Published in final edited form as:
Lang Learn Dev. 2009 ; 5(3): 147–171. doi:10.1080/15475440902824079.

Effects of language and similarity on comparison processing

Susan A. Gelman,
University of Michigan

Lakshmi Raman, and
Oakland University

Dedre Gentner
Northwestern University

Abstract

What factors promote conceptual (deep) processing in young children? In this research we examine two factors that seem likely to invite a focus on important conceptual information. The first is comparison processing: comparisons (such as "cats are like dogs") involve a *structural alignment* that highlights common relational structure as well as differences connected to that structure. The second factor is the use of generic language (such as "cats have sharp teeth"), which invites a construal organized around information that is relatively central to the represented item. We ask whether these two forces can combine to foster deep processing in four-year-olds, as well as in adults. Our secondary goal is to test whether the process of comparison operates in the same way in preschool children as in adults. In two studies (N = 132), we examined preschool children's and adults' comparison processing, by asking participants to produce either commonalities or differences for pairs of items while varying similarity (high vs. low) and wording (generic vs. specific). As predicted, for both ages, (1) high-similarity pairs generated both more commonalities *and* more alignable differences than low-similarity pairs; (2) generic wording differed from specific language in relatively more deep properties for both ages; and (3) the combination of generic language and high similarity was especially favorable for producing deep properties. The detailed parallels between age groups suggest that the same comparison processes hold for children as for adults. Most importantly, the results show that two ways of highlighting deep conceptual structure—generic language and structural alignment—can be combined to provide a source of insight for both children and adults.

One of the fundamental ways of learning about the world is by means of comparison processes: making use of available information to form new insights by means of noting commonalities and differences. Examples include category learning and induction (Gentner & Namy, 1999; Inagaki & Hatano, 1987); spatial learning (Gattis, 2002; Loewenstein & Gentner, 2001), mathematical insight (Bassok & Holyoak, 1989; Dixon & Dohn, 2003; Mix, 1999; Richland, Zur, & Holyoak, 2007), and scientific reasoning (Chen & Klahr, 1999; Gentner & Gentner, 1983). Past research shows that there is developmental change in such comparisons –a relational shift whereby young children focus on object commonalities, often perceptual, whereas adults can focus on common relational structure. For example, when asked "How is a cloud like a sponge?", a 4-year-old answers, "Both are round and fluffy", while a 9-year-old says, "Both hold water and later give it back" (Gentner, 1988). Nonetheless, there is also reason to believe that children's attention can be directed so as to yield more fruitful comparisons.

Correspondence concerning this article should be addressed to: Susan Gelman, Department of Psychology, University of Michigan, 530 Church St., Ann Arbor MI 48109-1043, gelman@umich.edu..

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In this work we investigate two well-researched but distinct ways of highlighting deep conceptual structure—structural alignment processes and generic language–to our knowledge the first such investigation to be conducted. We ask whether these two factors act to promote deep conceptual processing in 4-year-olds and in adults, and whether their combination is especially effective. To do this, we conducted a detailed examination of the kinds of commonalities and differences produced by children and adults when given pairs varying in similarity (and alignability), phrased in either generic or specific language. The combination of these manipulations yields five predictions, two from each of the frameworks individually and one stemming from their interaction. We first lay out the predictions (summarized in Table 1) and then present two studies: the first testing the predictions for adults, and the second asking whether the same patterns hold for children.

This brings us to the second motivation for this research: namely, to gain a better understanding of children's early comparison processes. Although, as noted above, there is general agreement on the relational shift phenomenon in development, it is unclear to what extent children's gain in relational insight betokens a change in the *process* of comparison—perhaps due to increasing processing capacity (Halford, 1992), as opposed to gains in domain knowledge (Gentner, 1988; Gentner & Rattermann, 1991; Goswami & Brown, 1989; Rattermann & Gentner, 1998). One way to gain evidence on whether children engage in the same kinds of comparison processes as adults (albeit with different degrees of domain knowledge) is to compare similarity processes at a relatively detailed level. In the present work, we compare young children with adults in their production of commonalities and differences. To obtain a detailed profile, we vary two further factors: (a) the degree of similarity between the pairs and (b) whether participants are linguistically encouraged to focus on surface information or more conceptually central information. In this way we should be able to gain insight as to whether the focus on surface information in children stems from knowledge differences between children and adults or whether the pattern indicates a different kind of comparison process.

Structure-mapping theory

Structure-mapping theory (SMT) (Gentner, 1983) states that comparisons ("cats are like dogs") involve an alignment of relational structures. In this process, matching relations are placed in correspondence and objects and entities (which need not match identically) are placed in correspondence based on their roles in the aligned relational structure (see Forbus, Gentner & Law (1995) for details). The structural alignment process highlights connected systems of relations that link the surface perceptual properties (Gentner & Namy, 1999; Markman & Gentner, 1993a). In addition to highlighting common relational structure, a structural alignment highlights differences connected to that structure (*alignable differences*): e.g., "Cats *meow* but dogs *bark*" (Markman & Gentner, 1993b).

One prediction of structure-mapping theory is that high similarity will yield both more commonalities and more differences, particularly alignable differences (Gentner & Markman, 1994; Markman & Gentner, 1993b). This is because high-similarity pairs (e.g., cat/dog) have larger alignable structures than low-similarity pairs (e.g., cat/fork). The prediction for commonalities is obvious and widely shared (e.g., Tversky, 1977); but the prediction of more differences for high-similarity pairs is somewhat paradoxical, given that low-similarity pairs are more different from one another and so would seem to offer more possibilities for differences to be expressed. Nonetheless, the fact that high-similarity pairs have larger common structures, coupled with the greater ease of aligning high-similarity pairs, should make it easier to generate alignable differences for high-similarity pairs. Research has borne out these predictions for adults: people generate more alignable differences (Markman & Gentner, 1993b, 1996) and are better able to produce differences for high-similarity than for low-similarity pairs (Gentner & Gunn, 2001; Gentner & Markman, 1994). Here we test how this pattern interacts with generic language. More importantly, we test whether it also holds for children.

A related prediction concerns the nature of the features one would expect to see when examining high- versus low-similarity pairs. Features can be broadly classified as either *surface* (observable properties such as size, color, texture) or *deep* (properties such as behavior, function, and internal parts, which are causally connected to each other and to other deep features)[1]. Deep features are of particular interest in that they causally relate to a larger system of knowledge (Ahn, 1998; Jameson & Gentner, 2007; Lombrozo & Carey, 2006; Rehder & Hastie, 2001). According to structure-mapping, human comparison processing naturally tends to highlight interconnected relational structure (Gentner, 1983). A relation held in common between two items is more likely to be prominent in their comparison if it is connected to other relations held in common (e.g., Clement & Gentner, 1990). Thus structure-mapping theory predicts that aligning two items will encourage a focus on systems of interconnected properties —that is, on deeper features. Because the high-similarity pairs afford a fuller alignment (in that they share considerably more conceptual structure, as well as more surface similarity, than the low-similarity pairs), we expect adults to produce more deep commonalities and differences for high-similarity than for low-similarity pairs[2].

If children follow the same process as adults, then they will show the same pattern. The high-similarity pairs not only yield a fuller alignment but are easier to align; children find it easier to achieve a structural alignment when there are surface commonalities that support the alignment (Gentner & Medina, 1998; Gentner & Toupin, 1986). Thus deep features may be most available to children when they compare high-similarity pairs, paralleling the expected adult pattern. On the other hand, children could show a very different pattern from adults. The high-similarity pairs present many more obvious surface similarities and differences than the low-similarity pairs. If young children's processes are focused on surface similarity, then they may preferentially produce surface commonalities and differences for such pairs.

Generic language

Generic noun phrases (e.g., “lions” in the sentence “Lions are fierce”) are the most common means of using language to refer to kinds (lions as a group; Gelman, Hollander, Star, & Heyman, 2000). By inviting speakers to think about kinds, generics correspondingly focus on central conceptual knowledge (Prasada & Dillingham, 2006). Information stated generically is hypothesized to be more central to a category than information stated specifically. Whereas individual instance representations by their very nature include idiosyncratic as well as core properties, kinds capture primarily core properties that participate in interconnected causal structures that explain the surface properties. For example, an individual cat has both idiosyncratic properties (a particular color, size, age, etc.) and core properties (e.g., 4 legs, ability to hunt, internal bodily structure), but cats in general have in common only the core properties. Because generic noun phrases (e.g., “*Cats* climb trees”) are kind-referring, they should therefore highlight central aspects of a category (Gelman, 2003). Prior research has indeed demonstrated that by 4 years of age, children treat properties expressed generically as less variable than properties expressed with an indefinite quantifier (“some”), though more variable than properties expressed with a universal quantifier (“all”) (Hollander, Gelman, & Star, 2002). Moreover, 4- to 5-year-olds judge that predicating a property of a depicted novel animal using generic language (e.g., “Bants have stripes”), rather than non-generic language

[1]For purposes of strict scoring, we will operationalize *deep features* as those that are not directly observable from a perceptual examination of the pictures.

[2]Of course, adults do focus on deep commonalities when given a surface-dissimilar (but relationally similar) pair such as “staircase/ education” (Gentner, 1988), but in the present study, the low-similarity pairs were not relationally similar.

(e.g., “This bant has stripes”) implies a more kind-relevant connection between category and property (Hollander, Gelman, & Raman, in press). We thus predicted that items presented in generic form would prompt deeper alignments than items presented in specific form.

Present studies

We conducted two studies—examining adults’ (Experiment 1) and preschool children's (Experiment 2) comparison processing, by asking participants to produce either commonalities or differences for pairs of items. We varied two primary factors: similarity (high or low) and wording (generic or specific), both between-subject. The predictions, summarized in Table 1, are as follows. High similarity pairs should show more commonalities and alignable differences (Hypothesis-1a, Hypothesis-2a) and deeper commonalities and alignable differences (Hypothesis-1b, Hypothesis-2b) than low similarity pairs. Generic wording should lead to deeper commonalities and alignable differences (Hypothesis-3, Hypothesis-4) than specific wording.

Our most novel predictions concern how these two forces (similarity and generic language) will combine in reasoning. Structural alignment and generic language are predicted to operate in different but complementary ways. Generic wording should encourage a deeper representational construal of the two items. High similarity should allow for fuller alignment of the two representations. Thus, when a high-similarity pair is described using generic language the resulting structural alignment should show an especially strong focus on deep relational structure. Thus, we should see an interaction between similarity and language, such that combining the two factors (generics used for high-similarity pairs) will lead to the strongest focus on deep commonalities and differences (Hypothesis-5).

One question that might arise is why generic wording and high similarity are predicted to *interact* (vs. simply showing additive effects). Our reasoning is as follows. Generic wording and high similarity are hypothesized to operate at distinct levels: Generic language invites a coherent representation (with more interconnected relational information), whereas high similarity leads to greater structural alignment. Thus, if the members of the pair are represented in a coherent way (due to generic language) and they are strongly aligned (due to high similarity), we would predict a rich set of deep alignable differences to result. In contrast, if the pair is represented shallowly (due to specific language) *or* if the alignment is poor (due to low similarity), we are unlikely to obtain deep alignable differences. Therefore, the combination of both generic wording and high similarity is predicted to result in more than an additive combination of these individual factors.

As noted previously, a key question is whether children will show the same patterns as adults. Of course, due to their relative lack of knowledge about the conceptual underpinnings of the concepts in the study, children are expected to show a lower overall proportion of deep features than adults. But if children engage in the same kinds of comparison processes as adults, then they should show the same overall patterns of deep and shallow features in their commonalities and differences as adults.

This work is novel in three respects. First, we provide the first study, to our knowledge, of children's performance on a commonality- and difference-listing task[3]. As discussed above, despite considerable past research on analogy and similarity in children, there has been little examination of the commonalities and differences that children produce. This developmental focus should offer insights into the developmental course of comparison processing. Second,

[3]There is an unpublished paper by Tanya Feinstein (1999) on children's listing of commonalities and differences, although without the manipulation of generic versus specific language.

this is one of the first studies to examine the broader implications of generic language for conceptual processing, in either children or adults. Although generics are thought to be cross-linguistically universal (Carlson & Pelletier, 1995) and commonly expressed in child-directed speech (Gelman, Goetz, Sarnecka, & Flukes, 2008), to this point little is known regarding the cognitive implications of these expressions. Most prior work on generics has focused on the more immediate question of their semantic interpretation (Gelman & Raman, 2002;Hollander et al., 2002), leaving open the broader extent of the implications that generic language may have in everyday reasoning. Third, this project is the first to bring together analyses of similarity and generic language in examining children's cognitive processing. The current approach permits us to go beyond debates that pit differing cues (perceptual, conceptual, linguistic) against one another, instead fostering an examination of how multiple cues work together in development.

STUDY 1: COMPARISON PROCESSING IN ADULTS

Study 1 examines our key predictions in a sample of adults. This study provides a foundation for Study 2, which addresses these ideas in children, our central interest.

Method

Participants—67 college students (mean age 18,5; range 18,0 to 21,6; 36 females and 31 males) participated. They were randomly assigned to four conditions: generic-commonalities (*N* = 17), generic-differences (*N* = 16), specific-commonalities (*N* = 17), and specific-differences (*N* = 17). In addition, a separate group of 10 college students provided similarity ratings (see “Deconfounding perceptual and conceptual similarity” in Results, below).

Materials—Thirty-two items were used in the experiment, including 16 animals and 16 artifacts. We included a realistic color drawing of each, depicting the target item placed against a white background. Materials were scanned into a computer, printed out in color, and laminated on cards. Items were paired to include both high-similarity pairs (e.g., dog/cat, spoon/fork) and low-similarity pairs (e.g., dog/spoon, cat/fork). High-similarity pairs consisted of items from the same superordinate-level category (animals, utensils, furniture, etc.). Half the high-similarity pairs were animals and half were artifacts. Low-similarity pairs consisted of the same items as in the high-similarity pairs, but rearranged such that each pair consisted of an animal and an artifact. For each participant, each item appeared in either a high-similarity pair or a low-similarity pair, but not both (i.e., none of the items were re-used for a given participant). Across participants, each item appeared in a high-similarity pair for half the participants, and in a low-similarity pair for the remaining participants. A full list of items is provided in Tables 2 and 3.

Procedure—Participants were randomly assigned either to the generic noun condition or to the specific noun condition. First participants received a brief introduction (“We are going to play a picture game in which I am going to ask you to tell me how some things are (the same) (different). Are you ready?”). For the remainder of the session, half the participants heard generic noun phrases (“I want you to think about dogs and cats. Can you tell me some things that are (the same) (different) about dogs and cats?”); and half heard specific noun phrases (“I want you to think about this dog and this cat. Can you tell me some things that are (the same) (different) about this dog and this cat?”).

Before receiving the test pairs, each participant first received a warm-up in which they saw four pairs of items with obvious commonalities and differences (2 circles, one wearing a hat; 2 squares, one with stripes and one without stripes; 2 hearts, one big and one small; 2 triangles, one wearing glasses), and were asked to provide commonalities (in the Commonalities condition) or Differences (in the Differences condition). Any errors on the warm-up were corrected.

Each participant then received 8 high-similarity and 8 low-similarity test pairs. The order of pairs was randomized with the constraint that no more than two item-pairs in a row could be of either low- or high-similarity.

Coding—All responses were transcribed and separated into distinct properties (e.g., “they both have ears, feet, and tail” was separated into “they both have ears”, “[they both have] feet”, and “[they both have] tail”). Each property was then coded according to the following system.

Responses were coded as: *commonalities* (“they both have four legs”); *alignable differences* (“this dog has a short tail, this cat has a long tail”); or *other* (“ducks have wings”).[4] Note that commonalities and alignable differences can each involve either a single property or a pair of properties: e.g., for commonalities, “Both are jewelry” or “Snake is cold blooded, and so is the refrigerator”; for alignable differences, “They are different shapes” or “A car can beep and a duck can quack.” If two properties combined formed a commonality or an alignable difference, they received just a single (combined) code. “Other” responses included non-alignable differences (“pencils don't have tails”), single-item descriptions (“a dog has ears on it”), thematic links (“a camel wears a necklace”), and irrelevant responses (“I don't know”; “nothing”). Irrelevant responses (constituting only 3.3% of adults’ responses) were dropped from all analyses and not considered further.

Commonalities and alignable differences were further subdivided into *surface* (perceptual properties such as “the dog is little, the spoon is big”; “they are both brown”), *deep* (actions, functions, traits, mental states, and taxonomic judgments such as “snakes slither, rabbits run fast”; “they are both animals”), and *other* (i.e., neither clearly perceptible nor clearly non-obvious; this includes location, substance, composition, or other; e.g., “dogs are outside and spoons are inside”). “Surface” features were defined as those that could be viewed directly in the pictures; “deep” features were defined as those that could not be viewed directly, and included inferred perceptual outcomes (e.g., type of movement). See Tables 4 and 5 for further coding examples.

Two coders each coded all responses independently then resolved disagreements by discussion. Cohen's kappas were .94 (commonalities, differences, other) and .93 (surface, deep, other), indicating high inter-rater agreement.

Results

We conducted three sets of analyses: (a) frequency of response types, to provide an overview of the data and validate the tasks, (b) percentage of properties that were commonalities (for the commonalities task), and (c) percentage of properties that were alignable differences (for the differences task). Here we examined three basic questions: are adults sensitive to similarity in their comparison processing, are they sensitive to generic language, and do similarity and generic language combine in their effects?

Frequency of Response Types—We first wished to establish the frequency of commonalities (on the commonalities task) and differences (on the difference task), to verify that participants were appropriately following directions. Results are shown in Table 6 (with “other” responses broken down into subtypes, just for informational purposes, although they are not subdivided in the analyses). We conducted an ANOVA involving task (commonalities,

[4]Examples provided include both child and adult responses.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

differences) and response type (commonalities, alignable differences, other). All main effects and interactions were significant, all *p*s < .05. Post-hoc tests using Bonferroni's correction reveal that commonalities were more common on the Commonalities task than on the Differences task (*M*s = 49.06 and 1.94), $p < .001$. In contrast, Alignable Differences and Other responses were more common on the Differences task than on the Commonalities task (*M*s = 82.23 and 4.79), *p*s < .001. Clearly adults were highly sensitive to the task instructions.

Commonalities—This set of analyses focuses on commonalities expressed on the commonalities task, separately for surface and deep properties (as these scores were not wholly independent of one another). For each property type (surface and deep), we conducted a 2-way ANOVA including wording (generic, specific) and similarity (high, low) as factors, with wording as a between-subjects factor and similarity as a within-subject factor. The dependent variable was the percentage of properties or property-pairs (excluding irrelevant properties) that expressed a commonality. Because the data are percentages, they were transformed with an arcsine transformation before undergoing analysis (the arcsine of the square root of the value). The results of the deep commonalities are presented first, as they involve the most direct predictions.

Deep commonalities: As predicted by H_1, high-similarity pairs (e.g., cat/dog) yielded more deep commonalities than low-similarity pairs (e.g., cat/fork) (see Figure 1), $F(1, 32) = 11.01$, $p = .005$. Moreover, consistent with H_3, generic wording yielded more deep commonalities than did specific wording, $F(1,32) = 21.47$, $p < .001$. Finally, as predicted by H_5, there was a similarity × wording interaction, $F(1,32) = 4.01$, $p = .054$, revealing that the combination of generic wording and high similarity yielded particularly high levels of deep commonalities. We conducted planned t-test comparisons to test whether the combination of generic wording and high similarity yielded higher scores than in the other three conditions. We discovered that, consistent with H_5, deep commonalities were most evident in the generic/high-similarity condition and least evident in the specific/low-similarity condition, $p < .01$. Furthermore, deep commonalities were higher in the generic/high-similarity condition than in the specific/high-similarity condition, $p < .001$, and were lower in the specific/low-similarity condition than in each of the other three conditions, all *p*s < .001. However, there was no significant difference between the generic/high-similarity condition and the generic/low-similarity condition.

Surface commonalities: The only significant effect for surface commonalities was a main effect of wording, $F(1,32) = 13.68$, $p = .001$, indicating that participants provided more surface properties when hearing specific wording than generic wording. Note that this result is the opposite of that obtained with deep properties (which were particularly frequent when hearing generic wording). Overall, the results reveal wholly different patterns for deep versus surface commonalities, indicating that the benefits of generic wording and similarity emerge for deep commonalities only. For example, high-similarity pairs did not exceed low-similarity pairs in the number of surface properties produced (*M*s = 46% and 50%; see Figure 1). This effect, though predicted, is striking, given that surface commonalities should be particularly easy to generate on this task, in which detailed pictures were present. Instead, the effect of similarity was focused on the non-obvious properties.

Fine-grained analysis of deep properties: In the preceding analyses, we defined "deep" properties as those that are not observable in the pictures that were presented to participants. However, even among the deep properties, some are potentially visible (although not in the pictures), such as "run fast", whereas others are wholly inferred and nonvisible (such as "are mammals"). In order to determine whether the effects hold for both kinds of "deep" properties, we conducted a more fine-grained analysis of the deep properties, comparing *actions* to *traits, functions, mental states, and taxonomic judgments* (referred to collectively as "nonvisible").

We conducted a property type (action vs. nonvisible) × similarity × wording ANOVA, and report only those effects involving property type. Results indicate a main effect of property type, $F(1,32) = 10.04$, $p < .005$, indicating that nonvisible deep properties were produced more often than action properties. There was also a property-type × similarity interaction, $F(1,32) = 26.48$, $p < .001$, indicating that similarity effects held for nonvisible deep properties only (greater among high-similarity items than low-similarity items, $p < .001$), with no significant similarity effects for action properties. There were no other significant effects involving property type, which means that the effects of wording were unaffected by whether the deep properties were of the action type or of the nonvisible type.

Deconfounding perceptual and conceptual similarity: In all the analyses above, the high-similarity parings consist of two exemplars from the same superordinate category, whereas the low-similarity pairings consist of two exemplars from different superordinate categories. This raises the question of whether the effects of similarity (noted above) are due chiefly to the inclusion of high-perceptual-similarity items. To discover whether the findings would be changed if such items were excluded, we conducted a further analysis of the items. First, we collected adult similarity ratings by a separate group of undergraduates ($N = 10$). For each of the 16 pairs of high-similarity items, each participant was asked to rate "how perceptually similar these pairs are -- that is, how similar they look to you" on a scale of 1 ("very different in appearance") to 7 ("highly similar in appearance"). Based on these ratings, the items were divided into two groups, items that were relatively lower in similarity (ranging from 1.2 to 2.8, $M = 1.99$) and those that were relatively higher in similarity (ranging from 2.9 to 5.1, $M = 3.69$), which significantly differ from one another, $t(9) = 7.96$, $p < .001$. The items with relatively higher similarity ratings are indicated with asterisks in Table 2. We refer to the high-similarity pairs with relatively lower ratings as low-perceptual-similarity/high-overall-similarity (or low-PS/high-OS, for short). These were then compared to the low-similarity pairs (which are uniformly low in perceptual similarity as well as in conceptual similarity; we refer to these as low-OS pairs, for short).

Next we conducted two t-tests, comparing the low-PS/high-OS items (e.g., camel-whale) to the low-OS items (e.g., duck-car), on two dependent measures: the percentage of surface properties and the percentage of deep properties. These analyses revealed significant similarity effects for both comparisons. For both surface and deep properties, low-PS/high-OS items yielded more commonalities than low-PS/low-OS items, ps < .001. Thus the same patterns hold even when high-perceptual similarity pairs are excluded.

Alignable Differences—This set of analyses focuses on alignable differences expressed in the differences task, separately for surface and deep properties. For each property type (surface and deep), we conducted a 2-way ANOVA including wording (generic, specific) and similarity (high, low) as factors, with wording as a between-subjects factor and similarity as a within-subject factor. The dependent variable was the percentage of properties or property-pairs (excluding irrelevant properties) that expressed an alignable difference. Because the data are percentages, they were transformed with an arcsine transformation before undergoing analysis (the arcsine of the square root of the value). The results of the deep alignable differences are presented first, as they involve the most direct predictions.

Deep alignable differences: As predicted by H_2, high-similarity pairs (e.g., cat/dog) yielded more deep alignable differences than low-similarity pairs (e.g., cat/fork) (see Figure 2), $F(1, 31) = 36.10$, $p < .001$. Recall that this result, although predicted, is counter-intuitive, and supports the structure-mapping theory of similarity. Turning to the effects of wording, we found that as predicted by H_4, deep alignable differences were more common in the Generic wording condition than in the Specific wording condition (Ms = 20% and 11%), $F(1, 31) = 7.81$, $p < .01$. This pattern is predicted by Gelman's (2004) theory of the effects of generic language.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Finally, we also found evidence that combining generic language with high alignability was particularly favorable for the highlighting of deep alignable differences, as predicted by H_5, and as shown by a similarity × wording interaction, $F(1,31) = 11.86$, $p < .005$ That is, the beneficial effect of generics on deep alignable differences held for the high-similarity pairs only (*M*s = 31% and 14%), $p < .001$; deep alignable differences showed no wording effect among low-similarity pairs (*M*s = 9% for each).

Surface alignable differences: As with the deep alignable differences, and as predicted by H_2, high-similarity pairs (e.g., cat/dog) yielded more surface alignable differences than low-similarity pairs (e.g., cat/fork) (see Figure 2), $F(1, 31) = 61.07$, $p < .001$. Once again, that high similarity favors the noticing of differences is a counter-intuitive result that supports the structure-mapping theory of similarity. Turning to the effects of wording, we found a sharp contrast to the effects with deep alignable differences; namely, surface alignable differences were more common in the Specific wording condition than in the Generic wording condition, $F(1, 31) = 9.13$, $p < .01$. Thus, the benefits of generic wording emerge for deep alignable differences only. Overall, the results reveal (1) that high similarity favors both deep and surface differences; and (2) that generic language favors deep alignable differences.

Fine-grained analysis of deep properties: As with the analysis of commonalities earlier, we again examined the effects of whether the “deep” properties were *actions* (and thus potentially visible, though not in the pictures) or *nonvisible* in a more principled way. We conducted a property type (action vs. nonvisible) × similarity × wording interaction. There was a main effect of property type, $F(1,31) = 24.07$, $p < .001$, indicating that nonvisible properties were produced more frequently than action properties. There were no other significant effects involving property type. Thus, the effects of wording and of similarity were unaffected by whether the deep properties were of the action type or of the nonvisible type.

Deconfounding perceptual and conceptual similarity: As with the analyses of commonalities above, we again examined the effects of varying overall similarity when controlling for perceptual similarity. We conducted two t-tests, comparing the low-PS/high-OS items to the low-OS items on two dependent measures: the percentage of surface properties and the percentage of deep properties. These analyses revealed significant similarity effects for both comparisons. For both surface and deep properties, low-PS/high-OS items yielded more alignable differences than low-PS/low-OS items, *p*s < .001.

Discussion

When adults receive a task that requires them to make comparisons (noting either commonalities or differences), they are highly sensitive to both item similarity and generic wording. Consistent with prior research (Gentner & Markman, 1994; Markman & Gentner 1993b, 1996), highly similar pairs (e.g., cat/dog) yielded more commonalities and more alignable differences across the board than highly dissimilar pairs (e.g., cat/fork). Also as predicted, generic wording (e.g., “cats”/“dogs”) yielded more deep commonalities and alignable differences—properties of the sort that cannot simply be “read off” the pictures—as compared to specific wording (e.g., “this cat”/“this dog”). Of particular interest was that the wording effect seemed boosted when combined with high-similarity items. Thus, the two factors of wording and similarity seemed to interact in adults’ judgments. We suggest, however, that the two factors operated in different ways. Generic wording was more likely than specific wording to encourage a deeper construal of the two items, and high-similarity allowed for aligning the two representations and deriving common structure. Thus the combination gave rise both to deep commonalities and to deep alignable differences.

The wording effects are striking, given the subtlety of the manipulation. From a logical perspective, all the properties that participants generated in the generic wording condition could also have been generated in the specific wording condition (e.g., the particular cat and dog are both animals, just as “cats” and “dogs” [generically] are both animals). Nonetheless, participants apparently adopted a different stance, depending on the manner in which the comparisons were framed.

The primary similarity manipulation involved a mixture of conceptual and perceptual similarity, in that the high-similarity items were pairs drawn from the same superordinate category, and the low-similarity items were pairs drawn from different ontological categories. However, even when the items that were highest in perceptual similarity were removed, we still obtain a strong effect of overall similarity.

STUDY 2: COMPARISON PROCESSING IN CHILDREN

Study 1 provided clear evidence for effects of both similarity and generic wording on comparison processing in adults. The purpose of Study 2 was to examine these effects in preschool children. This question is of interest for three primary reasons. First, given children's difficulty generating and appreciating appropriate comparisons on analogy tasks (Gentner & Toupin, 1986; Paik & Mix, 2006; Rattermann & Gentner, 1998; Richland, Morrison & Holyoak, 2006), it is of both theoretical and practical interest to determine whether children can be guided to make more meaningful comparisons under certain scaffolding conditions. Second, aside from one unpublished paper (Feinstein, 1999), the effects of similarity on children's generation of commonalties and differences have never been directly tested. The present study thus provides one of the first such tests. Third, although prior work has demonstrated that young children appropriately understand and produce generic nouns (Gelman & Raman, 2003; Hollander et al., 2002; Gelman et al., 2008), and that generics benefit inductive reasoning (Gelman, Star, & Flukes, 2002), to date no work has examined whether generics benefit children's reasoning more broadly. Thus, the present study permits a test of the effects of generics on an important aspect of human reasoning. We focused on 4-year-olds, as in our view this is the youngest age that could handle the high verbal demands of the task. Furthermore, prior research indicated that children 4-5 years of age distinguish generics from other quantifiers and can carry out a property-listing task (Hollander et al., 2002).

Method

Participants—66 four-year-olds (mean age 4,8; range 4,0 to 5,10; 35 boys and 31 girls) participated. Children were recruited from university and local area preschools in a midwestern U.S. community. They were randomly assigned to each of the four conditions: generic-commonalities (N = 17), generic-differences (N = 16), specific-commonalities (N = 17), and specific-differences (N = 16).

Materials, Procedure, Coding—Materials, procedure, and coding were identical to that of Study 1. As with the adults, irrelevant responses were rare, constituting only 6.7% of children's responses, and dropped from all analyses.

Results

We conducted three sets of analyses: (a) frequency of response types, to provide an overview of the data and validate the tasks, (b) percentage of properties that were alignable differences (for the differences task), and (c) percentage of properties that were commonalities (for the commonalities task). Our question is whether the hypotheses laid out in Table 1, which fit the adult pattern quite well, will also apply to children's processing. Most importantly, we ask whether children, like adults, would produce more deep commonalities and differences when given high-similarity pairs, generic language, or (especially) both of these.

Frequency of Response Types—As with the adults, we first examined the frequency of commonalities (on the commonalities task) and differences (on the difference task). This provides a check that participants were appropriately following directions and understood the task. Results are shown in Table 6 (with "other" responses broken down into subtypes, just for informational purposes, although they are not subdivided in the analyses). We conducted an ANOVA involving task (commonalities, differences) and response type (commonalities, alignable differences, other). All main effects and interactions were significant, all *p*s < .05. Post-hoc tests using Bonferroni's correction confirm that commonalities were more common on the Commonalities task than on the Differences task (*M*s = 22.88 and 1.19), *p* < .001. In contrast, Alignable Differences and Other responses were more common on the Differences task than on the Commonalities task (*M*s = 42.30 and 9.85), *p*s < .001. Clearly children--like adults in Study 1--were highly sensitive to the task instructions. These results license further examination of commonalities and alignable differences.

Commonalities—This set of analyses focuses on commonalities expressed on the commonalities task, separately for surface and deep properties (as these scores were not wholly independent of one another). For each property type (surface and deep), we conducted a 2-way ANOVA including wording (generic, specific) and similarity (high, low) as factors, with wording as a between-subjects factor and similarity as a within-subject factor. The dependent variable was the percentage of properties or property-pairs (excluding irrelevant properties) that expressed a commonality. Because the data are percentages, they were transformed with an arcsine transformation before undergoing analysis (the arcsine of the square root of the value). The results of the deep commonalities are presented first, as they involve the most direct predictions.

Deep commonalities: As predicted by H_{1a} and H_{1b}, high-similarity pairs (e.g., cat/dog) yielded more deep commonalities than low-similarity pairs (e.g., cat/fork) (see Figure 3), $F(1, 32) = 9.12$, $p < .005$. Although there was no significant main effect or interaction involving wording, Figure 3 shows that deep commonalities were more frequently produced when children were provided with the combination of generic wording and high similarity as compared to the other three conditions. We found that this predicted optimal condition was significantly higher than the generic/low-similarity condition ($p < .02$) and the specific/low-similarity condition ($p < .01$). (Because these were planned comparisons, we did not conduct post-hoc corrections to these tests.) None of the other conditions were significantly different from one another.

Surface commonalities: The only significant effect for surface commonalities was a main effect of similarity, $F(1,32) = 6.20$, $p < .02$, indicating that participants provided more surface properties for high- than low-similarity pairs.

Fine-grained analysis of deep properties: As with the analyses with adults, we again examined the effects of whether the deep properties were *actions* (and thus potentially visible, though not in the pictures) or *nonvisible* in a more principled way. We conducted a property type (action vs. nonvisible) × similarity × wording interaction. This analysis yielded no significant effects of property type. Thus, the effects of wording and similarity were unaffected by whether the deep property involved action or was wholly nonvisible.

Deconfounding perceptual and conceptual similarity: As with the analyses of the adult data, we again examined the effects of varying overall similarity within the low-perceptual similarity (low-PS) items. We conducted two t-tests, comparing low-PS/high-OS items to low-OS items on two dependent measures: the percentage of surface properties and the percentage of deep properties. These analyses revealed significant similarity effects for both comparisons. For both surface and deep properties, low-PS/high-OS items yielded more commonalities than low-PS/low-OS items, *p*s < .01.

Alignable Differences—This set of analyses focuses on alignable differences expressed in the differences task, separately for surface and deep properties. For each property type (surface and deep), we conducted a 2-way ANOVA including wording (generic, specific) and similarity (high, low) as factors, with wording as a between-subjects factor and similarity as a within-subject factor. The dependent variable was the percentage of properties or property-pairs (excluding irrelevant properties) that expressed an alignable difference, transformed with an arcsine transformation. The results of the deep alignable differences are presented first, as they involve the most direct predictions.

Figure 4 shows the pattern of responses.

Deep alignable differences: As predicted by H_2 and H_{2b}, high-similarity pairs (e.g., cat/dog) yielded more deep alignable differences than low-similarity pairs (e.g., cat/fork) (see Figure 4), $F(1, 30) = 8.83$, $p < .001$. Recall that this result--that high-similarity pairs yield more (deep) alignable differences—although predicted, is counter-intuitive. As predicted by H_4 and shown in Figure 4, deep alignable differences were more common in the Generic wording condition than in the Specific wording condition (*M*s = 11% and 5%), $p < .05$. Moreover, similarity interacted with wording in a significant 2-way interaction, F(1,30) = 11.80, p < .005. As predicted by H_5, the generic advantage for deep properties was significant only for the high-similarity pairs (*M*s = 16% and 4%), $p < .01$, Bonferroni's, and not the low-similarity pairs (*M*s = 6% and 5%). It therefore appears that deep alignable differences are most often produced in the context of high similarity and generic wording.

Surface alignable differences: The only significant effect for surface alignable differences was a main effect of similarity, F(1,30) = 32.78, p < .001. As predicted by H_2, high-similarity pairs (e.g., cat/dog) yielded more deep alignable differences than low-similarity pairs (e.g., cat/fork) (see Figure 4). In contrast to the results with the deep alignable differences, there was a non-significant tendency for surface alignable differences to be more common in the Specific wording condition than in the Generic wording condition (20% and 12%).

Fine-grained analysis of deep properties: As with the analysis of commonalities above, we again examined the effects of whether the “deep” properties were *actions* (and thus potentially visible, though not in the pictures) or *nonvisible* in a more principled way. We conducted a property type (action vs. nonvisible) × similarity × wording interaction. There was a main effect of property type, $F(1,30) = 6.47$, $p < .02$, indicating that children produced more action properties than nonvisible properties. Furthermore, there was a property type × wording interaction, $F(1,30) = 4.20$, $p < .05$. This result indicated that the wording effect was significant only for the actions (generic wording led to relatively more actions than specific wording, $p < .01$) and not for nonvisible properties. In contrast, the effects of similarity were unaffected by whether the deep property was an action or was wholly nonvisible.

Deconfounding perceptual and conceptual similarity: As with the analyses of commonalities above, we again examined the effects of varying overall similarity within the low-perceptual similarity (low-PS) items. We conducted two t-tests, comparing low-PS/high-OS items to low-OS items on two dependent measures: the percentage of surface properties and the percentage of deep properties. These analyses revealed significant similarity effects for both comparisons. For both surface and deep properties, low-PS/high-OS items yielded more commonalities than low-PS/low-OS items; surface properties: $p < .001$; deep properties: $p < .05$, one-tailed.

Comparing Children and Adults—In order to compare children's performance with that of adults, we conducted a further set of analyses, separately for the Commonalities and Differences tasks, that included age as a between-subjects factor. We conducted two 4-way ANOVAs including age (adult, preschool), wording (generic, specific), similarity (high, low), and property (surface, deep). The dependent variable was the percentage of properties or property-pairs (excluding irrelevant properties) that expressed either a commonality (for the Commonalities analysis) or an alignable difference (for the Alignable Differences analysis), transformed by an arcsine transformation (the arcsine of the square root of the value) to deal with percentages. Only those results involving age are reported below, as all other results are redundant with analyses reported previously.

Commonalities: As expected, college students provided commonalities at a higher rate than preschoolers (*M*s = 81% and 61%), $F(1,64) = 19.33$, $p < .001$. We also found a larger similarity effect for children than adults, as reflected in a similarity × age group interaction, $F(1,64) = 13.08$, $p = .001$. For children, high-similarity items consistently elicit more commonalities than low-similarity items (for specific surface properties, $p < .001$; for specific deep properties, $p = .078$; for generic deep properties, $p < .01$; Bonferroni's), whereas for adults, high-similarity items elicit more commonalities only for specific deep properties ($p = .001$), as reflected in a similarity × property × wording × age interaction, $F(1,64) = 4.19$, $p < .05$. Finally, we obtained a property × wording × age interaction, $F(1,64) = 5.58$, $p < .05$, indicating that children consistently provide more surface than deep properties, whereas adults do so in the specific wording condition only.

Alignable Differences: As expected, college students provided alignable differences at a higher rate than preschoolers (*M*s = 43% and 24%), $F(1,61) = 33.43$, $p < .001$. The only other significant effect involving age was a similarity × age interaction, $F(1,61) = 5.73$, $p = .02$, indicating a somewhat greater age effect for high-similarity than low-similarity items (although the age effect was significant at both levels of similarity). Thus, overall the alignable differences data were remarkably similar for children as for adults.

Discussion

The present findings demonstrate that preschool children, like college students, are sensitive to both similarity and generic wording when forming comparisons. We found clear support for the first four hypotheses outlined in Table 1, concerning structural alignment and generic wording. As predicted by the structural alignment hypotheses, children made more comparisons, and deeper comparisons, for high-similarity pairs than low-similarity pairs. As predicted by the generic wording hypotheses, children made deeper comparisons when given generic wording than specific wording. We also found some support for the fifth hypothesis, that similarity and generic wording would interact to influence children's judgments. That is, for both commonalities and alignable differences, deep properties were most often produced in the generic/high-similarity condition and least often produced in the specific/low-similarity condition. Moreover, generic language has a larger effect for high-similarity items, suggesting that the two factors interacted in children's judgments.

Children's performance paralleled that of adults in some important ways. Like adults, children were most likely to produce deep features when they encountered a combination of generic language and high similarity. This is particularly striking in that high-similarity pairs are just those for which surface commonalities were especially available. This pattern suggests that the 4-year-olds in our study engaged in similar basic processes as adults. These findings argue against an account of the relational shift whereby children's early reasoning is different in kind from that of adults, and limited to concrete perceptual information. Clearly even preschool children are capable of engaging in a process of structure-mapping that involves aligning like relational structures and considering hidden similarities and alignable differences. However, while these findings argue for continuity in processing between children and adults, they do not rule out other possible differences in relational reasoning (see Doumas, Hummel & Sandhofer, 2008; Halford, 1992; Richland et al., 2006).

As noted earlier, the primary similarity manipulation involved a mixture of conceptual and perceptual similarity, in that the high-similarity items were pairs drawn from the same superordinate category, and the low-similarity items were pairs drawn from different ontological categories. We also conducted a set of post-hoc analyses to examine whether similarity effects would be upheld even when all item-pairs are perceptually dissimilar (i.e., comparing high-OS/low-PS items with low-OS items). This further analysis indicated strong effects of similarity, suggesting that similarity effects are maintained even within this narrower perceptual range. One caveat is that this analysis does not fully control for perceptual similarity. Nonetheless, it does demonstrate that children respond to shared features even for items that lack a high degree of perceptual similarity.

Although for the most part children responded much the same as adults on these tasks, there were also some notable developmental differences. Most obviously, children generated comparisons at a lower rate than adults; this was true for both commonalities and alignable differences. Moreover, children generated relatively more surface properties than deep properties, as compared to adults. Both these effects may reflect the greater knowledge base that adults have about these categories. Additionally, adults were able to generate abstract commonalities (e.g., “They are physical objects”) and alignable differences (e.g., “Windows are made of glass, elephants are made of organic material”) that could apply to a wide range of items, perhaps reflecting a more analytic approach to the task.

The age effects were more pronounced on the Commonalities task than on the Differences task. This result was unanticipated, because on other measures the Differences task appeared to be more difficult overall (for both preschool children and college students). The relative difficulty of the Differences task can be seen in Table 6: both preschoolers and college students generated more commonalities on the Commonalities task than alignable differences on the Differences task. Moreover, whereas the majority of responses to the Commonalities task were commonalities, less than half of responses to the Differences task were alignable differences. Why, then, were children particularly prone to focus on surface features on the Commonalities task? We suspect that one reason is that, in order to express commonalities, one must generate properties that are quite abstract (especially for the low-similarity pairs; consider the commonalities between a camel and a necklace, for example), and so are harder to express. In contrast, when expressing an alignable difference, the component pieces can be quite concrete and so may be easier to articulate. For example, it is relatively straightforward to discuss the differences between a sheep and a cat (curly vs. straight fur; says ‘baa’ vs. ‘meow’, eats grass vs. fish, etc.). In contrast, to express the commonality one must articulate either a more abstract category (both are animals) or, even more challenging, list abstract properties that attach to this category (both breathe, both can move around, etc.) Despite the greater difficulty of articulating commonalities, children did often succeed in doing this, particularly when generic language had invited abstract representations of the items.

These findings are of interest not only for what they reveal about early comparison processing, but also for providing evidence that subtle variation in input wording (in this case, generic vs. specific) can have clear effects on the depth to which children process information they are asked to consider. These wording effects are consistent with arguments suggesting that language can provide subtle yet consistent effects on human reasoning (Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1996).

General Discussion

Our research examines two distinct means of highlighting deep conceptual structure and their interaction. First, subtle wording differences (generic vs. specific language) affect how people reason. We speculate that generic language invites thinking about kinds, and a concomitant focus on central conceptual knowledge. We also speculate that generic language invites deep representations, which will typically center on core causal or functional knowledge. The second factor is the structural alignment process, which preferentially highlights connected systems of relations that provide explanatory power. The combination of these two forces appears to provide a powerful source of insight. Together they lead to a substantial shift of focus from surface properties to deep properties. When children compare a highly alignable pair that is described using generic language, the resulting structural alignment shows an especially strong focus on deep relational structure.

Generic language and structure-mapping operate in complementary ways to promote a focus on deep knowledge: generic language at the representational level, and structure-mapping at the process level. Generic language influences the representations formed for the examples: it invites drawing upon more central or essential knowledge in construing the example. These construals will typically include causal or purposive relational systems. Structure-mapping operates during the comparison process to favor deeply interconnected systems of relations (by the systematicity principle, Gentner, 1983). Thus when these are combined, the person (child or adult) will typically respond to the generic description by calling forth deep knowledge about the exemplars, and the structural alignment process will favor this deep knowledge, even when the items compared are similar in many surface ways as well as in deeper ways. This pattern adds force to the claim that the central features of a concept are just those that are part of a deeply interconnected system of causal or functional relations (Jameson & Gentner, 2007).

Generic language—In this work, we found that generic wording (which highlights *kinds*) and specific wording (which highlights *individuals*) differentially affect the types of properties generated, for both age groups. Preschoolers as well as adults generated more deep properties in the generic wording condition, whereas they generated more superficial properties in the specific wording condition, in both the commonality and the difference task. However, for children this effect was limited to the high-similarity pairs, consistent with prior findings that children have difficulty achieving an alignment unless the relational commonalities are supported by surface commonalities (Gentner & Toupin, 1986; Gentner & Medina, 1998). This is the first demonstration we know of to show that generic versus specific language influences the kinds of information children consider. Importantly, this emphasis on deeper properties when given generic language was far stronger when the pairs were highly similar.

One important caveat is that we cannot be certain that generic language invites the formation of deeper commonalities, as it is instead possible that specific language directs children's attention to surface commonalities. That is, we would need to include a baseline control condition with neutral wording (neither generic nor specific) in order to interpret whether the effects are due more to generic wording raising performance or specific wording lowering performance.

To the extent that generics per se influence comparisons on this task, what is the mechanism by which they lead to wording effects? We propose that generics may be distinctive in that they call to mind a category abstraction (that is, a kind, as opposed to any set of individuals). This abstract representation highlights those features that are common across individual instances. Such features tend to be less transient, more stable, and more "essential" than features identified with particular instances (Carlson & Pelletier, 1995; Lyons, 1977). Thus, hearing a generic may emphasize the features we have referred to as "deeper".

In contrast, if the wording differences are due to the *specific* wording condition, this may reflect the ability of specific wording to point to an *individual* object or animal. Features of individuals can be transitory, accidental, or linked to a particular context, and thus relatively more superficial.

Structure-mapping and similarity—It is rather remarkable that high similarity leads to greater focus on deep properties among children. On a "least effort" account, one might have expected children to simply seize on any of the many easy surface commonalities available between, say, a kitten and a puppy. On the contrary, these data fit with the idea that the most natural process for high-similarity pairs is in fact structural alignment, which by its nature leads to the highlighting of common structure and the concomitant highlighting of alignable differences.

There is prior evidence for this claim. For example, as reviewed above, children are more likely to choose a same-category item (e.g., a banana) over a perceptually similar item (e.g., a round balloon) if they have compared two high-similarity examples (e.g., an apple and an orange) than if they have perceived either standard alone (Gentner & Namy, 1999; Namy & Gentner, 2002; see Liu, Golnikoff, & Sak, 2001 for related findings). As in the current studies, comparing these two perceptually similar standards did not seem to highlight the obvious common perceptual property (roundness), but rather to create a focus on their common causal relational structure. However, the current study provides the first direct investigation of the kinds of commonalities and differences children produce given high- versus low-similarity pairs.

Importantly, the similarity effects obtained here are robust even when we consider just those items that are relatively low in perceptual similarity. That is, both children and adults benefit from aligning items that share conceptual similarities, even when such items are perceptually highly dissimilar.

Implications for analogical development—As discussed above, one goal of the research was to discover whether children's comparison processing is similar to adults' (despite variation in amount of knowledge) or whether there are qualitative differences in how analogy is processed. In this research we tested five detailed predictions (seven, if sub-predictions are counted). We found that the results for children paralleled those for adults across variations in high versus low similarity and generic versus specific language, both for commonalities and differences. These findings support the idea that 4-year-olds engage in the same basic structural alignment process as adults, and that the age-related differences here can best be accounted for by increases in domain knowledge and corresponding differences in the richness of the representations being compared. This does not rule out other changes in analogical processing, such as increased ability to inhibit extraneous matches (Richland et al., 2006) or increases in working memory capacity (Doumas, Hummel, & Sandhofer, 2008; Halford, 1992).

Conclusions—The present findings demonstrate that how a task is framed can have clear implications for which information children and adults make use of, when engaged in comparison processing. Because the items were all familiar, these results are most appropriately interpreted as demonstrating that wording and similarity *highlight* certain kinds of relations more than others, thereby making them available for use in new situations. For example, it is unlikely that our participants—children or adults—had been asked previously to compare turtles and buses (or even windows and doors). Thus, the task required an analysis of familiar concepts in a novel reasoning task. An important question for the future is whether these factors can also provide a mechanism for *discovering* new, non-obvious properties. For example, when provided with wholly novel categories and asked to compare these under either generic or specific wording conditions, would similarity and wording lead to the detection of features that would otherwise have gone unnoticed? If so, this result could have important implications for educational contexts.

Acknowledgments

This research was supported by NICHD grant HD36043 to Gelman. We are grateful to the teachers, staff, and children of the University of Michigan Children's Centers, Generations Together, and Gretchen's House (Oak Valley and Dhu Varren) for participating in this research. We would also like to thank Jenna Salm for her able assistance with data collection.

References

Ahn WK. Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. Cognition 1998;69:135–178. [PubMed: 9894403]

Bassok M, Holyoak KJ. Interdomain transfer between isomorphic topics in algebra and physics. Journal of Experimental Psychology: Learning, Memory, and Cognition 1989;15:153–166.

Carlson, GN.; Pelletier, FJ., editors. The generic book. University of Chicago Press; Chicago: 1995.

Chen Z, Klahr D. All other things being equal: Acquisition and transfer of the Control of Variables strategy. Child Development 1999;70:1098–1120. [PubMed: 10546337]

Dixon JA, Dohn MC. Redescription disembeds relations: Evidence from relational transfer and use in problem solving. Memory and Cognition 2003;31:1082–1093.

Doumas LAA, Hummel JE, Sandhofer CM. A theory of the discovery and predication of relational concepts. Psychological Review 2008;115:1–43. [PubMed: 18211183]

Feinstein, T. The development of the ability to name commonalities and differences. Northwestern University; 1999. Unpublished ms.

Forbus KD, Gentner D, Law K. MAC/FAC: A model of similarity-based retrieval. Cognitive Science 1995;19:141–205.

Gattis M. Structure mapping in spatial reasoning. Cognitive Science 2002;17:1157–1183.

Gelman, SA. The essential child: Origins of essentialism in everyday thought. Oxford University Press; New York: 2003.

Gelman, SA. Learning words for kinds: Generic noun phrases in acquisition.. In: Hall, DG.; Waxman, SR., editors. Weaving a lexicon. MIT Press; Cambridge, MA: 2004.

Gelman, SA.; Hollander, M.; Star, J.; Heyman, GD. The role of language in the construction of kinds.. In: Medin, D., editor. Psychology of learning and motivation. Vol. 39. Academic Press; New York: 2000. p. 201-263.

Gelman SA, Goetz PJ, Sarnecka BW, Flukes J. Generic language in parent-child conversations. Language Learning and Development 2008;4:1–31.

Gelman SA, Raman L. Biology as a window onto cognitive development. Human Development 2002;45:61–68.

Gelman SA, Star J, Flukes J. Children's use of generics in inductive inferences. Journal of Cognition and Development 2002;3:179–199.

Gentner D. Structure-mapping: A theoretical framework for analogy. Cognitive Science 1983;7:155–170.

Gentner D. Metaphor as structure mapping: The relational shift. Child Development 1988;59:47–59.

Gentner, D.; Gentner, DR. Flowing waters or teeming crowds: Mental models of electricity.. In: Gentner, D.; Stevens, AL., editors. Mental models. Erlbaum; Hillsdale, NJ: 1983. p. 99-129.

Gentner, D.; Goldin-Meadow, S., editors. Language in mind: Advances in the study of language and thought. MIT Press; Cambridge, MA: 2003.

Gentner D, Gunn V. Structural alignment facilitates the noticing of differences. Memory and Cognition 2001;29:565–577.

Gentner D, Loewenstein J, Hung B. Comparison facilitates children's learning of names for parts. Journal of Cognition and Development. 2007

Gentner D, Markman AB. Structural alignment in comparison: No difference without similarity. Psychological Science 1994;5:152–158.

Gentner D, Medina J. Similarity and the development of rules. Cognition 1998;65:263–297. [PubMed: 9557385]

Gentner D, Namy L. Comparison in the development of categories. Cognitive Development 1999;14:487–513.

Gentner D, Toupin C. Systematicity and surface similarity in the development of analogy. Cognitive Science 1986;10:277–300.

Gumperz, JJ.; Levinson, SC., editors. Rethinking linguistic relativity. Cambridge University Press; New York: 1996.

Halford GS. Analogical reasoning and conceptual complexity in cognitive development. Human Development 1992;35:193–218.

Hollander MA, Gelman SA, Raman L. Generic language and judgments about category membership: Can generics highlight properties as central? Language and Cognitive Processes. in press.

Hollander MA, Gelman SA, Star J. Children's interpretation of generic noun phrases. Developmental Psychology 2002;38:883–894. [PubMed: 12428701]

Inagaki K, Hatano G. Young children's spontaneous personification as analogy. Child Development 1987;58:1013–1020.

Jameson, J.; Gentner, D. Conceptual centrality and the role of comparison.. In: McNamara, DS.; Trafton, JG., editors. Proceedings of the 29th Annual Cognitive Science Society; Austin, TX: Cognitive Science Society. 2007. p. 64-70.

Liu J, Golinkoff RM, Sak K. One cow does not an animal make: Children can extend novel words at the superordinate level. Child Development 2001;72:1674–1694. [PubMed: 11768139]

Loewenstein J, Gentner D. Spatial mapping in preschoolers: Close comparisons facilitate far mappings. Journal of Cognition and Development 2001;2:189–219.

Lombrozo T, Carey S. Functional explanation and the function of explanation. Cognition 2006;99:167–204. [PubMed: 15939416]

Lyons, J. Semantics. Vol. 1. Cambridge University Press; New York: 1977.

Markman AB, Gentner D. Structural alignment during similarity comparisons. Cognitive Psychology 1993a;25:431–467.

Markman AB, Gentner D. Splitting the differences: A structural alignment view of similarity. Journal of Memory and Language 1993b;32:517–535.

Markman AB, Gentner D. Commonalities and differences in similarity comparisons. Memory & Cognition 1996;24:235–249.

Mix KS. Similarity and numerical equivalence: Appearances count. Cognitive Development 1999;14:269–297.

Namy LL, Gentner D. Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. Journal of Experimental Psychology: General 2002;131:5–15. [PubMed: 11900103]

Paik JH, Mix KS. Preschoolers’ use of surface similarity in object comparisons: Taking context into account. Journal of Experimental Child Psychology 2006;95:194–214. [PubMed: 16904122]

Prasada S, Dillingham EM. Principled and statistical connections in common sense conception. Cognition 2006;99:73–112. [PubMed: 16443448]

Rattermann MJ, Gentner D. More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. Cognitive Development 1998;13:453–478.

Rehder B, Hastie R. Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. Journal of Experimental Psychology: General 2001;130:323–360. [PubMed: 11561914]

Richland LE, Morrison RG, Holyoak KJ. Children's development of analogical reasoning: Insights from scene analogy problems. Journal of Experimental Child Psychology 2006;94:249–273. [PubMed: 16620867]

Richland LE, Zur O, Holyoak K. Cognitive supports for analogies in the mathematics classroom. Science 2007;316:1128–1129. [PubMed: 17525320]

Tversky A. Features of similarity. Psychological Review 1977;84:31–48.

Vosniadou S. Analogical reasoning in cognitive development. Metaphor and Symbol 1995;10:297–308.

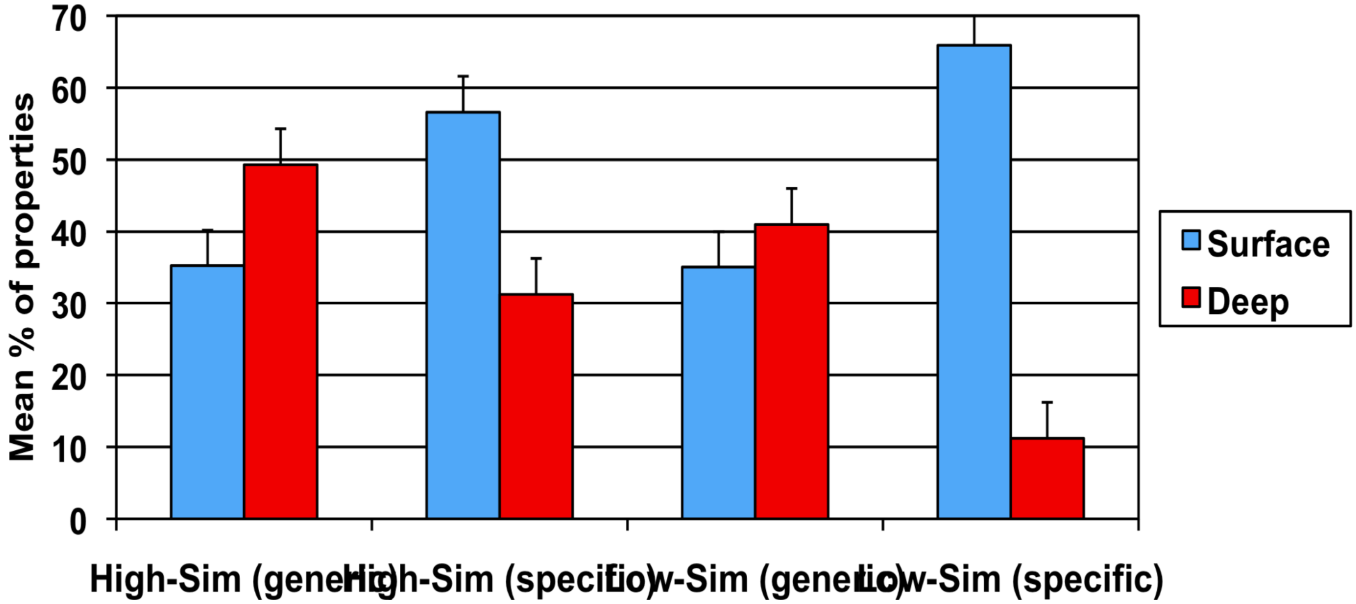

Figure 1.
Study 1 (Adults), Mean percentage of properties expressing commonalities, as a function of similarity, wording, and property type. Error bars indicate standard errors.

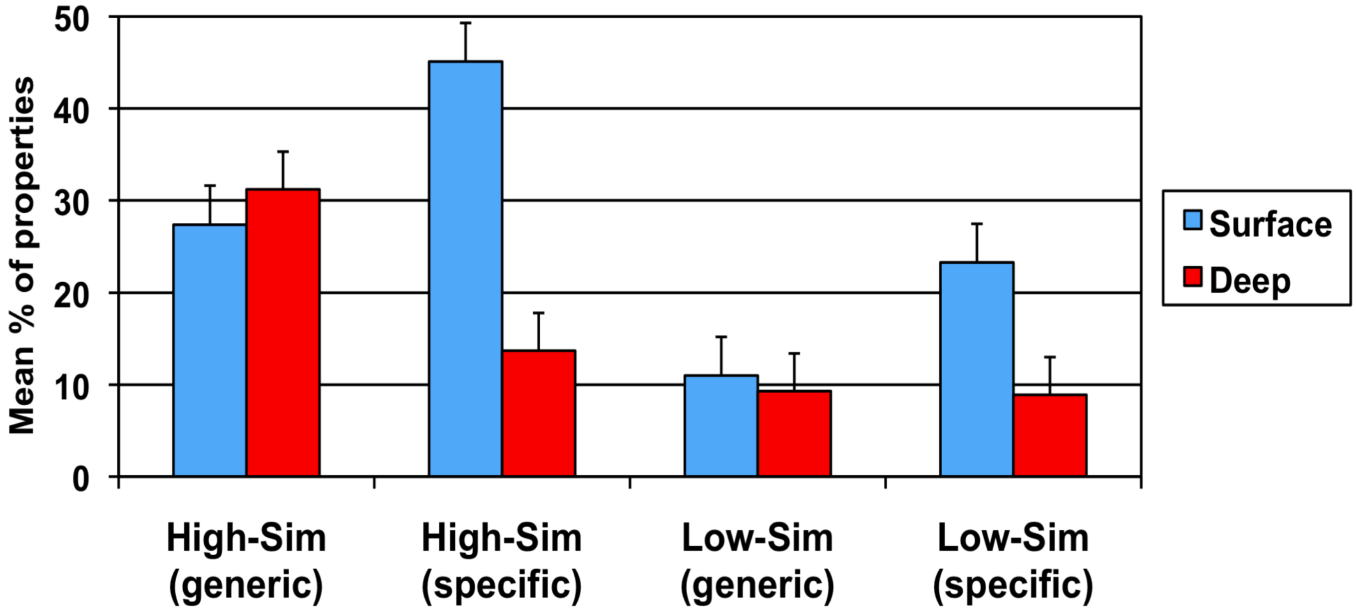

Figure 2.
Study 1 (Adults), Mean percentage of properties expressing alignable differences, as a function of similarity, wording, and property type. Error bars indicate standard errors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

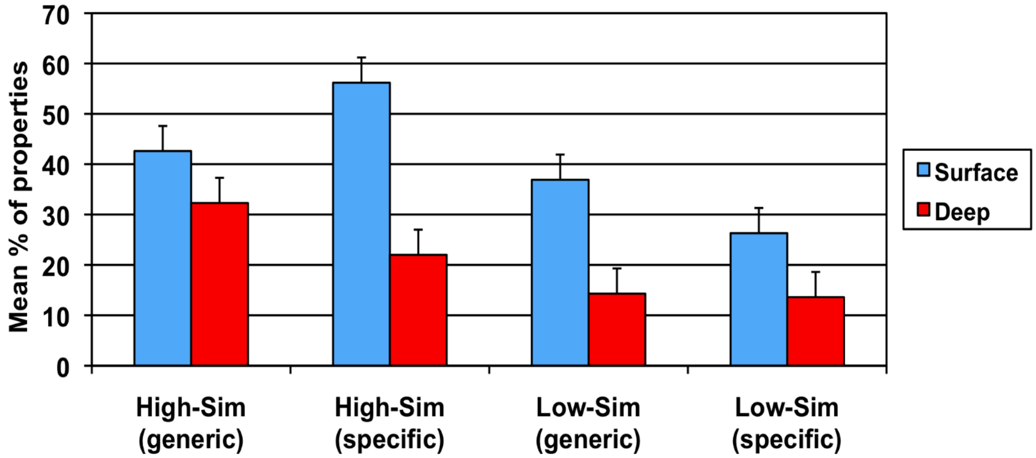

Figure 3.
Study 2 (Children), Mean percentage of properties expressing commonalities, as a function of similarity, wording, and property type. Error bars indicate standard errors.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

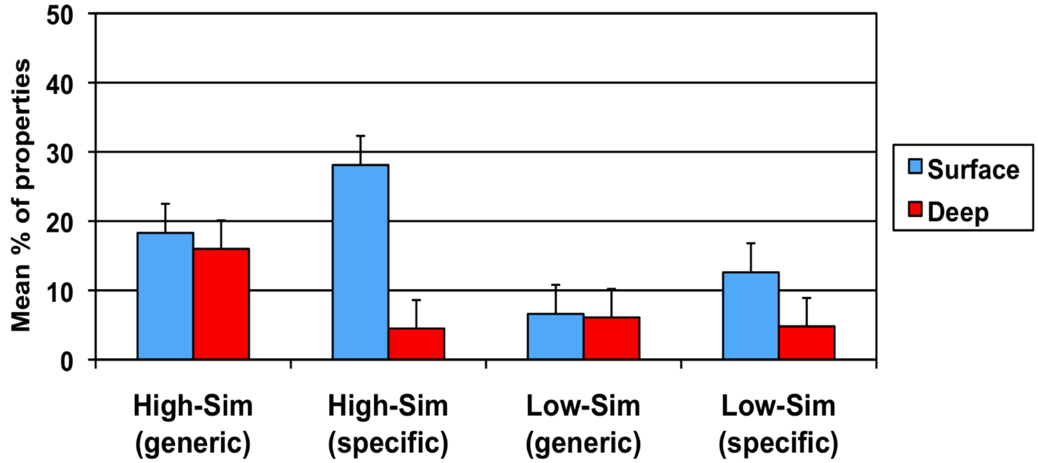

Figure 4.
Study 2 (Children), Mean percentage of properties expressing alignable differences, as a function of similarity, wording, and property type. Error bars indicate standard errors.

Table 1

Predictions, Studies 1 and 2.

Hypothesis #	Prediction
STRUCTURAL ALIGNMENT	
H_1	*(a) High-similarity items will result in more commonalities than low-similarity items.*
	(b) High-similarity items will result in more deep (causally connected) commonalities than low-similarity items.
H_2	*(a) High-similarity items will result in more differences, especially more alignable differences, than low-similarity items.*
	(b) High-similarity items will result in more deep (centrally connected) alignable differences than low-similarity items.
GENERIC WORDING	
H_3	*Generic wording will result in deeper (less obvious) commonalities than specific wording.*
H_4	*Generic wording will result in deeper (less obvious) alignable differences than specific wording.*
COMBINATION (STRUCTURAL ALIGNMENT AND GENERIC WORDING)	
H_5	*Deeper commonalities and alignable differences will be most evident with high-similarity items that are presented in generic wording (i.e., when structure-mapping theory and generic theory are combined), especially for children.*

Table 2

High-Similarity Items

Items with higher similarity ratings are indicated with an asterisk (*).

* Bus-car
* Cow-horse
* Crayon-pencil
* Fork-spoon
* Giraffe-zebra
* Plate-bowl
* Ring-necklace
* Window-door
Airplane-train
Camel-whale
Dog-cat
Refrigerator-oven
Sheep-pig
Snake-rabbit
Tiger-elephant
Turtle-duck

Table 3

Low-Similarity Items.

Camel-necklace
Cat-fork
Cow-airplane
Dog-spoon
Duck-car
Elephant-window
Giraffe-pencil
Horse-train
Pig-bowl
Rabbit-oven
Sheep-plate
Snake-refrigerator
Tiger-door
Turtle-bus
Whale-ring
Zebra-crayon

Table 4

Sample Coding: Commonalities, Alignable Differences, Other.

Commonalities	
	"Cats' nails are sharp, and the fork is sharp."
	"Elephants and windows are big."
	"They are both cold inside." [snake/refrigerator]
Alignable differences	
	"This dog has a short tail, this cat has a long tail."
	"Horses you can ride, cows you milk."
	"Dogs are outside, and spoons are inside."
Other	
	"Sheep have legs and pigs roll in the dirt."
	"You can't see out of elephants." [elephant/window]
	"This camel has two bumps on it." [camel/whale]

Table 5

Content Coding: Surface, Deep, and Other

Surface: part, color, shape, size, pattern, texture, properties of the picture or word
“They both have four legs.” [tiger-elephant]
“The door has pink and the tiger has orange.”
“The dog is little and the spoon is big.”
Deep: trait (strong, smart), action, function, psychological state, categorization
“Snakes slither, rabbits run fast.”
“Both are animals.” [dog-cat]
“They give people rides to grandma's house.” [airplane-train]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 6

Mean # of responses, as a function of age group, task, and response type.

Age	Task	Response Type*	Mean	SD
College	Commonality	C	49.06	24.09
		AD	1.91	2.16
		NAD	0.03	0.17
		SINGLE	2.06	3.79
		THEM	0.79	1.61
	Difference	C	1.94	4.65
		AD	35.54	13.39
		NAD	13.48	9.54
		SINGLE	33.06	24.37
		THEM	0.15	.44
Preschool	Commonality	C	21.88	13.36
		AD	1.56	2.15
		NAD	0.56	1.08
		SINGLE	6.38	7.20
		THEM	1.35	2.24
	Difference	C	1.19	1.57
		AD	9.37	5.46
		NAD	15.78	16.66
		SINGLE	16.56	15.52
		THEM	0.59	1.07

*C = commonalities; AD = alignable differences; NAD = non-alignable differences; SINGLE = single-item properties; THEM = thematic properties.