create the alignable stimuli, the car was replaced in the scene by either a truck or another robot. Because cars are more similar to trucks than to robots, the alignable-similar (AS) alternative consisted of the robot fixing the truck (see Figure 2B), whereas the alignable–different (AD) alternative consisted of the robot fixing another robot (see Figure 2C). Nonalignable stimuli were created by keeping constant the scene depicted in the standard and adding either a truck or a robot. Thus, in the nonalignable–similar (NS) alternative, the robot still fixed a car, but the scene additionally included the truck as a nonalignable difference (see Figure 2D). Likewise, the nonalignable–different (ND) alternative additionally included another robot (see Figure 2E). Each of the four alternatives was separately presented with the standard, and participants rated the similarity of each pair of scenes. Alignment theory predicts that because alignable differences are more salient, they should be more sensitive to variation. Alignment theory therefore predicts an interaction of alignability and variation, with variation affecting ratings more in alignable differences than in nonalignable differences. Indeed, this interaction obtained. The AS stimuli (see Figure 2B) were found to be much more similar to the standard than were the AD stimuli (see Figure 2C). However, the NS stimuli (see Figure 2D) were only slightly more similar to the standard than were the ND stimuli (see Figure 2E). In addition to this interaction, there was also a main effect of alignability such that nonalignable differences were rated more similar to the standard than were alignable differences. Thus, support for alignment theory was two-fold—alignable differences counted more against similarity, and variation in alignable differences exerted a greater effect on similarity.

Empirical Concerns

The experiments by Markman and Gentner (1996), described above, are paradigmatic examples of the evidence for structural alignment. This evidential support rests on the conclusion that alignable differences are more important for similarity than are nonalignable differences. On closer inspection, however, it is not clear that the experiments truly support this prediction of structural alignment theory. Two empirical concerns are outlined below.

First, Markman and Gentner (1996) actually demonstrated that in certain cases, the influence of alignable differences is not greater

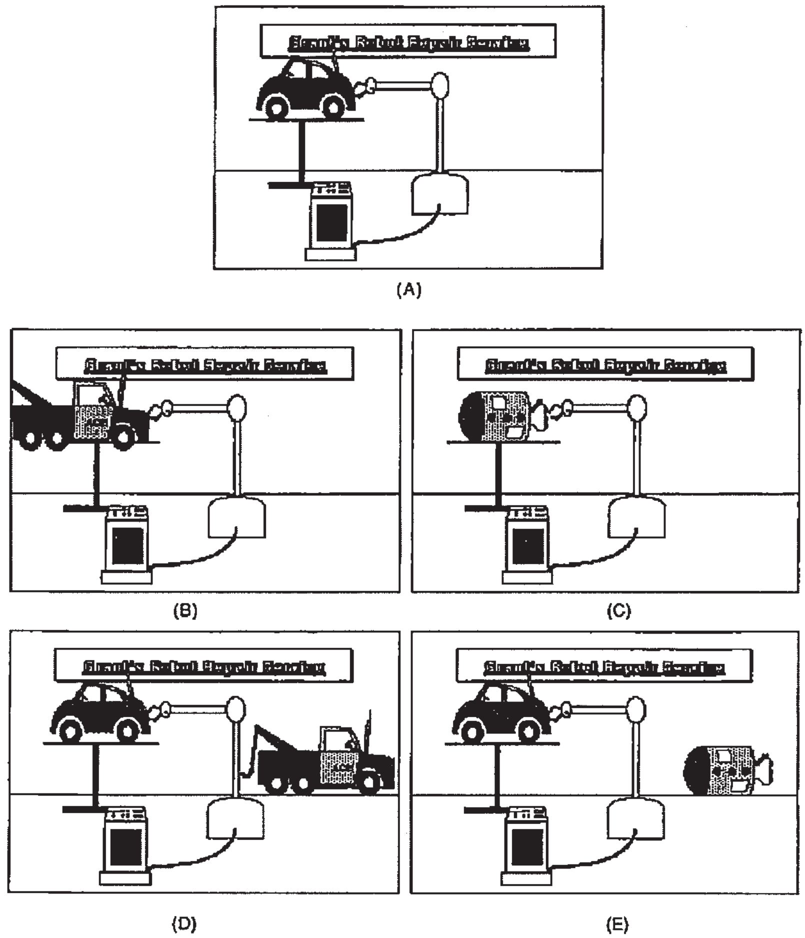

Figure 2. Examples of stimuli from Markman and Gentner (1996, Experiment 2). Panel A is the standard, Panel B is the alignable–similar alternative, Panel C is the alignable–different alternative, Panel D is the nonalignable–similar alternative, and Panel E is the nonalignable–different alternative. From "Commonalities and Differences in Similarity Comparisons," by A. B. Markman and D. Gentner, 1996, *Memory & Cognition, 24,* p. 244. Copyright 1996 by the Psychonomic Society. Reprinted with permission.

than that of nonalignable differences. In their Experiment 2, they found that in some cases, alignable and nonalignable differences affected similarity approximately equally. Specifically, although the AD stimuli (see Figure 2C) were rated less similar to the standard than were the ND stimuli (see Figure 2E), the AS stimuli (see Figure 2B) were not reliably less similar to the standard than were the NS stimuli (see Figure 2D).[2] Thus, in some cases, nonalignable differences were no less important than alignable differences. To provide another example, in comparing a rose and violet, the nonalignable property of possessing thorns is particularly diagnostic of roses. Thus, in this case, the nonalignable difference is inherently salient and therefore may be more important than any alignable difference (Costello & Keane, 2001; see also Wisniewski & Middleton, 2002).

Second, it is ambiguous whether participants' similarity judgments were based on the objects in the scenes, the actions depicted in those scenes, or both (see Davenport & Keane, 1999, for a similar critique.) That is, in their standard stimuli, there is a highly salient action—for example, an archer aiming at a target (see Figure 1) or a robot fixing a car (see Figure 2). It is important to note that the alignable differences fundamentally change the action of those scenes (e.g., aiming at a bird or fixing a truck). In contrast, the nonalignable differences are irrelevant to the action of the scene (e.g., a bird behind the target or a truck beside the robot). Thus, the greater impact of alignable differences may be attributable to the fact that these differences were not merely related to commonalities but were relevant to the semantics of the scene.

More generally, then, the data of Markman and Gentner (1996) raise some empirical concerns about the presumed relation between alignability and similarity. To begin with, the predicted effect of alignability did not always obtain. Furthermore, when the effect of alignability did obtain, it could alternatively be attributed to spurious semantic aspects of the stimuli. Below we report three

[2] If one conducts *t* tests on the means presented in Markman and Gentner's (1996) Table 2, then $t(7) = 4.25$, $p < .01$ for the AD and ND stimuli, but $t(7) = 1.33$, $p = .23$ for the AS and NS stimuli.

experiments that address these concerns. Before presenting those experiments, however, an important theoretical precaution must be noted.

Theoretical Precaution

Fundamental to structural alignment theory is its distinction between alignable and nonalignable differences. A serious theoretical concern, however, is the subjectivity of this distinction—whether a given difference is considered alignable or nonalignable is subject to the participant's own conception of the stimuli. Markman and Gentner (1996) stated that "the classification of differences as alignable or nonalignable is done *relative to the way in which the two [stimuli] are aligned*: that is, relative to the perceived common structure" (p. 237). It is of the utmost importance to realize that this subjectivity can potentially lead to circularity. For example, suppose that nonalignable differences were shown to be more influential than alignable differences in the judgment of similarity. Such a result would be contrary to the prediction of alignment theory. However, experimenters could simply cite the data as evidence that what they thought was nonalignable in fact must have been conceptualized by participants as alignable.

Structural alignment theory avoids this potential circularity by strictly adhering to its operational definitions. To this end, experimenters create stimuli for which alignable and nonalignable differences are clearly identifiable. It is for these stimuli that alignment theory makes unambiguous and testable predictions. In particular, Markman and Gentner (1996) espouse the use of visual stimuli because "picture pairs permit us to vary whether a difference is alignable or nonalignable, and thus lend themselves to separating the effects of alignable and nonalignable differences on subjects' judgments of similarity" (p. 238). Such was the approach of the present experiments. Namely, visual stimuli were created to be transparently analogous to Markman and Gentner's (1996) alignable and nonalignable stimuli, and the experiments tested predictions that were clearly stated by those researchers. Therefore, the obtained evidence cannot be attributed to faulty stimuli or infelicitous predictions.

Present Research

In the experiments reported below, we provided a direct test of alignment theory. In particular, we tested the claim that alignable differences are more important than nonalignable differences in similarity judgments. This claim is manifest in two predictions (Markman & Gentner, 1996), both of which are tested below. First, stimuli with an alignable difference should be judged less similar to a standard than should stimuli with a nonalignable difference. This prediction was tested in Experiment 1. Second, alignable differences ought to be more sensitive to variation than nonalignable differences. This prediction was tested in Experiment 2. Finally, in Experiment 3, we also tested whether contextual constraints affect the relative impact of alignable and nonalignable differences.

We used similarity judgments because this task is perhaps the most basic: Similarity is fundamental to a number of cognitive processes. We also used the simplest possible stimuli—geometric figures—to minimize the possibility of spurious semantic factors (e.g., the action of a scene) influencing the judgments. Hence, by using the simplest stimuli in the most basic task, we hoped to establish the generality of the result. If nonalignable differences were found to be more influential than alignable differences in the judgment of similarity, this result would have critical implications for alignment theory.

Experiment 1

Alignment theory predicts that alignable differences will be more important than nonalignable differences in similarity and difference judgments. As Markman and Gentner (1996) suggested, "The most straightforward test of this hypothesis is a forced-choice task that pits an item serving as an alignable difference against the same item serving as a nonalignable difference" (p. 242). In accordance with this decree, we elicited forced choice similarity judgments and difference judgments (between-participants) of geometric stimuli. The stimuli were modeled after previous investigations of alignment in similarity and difference judgments (e.g., Goldstone et al., 1991; Medin et al., 1990), and were presented as triads of one standard and two alternatives (e.g., Markman & Gentner, 1996; Medin et al., 1990).

According to structural alignment, stimuli are represented in terms of attributes, objects, and relations between those attributes and objects. For instance, consider a black square above a white circle. In this example, the attributes (e.g., black, white) describe the objects (e.g., square, circle), which serve as the arguments of a relation (e.g., above). To provide a thorough test of alignment theory, the experiments included alignable differences of all three types. These various alignable differences are illustrated in Figure 3. In comparison to the standard stimulus, the attribute stimulus constitutes an alignable difference because the white circle in the standard is aligned with, or corresponds to, the black circle in the attribute stimulus. The *relation* stimulus differs from the standard only in that the relation between the objects (i.e., above–below vs. left–right) has been changed. Thus, the structures of

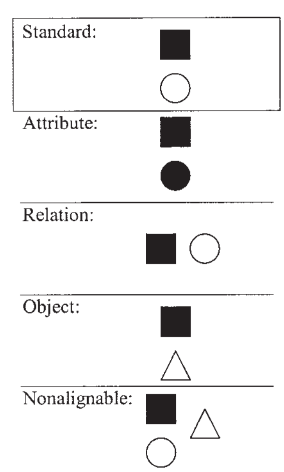

Figure 3. Examples of stimuli, Experiment 1. Standard = comparison standard; Attribute = alignably different attribute; Relation = alignably different relation; Object = alignably different object; Nonalignable = nonalignable difference.

those two stimuli are alignable, but have a different relation. In the case of the object stimulus, it is alignable with the standard, but it contains a different object—it has a triangle where the standard has a circle. Most critical, however, is the nonalignable stimulus. Here, an object (i.e., the triangle) that is present in the alternative stimulus has no correspondence in the standard stimulus.[3]

The critical trials occurred when the nonalignable alternative was presented with one of the three alignable alternatives (i.e., attribute, relation, or object). If alignable differences are more important than nonalignable differences in similarity and difference judgments, then stimuli with an alignable difference (i.e., attribute, relation, object) should be judged less similar to, and more different from, the standard than should a stimulus with a nonalignable difference (Gentner & Markman, 1997; Markman & Gentner, 1996).

Similarity judgments and difference judgments are generally thought to be inverses. That is, if 75% of participants judge *S* as more similar to *A* than to *B*, then approximately 75% of participants should choose *S* as more different from *B* than from *A*. However, Medin et al. (1990) provided evidence of an important exception to this generality (see also Gati & Tversky, 1984; Tversky, 1977). Medin et al. presented stimulus triads that consisted of a standard and two alternatives. As in the present experiment, participants were asked to choose the alternative more similar to the standard or the one more different from the standard. The two alternatives shared either a relation or an attribute with the standard. Medin et al. found that, relative to the attribute match, the relation match was chosen as both more similar to the standard and as more different from that same standard. Thus, similarity and difference choices were not inverses. Accordingly, in the present experiment, when an attribute alternative is presented with a relation alternative, similarity and difference judgments should reveal noninverse results. In all other alternative pairs, however, similarity and difference should be inverses of one another.

Method

Participants. Two hundred twenty-six Princeton University undergraduates voluntarily participated in Experiment 1.

Materials and procedure. Six standard stimuli were constructed. All standards consisted of two geometric shapes—one gray and one white—arranged in various orientations. For each of these six standard stimuli, two alternative stimuli were constructed. Thus, there were six pairs of alternative stimuli, one pair for each possible pairwise combination of the attribute, relation, object, and nonalignable stimuli (see the left column of Table 1). An example trial is presented in Figure 4. The six standards (and their alternatives) were presented in two different random orders in two different experimental lists. For example, the trial in Figure 4 appeared second in one list and fourth in the other list. The left–right order of the alternatives was also random. Two additional lists were created by reversing the order of those alternatives, such that each alternative appeared first in one list but appeared second in another list. Using the example in Figure 4 again, the nonalignable alternative appeared as the first (i.e., left) alternative in one list and as the second (i.e., right) alternative in another list. Thus, there were four experimental lists: 2 (standard orders) × 2 (left–right alternative orders). Participants were instructed to "please circle your answer for each of the 6 questions below." All six questions were presented on a single page. In the similarity condition, the questions were presented exactly as shown in Figure 4. The difference condition was identical, except that the phrase *similar to* was replaced with the phrase *different from.* Participants were randomly assigned to conditions, with an approximately equal number of participants in each condition.

Table 1
Proportions of Similarity and Difference Choices (Experiment 1)

Alternative	Judgment: Similarity ($n = 117$)	Judgment: Difference ($n = 109$)
Nonalignable & attribute	Attribute = .74[a]	Nonalignable = .72[a]
Nonalignable & relation	Relation = .69[a]	Nonalignable = .68[a]
Nonalignable & object	Object = .53	Nonalignable = .50
Attribute & relation[b]	Attribute = .74[a]	Relation = .57
Attribute & object	Attribute = .65[a]	Object = .70[a]
Relation & object	Relation = .55	Object = .58

[a] The proportion differs significantly from chance (i.e., $p < .004$, Bonferroni). [b] The similarity and difference proportions differ significantly from one another.

Results

Results are presented in Table 1. The table presents, for each pair of alternative stimuli (left column), the alternative that was judged more similar to the standard (middle column) and the alternative that was judged more different from the standard (right column). Next to each is the proportion of participants who chose the given alternative. The table reveals that the distinction between alignable and nonalignable differences is indeed one that has psychological significance. However, the table also shows that in contrast to the specific prediction of alignment theory, nonalignable differences were more influential than alignable differences in similarity and difference judgments. In no case was a nonalignable alternative judged more similar to or less different from the standard than the alignable alternative. This result is contrary to the prediction of alignment theory and contrary to the finding of Markman and Gentner (1996). Another result of interest is that, when the attribute and relation alternatives were compared, similarity and difference judgments were not inverses. This result replicates the previous finding by Medin et al. (1990) and therefore serves as validation of the materials.

Statistical analyses consisted of calculating the binomial probability for the outcomes of each of the six pairs of alternatives. This analysis determines the probability of obtaining the given proportion of choices by mere chance. To correct for multiple comparisons, the significance criterion was set at $p < .004$ (Bonferroni). Chi-square analyses were also used to test whether similarity and difference choices for each of the six pairs of alternative

[3] In what we refer to as a *relation* difference, both stimuli contain the same arguments (i.e., square, circle) but have a different relation (i.e., above–below, left–right). Such cases have been described elsewhere as an *attribute match* or as *attributionally similar* (Gentner & Gunn, 2001; Goldstone et al., 1991; Medin et al., 1990, 1993). In addition, note that what we refer to as an *object* difference (e.g., triangle vs. circle) might alternatively be considered an *attribute* difference in shape. However, in keeping with the distinction between attributes, objects, and relations, and for ease of exposition, we refer to our stimuli as *attribute, object,* and *relation* differences.

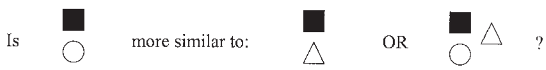

Figure 4. An example of the presentation of stimuli, Experiment 1. Presentation of stimuli in the difference condition was identical except that the phrase *similar to* was replaced by the phrase *different from.*

stimuli were inverses. Similarity and difference results are described separately and then conjointly below.

Similarity. The comparisons of most interest are those in which a nonalignable difference appeared as one of the alternative choices. When a nonalignable alternative appeared with an attribute alternative, participants reliably chose the attribute alternative as more similar to the standard ($p < .001$). Likewise, when a nonalignable alternative appeared with a relation alternative, participants reliably chose the relation alternative as more similar to the standard ($p < .001$). However, when a nonalignable alternative was presented with an object alternative, similarity judgments were at chance level ($p = .52$). The remaining comparisons indicated that not all alignable differences are equal (cf. Medin et al., 1990). Participants were reliably more likely to choose the attribute alternative than the relation alternative ($p < .001$) or the object alternative ($p = .002$). There was no reliable difference between the relation and the object alternatives ($p = .36$).

Difference. The pattern of difference judgments generally mirrored that obtained with similarity judgments. The nonalignable alternative was judged more different than the attribute alternative ($p < .001$) and the relation alternative ($p < .001$). Once more, when an object alternative appeared with a nonalignable alternative, choice was at chance level for difference judgments ($p = .50$), just as it was for similarity judgments. The only other reliable difference was that between the object and the attribute stimuli ($p < .001$).

Similarity and difference. Generally, similarity and difference were expected to be inverses. This inversion is apparent in Table 1. For instance, when the nonalignable alternative and the attribute alternative were presented together, the attribute alternative was chosen as more similar to the standard 74% of the time, and the nonalignable alternative was chosen as more different from the standard 72% of the time. However, a notable exception occurred in the choice between the attribute alternative and the relation alternative. In that comparison, similarity and difference judgments were not inverses (i.e., the difference between .74 and .57), $\chi^2(1, N = 226) = 6.90, p < .01$. This noninversion replicates the findings of Medin et al. (1990). Specifically, relations were more heavily weighted in similarity judgments than in difference judgments. Thus, in the similarity judgments, 74% of participants chose the attribute alternative as more similar to the standard. If difference were simply the inverse of similarity, then about 74% of participants should have chosen the relation alternative as more different from the standard. Instead, the relation alternative was chosen as more different only 57% of the time, not reliably more often than would be expected by random chance. Although this result failed to replicate the tendency for the same alternative to be chosen as both more similar to and more different from the standard (Medin et al., 1990), the present experiment nonetheless replicated the finding that similarity and difference judgments are not always inverses of one another.

Discussion

Structural alignment theory makes a novel distinction between alignable and nonalignable differences. The results of Experiment 1 supported the psychological validity of this distinction in that the status of a given difference as alignable or nonalignable affected its impact on participants' similarity choices. However, contrary to previous research, we found a reliable tendency for the nonalignable stimuli to be judged less similar to, and more different from, the standard. This greater salience of nonalignable differences directly contradicted the prediction of alignment theory, which asserts that alignable differences are more salient than nonalignable differences (Gentner & Markman, 1994, 1997; Markman & Gentner, 1996). Thus, the results generally supported alignment theory in its distinction between alignable and nonalignable differences but simultaneously contradicted its specific prediction that alignable differences will be more heavily weighted in similarity and difference judgments. Finally, Experiment 1 also corroborated the differential weighting of attributes and relations in similarity and difference judgments, as demonstrated by the noninversion effect of Medin et al. (1990).

Experiment 2

In Experiment 1, we found that, contrary to the prediction of alignment theory, nonalignable differences detracted from similarity (and added to difference) more than alignable differences. In Experiment 2, we tested an additional prediction of alignment theory. As described in the introduction, if alignable differences are more salient than nonalignable differences in similarity judgments, then similarity ratings should be affected more by variation in alignable differences than by variation in nonalignable differences. To test this, Markman and Gentner (1996, Experiment 2) presented stimuli that varied in terms of alignability (i.e., alignable vs. nonalignable difference) and variation (i.e., similar vs. different object). Recall the example of the robot fixing a car, a truck, or another robot (see Figure 2). By this methodology, pairs of stimuli are presented, and participants rate their similarity.

The present experiment used this similarity rating methodology, with geometric stimuli designed to be analogous to Markman and Gentner's (1996). For example, relative to the standard stimulus shown in Figure 5, the AS alternative includes an alignable object (i.e., an oval) that is similar to that in the standard (i.e., the circle). This is analogous to replacing a car with a truck in Markman and Gentner's stimulus set. Our AD alternative included an alignable object (i.e., the triangle) that was different from that in the standard (i.e., the circle), much like Markman and Gentner replaced the car with a robot. The NS alternative in Figure 5 differs from the standard in that the oval has been added as a nonalignable object. This is analogous to adding the truck next to the robot. Finally, our ND alternative consisted of the standard plus the triangle as a nonalignable object; this is just like adding another robot to the standard scene in Markman and Gentner (1996). Each of the alternative stimuli was presented with the standard from which it was derived, and participants rated the similarity of the two stimuli. Alignment theory makes two critical predictions: (a) alignable differences should be judged less similar to the standard than should nonalignable differences and (b) variation in alignable differences should have a greater effect than variation in nonalign-

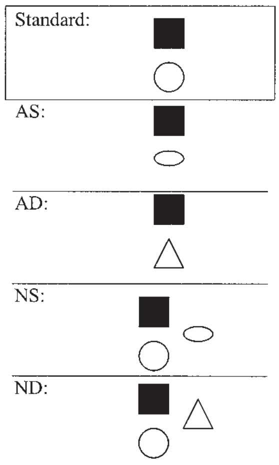

Figure 5. Examples of stimuli, Experiment 2. Standard = comparison standard; AS = alignable–similar; AD = alignable–different; NS = nonalignable–similar; ND = nonalignable–different.

able differences. Both of these predictions obtained in Markman and Gentner (1996).

Method

Participants. Eighty undergraduates at the University of Georgia participated as part of a course requirement.

Materials and procedure. Four standard stimuli were constructed. For each standard, four alternative stimuli were constructed: AS, AD, NS, and ND. One of the four stimulus sets is presented in Figure 5. The similar stimuli contained an object that differed only slightly from the standard, whereas the different stimuli contained an object that differed greatly from the standard. In addition, those similar and different objects served either as an alignable difference or as a nonalignable difference across stimuli.

Each standard was presented with each of its alternatives once. This created 16 experimental trials, four alternatives for each of the four standards. The order of the stimuli within each trial (i.e., standard–alternative or alternative–standard) was random. The order of trials was also random, with the constraint that stimuli from the same standard set did not occur on consecutive trials. Participants were instructed to circle a number on the scale indicating how similar the two objects are. The scale ranged from 1 (*highly dissimilar*) to 9 (*highly similar*), following Markman and Gentner (1996).

Results and Discussion

Mean similarity ratings are presented in Figure 6. Alignment theory predicts that (a) nonalignable differences will be judged more similar to the standard than will alignable differences and (b) alignable differences will be more sensitive to variation than nonalignable differences. Figure 6 reveals that the former prediction was not supported; rather, alignable differences were judged more similar to the standard. The latter prediction, however, was supported: Alignable differences were more sensitive to variation. Despite the fact that the present experiment was closely modeled after that of Markman and Gentner (1996, Experiment 2), the present results support only one of the two critical predictions of alignment theory.

Mean similarity ratings were analyzed via separate 2 (alignability: alignable, nonalignable) × 2 (variation: similar, different) analyses of variance (ANOVAs), one with participants as a random variable (F_p and t_p) and one with items random (F_i and t_i). Both factors were analyzed within-participants but between-items. To begin, the main effect of variation was significant, $F_p(1, 79) = 55.07$, $p < .001$, and $F_i(1, 12) = 28.96$, $p < .001$, such that stimuli designed to be more similar to the standard were in fact judged more similar to that standard. This result validates the experimental materials.

Critically, the main effect of alignability was also highly reliable, $F_p(1, 79) = 23.15$, $p < .001$, and $F_i(1, 12) = 32.64$, $p < .001$, but in the direction opposite to that predicted by alignment theory: Alignable differences ($M = 5.89$, $SE = .10$) were judged more similar to the standard than were nonalignable differences ($M = 5.25$, $SE = .10$). This result replicates the finding of Experiment 1 and is inconsistent with the claim that alignable differences count more against similarity. In no case was the nonalignable alternative rated more similar to the standard than was its alignable counterpart.

As shown in Figure 6, alignable differences were more sensitive to variation than were nonalignable differences. This interaction was significant, $F_p(1, 79) = 58.60$, $p < .001$, and $F_i(1, 12) = 25.67$, $p < .001$, thus replicating the result obtained by Markman and Gentner (1996). Specifically, the AS stimuli ($M = 6.47$, $SE = .13$) were judged significantly more similar to the standard than were the AD stimuli ($M = 5.31$, $SE = .13$), $t_p(79) = 8.75$, $p < .001$, and $t_i(6) = 6.50$, $p < .001$. However, the NS stimuli ($M = 5.28$, $SE = .14$) did not differ from the ND stimuli ($M = 5.23$, $SE = .15$), both $p > .60$. Thus, although alignment theory failed to predict the greater weighting of nonalignable differences (i.e., the main effect of alignability), it did correctly predict the greater sensitivity to variation of alignable differences (i.e., the interaction of alignability and variation).

Experiment 3

Markman and Gentner (1996) obtained an effect of alignability. Experiments 1 and 2 above, using stimuli analogous to theirs, also yielded an effect of alignability, but it was in the opposite direction. The question thus becomes how these disparate results may be understood together. A distinct possibility is that some as yet

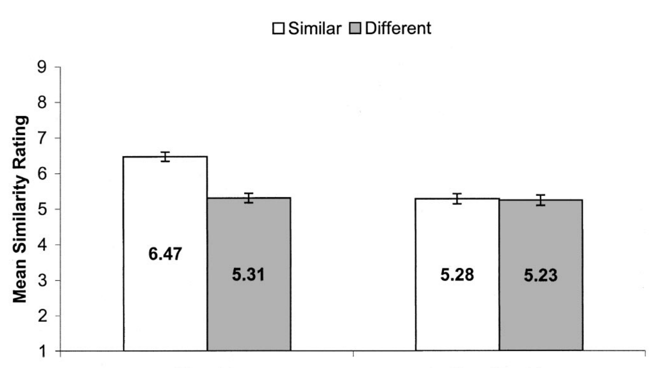

Figure 6. Mean similarity ratings, Experiment 2. Error bars represent one standard error of the mean.

untested variable, which differed between Markman and Gentner's (1996) experiments and our own, may explain the opposite results obtained across studies. In the introduction, we suggested that Markman and Gentner's stimuli might have induced participants to focus on the action of the scenes (e.g., a robot fixing a car). Critically, their alignable differences resulted in a change of the action (e.g., a robot fixing a truck), whereas their nonalignable differences did not change the action (e.g., a robot fixing a car with a truck in the background). Focusing on the action of the scene would therefore equate to focusing on the alignable difference thereby causing alignable differences to detract from similarity more than nonalignable differences. In our own stimuli, however, there was no explicit action or semantic information of any sort. In this case, nonalignable differences detracted from similarity more than alignable differences. This result suggests that, in the absence of semantic information, participants' default strategy is to weigh nonalignable differences more heavily than alignable differences. Thus, which type of difference (i.e., alignable or nonalignable) is more heavily weighted may be contingent on the current focus of the participants' attention. We tested this possibility in Experiment 3 by manipulating participants' relative focus on alignable or nonalignable differences.

In this experiment, an alignable condition was intended to increase the relative focus on the alignable differences, whereas a nonalignable condition was intended to focus participants on the nonalignable differences. The relative focus on the alignable and nonalignable differences was manipulated (between-participants) by explicitly emphasizing, in the experiment instructions, their respective occurrence. Thus, with the exception of this instructional manipulation, Experiment 3 was identical to the similarity condition of Experiment 1. By focusing on the nonalignable differences, the nonalignable condition should replicate the tendency to choose the alignable stimuli as more similar to the standard, as in Experiment 1. In contrast, by emphasizing the alignable differences, the alignable condition should increase the weight of those alignable differences and hence should diminish or possibly reverse the effect obtained in Experiment 1 (cf. Markman & Gentner, 1996).

This possibility that the relation between alignability and similarity is mediated by participants' focus is consistent with the intuition that similarity is quite sensitive to individual and contextual factors. Indeed, similarity is a notoriously vague theoretical construct (Goodman, 1972; see also Goldstone, 1994; Medin et al., 1993). In Experiments 1 and 2, for instance, different participants may have adopted different strategies for their similarity choices—after all, similarity choices were far from unanimous. Furthermore, the same participant may have adopted different strategies on different trials of the experiment. Thus, a potential criticism of such experiments is that they are not sufficiently constrained to infer how people compute similarity. By differentially constraining participants' similarity judgments, however, Experiment 3 addressed this criticism and may illuminate the effect of contextual constraints on similarity judgments.

Method

Participants. Seventy-eight undergraduates at the University of Georgia participated for partial course credit.

Materials and procedures. Materials and procedures were identical to those of the similarity condition of Experiment 1 with one exception. In the alignable condition, the instructions stated, "Please circle your answer for each of the 6 questions below. Each question has three figures that differ in various ways. For instance, some of the figures consist of different shapes and colors. Your task is to judge how similar the figures are." The nonalignable instructions were identical, except that the phrase *different shapes and colors* was replaced with *different numbers of objects.* This instructional manipulation was intended to emphasize either alignable or nonalignable differences, respectively. Participants were randomly and evenly assigned to conditions such that both conditions included 39 participants.

Results

Results are presented in Table 2. Binomial probabilities were calculated to determine whether participants exhibited nonrandom similarity choices, as in Experiment 1. To correct for the number of such binomial tests, the significance criterion was adjusted to $p < .004$ by the Bonferroni method.

Consider first the nonalignable condition, which yielded a pattern of results identical to that of Experiment 1. The attribute alternative ($p < .001$) and the relation alternative ($p < .004$) were both reliably more likely to be chosen than the nonalignable alternative. And again, the choice between the object alternative and the nonalignable alternative was random ($p = .34$). In the nonalignable condition, then, nonalignable differences tended to count more against similarity than alignable differences, just like in Experiment 1. The attribute alternative was more likely to be chosen than either the relation alternative or the object alternative (both $p < .001$), and the relation alternative was marginally more likely to be selected than the object alternative ($p = .06$).

Consider now the alignable condition. It is interesting to note that this condition yielded a pattern of results somewhat different from that of Experiment 1. In this condition, only the attribute alternative was chosen reliably more often than the nonalignable alternative ($p < .001$). The relation alternative was again more likely to be chosen than the nonalignable alternative, but this preference did not differ reliably from chance ($p = .11$). Moreover, unlike in Experiment 1, here in Experiment 3, the nonalignable alternative was chosen more frequently than the object alternative ($p = .01$), although this preference was only marginal when the Bonferroni correction is taken into account. The alignable condition thus provided the only instance, across three experi-

Table 2
Proportions of Similarity Choices by Condition (Experiment 3)

	Condition	
Alternative	Alignable ($n = 39$)	Nonalignable ($n = 39$)
Nonalignable & attribute	Attribute = .82[b]	Attribute = .92[b]
Nonalignable & relation	Relation = .64	Relation = .74[b]
Nonalignable & object	Nonalignable = .72[a]	Nonalignable = .59
Attribute & relation	Attribute = .85[b]	Attribute = .77[b]
Attribute & object	Attribute = .90[b]	Attribute = .92[b]
Relation & object	Relation = .72[a]	Relation = .67

[a] The proportion differs marginally from chance (i.e., $p \leq .01$). [b] The proportion differs significantly from chance (i.e., $p < .004$, Bonferroni).

ments, in which an alignable difference was more heavily weighted than the nonalignable difference in the similarity computation (cf. Markman & Gentner, 1996). Finally, as in Experiment 1, the attribute alternative was more likely to be chosen than either the relation alternative or the object alternative (both $p <$ .001). The relation alternative was marginally more likely to be selected than the object alternative ($p = .01$).

Table 2 shows a consistent trend for the alignable alternatives to be chosen more frequently in the nonalignable condition than in the alignable condition. That is, the attribute alternative was chosen over the nonalignable alternative 92% of the time in the nonalignable condition but only 82% of the time in the alignable condition. Similarly, the relation alternative was also chosen over the nonalignable alternative more frequently in the nonalignable condition (74%) than in the alignable condition (64%). The object alternative produced the same pattern of similarity judgments, being chosen over the nonalignable alternative more often in the nonalignable condition (41%) than in the alignable condition (28%). Thus, across all three critical comparisons, the alignable alternatives were apparently more heavily weighted in the alignable condition than in the nonalignable condition, and the nonalignable alternatives were more heavily weighted in the nonalignable condition than in the alignable condition. This trend suggests that the instructional manipulation did indeed tend to increase participants' weighting of either alignable or nonalignable differences, as expected. Although none of these individual comparisons reached significance (all ps $> .15$ by chi-square), a directional test collapsing across these three critical trials did indicate reliability. For the latter analysis, we simply counted for each participant the number of times (out of 3) that they chose the alignable option over the nonalignable option. The difference in choices between the alignable ($M = 1.74$, $SE = .15$) and nonalignable ($M = 2.08$, $SE = .13$) groups was significant in the predicted direction, $t(76) = 1.65$, $p = .05$ (one-tailed). That is, alignable differences were more heavily weighted in the alignable group than in the nonalignable group, and hence participants in the alignable group were less likely to choose the alignable option.

Discussion

The alignable and nonalignable conditions yielded different patterns of results: Alignable differences were more heavily weighted when they were emphasized in the instructions, whereas nonalignable differences were more heavily weighted when they were emphasized. In other words, the relative weight accorded a given difference was not predicted just by its status as an alignable or nonalignable difference, it was also affected by contextual constraints. The differential focus on alignable or nonalignable differences predicted the differential pattern of similarity ratings. Hence, it appears that the discrepancy between our results (Experiments 1 and 2 above) and those of Markman and Gentner (1996) may be attributable to a differential focus on alignable and nonalignable differences.

Results of the nonalignable condition resembled those of Experiment 1: The standard tended to be judged more similar to the alignable alternatives than to the nonalignable alternative. This result is apparent in the first two rows of Table 2—the attribute alternative and the relation alternative were both judged more similar to the standard than was the nonalignable alternative. In the alignable condition, contrarily, only the attribute alternative was more likely to be chosen than the nonalignable alternative. Notably, in this condition, the nonalignable alternative was chosen more frequently than the object alternative. This result was not observed in Experiment 1: The result replicates the finding of Markman and Gentner (1996) instead.

The finding that the nonalignable condition closely replicated Experiment 1 suggests that, as a default strategy, participants may have weighted nonalignable differences more heavily than alignable differences in Experiment 1. Together, then, the present experiments provide corroborative evidence suggesting that participants use a default strategy of weighting nonalignable differences more heavily than alignable differences in similarity judgments. Stated conversely, because results from the alignable condition were qualitatively different from those of Experiment 1, weighting alignable differences more heavily than nonalignable differences does not appear to be the default in similarity computations. It is apparent that it is only when alignable differences are somehow emphasized, whether via contextual constraints (i.e., the alignable condition) or semantics (Markman & Gentner, 1996), that those alignable differences are given more weight.

General Discussion

Structural alignment theory makes a unique and important theoretical distinction between alignable and nonalignable differences. All three experiments reported above support the psychological validity of this distinction. The novel contribution of the present experiments is the demonstration that nonalignable differences are clearly more influential in the judgment of similarity than has previously been supposed. In the similarity and difference judgments of Experiment 1, nonalignable differences were weighted at least as heavily as, and often more heavily than, alignable differences. Thus, Experiment 1 failed to support the prediction of alignment theory. In validation of those results, however, Experiment 1 did successfully replicate the noninversion effect of similarity and difference judgments (Medin et al., 1990). Experiment 2 corroborated the greater weight accorded nonalignable differences in similarity judgment (cf. Experiment 1) and simultaneously replicated the finding that alignable differences are more sensitive to variation than are nonalignable differences (Markman & Gentner, 1996). Finally, Experiment 3 demonstrated that the relative weighting of alignable and nonalignable differences is sensitive to contextual constraints.

The obtained data were discordant with an important prediction of structural alignment theory. According to Markman and Gentner (1996), "if the structural alignment view is correct, then alignable differences should count more against similarity than should nonalignable differences" (p. 238). In the present experiments, however, alignable differences did not count more against similarity than nonalignable differences. Following their logic then, structural alignment theory must be incorrect. Critically, however, this is not to suggest that structural alignment theory should be abandoned as a theory of similarity; failure to support the prediction of a theory is not grounds for rejecting its constructs altogether. Alignment theory appears to be correct in its distinction between alignable and nonalignable differences but incorrect in its specific prediction that alignable differences will be more salient than nonalignable differences in similarity judgment.

At face value, similarity judgments appear to be inherently unconstrained (Goodman, 1972). That is, it seems likely that participants employ idiosyncratic strategies in similarity judgment tasks. However, more generally, similarity judgments do exhibit some systematicity (Goldstone, 1994; Medin et al., 1993). In support of this systematicity, Goldstone (1994) identified three distinct constraints on similarity: (a) perceptual processes, (b) comparison processes, and (c) task factors. Because the perceptual processes evoked by our geometric stimuli were presumably relatively simple and were held constant across experimental conditions (particularly in Experiment 3), we do not discuss this factor of similarity. Of more relevance to the present research are Constraints B and C. Essentially, alignment theory emphasizes the comparison process as a primary determinant of similarity (Medin et al., 1993). Alignment theorists make the strong assertion that the alignability of a given difference will determine its influence on similarity, with alignable differences being more heavily weighted (Gentner & Markman, 1994, 1997; Markman & Gentner, 1996). The present research indicates, via simple empirical demonstrations, that this assertion is not supported. The greater weight predicted for alignable differences did not obtain in Experiments 1 and 2. Moreover, in Experiment 3, the constraint imposed by task factors affected similarity judgments. The relative weighting of a given difference was partially contingent on the instructional manipulation of participants' focus. Thus, the current experiments serve to underscore contextual constraints as an important determinant of similarity.

One may wonder whether the present results are merely a special, anomalous case. That is, perhaps alignment theory accounts for the vast majority of comparative cognitive processes, just not similarity and difference judgments of nonsemantic stimuli. Note first that alignment theory explicitly has been offered as an explanation of similarity and difference judgments of nonsemantic stimuli (e.g., Goldstone et al., 1991; Medin et al., 1990, 1993; see also Markman & Gentner, 1996). Moreover, nonalignable differences exert a greater influence than alignable differences in other tasks with semantically rich stimuli as well (e.g., Costello & Keane, 2001; Wisniewski & Middleton, 2002).

In the introduction we noted that the potential subjectivity of defining alignable and nonalignable differences brings with it the theoretical danger of circularity. For instance, given that the present results indicated that nonalignable differences were more heavily weighted than alignable differences, one is tempted to point to the data as evidence that our nonalignable differences must have actually been perceived by participants as alignable instead. (Indeed, we entertain this possibility below.) Of course this argument is circular: It is not theoretically permissible to use the obtained data to redefine which differences are alignable and which are nonalignable. Rather, operational definitions must be established a priori and upheld empirically. Nevertheless, that being said, there are a couple ways in which structural alignment theory might attempt to accommodate the present results post hoc. Both alternative accounts take the approach of reclassifying our nonalignable stimuli as alignable, as alluded to above. Notice that this reclassification of stimuli is tantamount to rejecting the operational definitions of structural alignment theory because our stimuli were constructed according to those definitions.

Alternative 1: Nonalignable Difference as Alignable Difference

One possibility is to argue that, although our stimuli satisfied the operational definition of a nonalignable difference, participants conceived of that nonalignable stimulus as an alignable difference in the number of objects present in the stimuli (see Markman & Gentner, 1996, p. 242–243). That is, whereas the standard consists of two objects (e.g., square and circle), the nonalignable stimulus consists of three objects (e.g., square, circle, and triangle). If participants did in fact conceive of the nonalignable stimuli as an alignable difference in the number of objects, then the present experiments would not constitute a test of alignable and nonalignable differences. However, as explained below, this account of our experiments is not supported by the present results.

Alternative 2: Nonalignable Difference as Alignable Absence

Structural alignment theory might alternatively accommodate the present results by positing that the nonalignable differences used in our experiments were in fact related to structural commonalities. To illustrate, alignment of the thorny stem of a rose with the thornless stem of a violet constitutes a nonalignable difference (i.e., the presence and absence of thorns). However, critically, such nonalignable differences appear to be related to commonalities (e.g., common stem structures). Thus, Wisniewski and Middleton (2002, p. 20) suggested that these differences should be considered alignable rather than nonalignable (see also Costello & Keane, 2001). The same argument may apply to the present study. After all, the standard and the nonalignable stimulus have structures that are easily aligned (see Figure 3). Once the square and circle are aligned, the triangle may be considered alignable, because it is related to the common square–circle structure of the two stimuli. One may refer to such differences as alignable absences, because they involve a feature that is present in one concept but absent from the other, and importantly, the concepts have alignable structures.[4] Thus, having and not having a triangle may be an alignable absence between the standard and the nonalignable stimuli in Figures 3 and 5.

Evaluation of Alternatives

The claim that participants interpreted our nonalignable stimuli as an alignable difference in the number of objects (i.e., Alternative 1) is not supported by the current data. Specifically, the results of Experiment 2 contradict this interpretation of our stimuli. Recall that the AS and AD stimuli exhibited a reliable effect of variation. If the NS and ND alternatives were also conceived as alignable, then they should have exhibited an effect of variation. However, they didn't: The alignable and nonalignable alternatives exhibited qualitatively different patterns of results (see Figure 6). Thus, there is no reason to believe that those stimuli are of the same type. Thus, we find it untenable, on empirical grounds, that our non-

[4] This terminology is not our own. We became acquainted with the phrase *alignable absence* at the 2001 23rd Annual Conference of the Cognitive Science Society in Edinburgh, Scotland. The source of the phrase, unfortunately, is unknown to us.

alignable differences were interpreted as just another variety of alignable difference (see also Markman & Gentner, 1996). More empirically tenable is the conjecture that our nonalignable differences were interpreted as alignable absences (i.e., Alternative 2). Assuming that they are qualitatively distinct from alignable differences, alignable absences need not produce the same pattern of results as the alignable differences.

The present experiments do not address the possibility of alignable absences. We believe alignable absences to be an interesting possibility that warrants more direct research. However, we emphasize that if alignable absences were accepted as the explanation of these results, two implications would follow: (a) Structural alignment theory currently offers inadequate operational definitions of stimuli and therefore (b) prior investigations of alignable and nonalignable differences are called into question. These implications follow from the fact that our nonalignable stimuli were modeled after those of previous studies and were explicitly constructed according to the operational definition of nonalignable differences (i.e., the former implication above). Moreover, if our nonalignable differences were interpreted as alignable absences, then the presumed nonalignable differences in prior studies may also be interpreted as alignable absences, and hence it is no longer clear whether those prior studies were indeed contrasting alignable and nonalignable differences as they claimed (i.e., the latter implication). Thus, the latter implication would potentially negate much prior empirical support for structural alignment theory, and the former implication would constitute a major revision of the theory.

Conclusions

Alignment theory has successfully explained an impressive range of cognitive phenomena, arguing that alignable differences are more salient than nonalignable differences. The contribution of the present investigation has been to demonstrate that nonalignable differences are also important, much more so than has been suggested by prior research or indeed by structural alignment theory itself. This research suggests a need for further investigation into the distinction between alignable and nonalignable differences, as well as the contextual constraints that affect the relative weighting of these differences.

References

Costello, F. J., & Keane, M. T. (2001). Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 255–271.

Davenport, J., & Keane, M. T. (1999). Similarity and structural alignment: You can have one without the other. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 132–137). Mahwah, NJ: Erlbaum.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16,* 341–370.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition, 29,* 565–577.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5,* 152–158.

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.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52,* 125–157.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology, 23,* 222–264.

Goodman, N. (1972). Seven strictures on similarity. In *Problems and projects* (pp. 437–446). New York: Bobbs-Merrill.

Lassaline, M. E., & Murphy, G. L. (1998). Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 144–160.

Markman, A. B. (1996). Structural alignment in similarity and difference judgments. *Psychonomic Bulletin & Review, 3,* 227–230.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24,* 235–249.

Markman, A. B., & Gentner D. (1997). The effects of alignability on memory. *Psychological Science, 8,* 363–367.

Markman, A. B., & Medin, D. L. (1995). Similarity and alignment in choice. *Organizational Behavior and Human Decision Processes, 63,* 117–130.

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 54–70.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science, 1,* 64–69.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

Medin, D. L., Goldstone, R. L., & Markman, A. B. (1995). Comparison and choice: Relations between similarity processing and decision processing. *Psychonomic Bulletin & Review, 2,* 1–19.

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

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review, 4,* 167–183.

Wisniewski, E. J., & Middleton, E. L. (2002). Of bucket bowls and coffee cup bowls: Spatial alignment in conceptual combination. *Journal of Memory and Language, 46,* 1–23.

Received March 12, 2003
Revision received January 29, 2004
Accepted February 11, 2004 ■