THINKING & REASONING, 2010, 16 (4), 372–401

Psychology Press
Taylor & Francis Group

When learning to classify by relations is easier than by features

Marc T. Tomlinson and Bradley C. Love
University of Texas at Austin, TX, USA

Relational reasoning is often considered more resource intensive than feature-based reasoning. This view implies that learning categories defined by relational regularities should be more difficult than learning categories defined by featural regularities. Unfortunately previous studies do not ground featural and relational information in a common perceptual substrate. After addressing this concern, a series of experiments compare learning performance for relation- and feature-based categories. Under certain circumstances we find faster learning for relation-based categories. The results suggest that mechanisms rooted in relational processes (e.g., relative stimulus judgement, analogical comparison) facilitate or hinder learning depending on whether the relational processes highlight or obscure the underlying category structure. Conversely, category learning affects relational processes by promoting relational comparisons that increase the coherence of acquired categories. In contrast to the largely independent research efforts in category learning and analogy research, our findings suggest that learning and comparison processes are deeply intertwined.

Keywords: Category learning; Relations; Features; Analogy.

The ability to grasp complex relations is a hallmark of human intelligence (Penn, Holyoak, & Povinelli, 2008). Evaluative tests such as the SAT reasoning test, graduate record examination, or Raven's Progressive Matrices stress the importance of relational thinking. Most research comparing relational and featural performance supports the view that relational processing is a more advanced competency. Children learn concepts defined by features earlier than those defined by relations

Correspondence should be address to Marc T. Tomlinson, Department of Psychology, The University of Texas at Austin, Austin, TX 78712, USA.
E-mail: mtomlinson@love.psy.utexas.edu

This work was support by ASFOR grant FA9550–10–1–0268.

http://www.psypress.com/tar DOI: 10.1080/13546783.2010.530464

(Gentner, 1978) and experts differentiate themselves from novices by organising knowledge within their domain along relational lines (Chi, Feltovich, & Glaser, 1981; Danovitch & Keil, 2004).

Despite strong interest in understanding these human competencies, there is a lack of work that examines feature- and relation-based learning under comparable conditions. Indeed, fundamental differences in the nature of feature- and relation-based categories make constructing apropos comparisons challenging. Membership in feature-based categories is determined by concrete feature attributes, such as size, shape, and colour, whereas membership in relation-based categories is determined by fulfilling an abstract relational role, as in concepts like *predator* and *prey* (cf. Markman & Stillwell, 2001). Additionally, relation-based categories generally exhibit greater perceptual variability across members than do feature-based categories (Gentner, 1981).

In laboratory studies comparing feature- and relation-based learning, acquiring the relation-based categories often requires processing more attributes (Kittur, Hummel, & Holyoak, 2004; Waltz, Lau, Grewal, & Holyoak, 2000). Furthermore, no study comparing feature- and relation-based learning has used stimuli that rely on the same perceptual substrate. In all cases the perceptual features carrying relational and featural information are different. In other domains, such as same/different discrimination (Love, Rouder, & Wisniewski, 1999) and change detection (Kroger, Holyoak, & Hummel, 2004), this concordance has been achieved, but never in a learning study. Placing features and relations in the same perceptual substrate would make it possible to identify factors that differentially affect the learning rates of categories defined by features and relations.

Research in perceptual learning and discrimination suggests that there are many cases in which relational processing is advantaged over featural processing. For example, there is a consensus that children and adults are more proficient at making relative judgements (Garner, 1954; Huttenlocher, Duffy & Levine, 2002) than absolute judgements. Difficulties with absolute (i.e., featural) judgement are so extreme that models of absolute identification have been proposed that rely on sequential effects arising from relative judgements between stimulus pairs (Stewert, Brown, & Chater, 2005). Accordingly, relative judgements of sequential stimuli strongly affect categorisation performance in tasks that ostensibly should and were once thought to rely solely on memory of featural information (Jones, Love, & Maddox, 2006; Stewart & Brown, 2004). Difficulties in making absolute judgements should cause difficulties in learning categories defined by absolute values of the stimuli.

In addition to effects arising from the nature of absolute and relative judgements, more elaborate comparison processes, such as those discussed in the analogy literature, suggest differences in how categories will be learned depending on whether learning mechanisms engage featural or relational processes. Below we discuss predictions for how these more elaborate comparison processes should impact the learning of feature- and relation-based categories.

ONLINE RELATIONAL COMPARISONS

Accounts of relational comparison allow for the basis of comparison to be determined in an online manner. Online comparison can be contrasted with the inflexibility inherent in accounts that do not differentiate between features and relations. In these inflexible accounts, correspondences between features (e.g., Tversky, 1977) or stimulus dimensions (e.g., Shepard, 1964) are typically predetermined (i.e., determined offline). For instance, in multidimensional models of similarity, each stimulus is represented by a point in a common space and the basis for comparison is determined by the dimensions of the space. For example, the value of one attribute (e.g., size) is never contrasted with the value of another attribute (e.g., luminance), nor do the values interact; the stimulus attributes are assumed to be independent. The independence assumption stipulates that changes in one attribute should not affect the way changes in another attribute affect similarity (this independence assumption is separate from that of integral/separable dimensions).

Violations of these assumptions have been observed when stimuli include prominent relational content. For example, Goldstone (1996) found that increasing the number of relational changes between two scenes could increase the rated similarity of the two scenes instead of decreasing it. In addition, Love and Markman (2003) showed that even classical features, such as shape, exhibit signs of relational processing by interacting with other features such as size, causing conjunctive rules involving shape and size to be harder to learn. In particular, they found that shape served as the argument to dimension predicates such as size and colour. In other words, a stimulus is conceptualised as a large triangle, not as a feature set containing large and triangle. These findings suggest that comparison processes engaged by learners may differ from standard accounts when the stimuli contain relations and that category structure (e.g., which dimensions are relevant to classification) plays a critical role in modulating the outcomes.

Several popular accounts of relational comparison allow for online comparison processes, for example, structural alignment, and transformation. The structural alignment view holds that features, objects, and relations in compared scenes are put into correspondence (i.e., analogical mappings are found) according to a number of structural constraints (Gentner, 1983). Similarity is proportional to the soundness of the mapping between the two scenes. For instance, similarity is higher when corresponding objects play like roles in corresponding relations (e.g., the causal agent in one scene corresponds to the causal agent in the other scene). The transformational view holds that people apply operators to one scene to transform it to be identical to the other scene (Hahn, Chater, & Richardson, 2003). According to the transformational view, similarity is inversely related to the number of transformations that are required.

Both alignment and transformation accounts are consistent with the idea that relations enable online comparison processes whose basis is not predetermined. Structural alignment establishes correspondences among stimulus pairs in an online fashion, as opposed to relying on predetermined alignments. Relations determine these alignments, although features can also play a role. For example, in comparing a scene in which a car tows a boat with a second scene in which a tow-truck tows a car, the car from the first scene may be aligned (i.e., cross mapped) with the tow-truck from the second scene because these entities play the same relational role even though the cars in the scenes are more featurally similar to one another (Markman & Gentner, 1993). Likewise, according to the transformational view, the presence of relations can suggest additional operators (i.e., relational transformations) that can be applied opportunistically to stimulus pairs.

These online views hold that the attributes of a stimulus are not independent and that changes on one attribute in a complex scene (e.g., head colour) can affect how differences between other attributes (e.g., tail colour) are calculated because they can result in different alignments between the scenes. This can cause non-monotonicities in the similarity ratings, where the introduction of additional changes between two stimuli increases the similarity between the items (Goldstone, 1996). These alignment driven non-monotonicities should play a role in category learning as well, suggesting that the difficulty of learning a given category structure could change depending on whether the stimuli exhibit interactions between their attributes.

Lassaline and Murphy (1997) provided initial evidence that alignment processes are active during category learning. They found that the number of Matches In Place (MIPS, e.g., two birds having the same colour head) and Matches Out of Place (MOPS, e.g., a bird having the same colour tail as another bird's head) relate to the difficulty of learning a category. Categories are easier to acquire when members have many MIPS in common compared to when they have many MOPS in common. Lassaline and Murphy's results provide support for the notion that dimensional correspondences (Goldstone, 1994) are critical in category learning. However, their results do not speak to the role of relational commonalities (i.e., relationships across dimension values). For example, two birds could share no MIPS or MOPS, yet share the relational commonality of having their heads be brighter than their tails.

In this contribution we consider how people learn categories defined by such relational matches. We suggest that people use relational comparison processes (such as alignment) during category learning to match the current stimulus to stored category examples. Such processes may explain people's ability to readily learn purely relational categories from a small set of examples (e.g., Rehder & Ross, 2001). Indeed, computational models that incorporate these processes successfully explain how infants, adults, and animals learn seemingly abstract concepts based on a small set of training examples (Tomlinson & Love, 2006).

A number of predictions fall out of the view that category learning involves online relational comparisons to stored examples. One prediction that we test here is that online relational processing can benefit or hinder learning depending on the relationship between presented stimuli and previous exemplars. When the preferred mapping between the current stimulus and stored stimuli serves to increase the coherence of a category, learners should benefit from relational processing. However, in other cases, relational matching processes can actually reduce category coherence by enabling discovery of non-obvious similarity relations among members of contrasting categories. In such cases, learning should be retarded and error patterns indicative of interference arising from relational comparison processes should be observed. Rather than being a fixed process, one possibility is that the preferred alignments themselves can change over the course of learning in order to increase category coherence.

STIMULUS DESIGN

The stimuli used in our experiments consist of simple scenes that were designed to provide an informative comparison of relation- and feature-based category learning. To this end, both categories were defined over, and required processing the same perceptual attributes (i.e., size, luminance) and the same number of perceptual attributes (two attributes). Because features and relations are psychologically distinct and are associated with different processes (see Goldstone, Medin, & Gentner, 1991), in principle, one cannot construct a stimulus set that does not bias in favour of featural or relational learning. Fundamentally, relations are defined over distinct entities whereas features integrate over some range. For example, determining the overall luminance (a feature) of a scene requires integrating over the entire scene. It stands to reason that complex scenes consisting of multiple entities would make such a computation more difficult. Indeed, Duncan (1984) finds that specifying features across multiple objects increases the time required to identify the features. Conversely, relational regularities are likely to be easier to detect when entities are readily individuated. To appreciate A causes B, one must be able to clearly discern A and B as distinct entities. In summary, stimuli that contain a single entity are likely to favour featural processing, whereas stimuli that contain multiple entities are likely to favour relational processing.

The norm in the field (e.g., Lassaline & Murphy, 1997) is for each stimulus to consist of a single entity, thus favouring featural processing. Because our primary interest is in exploring relational influences on category learning, we used a stimulus set consisting of scenes composed of two entities. Although it is not our focus, one interesting question is whether relational category learning is actually favoured under such conditions compared to featural category learning. To foreshadow, our results indicate that it is. Although not discussed in this contribution, we find the same overall pattern of results, with a bias in favour of featural learning, when the present studies are conducted using stimuli that consist of a single entity.

In the studies reported here each stimulus consisted of two circles appearing side-by-side. Across trials, these two circles varied in their size (small, medium, large) and brightness (light, moderate, dark). These circles were combined to give two overall relation attributes (which side was bigger and which was brighter) and two feature attributes (overall size and overall brightness). The medium and moderate values were always manifested once in a scene (see Figure 1).

In summary, in order to balance the perceptual requirements, the relational attributes defined for these scenes, which side is bigger and which side is brighter, require the processing of the relationship between the two objects in the scene to correctly identify the attribute. In contrast, the evaluation of featural attributes does not require any consideration of the relational role a specific circle plays in determining an attribute's value. The features and relations differ only in how the participant combines the information about the two circles. For the relations, consistent with definitions of relational categories (Markman & Stilwell, 2001), information from each circle plays a distinct role in determining the value of each relation, whereas for the features, information from each circle is combined independently in determining the value of each feature. Additionally, the requirement that the features are separated across the two circles could bias learning rates towards the relations because features become more difficult to process when spread across multiple objects (Duncan, 1984).

To foreshadow our results, in Experiment 1 pair-wise similarity ratings are collected for the stimuli. The collected similarity ratings between stimulus pairs are consistent with an online comparison process that manifests itself as a non-monotonic relationship between the number of changes across the attributes and their rated similarity. In Experiment 2

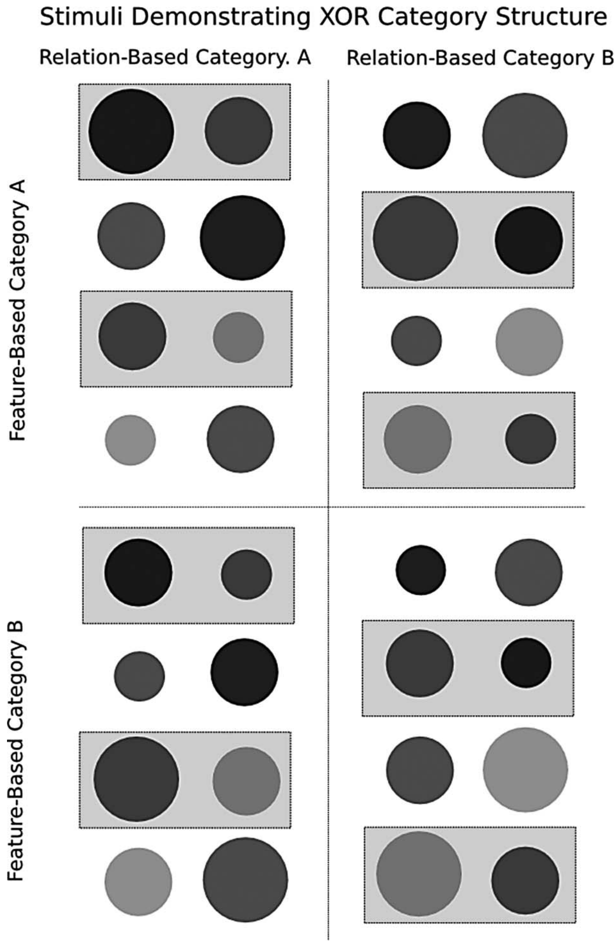

Figure 1. The 16 stimuli arranged according to the relational (left vs right) and featural (top vs bottom) XOR category structure. The circles vary on four attributes: two features and two relations. The features are overall size and overall brightness (defined over both circles). The relations are which circle (by left/right spatial position) is bigger and which circle is brighter. The size-relation based one-dimensional category groups stimuli where the larger circle is on the left in category A, and those with the larger circle on the right in B. For the size feature, stimuli with large circles are in one category, while those with small circles are in the other. The feature-based XOR groups large dark stimuli with small light stimuli, while the relation-based XOR groups stimuli with darker circles on the left and smaller circles on the right with stimuli that have lighter circles on the left and larger circles on the right. The grey boxes are not part of any stimulus and are intended to promote the clarity of the figure by grouping constituent stimulus elements together.

we find that simple relational categories are learned faster than simple feature-based categories. Although this result may be surprising on the surface, it follows from our stimulus design (two entities as opposed to one,

TABLE 1
Predicted learning biases and their effect on the learning rate of the category structures used in Experiments 2, 3, and 4

	Rule	*Relative processing*	*Online correspondence*
Exp. 2	1-D	Bias in favour of relations	No effect
Exp. 3	XOR	Bias in favour of relations	Bias in favour of relations
Exp. 4	Four Category	Bias in favour of relations	Bias against relations

same perceptual substrate for featural and relational information) and previous findings in relative vs absolute stimulus judgements. Using more complex category structures, Experiments 3 and 4 consider conditions under which online relational comparisons promote or hinder learning (see Table 1). In Experiment 3 we find that relation-based category learning is advantaged over feature-based category learning under conditions in which relational comparisons allow learners to increase category coherency. In Experiment 4, using category structures in which relational comparisons should not increase coherency, we find no advantage for learning relation-based categories over feature-based categories. Indeed, we observe error patterns indicative of comparison processes interfering with relation-based category learning. While Experiments 1–4 support the notion that relational comparison can shape learning, Experiment 5 finds support for the complementary position that learning can shape relational comparison.

EXPERIMENT 1

In Experiment 1 we investigate the nature of the comparison process by looking at similarity rating data for the stimuli. According to relational accounts of processing, the stimuli should exhibit the non-metrical effects of an online comparison process. In contrast to metrical views of comparison and similarity, where distances between stimuli follow a set of strict axioms and similarity decreases with each difference between the stimuli, transformation- and alignment-based approaches suggest that stimuli that differ along both relations might be rated as highly similar (Goldstone, 1996). We predict that ratings will be non-metrical, in that similarity will not substantially decrease (and may in fact increase) for stimulus pairs mismatching on both relations compared to stimulus pairs mismatching on only one relation. Additionally, overall influences of featural and relational attribute matches can be assessed in the similarity rating data. If the features or relations of the stimuli play a larger role in determining similarity then category distinctions on those dimensions could be learned easier because differences between the stimuli on those dimensions are more salient.

Method

Participants. A total of 22 undergraduate students from the University of Texas at Austin participated for course credit.

Stimuli. The stimuli were the same as those described above. They varied along four binary attributes: two relational attributes, which side was bigger and which side was brighter, and two featural attributes, overall brightness (both circles combined) and overall size (both circles combined).

Procedure. Participants were instructed to rate the similarity of two presented stimuli on a scale from 1 to 9. Participants were instructed that each stimulus varied along four binary-valued attributes (overall brightness, overall size, which circle was brighter, which circle was bigger). On each trial two stimuli were simultaneously presented on screen with text designating pair 1 and pair 2, as well as text asking for their similarity on a scale of 1–9. One pair was displayed on the top of the screen and the other on the bottom. A line separated the pairs. Participants responded by pressing key 1 through 9. Following the participant's response, the screen blanked for 500 ms and the next trial began. Each participant rated 136 pairs of stimuli [(16 * 15)/ 2 + 16]: each stimulus paired with every other stimulus, plus each stimulus paired with itself. The overall order of the trials and the assignment of pairs to the top or bottom of the screen were randomised.

Results and discussion

For the purposes of analyses, the similarity ratings were grouped according to how many features or relations were different within the comparison. Figure 2 illustrates the nine means resulting from this aggregation. A 3 (0, 1, or 2 relation differences) $\times$ 3 (0, 1, or 2 feature differences) within-participant ANOVA revealed a main effect of both the number of different relations, $F(2, 42) = 33.68$, $p < .001$, and the number of different features, $F(2, 42) = 204.81$, $p < .001$, as well as a significant interaction between the number of feature and relational differences, $F(4, 84) = 36.47$, $p < .001$.

The above interaction is indicative of a non-metrical similarity space arising from relational processes. To test the predictions of the alignment account more precisely, a 2 (relation or feature) $\times$ 2 (one or two differences) ANOVA was performed to compare the effects of mismatching on one or both relations (with both features matching) with the effects of mismatching on one or both features (with both relations matching). The strong interaction predicted is shown in Figure 3. An ANOVA revealed a significant main effect for feature or relation difference, $F(1, 21) = 26.32$, $p < .001$, and a main effect for the number of differences, $F(1, 21) = 23.85$,

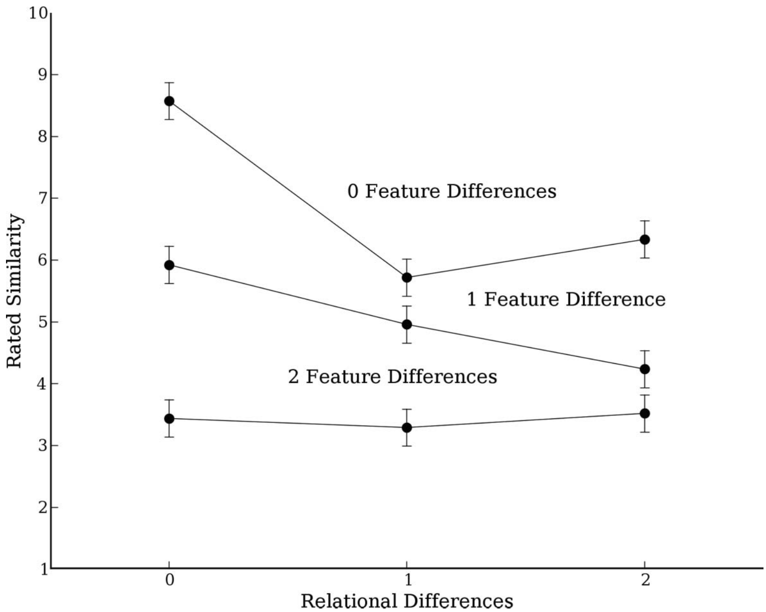

Figure 2. Experiment 4's mean similarity ratings as function of number of feature and relation differences. Error bars represent approximate 95% confidence intervals.

$p < .001$, as well as a significant interaction between the type of difference and the number of differences, $F(1, 21) = 96.05$, $p < .001$. Planned t-tests revealed that rated similarity was higher when both relations differed than when only one relation differed, $t(21) = 2.13$, $p < .05$, whereas rated similarity was lower when both features differed than when only one feature differed, $t(21) = 13.68$, $p < .001$. As predicted by the online relational processing hypothesis, similarity ratings were non-metrical in that stimulus pairs differing on both relations were rated as more similar than stimulus pairs differing on only one relation. These similarity data are consistent with the learning results from Experiments 1 and 2.

Importantly, as predicted by relational processing accounts, this effect is constrained to the case of two relational changes and no feature changes (see Figure 2). For this case, the two stimuli consist of identical circles that have been transposed (see Figure 1). For the other cases, comparisons involve stimuli that differ in terms of both features and relations, which complicates the alignment or transformation process.

To test for differences in salience between the relations and the features, a regression model was fitted to each participant's similarity ratings, with the number of relational differences and the number of featural differences as independent predictors. A paired t-test was then conducted on the fitted

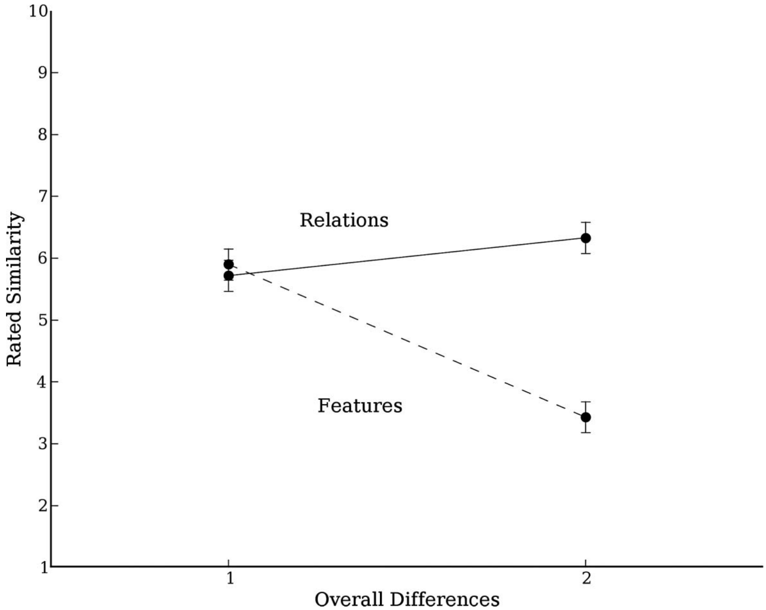

Figure 3. A subset of Experiment 4's mean similarity ratings reveals the strong interaction consistent with relational flexibility. Mismatching on both relations (with both features matching) increases similarity, whereas mismatching on both features (with both relations matching) decreases similarity. Error bars represent approximate 95% confidence intervals.

weights for the relational and featural terms across the participants. This test showed a significantly larger effect for feature differences on rated similarity, mean coefficient of 1.68, compared to relational differences, mean of .77, $t(21) = 5.01$, $p < .001$. This test suggests that the features play a larger role in determining similarity than the relations.

EXPERIMENT 2

Experiment 1 found the expected non-monotonicity in participants' similarity ratings of the stimuli, evidence of relational comparison processes. In addition it showed that the features were more salient than the relations, e.g., differences in features contributed more to the rated similarity. The next several experiments look at the implications of relational-processes on the difficulty of learning various category structures. Experiment 2 examines participants' ability to learn simple category structures defined by a single attribute. Participants learned to classify stimuli as members of one of two contrasting categories based on a single featural attribute (e.g., items with big circles are in one category, whereas items with small circles are in the other category) or learned to classify based on a single relational attribute (e.g., items in which circle on the right is bigger are in one category, whereas items in which the circle on the left is bigger are in the other category).

Whereas in Experiment 1 the relevant stimuli to compare were presented simultaneously, stimuli in Experiment 2 are presented in isolation and sequentially. The standard category-learning paradigm used in Experiment 2 requires that any stimulus comparisons be made to representations of items stored in memory. Because absolute feature judgements are harder to make than relative judgements (Garner, 1954; Huttenlocher et al., 2002), one prediction is that relation-based categories will be easier to learn than feature-based categories despite the fact that features were more salient than relations, as measured by Experiment 1's similarity ratings. Participants learning relation-based categories should have an advantage as attribute values can be determined by within-stimulus comparisons, as opposed to comparisons to other representations stored in memory.

Method

Participants. A total of 53 undergraduate students from the University of Texas at Austin participated for course credit. Participants were randomly assigned to the brightness-relation relevant ($n = 13$), the size-relation relevant ($n = 13$), the brightness-feature relevant ($n = 13$), or the size-feature relevant conditions ($n = 14$).

Stimuli and category structure. The stimuli were the same as those used in Experiment 1. Participants learned to classify the stimuli according to a unidimensional rule that was defined over one of the four stimulus attributes (brightness-relation, size-relation, brightness-feature, or the size-feature). In Figure 1 the coloured background distinguishes the two categories learned for the size-relation condition. Those with a grey background belong to one category (bigger-left), while those with a white background belong in the opposite category (bigger-right). Similarly for the size-feature condition, the top four stimuli and the bottom four stimuli of Figure 1 would belong to one category (overall big), while the middle eight stimuli would belong to the opposite category (overall small).

Procedure. Participants were presented with a screen of detailed instructions informing them that they were going to learn to categorise pairs of circles into two categories, A and B. Participants were instructed that each stimulus varied along four attributes: overall brightness, overall size, which circle was brighter, and which circle was bigger. They were told to look for a rule involving one of those attributes. For each participant the labels A and B were randomly assigned to the two categories.

On each learning trial, two circles were presented in the centre of the computer screen. The stimulus was accompanied by the text prompt "Category A or B?". Participants freely responded with an A or B key press and immediately received either a brief low (wrong) or high (right) pitched auditory tone concurrent with text containing "WRONG" or "RIGHT" and the correct category label for the stimulus. The correct category label and the stimulus were presented for 1250 ms, followed by a blank screen. After 500 ms the next trial began.

The trials were blocked in groups of 16. Each block consisted of a random ordering of the 16 stimuli. Participants were not made aware of transition between blocks. Category training terminated when participants reached a learning criterion of correctly classifying 12 stimuli in a row or completed 18 blocks (288 trials) without reaching the criterion.

Results and discussion

The proportion of trials correct for each participant was calculated. Remaining trials for participants reaching the learning criterion were scored as correct. Statistical tests found no significant differences between size and brightness for learning feature- or relation-based categories. Therefore analyses collapse across size and brightness sub-conditions and focus on the distinction between feature- and relation-relevant category learning.

The results are displayed later in Table 3. In accord with studies suggesting an advantage for relative judgements (e.g., Garner, 1954; Huttenlocher et al., 2002), participants in a relation-based category were significantly more accurate (.95 vs .76) then those in a feature-based condition, $t(51) = 4.56$, $p < .001$.[1] All 26 participants in the relation-based condition reached the criterion, while only 21 of the 27 participants in the feature condition did so, this difference is significant, Yates' $\chi^2(1,\ N = 53) = 4.49$, $p < .05$. These results confirm the hypothesis that relation-based categories can be easier to learn than feature-based categories.

One explanation for the relation advantage is that relevant commonalities and differences across stimuli could be determined by relative

[1]One possible concern with this performance measure is that it could inflate estimates for conditions in which a large proportion of participants reach criterion, as it assumes that participants reaching criterion would be correct on future trials if they could maintain attention levels on the mastered task. An alternative analysis is to consider only those trials prior to reaching criterion. However, while this analysis suffers from the opposite concern, it yields the same pattern of results when applied to those cases in which the proportion of participants reaching criterion varies widely across conditions. The significant differences found in Experiments 2, 3, and 5 persist, $t(51) = 5.89$, $p < .001$, $t(50) = 3.42$, $p = .001$, and $t(21) = 3.64$, $p = .002$, respectively.

comparisons for relation-based categories, whereas absolute judgements were necessary to learn feature-based categories. Note that this explanation is not rooted in saliency. In fact, Experiment 1's results suggest that feature differences are more salient than relation differences when stimuli are presented simultaneously, which alleviates difficulties in absolute judgement and comparison to memory representations. Using a common stimulus set, Experiment 2's results stand in contrast to previous findings (e.g., Kittur et al., 2004; Waltz et al., 2000) that suggest that relation-based categories are always harder to learn than feature-based categories.

EXPERIMENT 3

The combined results from Experiments 1 and 2 suggest that relation-based categories can be advantaged, because of relative comparisons. Experiments 3 and 4 use more complex category structures, consisting of multiple relevant attributes, which allow for consideration of a broader set of theoretical issues concerning comparison and category learning. Experiments 3 and 4 probe interactions across attributes and the use of alignment processes during relation-based category learning.

Relation-based category structures that are supported by alignment processes should prove easier to learn than their feature-based compatriots, whereas relation-based category structures that conflict with preferred stimulus alignments should hinder learning. Table 1 outlines how the two factors, relative processing (explored directly in Experiment 2) and relational comparison (explored directly in Experiments 3 and 4) predict the ease of learning relation- and feature-based variants of the different category structures. Experiment 3 utilises a non-linear category structure (an XOR on the two relevant dimensions) in which items that are opposite in every respect are members of the same category (see Table 2). The relation-

TABLE 2
Category structures

Attr. 1	*Attr. 2*	*Attr. 3*	*Attr. 4*	*XOR*	*Four-Category*
0	0	0 or 1	0 or 1	A	A
0	1	0 or 1	0 or 1	B	B
1	0	0 or 1	0 or 1	B	C
1	1	0 or 1	0 or 1	A	D

Either both features or both relations are relevant to determining category membership for both the XOR (used in Experiment 1) and Four-Category (used in Experiment 2) structures. For example, with features relevant in the Four-Category structure: large and bright would be A; large and dark, B; small and bright, C; and small and dark, D. In the XOR structure, A and D form one category, and B and C form the other category.

based version of this category structure should be supported by preferred stimulus alignments.

In the XOR category structure used in Experiment 3, the most dissimilar items, those that differ on both relations or both features (see Figure 1), are placed within the same category. However, it is easy to see that the initial application a swapping transformation, or by mapping the stimuli based on role instead of position, during the comparison process makes the top and bottom stimuli identical. The similarity data from Experiment 1 support the idea that participants naturally apply this when comparing those pairs. This means that the within-category similarity for the relational XOR is probably much higher than it is for the feature-based equivalent, and this should speed learning (Lassaline & Murphy, 1997; Rosch & Mervis, 1975). Swapping transformations and cross-mapping operations are not advantageous for the feature-based categories; this is shown by the monotonic decrease in similarity with the number of feature changes found in Experiment 1.

The congruence between the category structure and the participants' preferred manner of mapping the stimuli, combined with advantages for relative judgements, should result in categories defined by relations being acquired more readily than comparable categories defined by features.[2] This avenue for boosting coherency should not be available to learners of feature-based categories. Finally, overall performance should be lower in Experiment 3 than in Experiment 2, because the categories in Experiment 3 are defined by two attributes whereas categories in Experiment 2 are defined by a single attribute.

Method

Participants. A total of 52 undergraduate students from the University of Texas at Austin participated for course credit. Participants were randomly assigned to the feature- ($n = 27$) or relation-relevant ($n = 25$) condition.

Stimuli and category structure. The stimuli were the same used in Experiment 1, pairs of circles varying along two relational and featural attributes. Two of the four stimulus attributes were relevant to determining category membership. An exclusive disjunction (XOR) rule involving the two relevant stimulus attributes defined the category structures. XOR is a non-linear classification rule that requires attention to both of the relevant

[2]Experiment 4 will provide a counterpoint in which this flexibility impedes learning of relation-based categories.

TABLE 3
Summary of results from Experiments 2, 3, and 4

		Relations rlevant			*Features relevant*			
	Rule	*Accuracy*	*SE*	*Criterion*	*Accuracy*	*SE*	*Criterion*	*Accuracy Differences*
Exp. 2	1-D	.95	.02	26/26	.76	.04	21/27	.19***
Exp. 3	XOR	.73	.04	14/25	.54	.02	4/27	.19***
Exp. 4	Four Category	.78	.04	22/26	.74	.04	22/27	.04

***$p < .001$.

attributes (see Table 2). Stimuli that are opposite one another on both relevant attributes are placed in the same category. In the relation condition, the two relation-based attributes were relevant and the features were irrelevant. The opposite was true for the feature condition.

Procedure. Training followed the same pattern as in Experiment 2.

Results and discussion

The proportion of trials correct for each participant was calculated. Remaining trials for participants reaching the learning criterion were scored as correct. The results are summarised in Table 3. As expected, inspection of Table 3 reveals that Experiment 3's more complex (defined by two attributes) category structures were more difficult to acquire than Experiment 2's simple (defined by one attribute) category structures.

Participants were significantly more accurate (.73 vs .54) in the relation relevant than in feature-relevant condition, $t(50) = 5.19$, $p < .001$. A significantly greater proportion of participants (14/25 vs 4/27) reached the learning criterion in the relation relevant condition than in the feature-relevant condition, Yates' $\chi^2(1, N = 52) = 8.00$, $p < .01$. Experiment 3's results demonstrate that more complex categories defined by the relations can be easier to learn than categories defined by features.

EXPERIMENT 4

The results from Experiments 2 supports the idea that relation-based categories can be easier to learn than feature-based ones when the categories are balanced on the number of perceptual attributes and those attributes are spread across multiple objects. Experiment 3 further suggested that the flexibility afforded by online, relational comparisons can benefit relational learners, particularly those who acquired the complex XOR category structure used in Experiment 3. Experiment 4 tests this hypothesis more fully, by attempting to match the speed of relation-based learning and feature-based learning by utilising a category structure that is not amenable to relational alignment.

Experiment 4 provides further evidence for this online comparison account by training participants on a category structure in which alignment processes should not be advantageous to relational learners. Compared to Experiments 2 and 3, differences in performance between featural and relational learners are predicted to compress under these conditions. The category structure used in Experiment 4 is the four-category structure specified in Table 2. Unlike the XOR category rule used in Experiment 3, in the four-category structure items that differ on both relevant attributes are members of different categories.

Unlike Experiment 3, regularising relational differences in Experiment 4 through online relational processes will not increase within category similarity, because stimuli that differ on both relations are now in separate categories. As a consequence of this, relational learners engaging in such relational processes may in fact increase confusions between categories that differ on both relations, leading to opposite classification errors (e.g., confusing members of categories A and D, or B and C in Table 2).

While it is difficult to make cross-experimental comparisons between the three experiments, because they contain a different number of categories, the key predictions for Experiment 4 are that the difficulty of featural and relational learning should converge. Instead of boosting performance as in Experiment 3, online relational comparison in Experiment 4 should manifest itself in more errors to the opposite category for relational learners.

Method

Participants. A total of 53 undergraduate students from the University of Texas at Austin participated for course credit. Participants were randomly assigned to the relation- ($n = 26$) or feature-relevant ($n = 27$) condition.

Stimuli and category structure. The stimuli were the same as those used in Experiments 1–3. Participants learned to classify each stimulus as a member of one of four different categories. The categories were the four unique combinations of the two values of the two relation attributes in the relation condition and the two feature attributes in the feature condition (see Table 2).

Procedure. The procedure was identical to that used in Experiments 2 and 3, except that participants had to learn to classify the circles as belonging to one of four categories: A, B, C, or D, by pressing the corresponding key.

Results and discussion

The proportion of trials correct for each participant was calculated. Remaining trials for participants reaching the learning criterion were scored as correct. The results are summarised in Table 3. Accuracies were comparable (.78 vs .74) in the relation- and feature-relevant conditions, $t < 1$. A comparable proportion of participants (22/26 vs 22/27) reached the learning criterion in the relation- and feature-relevant conditions, Yates' $\chi^2(1, N = 53) = .0039$, $p = .95$.

The pattern of participants' errors was also analysed. Each incorrect response was classified as either a mistake to an adjacent category (e.g., $A \rightarrow B$ or C) or as a mistake to the opposite category (e.g., $A \rightarrow D$). As predicted, participants in the relation-relevant condition made a larger (34% vs 27%) proportion of errors to the opposite category than did participants in the feature-relevant condition, $t(51) = 2.77$, $p < .01$. Relational learners who tended to make a higher proportion of opposite category errors relative to adjacent category errors had lower overall accuracy levels, $R^2 = .30$, $F(1, 24) = 10.49$, $p < .01$ whereas no such relationship held for feature learners, $R^2 = 0$.

As predicted, the relational advantage observed in Experiment 3 was not observed in Experiment 4. One question is whether this difference across experiments is statistically significant. Directly comparing Experiments 3 and 4's results is difficult, because chance performance levels differ across experiments (50% vs 25%). However, one simple correction for guessing is to model correct responses as arising from either knowing the correct response or correctly guessing:

$$p(correct) = p(know) + [1 - p(know)] * p(correct|guess). \tag{1}$$

In Equation 1 the second term accounts for the difference in guessing between the two experiments. Equation 1 can be solved to estimate the probability that a participant knows the correct response on a trial:

$$p(know) = \frac{p(correct) - p(correct|guess)}{1 - p(correct|guess)}. \tag{2}$$

Using this adjusted measure, a 2 (relation or feature) × 2 (XOR or Four-category) ANOVA was conducted. The interaction between condition and experimenter was significant, $F(1, 101) = 6.37$, $p = .01$. A learning criterion analysis, which is less susceptible to guessing artefacts, is also supportive of an interaction between the two variables. A Poisson-based general-linear-model analysis of the counts of participants reaching criterion indicates a marginally significant three-way interaction (condition, category type, and criterion) across experiments, $z = 1.77$, $p = .08$.

The results of Experiment 4 suggest that when a category structure is used in which online comparison processes are not beneficial to learning relation-based categories, featural and relational learning are of equal difficulty. Indeed, online comparison processes were manifested as a greater proportion of opposite category errors in the relation relevant condition and relational learners who showed stronger markers of online comparison processes were less accurate overall.

EXPERIMENT 5

The previous studies support the notion that relational comparison processes can affect category learning by increasing (Experiment 3) or decreasing (Experiment 4) category coherency. We hypothesised that these effects arise because two stimuli differing in both relational attributes can be made more similar by a relational alignment (or swapping transformations) that puts circles that differ in left/right spatial position into correspondence. This view is also supported by the similarity data from Experiment 1 (see Figure 2).

A contrasting view is that these results are not driven by a relational comparison processes, but are a by-product of the participants' default representation of the stimuli. In Experiment 5 we probe participants' preferred correspondences of compared stimuli to more directly test whether these relational alignments (or swapping transformations) occur. Additionally, we examine how category learning can affect these comparison processes. Whereas the previous studies support the notion that alignment affects coherency, in Experiment 5, we examine whether preferred alignments are altered to maximise category coherency.

To answer this question, as in Experiment 3, participants first were trained on either feature-based or relation-based categories with an XOR structure. Following training, participants compared two stimuli and reported which circles corresponded across the two stimuli. We predict an interaction such that participants in relation-relevant condition will be more likely than participants in the feature-relevant condition to put circles in correspondence that differ in spatial position when stimuli differ in both relations compared to when the stimuli match on both relations. This prediction can also be viewed as testing whether attention shifting-like phenomena (e.g., Kruschke, 1992) extend to relational stimuli (e.g., Tomlinson & Love, 2006).

Method

Participants. A total of 21 undergraduate students from the University of Texas at Austin participated for course credit. Participants were randomly assigned to the feature- ($n = 11$) or relation-relevant ($n = 10$) condition.

Stimuli. The stimuli were the same as those used in Experiments 1–4.

Procedure. The learning phase of this experiment was conducted exactly as Experiment 2, with the same instructions detailing the manner in which the stimuli varied. Participants learned either a relation-based XOR or the feature-based XOR category structure. After the participants could correctly categorise 12 stimuli in a row, or after 288 trials, the participants were transferred to the second phase of the experiment.

In the second phase the participants were instructed that they would see two stimuli and that one of the circles in one of the stimuli would be highlighted in red. The participant's task was to determine which circle in the other stimulus corresponded to the highlighted circle. Following this judgement, participants were instructed that they would then be asked to rate the similarity of the two presented stimuli on a scale from 1 to 9. Participants were alerted that they could take short breaks between responses to help maintain concentration.

On each trial, two stimuli were simultaneously presented on screen with text designating pair 1 and pair 2, along with text asking them to pick the circle that went with the highlighted circle. One pair was displayed on the top of the screen and the other on the bottom. A line separated the pairs. One of the circles (randomly determined each trial) was displayed with a red box around it. The participant used the mouse to select which of the two circles from the other stimulus corresponded to the highlighted circle. Immediately following this judgement, text appeared asking the participant to rate the similarity of the two stimuli on a scale of 1–9. Participants responded by pressing a key, 1 through 9. Following the participant's response, the screen blanked for 500 ms and the next trial began. The stimuli were presented in a random order and no participant saw a stimulus paired with itself, nor the same pairing twice. Unfortunately, due to a coding error, each participant saw only 105 of the 120 possible stimulus pairs (randomly determined for each participant).

Results and discussion

For the learning phase the proportion of trials correct for each participant was calculated. Remaining trials for participants reaching the learning criterion were scored as correct. The results replicated those from Experiment 2. Participants were significantly more accurate (.89 vs .57) in the relation-relevant than in feature-relevant condition, $t(19) = 4.52$, $p < .001$. Similarly, a significantly greater proportion of participants (9/10 vs 4/11) reached the learning criterion in the relation relevant condition than in the feature-relevant condition, Yates' $\chi^2(1, N = 21) = 4.32$, $p < .05$.

Turning our focus to the correspondence judgements, we calculated the proportion of times each participant selected circles as corresponding that matched in spatial position (left or right). The overall proportions (.45 vs .46) for the relation- and feature-based participants was not significantly different, $t < 1$.

To test our hypothesis that participants in the relation-based condition should favour positional correspondences when stimuli are relationally similar, but disfavour such correspondences when stimuli are relational dissimilar, we calculated the proportion of positional correspondences for stimulus pairs that matched or mismatched on both relations for both the relation- and feature-based condition participants. These four means are shown in Figure 4. A 2 (relational match/mismatch) × 2 (Condition) mixed ANOVA found that the predicted interaction was significant, $F(1, 19) = 11.05$, $p < .01$. Additionally there was a main effect of relational match/mismatch on the probability of a participant aligning the circles based on position, $F(1, 19) = 48.77$, $p < .001$. There was not a significant main effect of condition, $F(1, 19) = 2.36$, $p = .14$. Further investigating the interaction, relation-based learners were marginally more likely to map by position when no relations changed than were the feature-based learners, $t(19) = 1.80$, $p = .08$, whereas they were significantly less likely to map by position than feature-based learners when the stimuli differed on both relations, $t(19) = -4.5$, $p < .001$.

The previous analysis confirms our predictions, but two questions remain in regards to the feature-based condition, namely whether participants are choosing correspondences at random and whether shared category membership affects correspondence judgements. To evaluate these, a supplementary ANOVA was conducted using the correspondence judgements from the feature-based participants. A 2 (relational match/mismatch) × 2 (category match/mismatch) within-subject ANOVA was conducted and found no interaction, $F < 1$. nor a main effect of the category variable, $F(1, 10) = 2.00$, $p = .19$. Although it is dangerous to over interpret null effects, the lack of any significant effects involving the category variable do ease concerns in our interpretation of the data shown in Figure 4. The ANOVA did indicate

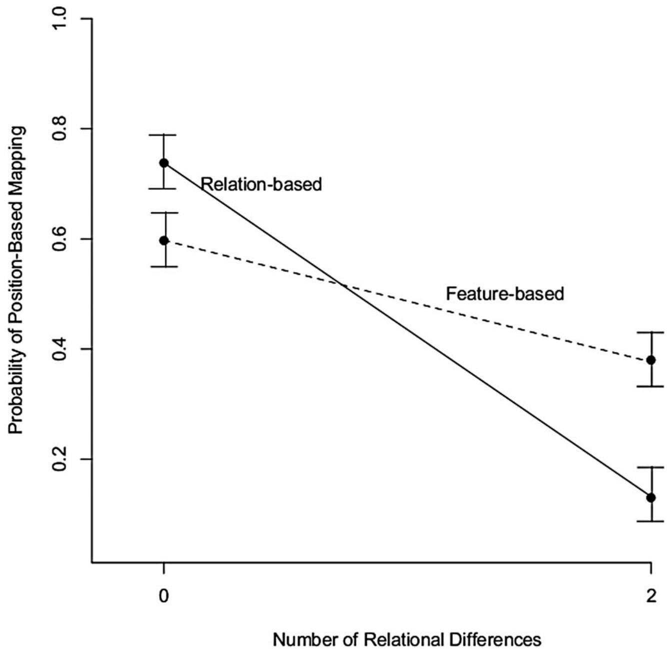

Figure 4. The mean probability that participants in the feature- and relation-based learning conditions establish stimulus correspondences based on position as a function of whether there were 0 or 2 relational differences between the stimuli. Error bars represent approximate 95% confidence intervals.

a main effect for relational match, $F(1, 10) = 13.67$, $p < .01$. Rather than determining correspondences randomly, feature-based condition participants were more likely (.60 vs .40) to prefer position matches when stimuli matched than mismatched in their relations, $t(10) = 3.70$, $p < .01$.

In summary, these analyses support our notion that participants do align stimuli that mismatch on both relations in a non-positional manner (or alternatively perform a relational, swapping transformation). Additionally, as we predicted, participants are more likely to do so after learning relation- than feature-based categories. This finding is consistent with the notion that attention-shifting phenomena extend to relational processing. As in strictly feature-based models of attention (e.g., Kruschke, 1992), the present findings indicate that attributes (featural and relational) that are more predictive of category membership receive greater weight. Experiment 5's results indicate that these weightings can guide the preferred alignments when comparing stimuli (see Tomlinson & Love, 2006, for a related modelling approach).

GENERAL DISCUSSION

In accord with previous studies, Experiment 1 showed that stimuli containing relations can exhibit alignment-based non-monotonicities in their similarity structure. Contrary to accepted wisdom, the results of Experiment 2 demonstrated that learning to classify by relations can be easier than by features when the categories are balanced on the number of perceptual attributes and those attributes are spread across multiple objects. Experiment 3 showed that this effect persists in a more complicated category structure, when the category structure is supported by the preferred alignments between the stimuli based on the relations. Experiment 4 employed a category structure for which online comparisons could be detrimental to relational learners and this structure tempered any inherent advantage for relations, as relation- and feature-based categories were acquired at the same rate. Indeed, relational learners who showed stronger markers of online comparison processes were less accurate overall. Experiment 5 supported this conclusion by demonstrating that learners in the relation-based condition preferred different correspondences for stimulus pairs than those in the feature-based condition, suggesting that the stimulus representations are flexible, and change to increase category coherency.

The combined results from Experiments 1–5 advance our understanding of the role of online, comparison processes during learning and preclude alternative explanations based on general biases in favour of features or relations. These results are important because they suggest revisiting findings demonstrating relational deficits, not just in adults and children, but in special populations, such as those suffering from schizophrenia (Johnson, Lowery, Kohler, & Turetsky, 2005) and Alzheimer's disease (Waltz et al., 2004), using the methods and well-matched stimulus set developed here.

These results suggest a developmental progression from appreciating feature matches to grasping more complex relational matches (Gentner, 1988; Gentner & Ratterman, 1991; Keil & Batterman, 1984; Kotovsky & Gentner, 1996; Medin, Goldstone, & Gentner, 1990; Oakes & Cohen, 1990; Richland, Morrison, & Holyoak, 2006). Even though relational processing generally requires more cognitive resources, its flexibility offers potential advantages over purely featural approaches.

Here we strengthen the link between work in relational processing and category learning. Earlier work, such as Lassaline and Murphy (1997), demonstrated the importance of considering feature matches across attributes (through alignment) in determining the difficulty of learning feature-based categories. Good alignments led to faster learning. We extend Lassaline and Murphy's findings to multi-place predicates, which allows for exploration of relation-based category learning. Our results support the view that the flexibility of relational processing can help or hinder learning with benefits observed when online comparison processes boost category coherency.

The vast majority of category-learning models only consider cases in which comparison processes are predetermined (e.g., Kruschke, 1992; Love, Medin, & Gureckis, 2004; although for exceptions see Goldstone & Medin, 1994; Kuehne, Genter, & Forbus, 2000; Tomlinson & Love, 2006). Thus the present work bears on the necessity of making provisions for relational processing in category learning theories and models. In addition, this work demonstrates the necessity for analogical models that include a learning component. Experiment 5 provides results that cannot be accounted for by static models of analogy, as the two groups mapped the stimuli differently depending on the learning conditions. We believe that our results suggest integration of research in analogy and category learning (e.g., Tomlinson & Love).

Our results do not uniquely favour alignment or transformational views of relational comparison. However our results, as well as response time measures (collected unbeknown to the participants), do constrain the specific form successful theories can take. One interesting finding is that response time in Experiment 1's similarity rating task was highly correlated ($r = -.93$) with rated similarity (see Figure 5). According to the alignment view, more readily aligned stimuli result in a feeling of fluency, which influences rated similarity (cf. Johnston, Dark, & Jacoby, 1985). According to the transformational view, as the number of transformations performed (or their difficulty) increases, response time should increase and rated similarity decrease (cf. Shepard & Metzler, 1971).

Response time was also collected during category learning in Experiments 2, 3, and 4. Unfortunately, large differences in learning accuracy and in the proportion of participants reaching the learning criterion for Experiment 2 and 3's relation and feature conditions rendered meaningful analysis of response times untenable. In Experiment 4 accuracy levels for the two conditions were roughly equal and, thus, comparing response times is meaningful. The mean of each participant's correct median response time was significantly greater (2501 ms vs 1892 ms) in the relation-relevant than in the feature-relevant condition, $t(42) = 2.72$, $p < .01$. According to the alignment view, this difference follows from the complexity of determining relational correspondences. According to the transformational view, this difference is interpreted as a greater difficulty in executing relational transformations.

The above response time biases also contribute an interesting data point with regard to whether our stimuli were biased against the features. Duncan (1984) suggests that processing time for attributes is slowed when the

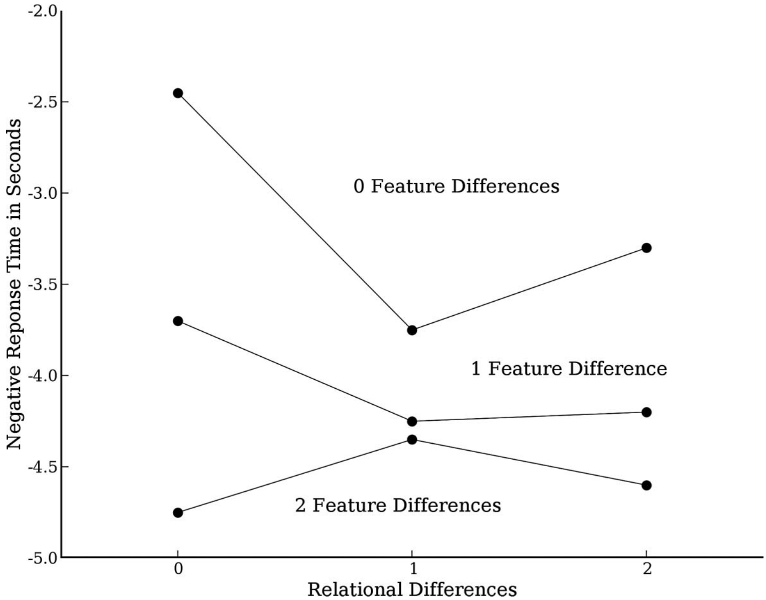

Figure 5. The mean of each participant's median response time was calculated for each cell and its negative is displayed to ease comparison with the strongly correlated similarity rating data shown in Figure 2. To reduce visual cluster, error bars are not included, but 95% confidence intervals on the means are approximately $\pm$ .55 seconds.

attributes are spread across multiple entities; however the above reaction times suggest that the participants were still significantly faster at processing the features than the relations, even when operating at the same level of accuracy. While this is not conclusive, it certainly warrants further investigation. One possibility is that the majority of response time is due to comparison processes rather than to the perception of attribute values.

Overall, our findings add to a growing body of evidence that suggests a central role for relational processing in categorisation. Many real-world categories have a strong relational basis (Gentner & Kurtz, 2005; Markman & Stilwell, 2001), as do many of the features that constitute categories that we do not view as relational (Jones & Love, 2007). The present work complements this line of inquiry by examining relation-based learning of novel categories in a rigorously controlled experimental setting. While limited to consideration of simple perceptual relationships, we hold hope that the present findings will inform future work in other domains, as previous work in relational comparison has revealed common underpinnings in perceptual and conceptual domains (cf. Falkenhainer, Forbus, & Gentner, 1989; Kotovsky & Gentner, 1996). We predict that relational learning in many domains will be advantaged in situations in which learners can exploit mappings or relational transformations that increase category coherence.

The interplay of featural and relational processes

Although we have framed featural and relational processes as distinct, these processes likely interface to the benefit of the overall cognitive system. One possibility is that useful relational comparisons give rise to new primitive features (cf. Doumas, Hummel, & Sandhofer, 2008). Such a division of labour is desirable given the distinct strengths of relational and featural processing. Relational processes can be characterised as powerful and flexible, but require capacity-limited cognitive resources (cf. Halford, Wilson, & Phillips, 1998; Waltz et al., 2000). In contrast, feature-based comparison is somewhat inflexible, but less costly in terms of cognitive resources. Even for our simple perceptual learning tasks, response times for relationally guided judgements were greater. Creating new features to recode frequent relational correspondences would allow the cognitive system to appreciate relational commonalities without straining limited cognitive resources. Furthermore, these new features can serve as inputs to relational processes that in turn lead to the discovery of further category regularities. Productions systems that compile operations into new chunks follow a similar logic (e.g., Rosenbloom & Newell, 1983).

Consideration of metaphor use and concepts rarified in language support our account of the interplay of featural and relational processes. Certain productive metaphors, such as understanding time in terms of space (e.g., "They moved the meeting forward two hours"), appear to be processed in an online manner (Casasanto & Boroditsky, 2008). In contrast, other metaphors do not appear to be processed online, but rather are polysemies or frozen idioms (Keysar & Bly, 1995; Murphy, 1996, 1997). According to our proposal for the interface of featural and relational processes, these "dead metaphors" are akin to the features generated by relational processes (i.e., repeated application of comparison processes). Likewise, many relational concepts (e.g., island, thief, predator) are reflected in the lexicon and perhaps the existence of these labels eases processing requirements and allows for appreciation of more complex relational regularities. Paralleling our proposal, this progression from online processing via analogical processes to operations that rely on static (i.e., pre-aligned) representations has been used to explain differences seen in the processing of conventional (e.g., "A soldier is a pawn") and novel (e.g., "The mind is a kitchen") metaphors (Bowdle & Gentner, 2005).

Our experimental methods may prove useful for bringing these ideas into the laboratory for systematic and controlled evaluation. In Experiment 3, Category A in the relation-relevant condition was defined by the bigger circle being on the left and the darker circle also being on the left, or the smaller circle being on the left and the lighter circle also being on the left (see Figure 1). As previously discussed, stimulus items that were opposite on these two relational attributes can be made more similar by a cross-mapping or swapping operation. Across studies, our results strongly suggest that participants engaged in these operations. Alternatively one could learn these categories by creating the new relational feature ''the bigger circle of the pair is also the darker one'' to unite all of Category A's members. Essentially, such a feature transforms the relational XOR category structure into a single attribute category structure, much like the category structures used in Experiment 2. Unfortunately, it is not clear whether any participants in the present studies discovered such emergent, relational features. Certainly participants did not quickly discover such a feature as the relational XOR in Experiment 3 proved much more difficult for participants to learn than Experiment 2's single-attribute relational category structure.

Systematic exploration of feature creation operations (cf. Schyns, Goldstone, & Thibaut, 1998) through relational mechanisms awaits future research. Our basic framework might prove useful in tackling this important issue. Following extensive training, we predict that participants will create new relational features in situations in which such features regularise category structure (e.g., Experiment 2, relational XOR category structure), but not in cases in which such features would highlight commonalities across contrasting categories (e.g., Experiment 4, relational four-category structure). The overarching idea is that relational processes can create new features that ease processing requirements and promote category coherency through relational comparisons (Kurtz, Miao, & Gentner, 2001). Whereas Rosch and Mervis (1975) focused on how the structure of the environment biases acquisition towards categories that have high within- and low between-category similarity, our findings suggest the cognitive machinery provided for online comparisons can exert a strong influence in regularising categories to conform to the Rosch and Mervis ideal.

Manuscript received 20 May 2010
Revised manuscript received 21 September 2010

REFERENCES

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*, 193–216.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*, 579–593.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist?: Developmental shifts in ways of clustering knowledge. *Child Development*, *75*(3), 918–931.

Doumas, L. A. A., Hummel, J., & Sandhofer, C.M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*, 1–43.

Duncan, J. (1984). Selective attention and the organisation of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517.

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The Structure Mapping Engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1–63.

Garner, W. R. (1954). Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, *48*(3), 218–224.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, *49*, 988–998.

Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, *4*, 161–178.

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

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*, 47–59.

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorisation inside and outside the laboratory: Essays in honor of Douglas l. Medin* (pp. 151–175). Washington, DC: American Psychological Association.

Gentner, D., & Ratterman, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London: Cambridge University Press.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 3–28.

Goldstone, R. L. (1996). Alignment-based nonmonotonicities in similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 988–1001.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 29–50.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgements. *Cognitive Psychology*, *23*, 222–262.

Hahn, U., Chater, N., & Richardson, L. (2003). Similarity as transformation. *Cognition*, *87*, 1–32.

Halford, G. S., Wilson, W. H., & Phillips, W. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral Brain Sciences*, *21*(6), 803–831.

Huttenlocher, J., Duffy, S., & Levine, S. (2002). Infants and toddlers discriminate amount: Are they measuring? *Psychological Science*, *13*, 224–249.

Johnson, S., Lowery, N., Kohler, C., & Turetsky, B. (2005). Global-local visual processing in schizophrenia: Evidence for an early visual processing deficit. *Biological Psychiatry*, *58*(12), 937–946.

Johnson, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 3–11.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 316–332.

Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology*, *51*, 196–231.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior*, *23*, 221–236.

Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*, *34*, 89–109.

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2004). Feature- vs. relation-defined categories: Probab(alistic)ly not the same. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual conference of the Cognitive Science Society* (pp. 696–701). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorisation in the development of relational similarity. *Child Development*, *67*, 2797–2822.

Kroger, J. K., Holyoak, K. J., & Hummel, J. E. (2004). Varieties of sameness: The impact of relational complexity on perceptual comparisons. *Cognitive Science*, *28*, 335–358.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Kuehne, S., Gentner, D., & Forbus, K. (2000). Modeling infant learning via symbolic structural alignment. In *Proceedings of the 22nd conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Kurtz, K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, *10*(4), 417–446.

Lassaline, M. E., & Murphy, G. L. (1997). Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 144–160.

Love, B. C., & Markman, A.B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition*, *31*(5), 790–799.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Love, B. C., Rouder, J., & Wisniewski, E. J. (1999). A structural account of global and local processing. *Cognitive Psychology*, *38*, 291–316.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence*, *13*, 329–358.

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

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*, 173–204.

Murphy, G. L. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition*, *62*, 99–108.

Oakes, L., & Cohen, L. (1990). Infant perception of causal events. *Cognitive Development*, *5*, 193–207.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between humans and non-human minds. *Behavioral and Brain Sciences*, *31*, 109–178.

Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *27*(5), 1261–1275.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*, 249–273.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rosenbloom, P. S., & Newell, A. (1983). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Proceedings of the International Machine Learning Workshop* (pp. 183–197), Urbana-Champaign, IL.

Schyns, P. G., Goldstone, R. L, & Thibaut, J. (1998). Development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87.

Shepard, R. N., & Metzler J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in categorizing tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 416–430.

Stewart, N., Brown, G. D., & Chater, N.(2005). Absolute identification by relative judgement. *Psychological Review*, *112*, 881–911.

Tomlinson, M. T., & Love, B. C. (2006). From pigeons to humans: Grounding relational learning in concrete exemplars. In *Proceedings of the 21st National Conference on Artificial Intelligence* (pp. 199–204). Menlo Park, CA: AAAI Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(2), 327–352.

Waltz, J. A., Knowlton, B., Holyoak, K., Boone, K., Back-Madruga, C., McPherson, S., et al. (2004). Relational integration and executive function in Alzheimer's disease. *Neuropsychol-gia*, *18*(2), 296–305.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*, 1205–1212.