In press, Journal of Experimental Psychology: General

Learning Relational Concepts from Within- vs. Between-Category Comparisons

Daniel Corral
University of Colorado Boulder

Kenneth J. Kurtz
Binghamton University

Matt Jones
University of Colorado Boulder

This article examines relational category learning in light of two influential theories of concept acquisition: the structure-mapping theory of analogy and theories of feature-based category learning. According to current dominant theories of analogy, comparing two instances of a relational concept enables alignment of their elements and reveals their shared relational structure. Therefore, learning relationally defined categories should be faster when comparing items of the same category than when comparing items of different categories. By contrast, feature-based theories predict a benefit of between-category comparisons, because such comparisons direct attention to the features that discriminate the categories. The present experiments tested these predictions using a two-category classification-learning task in which two items are presented on every trial: either in the same category (match condition) or in different categories (contrast condition). Subjects in the contrast condition outperformed those in the match condition for feature-based categories (Experiment 1) and across four different types of relational categories (Experiments 1-4). Although theorists have posited that structure-mapping theory is directly applicable to relational category learning, the present findings pose a distinct challenge to this assertion. We propose that many relational categories are learnable based solely on which relations are present in the stimulus, rather than requiring explicitly compositional representations based on role-filler binding. This process would be akin to feature processing and would not require structural alignment. This theoretical proposal together with the empirical results may lead to a better understanding of when people do and do not engage in the cognitively demanding process of structural alignment.

Keywords: Analogy, Relational Category Learning, Comparison, Concept Representation, Compositional Representation

Structured relational concepts have been argued to be central to the power of human cognition (Gentner, 2010; Gentner & Kurtz, 2005; Penn, Holyoak, & Povinelli, 2008; see also Fodor & Pylyshyn, 1988). The most impressive feats of the human mind, including scientific and technological innovation, mathematics, problem solving, natural language and formal logic, all make use of representations that bind objects to roles within relational systems, as opposed to simpler associative mechanisms. An important question is thus how relational concepts are learned. Two influential domains of research that bear on this question are those on feature-based category learning and analogical learning and reasoning. These two traditions offer fundamentally different views on concept acquisition, the former based on feature vectors and global similarity (Estes, 1986; Nosofsky, 1986), and the latter on internal relations and one-to-one correspondence among constituent elements (Gentner, 1983).

Recently there has been increasing interest in relational category learning, an empirical domain lying at the intersection of category learning and analogy (Corral, 2017; Corral & Jones, 2014, 2017; Dietrich, 2010; Foster, Cañas, & Jones, 2012; Goldwater & Gentner, 2015; Goldwater, Don, Krusche, & Livesey, 2018; Jung & Hummel, 2011; Kittur, Hummel, & Holyoak, 2004; Kurtz, 2015; Patterson & Kurtz, under review; Lassaline & Murphy, 1998). Instead of categories defined by features, rules, similarity, or family resemblance, relational category learning is concerned with categories defined by shared relational structure. For example, Foster et al. (2012) trained people to discriminate two types of outcome for an alien spaceship tournament. Each tournament comprised three ships racing in pairs. In one category, the results formed a cycle (e.g., A beats B, B beats C, C beats A), and in the other they formed a hierarchy (e.g., A beats B, B beats C, A beats C). Every trial contained the same elements (three ships, three pairwise races) and the ships varied randomly across trials, and thus the only information shared within each category was how these elements fit together into a relational system. The experiment followed a standard classification paradigm, in which the subject viewed a stimulus (a tournament), classified it as category 1 or 2, and then received corrective feedback. After several dozen trials most subjects were able to induce the correct concepts and reliably classify the stimuli.

The goal of the present work is to test differential predictions of accounts of relational category learning derived from theories of feature-based category learning and from theories of analogy. The starting point is the role of comparison between stimuli. Comparison has been shown to facilitate many high-level cognitive tasks, including concept learning, analogical reasoning, problem solving, and decision-making (for a review and meta-analysis, see Alfieri, Nokes-Malach, & Schunn, 2013). Classic research on

Daniel Corral and Matt Jones, Department of Psychology and Neuroscience and Institute of Cognitive Science, University of Colorado Boulder. Kenneth Kurtz, Department of Psychology, Binghamton University. This research was supported by AFOSR grants FA9550-10-1-0177 and FA9550-14-1-0318. Part of this work was presented at the 54th Annual Meeting of the Psychonomic Society, Toronto, Canada. All data are available on the Open Science Framework at https://osf.io/wudkf/?view_only=8e55a81929514a8ba474f0f5a9f0bce3. Address correspondence to Matt Jones at mcj@colorado.edu.

analogical transfer showed that comparing analogous story problems facilitates learning a solution schema that can be applied to future problems (Gick & Holyoak, 1983). Related work has found that examining side-by-side examples can improve mathematical problem solving (Ward & Sweller, 1990). Work in the field of education has shown that comparison can improve student learning in the classroom (Bransford & Schwartz, 1999; Schwartz & Bransford, 1998). Likewise, research on conceptual development has shown that object comparison can aid conceptual understanding and expedite the process of category formation (Gentner & Namy 1999; Namy & Gentner, 2002), providing a route to acquiring increasingly abstract concepts (Kotovsky & Gentner, 1996). Thus comparison seems to play a critical role in discovering and highlighting concepts that are relevant to a given task.

Comparison plays a critical role in theories of analogy, as exemplified by structure-mapping theory (Gentner, 1983), the dominant framework in that domain. Structure-mapping theory holds that stimuli are represented compositionally, as systems of objects and relations linked by role-filler bindings (Markman & Gentner, 2000), such as *under*(*object*$_1$, *object*$_2$). Comparing two instances of the same relational concept triggers an alignment process, whereby the instances' elements are put into correspondence in a way that preserves and highlights their shared relational structure (Gentner, 1983, 2003; Hummel & Holyoak, 2003). This alignment process allows for schema induction (Gick & Holyoak, 1983; Holyoak & Thagard, 1997), whereby superfluous properties from each scenario are stripped away and their shared structure is abstracted and represented as a new concept. For example, consider two scenarios, one in which a man uses an umbrella to shelter himself from the rain, and another wherein a cat runs under a tree to stay dry during a storm. These two scenarios are analogous because in both cases an agent places itself under an object to be protected from the rain. Thus there is a common relational structure that both scenarios exemplify. By comparing these scenarios and putting their elements into correspondence (*man* ↔ *cat*, *umbrella* ↔ *tree*), this common structure can be aligned and abstracted (e.g., *cause*[*under*(*agent*, *object*), *protected*(*agent*, *rain*)]), leading to the removal of idiosyncratic features from the original representations (i.e., features specific to men, cats, umbrellas, or trees) and yielding a representation of the abstract relational concept. Thus structure-mapping theory makes the strong prediction, supported in many experiments (Bowdle & Gentner, 1997; Clement & Gentner, 1991; Gentner & Markman, 1994; Gentner, Ratterman, & Forbus, 1993; Gick & Holyoak, 1983; Holyoak & Koh, 1987; Markman & Gentner, 1993a, 1993b), that relational concepts are best learned by comparing items that share a relational structure. Such comparisons allow for alignment and schema induction, whereas alignment is by definition not possible when two items have different relational structures.

Structure-mapping theory and related theories of analogy were originally formulated for cases of reasoning from just one or a few comparisons (Doumas, Hummel, & Sandhofer, 2008; Falkenhainer, Forbus, & Gentner, 1989; Forbus, Gentner, & Law, 1995; Gentner, 1983; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997, 2003; Kokinov, 1988, 1994; Larkey & Love, 2003). This is in contrast to traditional category learning paradigms, in which subjects are typically expected to learn only after scores or hundreds of trials (e.g., Ashby & Lee, 1991; Goldstone, 1994; McKinley & Nosofsky, 1995). Nevertheless, analogical reasoning has been posited to share important psychological commonalities with relational category learning (Dietrich, 2010; Gentner & Namy, 1999; Goldwater & Schalk, 2016; Ramscar & Pain, 1996). Indeed, a relational category can be defined as a set of items sharing a common relational structure (Gentner & Kurtz, 2005; Markman & Stilwell, 2001). Thus judging that two scenarios are analogous or alignable amounts to indicating they are members of the same relational category. Building on this connection, many researchers have proposed extending structure-mapping theory to relational category learning, by assuming that classification involves aligning the current stimulus to previous stimuli or to learned schemas for the categories (Corral & Jones, 2014; Kittur et al., 2004; Kuehne, Forbus, Gentner, & Quinn, 2000; Kurtz, Boukrina, & Gentner, 2013; Lassaline & Murphy, 1998; McLure, Friedman, & Forbus, 2010). This account leads to the prediction that relational category learning should be facilitated by leading subjects to compare items from the same category, whereas between-category comparisons should be of little benefit (Higgins, 2012; Higgins & Ross, 2011).

This predicted advantage for same-category comparison also follows from computational models of relational category learning based on structure-mapping (Barbella & Forbus, 2013; Chang & Forbus, 2013, 2014; Corral & Jones, 2014; McLure et al., 2010; McLure, Friedman, Lovett, & Forbus, 2011; Taylor, Friedman, Forbus, Goldwater, & Gentner, 2011; Tomlinson & Love, 2006). For example, SEQL (Kuehne et al., 2000; Skorstad, Gentner, & Medin, 1988) compares new items to previously stored exemplars or schemas and attempts to align their structures. When the alignment process is successful, meaning the items belong to a common relational category, the model induces a new schema that represents the shared substructure of the items that were aligned, which then serves as a representation of the category. Learning in these models is thus driven by within-category comparisons. In contrast, current models do not have mechanisms that would enable efficient learning from between-category comparisons.

Despite the attempts to link category learning and analogical reasoning, there are important distinctions between them. Learning relational categories over hundreds of trials is a substantially different task than analogical reasoning from as few as two items. Analogical reasoning also supports many functions beyond classification. Partial analogy between two items enables a powerful bootstrapping process, termed analogical transfer, in which a person can build on the alignment to draw further inferences from one item to the other (Gentner, 1983; Gentner & Markman, 1997; Holyoak & Thagard, 1997; Krawczyk, Holyoak, & Hummel, 2005; Markman & Gentner, 2000; Spellman & Holyoak, 1992, 1996). For example, recognizing that two scenarios both contain instances of *support* could lead a learner to infer that properties of one (e.g., the supporting object is sturdy and has greater mass than the supported object) also hold in the other. Although structural alignment enables this form of inferential transfer, the latter is not necessary for recognition of the initial commonality between two scenarios. This is an important point because structural alignment is posited to be cognitively expensive (Forbus et al., 1995) and can strain working memory (Kintsch & Bowles, 2002; Waltz, Lau, Grewal, & Holyoak, 2000). Thus, people might be less likely to engage in careful comparison and alignment over hundreds of trials during a classification task than they are during the sort of one-shot learning tasks that form the core domain of structure-mapping theory.

An alternative to the position of relational category learning as a form of analogy is that relational category learning is best explained by the same principles that have been successful with feature-based categories. The field of feature-based category learning has a long history marked by well-developed models capable of highly accurate quantitative predictions (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Lee, 1991; Gluck & Bower, 1988; Kruschke, 1992; Nosofsky, 1986; Nosofsky & Palmeri, 1997; Smith & Minda, 1998; Tenenbaum & Griffiths, 2001). One principle that is central to many of these models, and that has received extensive empirical support (Goldstone, 1994; Jones, Maddox, & Love, 2005; Kruschke, 1992; Maddox, 2002; Nosofsky, 1986, 1989, 1991; Rehder & Hoffman, 2005), is that of selective attention among stimulus dimensions. Under this principle, an important component of learning is identification of which features or dimensions are most diagnostic of category membership, and shifting attention to these dimensions such that they contribute more to similarity or associative learning (Kruschke, 1992; Mackintosh, 1975; Nosofsky, 1986). Although comparison between stimuli is not explicitly part of this theory, a natural prediction is that discovery of diagnostic dimensions will be facilitated by between-category comparisons, which highlight the way in which members of opposing categories differ. Indeed, existing work using feature-based categories suggests that learning is superior from between-category comparisons relative to within-category comparisons (Andrews, Livingston, & Kurtz, 2011; for related work see Hammer, Hertz, Hochstein, & Weinshall, 2009; Higgins, 2012; Higgins & Ross, 2011). Thus, to the extent that the mechanisms of feature-based category learning also apply to relational categories, relational category learning should be best from between-category comparisons. This prediction directly opposes that derived from structure-mapping theory.

The present work tests between these opposing predictions by building on recent research on the role of comparison in relational category learning. This recent work has modified the standard classification-learning paradigm, so that instead of presenting the subject with a single item at a time to classify, stimuli are presented in pairs and the subject is invited to compare the co-presented stimuli as part of deciding their category memberships. This paired stimulus presentation has been shown to produce better learning and generalization of relational categories (relative to an equal number of item exposures presented one at a time) using a mix of same-category and cross-category pairs (Kurtz et al., 2013) and using uniform pair types under a supervised observational learning mode (Patterson & Kurtz, 2015, 2016, under review).

The experiments reported here extend this approach, to investigate whether it is more beneficial to present item pairs from the same category (to highlight within-category commonalities) or to present item pairs from contrasting categories (to highlight between-category differences). The experiments used a supervised classification task with two categories (as opposed to the three-way classifications used in the work just cited), with two stimuli presented simultaneously on each trial. Subjects in the *match condition* were always shown pairs of stimuli in the same category and were asked whether the items were both members of Category A or both members of Category B (or nonmembers of A, in Experiment 4). Subjects in the *contrast condition* were always shown pairs of stimuli from different categories and asked which item was in Category A and which was in Category B (or which was not in Category A, in Experiment 4). In both conditions, the correct answer was shown after the subject responded. In order to control for any inherent differences in task difficulty and to allow for an equitable comparison of learning between conditions, on every fifth trial subjects were asked to classify a single item presented alone; no feedback was presented on these trials. Comparing performance on these single-item trials provides a direct test of the learning benefits of within- versus between-category comparison.

The novelty of our experiment design, relative to other recent research on categorization of co-presented items, is that it is (a) the first to contrast same-category and different-category comparison in learning of relational categories, and (b) the first to evaluate the effect of comparison in a diverse set of relational category domains. The manipulation of same- versus different-category comparison is critical because of its theoretical connection to posited learning mechanisms within competing accounts of relational category learning. Accounts based on structure-mapping theory predict superior learning in the match condition, because aligning co-presented items from the same category enables their shared relational structure to be discovered and abstracted as a schema that represents the category. Structural alignment should be less useful in the contrast condition, because attempting to align co-presented items from opposing categories will either fail or else yield some partial alignment that serves only to identify the structure common to both categories and hence is useless for discriminating them. On the other hand, accounts based on feature-based theories of category learning predict superior learning in the contrast condition, because comparing co-presented stimuli from opposing categories should aid discovery of diagnostic differences, thus helping subjects to allocate attention in a way that facilitates learning.

We conducted four experiments using five different category structures, all following the above design. Experiment 1 used a feature-based category structure and a relation-based category structure, constructed from the same set of stimuli, manipulated between subjects and crossed with the match/contrast manipulation. For the feature-based category structure, a learning advantage was found for the contrast group over the match group, confirming the prediction from selective attention. In addition, we observed a strong trend in the same direction for relational category structure, supporting the feature-based account of relational category learning over the structure-mapping account. Spurred by this latter result, Experiments 2-4 investigated the effect of comparison type on a wide range of other relational category structures with varying types of stimuli. Despite a concerted effort to find a relational category structure that yielded an advantage for the match condition, all studies demonstrated the opposite, namely that between-category comparisons led to superior learning. Thus the results provide robust support for the feature-based account and challenge the extension of structure-mapping theory. Following presentation of these findings, we discuss their connections to other work in feature-based categorization, including blocking/interleaving effects, as well as implications for structure-mapping theory and its application to category learning. In particular, we propose an explanation of how relational categories might be learned in a feature-based manner, which in turn suggests a natural unification between feature representations and the compositional representations that underpin theories of analogy. We hope this new proposal regarding concept representation, together with the findings of the present experiments, will provide a starting point for better understanding the limits on humans' use of structural alignment as a strategy for learning and reasoning.

Experiment 1

Experiment 1 examined how between- and within-category comparison affect the learning of featural and relational categories. Each stimulus was a pair of geometric objects, with each object characterized by values along three dimensions (size, brightness, and tilt of radius; see Figure 1). Orthogonal to the match/contrast manipulation described above, subjects learned to classify the objects into categories defined either by features or by relations. For subjects in the feature condition, the categories were defined by the objects' separate values on one dimension (e.g., Category A: both objects large, Category B: both objects small). For subjects in the relation condition, the categories were defined by the objects' relative values on one dimension (e.g., Category A: right object larger, Category B: left object larger). Within all four cells of the 2×2 design (feature-match, feature-contrast, relation-match, relation-contrast), the dimension defining the categories was

Figure 1. Example stimulus for Experiment 1 (on two-item trials, two of these were co-presented). The stimulus contains two objects (each a semicircle with a radius, adapted from Shepard, 1964). Each object is defined by values on three dimensions (size, brightness, and radius tilt), but for any given subject only one dimension was relevant for defining the categories. In the feature condition, categories were defined by the absolute values of both objects on the relevant dimension (e.g., both objects large vs. both objects small). In the relation condition, categories were defined by the objects' relative values on the relevant dimension (e.g., left object larger vs. right object larger).

counterbalanced across subjects. The comparisons of interest were between the feature-match and feature-contrast groups, and between the relation-match and relation-contrast groups, to test the effect of comparison type separately on feature-based and relational category learning. (No predictions were made comparing the feature groups to the relation groups, because the relative difficulty of learning featural versus relational categories was not a focus here.) A contrast advantage was predicted for the feature-based categories (i.e., the feature-contrast group outperforming the feature-match group), based on feature-based theories of category learning and selective attention (Nosofsky, 1986). The critical comparison was for the relational categories: If relational categories are learned by similar mechanisms to those governing feature-based learning, then the relational groups should also exhibit a contrast advantage. On the other hand, if relational categories are learned by structural alignment, then the relational groups should exhibit a match advantage. Thus, applying feature-based theories to relational category learning predicts a main effect of comparison, such that the contrast groups should outperform the match groups (irrespective of category type), whereas viewing relational category learning as a form of analogical reasoning predicts an interaction between comparison type (match vs. contrast) and category type (featural vs. relational).

Method

Subjects

One hundred seventy-one undergraduate students from the University of Colorado Boulder participated for course credit in an introductory psychology course. This and all subsequent experiments were approved by the institutional review board at the University of Colorado Boulder.

Stimuli

Each stimulus contained two objects as shown in Figure 1. Each object was defined by values on three dimensions (size, brightness, and radius tilt), and each dimension had four possible values, which were easily discriminable. Thus there were 64 (4^3) possible objects. The objects' brightness values (on a 0-255 gray scale on a standard LCD monitor) were 15, 60, 80, 250; their sizes (in radius) were 1.04, 1.73, 2.07, and 3.20 cm; and their radius tilts were 23°, 49°, 61°, and 85°. The values of each dimension were jointly assigned to the two objects within a stimulus as (1,2), (2,1), (3,4), or (4,3). For example, the pair (1,2) indicates a stimulus in which the left object had brightness level 1 and the right object had brightness level 2. Thus there were 64 possible stimuli (i.e., object pairs), obtained by crossing these 4 value pairs on all 3 dimensions. These assignments enabled construction of feature-based and relational categories from the same set of stimuli, as shown in Figure 2 and described next in the *Design* section.

Design

Subjects were randomly assigned to four conditions that crossed featural versus relational categories with contrast versus match learning: feature-match (N = 43), feature-contrast (N = 44), relation-match (N = 40), and relation-contrast (N = 44). For each subject, categorization depended only on the stimulus values on one dimension (which was counterbalanced within each condition), according to the scheme in Figure 2. In the feature conditions, stimuli in Category A had objects with values 1 and 2 on the relevant dimension (top row of Figure 2), and stimuli in Category B had objects with values 3 and 4 (bottom row). In the relation conditions, stimuli in Category A always had a greater value for the right object than for the left object (left column of Figure 2), and stimuli in Category B always had a greater value for the left object than for the right object (right column). For both the feature conditions and the relation conditions, the values on the other two dimensions were always irrelevant and could be chosen from any of the assignments (1,2), (2,1), (3,4), or (4,3). Thus by this construction the feature-based and relational conditions both used the same set of stimuli, the only difference being how the 64 stimuli were partitioned into two categories of 32 stimuli each.

Procedure

As a cover story, subjects were told that two alien species (Alkins and Bafsters) created different patterns of crop circles, and their task was to learn the difference. Full instructions are presented in Appendix C.

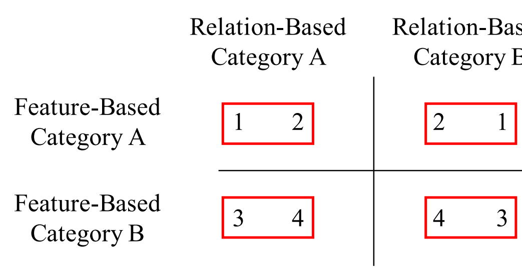

Figure 2. Stimulus values for the relevant dimension in Experiment 1. Each red border indicates a single stimulus, with numerals indicating values of the two objects in that stimulus. For example, if the relevant dimension for a subject were brightness, then the upper-left quadrant would represent stimuli with brightness level 1 for the left object and brightness level 2 for the right object (there are 16 such stimuli, differing in their values on the other two dimensions). Feature-based categories are separated by the horizontal partition (i.e., top vs. bottom rows) and are defined by the absolute values of both objects. Thus, when the relevant dimension was brightness, any stimulus from Category A had two dim objects (brightness levels 1 and 2), whereas any stimulus from Category B had two bright objects (brightness levels 3 and 4). Relation-based categories are separated by the vertical partition (i.e., left vs. right columns) and are defined by the relative values of the objects in each stimulus. Thus, when the relevant dimension was brightness, any stimulus in Category A had a dimmer object on the left than on the right (1 & 2, or 3 & 4), whereas any stimulus in Category B had a brighter object on the left than on the right (2 & 1, or 4 & 3).

Each subject completed 700 classification trials. The majority of these trials were two-item trials, in which two stimuli were generated at random (from the 64 possible stimuli), subject to the constraint that they were in the same category but not identical (match condition) or in opposing categories (contrast condition). Figure 3 shows the display for an example two-item trial in the match (Figure 3A) and contrast (Figure 3B) conditions. In the match condition, each two-item trial included a response prompt, “Type "R" if BOTH crop circles are from Alkins; Type "C" if BOTH crop circles are from Bafsters.” In the contrast condition, the prompt read, “Type "R" if the crop circles are from the Alkins (left) and Bafsters (right); Type "C" if the crop circles are from the Bafsters (left) and Alkins (right).” Responses were self-paced. After each response, the correct labels were presented directly beneath both stimuli, together with the word “Correct” or “Wrong” in the center of the screen. This feedback and the stimuli were displayed together for 800 ms, and the screen was then cleared for 400 ms before the start of the next trial.

Every fifth trial (starting on Trial 5) was a one-item test trial, in which a single randomly selected stimulus was displayed below the response prompt, “Test: Type "A" for Alkins, or "B" for Bafsters.” No feedback was provided on one-item trials. After the response, the stimulus was removed and “Thank You” was displayed for 500 ms in the center of the screen. The purpose of the one-item trials was to provide a direct comparison of learning between the match and contrast conditions, controlling for any possible difference in difficulty of performing the two-item match and contrast classification tasks.

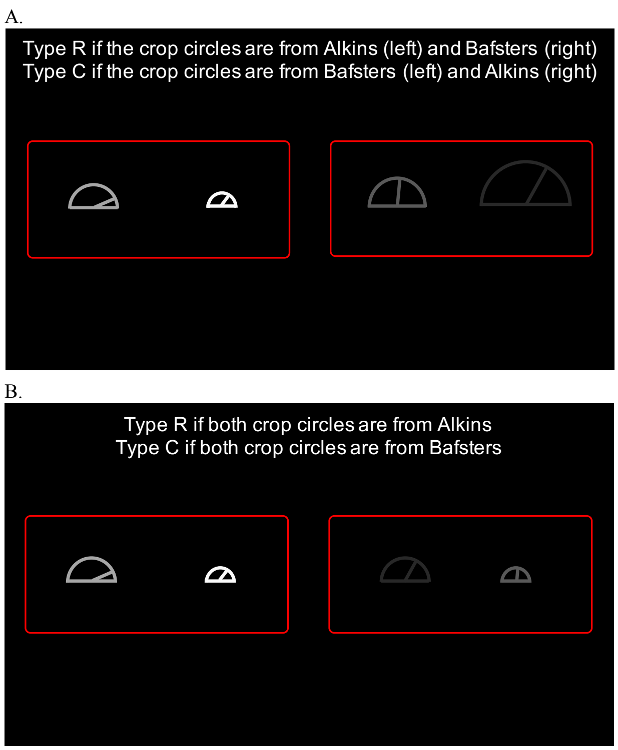

Figure 3. Examples of two-item trials from the match and contrast conditions in Experiment 1. This figure applies to both the relational and feature-based conditions, and in all cases the relevant dimension for this example is size. A: Match conditions. For the match-feature condition, the category is defined by both objects being small, whereas for the match-relation condition the category is defined by the object on the left side of the stimulus being larger than the object on the right side of the stimulus. B: Contrast conditions. For the contrast-feature condition, the category for the left stimulus is defined by both objects being small, whereas the category for the right stimulus is defined by both objects being large. For the contrast-relation condition, the category for the left stimulus is defined by the left object being larger than the right object, whereas the category for the right stimulus is defined by the right object being larger than the left object.

After every 50 trials, subjects were given a self-paced rest break, during which they were shown their percentage correct on those 50 trials and the number of remaining trials in the experiment. The experiment lasted 30-50 minutes.

Results and Discussion

Figure 4 presents individual learning curves, which show the proportion correct by each subject on each block of 50 trials (including both one- and two-item trials). Most subjects show approximately all-or-none learning, with performance at chance until it jumps to being nearly perfect.[1] Although individual curves are difficult to trace, the critical pattern in the figure is the strong bimodal clustering around 50% and 100%. This pattern was found for all other experiments reported in this paper. Although such data can be analyzed based on the proportion of subjects in each condition that are deemed to have learned the categories (by defining an appropriate learning criterion), dichotomizing a quasi-continuous dependent measure can mask information that is present in the data and lead to a reduction in power (DeCoster, Iselin, & Gallucci, 2009). Thus, all data reported in this paper were analyzed based on each subject’s proportion of correct responses throughout the experiment. This measure is sensitive both to whether a subject learned the task and to the approximate point in the trial sequence at which learning occurred; for example, a subject who learned around trial 50 should have a higher proportion correct than a subject who learned around trial 400, who in turn should have a higher proportion correct than a subject who never learned.

Figure 5A shows average learning curves for all four subject groups on one-item trials, and Figure 5B shows the corresponding curves for two-item trials. Because of the nature of the individual learning curves, the group curves are best thought of as indexing

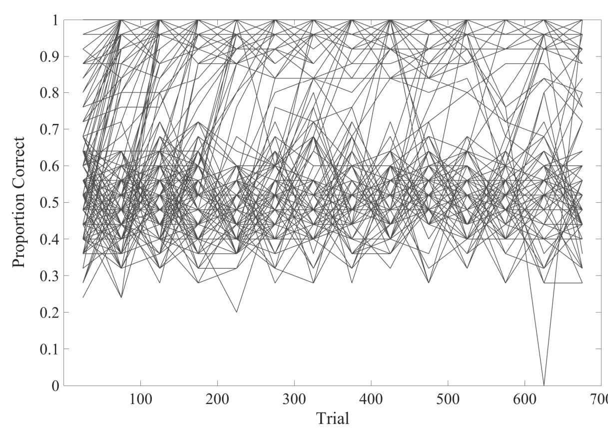

Figure 4. Individual learning curves for all subjects in Experiment 1, based on blocks of 50 trials (one- and two-item trials combined). The strong clustering of the data around 50% and 100% indicates nearly all-or-none learning.

[1] A few subjects reverted to chance performance later in the experiment (e.g., the downward-sloping segments at trials 550 and 650), and one subject gave all incorrect responses for a period (trials 601-650). These indications of lack of cooperation could be taken as grounds for exclusion, although we chose not to do this in order to keep the analysis as unbiased as possible.

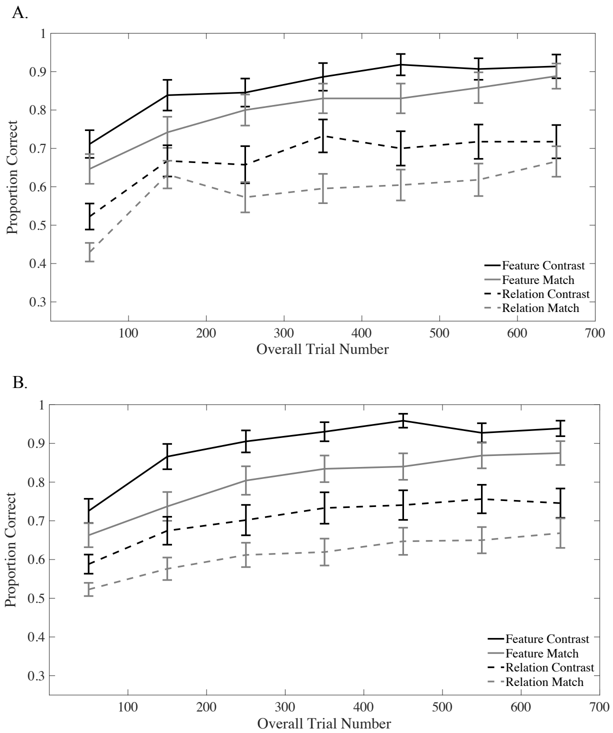

Figure 5. Average learning curves and standard errors for each group in Experiment 1. A: One-item trials. Each data point represents an average over 20 one-item trials (spanning 100 trials total). B: Two-item trials. Each data point represents an average over 80 two-item trials (spanning 100 trials total).

the proportions of subjects who have learned by each point in the experiment. The primary statistical analysis addressed accuracy on one-item trials, because the task on those trials was identical for the match and contrast groups, and thus any differences are due to differential learning of the categories and not to any differential difficulty of carrying out the two-item match and contrast tasks. In line with the prediction that follows from feature-based learning and theories of attention, a two-way ANOVA revealed a main effect of comparison on one-item trials, $F(1, 167) = 5.26$, $p = .023$, $MSE = .038$. Contrary to the prediction that follows from structure-mapping theory, there was no evidence of an interaction between category type and comparison type, $F(1, 167) = 0.17$, $p = .69$. As can be seen in Figure 5, subjects in the contrast conditions achieved higher accuracy than subjects in the match conditions, for feature-based categories and also for relational categories. To further test the reliability of the contrast advantage observed for both category types, separate planned t-tests were conducted to compare the feature-contrast and feature-match groups, as well as the relation-contrast and relation-match groups, both restricted to one-item trials. In the feature condition, a non-significant trend was found such that the contrast group outperformed the match group, $M_{\text{feature-contrast}} = .872$, $M_{\text{feature-match}} = .816$, $t(85) = 1.41$, $p = .164$, $MSE = .035$, $d = .306$. In the relation condition, a marginally significant advantage of the contrast group was found, $M_{\text{relation-contrast}} = .676$, $M_{\text{relation-match}} = .597$, $t(82) = 1.82$, $p = .073$, $MSE = .041$, $d = .402$.

Parallel analyses were conducted on two-item trials, which revealed a similar, but stronger set of results. The main effect of comparison type was significant, $F(1, 167) = 12.19$, $p = .0006$, $MSE = .032$, and there was no evidence of an interaction with category type, $F(1, 167) = .002$, $p = .97$. In planned t-tests, the feature-contrast group reliably outperformed the feature-match group, $M_{\text{feature-contrast}} = .904$, $M_{\text{feature-match}} = .810$, $t(85) = 2.63$, $p = .011$, $MSE = .028$, $d = .571$, and the relation-contrast group reliably outperformed the relation-match group, $M_{\text{relation-contrast}} = .711$, $M_{\text{relation-match}} = .615$, $t(82) = 2.33$, $p = .023$, $MSE = .036$, $d = .515$.

In order to more fully assess the effect of comparison type within each type of category structure, we merged the data from Experiment 1 with data from two pilot studies that had been conducted to calibrate the stimulus values (Ns = 65 and 166).[2] These pilot studies were identical to Experiment 1 in all respects except for the physical values of the stimuli on the three dimensions. Combining these datasets afforded more power to the primary analysis of performance differences on one-item trials. The data from all three studies ($N_{\text{feature-match}} = 98$, $N_{\text{feature-contrast}} = 103$, $N_{\text{relation-match}} = 104$, $N_{\text{relation-contrast}} = 97$) were subjected to an ANOVA restricted to one-item trials, with between-subjects factors of comparison type (match vs. contrast), category type (relational vs. featural), and study (Experiment 1 vs. Pilot 1 vs. Pilot 2, to control for any differences in difficulty across the three stimulus sets). This analysis revealed a main effect of comparison, $F(1, 398) = 8.49$, $p = .004$, $MSE = .038$, as subjects in the contrast groups ($M = .759$) outperformed subjects in the match groups ($M = .696$), and no interaction between comparison and category type, $F(1, 398) = .001$, $p = .946$, $MSE = .038$. Within the two category types, the feature-contrast group reliably outperformed the feature-match group, $M_{\text{feature-contrast}} = .852$, $M_{\text{feature-match}} = .797$, $t(199) = 2.03$, $p = .044$, $d = .288$, and the relation-contrast group reliably outperformed the relation-match group, $M_{\text{relation-contrast}} = .661$, $M_{\text{relation-match}} = .603$, $t(199) = 2.09$, $p = .038$, $d = .296$. Critically, no interaction was found between study and comparison type, $F(1, 390) = .35$, $p = .707$, $MSE = .038$, supporting the validity of the combined analysis.

[2] The primary goal of the pilot studies was to find stimulus values that yielded equivalent performance for the feature-based and the relational categories. As illustrated by Figure 2, the difficulty of the feature-based categories in this design is determined by the difference between the lower values (1 & 2) and the upper values (3 & 4) on each dimension, whereas the difficulty of the relational categories is determined by the differences between values 1 and 2 and between values 3 and 4. In Pilot Study 1, the objects' brightness values (on a 0-255 gray scale) were 15, 45, 120 and 250; their sizes (radii) were .85, 1.36, 2.30, and 3.20 cm; and their radius tilts were 5°, 23°, 59°, and 88°. This study showed a strong main effect of task type, such that the relation condition ($M = .643$) was considerably more challenging than the feature condition ($M = .903$), $F(1, 61) = 31.76$, $p < .0001$. Therefore, in Pilot Study 2 we decreased the differences between values 2 and 3 and increased the differences between values 1 and 2 and between values 3 and 4, on all three dimensions; values 2 and 3 were collapsed to be identical on each dimension, as a strategy to make the feature condition as difficult as possible. In Pilot Study 2, the objects' brightness values were 15, 65, 65, and 250; their sizes were 1.04, 1.82, 1.82, and 3.20 cm; and their radius tilts were 23°, 55°, 55°, and 85°. These changes reduced but did not eliminate the differences in performance between the relation ($M = .645$) and feature conditions ($M = .805$), $F(1, 161) = 37.03$, $p < .0001$. Both pilot studies showed non-significant trends of an overall contrast advantage (collapsing across the relation and feature categories). Because we were unable to eliminate the difference in difficulty between the two category structures, and because using identical or overlapping values (i.e., level 2 ≥ level 3) might introduce unanticipated complications in the feature condition, the dimension values in Experiment 1 were chosen as a middle ground between the values used in Pilot Studies 1 and 2. Note that equivalent performance in the feature and relation conditions is not strictly necessary in this design, because the analysis concerns the effect of the match-contrast manipulation separately for each category type.

Finally, to ensure the present findings reflect true concept learning and not memorization of individual items, an analysis was conducted of Experiment 1 restricted to trials that included only novel stimuli (i.e., ones not presented on any previous trial). Once again, a main effect of comparison was found for one-item trials, $F(1, 167) = 9.69$, $p = .002$, $MSE = .056$, as subjects in the contrast condition outperformed subjects in the match condition, $M_{\text{contrast}} = .644$, $M_{\text{match}} = .525$. Follow-up analyses showed this difference was reliable for the relation groups, $M_{\text{relation-contrast}} = .536$, $M_{\text{relation-match}} = .412$, $t(82) = 2.63$, $p = .010$, $MSE = .045$, $d = .581$, and marginally significant for the feature groups, $M_{\text{feature-contrast}} = .743$, $M_{\text{feature-match}} = .640$, $t(85) = 1.87$, $p = .065$, $MSE = .066$, $d = .406$. No interaction was found between type of comparison and type of category, $F(1, 167) = .07$, $p = .79$, $MSE = .056$. The analysis of two-item trials also showed evidence of a contrast advantage, albeit weaker: The main effect of comparison was marginally significant, $F(1, 167) = 2.90$, $p = .091$, $MSE = .033$, $M_{\text{contrast}} = .625$, $M_{\text{match}} = .575$, with no interaction between type of comparison and type of category, $F(1, 167) = .001$, $p = .998$, $MSE = .033$, but separate paired comparisons were non-significant (relation: $M_{\text{relation-contrast}} = .574$, $M_{\text{relation-match}} = .526$, $p = .15$; feature: $M_{\text{feature-contrast}} = .672$, $M_{\text{feature-match}} = .625$, $p = .29$).[3] This finding of a contrast advantage with novel items, which was reliable in the primary analysis of one-item trials, shows that the effect holds when subjects generalize to new stimuli and is hence a product of concept learning and not memorization.

In conclusion, we find evidence of a contrast advantage for the feature-based categories that is consistent with previous findings (Higgins, 2012) and with theories of selective attention in feature-based categorization (Nosofsky, 1986). More importantly, we also find evidence of a contrast advantage for the relational categories. The latter finding is consistent with the hypothesis that relational categories are learned similarly to feature-based ones, and it challenges the hypothesis that relational category learning depends on structural alignment. The evidence for the relational contrast advantage spans several different analyses, using one-item and two-item trials, the main experiment and two pilot studies, and trials with novel items. However, the effect was statistically significant in only a subset of these analyses, and the primary planned analysis (of all one-item trials in the main experiment) showed only a marginally significant effect. Therefore we maintain a cautious interpretation of the results and pursue further experiments to clarify and extend this finding.

Experiment 2

If there is an advantage to learning relational categories from between-category comparisons, one question is whether this advantage is limited to simple perceptual relations, such as those used in Experiment 1, or whether it extends to more conceptual or semantic relations. Thus, the goal of Experiment 2 was to evaluate the contrast advantage for relational categories using verbal stimuli. The stimuli differed from those in Experiment 1 in that they consisted only of semantic information and were thus more abstract. Specifically, stimuli in Experiment 2 were pairs of words, and each category was defined by a relation that held in each of its member pairs. Stimuli from one category were instances of the *contain* relation, such that one object can contain the other (e.g., *jug, milk*). Stimuli from the other category were instances of the *support* relation, such that one object can support the other (e.g., *legs, table*). Figure 6 shows how the word pairs were presented to subjects. Both categories included concrete (spatial) and more abstract instances (e.g., *movie, actors* for containment).

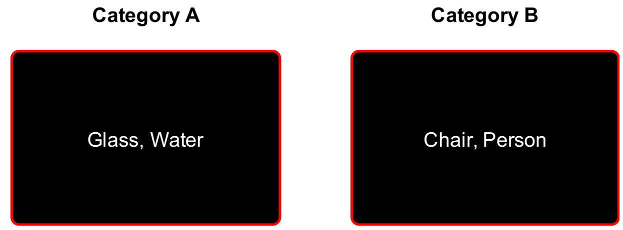

Figure 6. Example word pair from each category. The word pair in Category A is an instantiation of the *contain* relation. The word pair in Category B is an instantiation of the *support* relation. The positioning of words within all pairs was the same, such that the supporting or containing object was always displayed on the left.

Method

Subjects

One hundred forty undergraduates from the University of Colorado Boulder participated for course credit. Subjects were randomly assigned to match ($N = 69$) and contrast ($N = 71$) conditions.

Stimuli

The stimuli are listed in Appendix A. Eighty word pairs were constructed for each category. Each word pair was displayed as in Figure 6 with the containing or supporting object always on the left.

Design and Procedure

Every trial was a two-item trial, containing two word pairs (the one-item trials used in the other experiments were omitted due to a programming oversight). Subjects completed 400 trials, logically divided (i.e., for the purpose of defining the experiment design) into 5 blocks of 80, with each of the 160 stimuli appearing exactly once in each of these blocks. Assignment of stimuli to trials within each block was random, under the constraint that each trial contained two stimuli from the same category (match condition) or from opposite categories (contrast condition).

Subjects were given the cover story that rogue secret agents were using word pairs to communicate in two different types of code (Code A and Code B). Full instructions are presented in Appendix C. The categories (i.e., containment and support) corresponding to Codes A and B were counterbalanced within each condition. It was the subject's task to identify which word pairs were from Code A and which were from Code B. Subjects in the match condition were asked to type 'A' if both items were from Code A or 'B' if they were both from Code B. Subjects in the contrast condition were asked to type 'A' if the item on the left was from Code A and the item on the right was from Code B, or 'B' for the reverse. The rest of the design and procedure (i.e., presentation and timing of feedback, rest breaks, and approximate duration of the study) were identical to Experiment 1.

Results and Discussion

Figure 7 shows average learning curves for the contrast and match groups. These curves show an advantage for the contrast condition, which was significant by a t-test, $M_{\text{contrast}} = .809$, $M_{\text{match}} = .746$, $t(138) = 2.28$, $p = .024$, $MSE = .027$, $d = .39$. To ensure that the effect was driven by learning of the concepts underlying the categories, and not by memorization of individual items, a second *t*-test was conducted using only the first 80 trials for each subject, which contained only the first presentation of each item. This test also showed a significant effect, $M_{\text{contrast}} = .687$, $M_{\text{match}} = .629$,

[3] Because a two-item trial is more likely to contain a repeated item, this analysis contained fewer trials per subject, coming from earlier in learning, than did the analysis of novel one-item trials. These differences likely explain why the two-item analysis showed less reliable results.

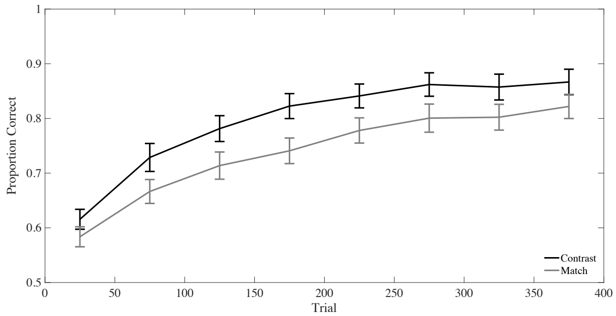

Figure 7. Average learning curves and standard errors across blocks of 50 trials for each condition in Experiment 2.

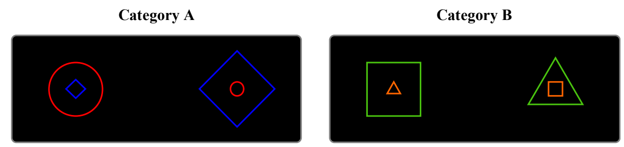

Figure 8. Examples of stimuli from both categories in Experiment 3. Each grey rectangle encloses a single stimulus. In Category A stimuli, objects of the same color always match in shape (two red circles and two blue diamonds in this example). In Category B stimuli, objects of the same color always match in size (two large green objects and two small orange objects in this example). See the online article for the color version of this figure.

$t(138) = 2.12$, $p = .036$, $MSE = .026$, $d = .361$. These findings more definitively establish the contrast advantage in relational category learning seen in Experiment 1, and they extend the result by showing the effect with verbal stimuli.

Experiment 3

One possible explanation for the contrast advantage observed in Experiments 1 and 2—elaborated in the General Discussion—is that structural alignment is not necessary for learning relational concepts that are defined by a single relation, and that structure-mapping's predicted learning advantage for within-category comparison may arise only for richer relationally structured concepts. To test this possibility, Experiment 3 used more complex relational stimuli, such that each category was defined by a configuration of multiple relations operating on a shared set of objects. Every stimulus in Experiment 3 contained four simple geometrical objects, arranged as in the examples in Figure 8. The objects within a stimulus varied on three dimensions—size, color, and shape—and the categories were defined by the pattern of agreement among these dimensions. For a stimulus in Category A, the objects that matched in color also matched in shape but mismatched in size. Thus the example Category A stimulus in Figure 8 contains two red circles (same color and shape but different sizes) and two blue diamonds. For stimuli in Category B, the objects that matched in color also matched in size but mismatched in shape. Thus the example in Figure 8 contains two large green objects (same color and size, different shapes) and two small orange objects.

There are multiple ways these stimuli might be represented (an issue we highlight in the General Discussion), but one simple possibility in line with previous work (e.g., Kotovsky & Gentner, 1996) is to assume *same-color*, *same-shape*, and *same-size* relations representing whether objects match on each of these dimensions. Thus members of Category A would all satisfy sets of relations such as:

$$\textit{same-color}(\textit{object}_1, \textit{object}_2)$$
$$\textit{same-shape}(\textit{object}_1, \textit{object}_2)$$
$$\textit{different-size}(\textit{object}_1, \textit{object}_2)$$

(where *different-size* stands for the negation, ¬*same-size*). Likewise, members of Category B would all satisfy sets of relations such as:

$$\textit{same-color}(\textit{object}_1, \textit{object}_2)$$
$$\textit{same-size}(\textit{object}_1, \textit{object}_2)$$
$$\textit{different-shape}(\textit{object}_1, \textit{object}_2).$$

According to structure-mapping theory, these systems of predicates should be alignable between members of the same category (and not between members of opposing categories), which should enable abstraction of a category-defining schema that could be used to identify future category members.

Importantly, the categories cannot be distinguished merely on the basis of the individual objects present in a given stimulus, because any given object (e.g., a large red square) is equally likely to occur in a stimulus from either category. Likewise, a stimulus cannot be categorized only on the basis of the pairwise relations between its objects, because those relations are the same for stimuli in both categories. Specifically, every stimulus contains two objects of one color and two objects of another color (e.g., two red objects and two blue objects), regardless of which category it is in. Thus there are always two instances of *same-color* relations (one between the two red objects, another between the two blue objects) and four instances of *different-color* relations (between either of the red objects and either of the blue objects) within any stimulus. The same holds for the size and shape dimensions. What distinguishes the categories is how the different relations are linked by shared objects—for example that *same-shape* and *same-color* relations can operate on the same pair of objects in Category A stimuli, but not in Category B stimuli. That is, every Category A stimulus contains pairs of objects that match on both shape and color (e.g., two red circles and two blue diamonds, as in Figure 8), whereas in a Category B stimulus no two objects can match on both shape and color. Likewise, every Category B stimulus contains pairs of objects that match on both size and color (e.g., two large green objects, and two small orange objects, as in Figure 8), whereas in a Category A stimulus no two objects can match on both size and color. This pattern of role binding across multiple relations with shared objects (i.e., a relational system as opposed to a single relation) is exactly the information that structural alignment and schema induction operate to discover (Corral & Jones, 2014).

Other than the shift from single to multiple relations, Experiment 3 matched the approach of Experiments 1 and 2. The use of categories defined by complex relational structures was predicted to engage structural alignment more reliably than in the previous experiments and, according to structure-mapping theory, should be more likely to produce a match advantage.

Method

Subjects

One hundred sixteen undergraduates from the University of Colorado Boulder participated for course credit. Subjects were randomly assigned to contrast (N = 56) and match (N = 60) conditions.

Stimuli

Each stimulus consisted of four geometric objects on a black background enclosed by a grey rectangle, as shown in Figure 8. The four objects were always arranged in two concentric pairs: a small object inside a large object on the left, and a small object inside a large object on the right. The objects also varied in shape (circle, square, diamond, triangle, or hexagon) and color (red, green, blue, yellow, or white) with exactly two shapes and two colors present within any one stimulus. For every stimulus in both categories, the large object on the left and the small object on the right matched in shape, as did the small object on the left and the large object on the right. Thus objects matching in size always mismatched in shape and vice versa. The two categories differed in how the color dimension related to the size and shape dimensions. In Category A, the large object on the left and the small object on the right matched in color, as did the small object on the left and large object on the right. Thus the objects matching in shape also matched in color, whereas objects matching in size always mismatched in color. In Category B, the two large objects matched in color, as did the two small objects. Thus the objects matching in size also matched in color, whereas objects matching in shape always mismatched in color.

Design and Procedure

Subjects were given the cover story that two alien species were creating different types of symbols, and it was the subject's job to figure out which species created each symbol. Full instructions are presented in Appendix C. Every subject performed 400 trials. Every fifth item was a one-item trial, starting with Trial 5, and the remainder were two-item trials. For each two-item trial, two stimuli were generated at random under the constraint defining the experimental condition (i.e., being in the same or opposing categories). For each one-item trial, the stimulus was generated fully at random. Response keys, feedback, and timing all matched the procedures of Experiment 1. The experiment lasted approximately 25 minutes.

Results and Discussion

Figure 9 shows average learning curves for the contrast and match conditions on one- and two-item trials. The primary analysis on one-item trials revealed that subjects in the contrast condition reliably outperformed those in the match condition, M_{contrast} = .879, M_{match} = .801, $t(114)$ = 1.99, p = .048, MSE = .035, d = .373. This same pattern of results was also found on two-item trials, although

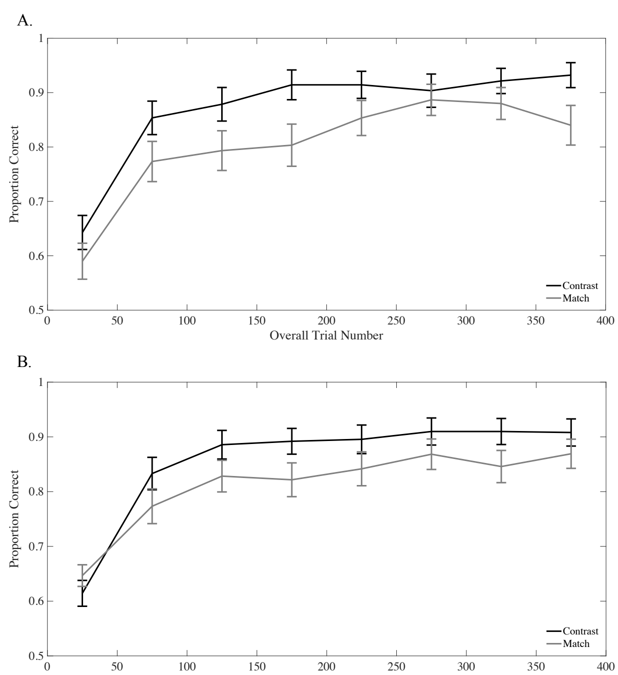

Figure 9. Average learning curves and standard errors for each condition in Experiment 3. A: One-item trials. Each data point represents an average over 10 one-item trials (spanning 50 trials total). B: Two item trials. Each data point represents an average over 40 two-item trials (spanning 50 trials total).

it was not significant, M_{contrast} = .870, M_{match} = .821, $t(114)$ = 1.49, p = .144, MSE = .032, d = .28. A similar pattern of results was observed when the analysis was restricted to trials that did not contain repeated stimuli: A reliable contrast advantage was observed on one-item trials, M_{contrast} = .867, M_{match} = .795, $t(114)$ = 2.06, p = .042, MSE = .035, d = .386, and a trend of a contrast advantage was observed on two-item trials, M_{contrast} = .840, M_{match} = .795, $t(114)$ = 1.30, p = .195, MSE = .031, d = .244. Taken together, these results demonstrate that the learning advantage of between-category comparison extends to perceptual concepts that are defined by relational structures composed of more than a single relation.

Experiment 4

A fourth experiment was conducted to test whether the contrast advantage observed across Experiments 1-3 holds for more abstract relational concepts. Previous work suggests that many abstract concepts are determined by people's prior knowledge of the relationships that exist among a concept's features (Rehder & Ross, 2001). Experiment 4 adopted stimuli developed by Rehder and Ross that form categories defined by the abstract coherence among each stimulus' components. Each stimulus consists of three short sentences describing features of a machine that works to remove waste products. The first sentence describes the location in which the machine operates, the second describes the waste product the machine works to remove, and the third describes the implement the machine uses. The machines are divided into two categories determined by how these three features relate to each other. For an item in the *coherent* category, the machine's implement is well-suited for collecting the machine's target waste material, and that waste material is typically found in the machine's operating location. For example, one member of the coherent category is defined by the properties "Operates on the surface of the water, works to clean spilled oil, and has a spongy material." This item is coherent because of additional relations (presumably known by the subject) that oil slicks are found on water and that a sponge can absorb oil. In contrast, an item in the *incoherent* category has features that do not share such relations with one another: the machine's implement cannot collect the material being cleaned nor can that material be found in the machine's operating location. Figure 10 illustrates how stimuli from the two categories can be represented in the framework of structure-mapping theory. The additional relations among the features of a coherent item yield a different (richer) relational structure than that of an incoherent item, which should be discoverable through alignment of two items from the same category. Indeed, Higgins (2012) found an advantage of within-category comparison using a subset of these stimuli (Items 1-9 from both categories in Appendix B), using a more elaborate

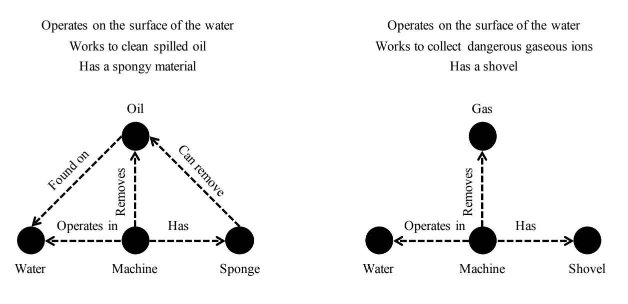

Figure 10. Illustration of how a coherent item and an incoherent item from Experiment 4 can be represented in the compositional role-binding framework of structure-mapping theory. The text at the top shows the descriptions presented to the subjects. Partially recreated from Corral & Jones (2014).

training procedure that required subjects to carefully consider the substructure of the stimuli. Therefore a goal of Experiment 4 was to test whether a match advantage would be found under the more neutral conditions of the present paradigm (see the *Implications for Structure-Mapping Theory* section for further discussion).

As in the previous experiments reported here, subjects' task was to learn to classify the items (in pairs) to the two categories. An initial version of the experiment followed previous experiments with these stimuli (Rehder & Ross, 2001) by using labels for both categories (Morkel and Krenshaw), but subjects in both the match and contrast conditions performed at chance. One possibility is that subjects were searching for and could not find a relational rule that defined the incoherent category (see Jung & Hummel, 2011, for a similar result). We therefore switched from an A/B task to an A/¬A task. That is, only the coherent category was given a label (Morkel), and subjects' task was to determine whether each stimulus was or was not a member of that category. Subjects were not told anything about what constituted a Morkel, just that some machines were Morkels and others were not. This framing of the task was expected to encourage subjects to learn the category rule for the coherent items, without attempting to discover a separate category rule defining the incoherent items.

Method

Subjects

One hundred twenty-nine undergraduates from the University of Colorado Boulder participated for course credit. They were randomly assigned to contrast (N = 62) and match (N = 67) conditions.

Stimuli

The stimuli are listed in Appendix B. Half were taken from Rehder and Ross (2001) and Higgins (2012). Rehder and Ross made three coherent items and three incoherent items; the latter set was generated by shuffling the features of the former set such that each incoherent item took one property from each of the three coherent items. Higgins used a similar method to create an additional 12 items (six coherent and six incoherent). A further 18 items (nine coherent and nine incoherent) were created by the present authors using the same shuffling method as Rehder and Ross. There were thus a total of 36 stimuli, 18 from each category. Each stimulus was presented to subjects as three lines of text bounded by a red border, as shown in the example display for a two-item trial in Figure 11.

Design and Procedure

Subjects were told they would learn about machines named Morkels, specifically that their task was to learn to identify which machines were Morkels and which were not. They were given no prior information about what qualified a machine as a Morkel. Full instructions are presented in Appendix C. Each subject completed 300 trials, with self-paced rest breaks every 20 trials in the same format as in previous experiments. Every fifth trial was a one-item trial and the others were two-item trials. Each of the 36 stimuli appeared exactly once in each block of 20 trials (16 two-item trials and 4 one-item trials), with the assignment made randomly subject to the constraint defining the subject's condition (match or contrast). On two-item trials, two stimuli were displayed side by side, labeled Machine A and Machine B (see Figure 11). In the match condition, subjects were asked to type '2' if both machines were Morkels and '0' if they were not. In the contrast condition, subjects were asked to type '2' if Machine A was a Morkel and Machine B was not, and '0' if Machine A was not a Morkel and Machine B was. After each trial, corrective feedback (provided as in the previous experiments) was shown directly under the stimuli for three seconds. On one-item trials, subjects were asked to type 'M' if the machine was a Morkel and 'K' if it was not. The rest of the design and procedure were identical to those of Experiments 1 and 3.

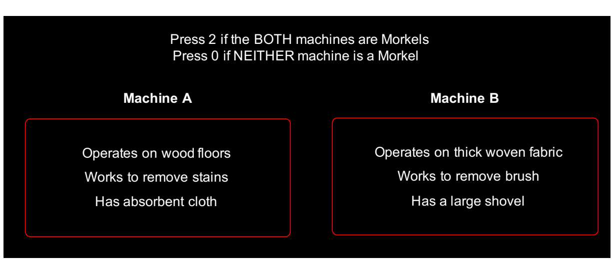

Figure 11. An example of a two-item trial from the match condition in Experiment 4.

Results and Discussion

Figure 12 shows average learning curves on one- and two-item trials for both conditions. As in the previous experiments, a learning advantage was found for the contrast group. This effect was reliable on one-item trials, M_{contrast} = .825, M_{match} = .765, $t(127)$ = 2.50, p = .015, MSE = .020, d = .44, and on two-item trials, M_{contrast} = .844, M_{match} = .774, $t(114)$ = 2.71, p = .008, MSE = .021, d = .481. An analysis restricted to trials with only novel stimuli (Trials 1-20) showed non-significant contrast advantages for one-item trials, M_{contrast} = .552, M_{match} = .489, $t(127)$ = 1.42, p = .158, MSE = .065, d = .252, and for two-item trials, M_{contrast} = .550, M_{match} = .506, $t(127)$ = 1.55, p = .124, MSE = .026, d = .244. The raw effect size for novel stimuli is similar to that for all trials, but the novel-item analysis is noisier due to the small number of stimuli in this experiment: The one-novel-item analysis includes only four trials per subject, and the two-novel-items analysis includes 16. In sum, the results for Experiment 4 replicate the findings from Experiments 1-3 and show that the contrast advantage extends to the learning of abstract, richly structured relational concepts.

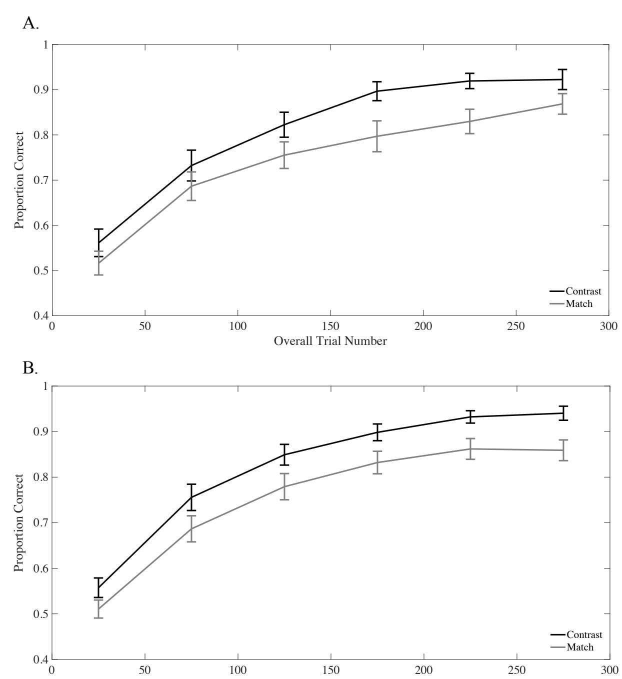

Figure 12. Average learning curves and standard errors for each condition in Experiment 4. A: One-item trials. Each data point represents an average over 10 one-item trials (spanning 50 total trials). B: Two-item trials. Each data point represents an average over 40 two-item trials (spanning 50 total trials).

General Discussion

The present series of studies examined how comparing items from the same or different categories affects relational concept acquisition. The difference in learning performance between match and contrast conditions in this paradigm is theoretically significant because it discriminates between two fundamentally different accounts of relational category learning. Characterizing relational category learning as a form of analogical reasoning predicts a match advantage, based on the central role of comparing items with common structure within theories of analogy (Gentner, 1983). Treating relational categories as akin to feature-based categories predicts a contrast advantage, based on the central role of selective attention to diagnostic dimensions in theories of feature-based category learning (Nosofsky, 1986). The present results demonstrate strong and robust support for the latter prediction. The relational categories tested here consisted of both geometric stimuli with visuospatial relations (Experiments 1 & 3) and abstract verbal stimuli with conceptual relations (Experiments 2 & 4). They were defined both by single relations (Experiments 1 & 2) and by interconnections among multiple relations (Experiments 3 & 4). The tasks were cast as two-category (A/B, Experiments 1-3) and single-category (A/¬A, Experiment 4), which recent work suggests are learned in different ways (Hendrickson, Perfors, Navarro, & Ransom, in press). Moreover, the analyses of novel items in all four experiments show the observed contrast advantage reflects true concept learning, not memory for individual items.

It might be argued that the contrast condition provides more information than the match condition, because a contrast trial affords learning about both categories simultaneously, whereas a match trial affords learning about only one category. However, it is important to note that subjects in both conditions were shown the same set of stimuli. After *n* trials, the average subject will have been presented with *n* stimuli from each category, regardless of condition. More importantly, this explanation does not accord with structure-mapping theory, which holds that learning simultaneously from two alignable stimuli (i.e., two stimuli in the same category) is more effective than learning from both stimuli separately.

The present findings thus stand as a challenge to recent attempts to extend theories of analogy to relational category learning (e.g., Corral & Jones, 2014; Kittur et al., 2004; Kuehne et al., 2000; Kurtz et al., 2013; Lassaline & Murphy, 1998). Some manner of comparison was taking place in these experiments (if subjects were processing the stimuli individually, then there should have been no effect of the match/contrast manipulation at all), but evidently it did not rely on discovery of shared structure in the way structure-mapping theory assumes. Instead, the pattern of results is consistent with feature processing and selective attention, and indeed the same finding was obtained with feature-based categories (Experiment 1). However, it would be wrong to conclude from the present results that relational categories are psychologically no different from feature-based ones. There is abundant evidence that relational structure matters to human concept representations (Jones & Love, 2007; Markman & Gentner, 1993a, 1993b; Sloman, Love, & Ahn, 1998). Thus the challenge is to understand how a process akin to feature-based learning might operate in the presence of structured representations, and to delineate the circumstances that lead people to engage in structural alignment versus less cognitively demanding strategies based on features. We discuss both issues in detail below.

Our findings are related to previous work examining how interleaved versus blocked stimulus presentation affects category learning (Carvalho & Goldstone, 2014a, 2014b, 2015; Goldstone, 1996; Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011). Building on earlier work by Goldstone (1996), Carvalho and Goldstone (2014a) compared performance on a feature-based classification task (with one stimulus per trial) between two conditions: a blocked condition in which the correct category matched the category from the previous trial on 75% of trials, and an interleaved condition in which the correct category matched that of the previous trial on 25% of trials. To the extent that subjects compare stimuli on successive trials, blocked training emphasizes within-category comparison (similar to our match condition), whereas interleaved training emphasizes between-category comparison (similar to our contrast condition). Carvalho and Goldstone's main finding was that the relative advantage of these two conditions depended on the degree of within-category versus between-category similarity among the stimuli. Specifically, they found that blocked training was superior to interleaved training when the stimuli varied on a large number of irrelevant features, whereas interleaved training was superior when there were few irrelevant features. This finding might suggest that the contrast advantage found in the present experiment reflects the patterns of similarity in the stimuli used here, rather than constituting evidence for feature processing over analogy. However, the nature of our materials makes this explanation unlikely, and in fact Carvalho and Goldstone's interpretation of their results reinforces our conclusions regarding feature processing.

First, the degree of irrelevant stimulus variation differed significantly across our studies. In Experiments 1 and 3, the stimuli in both categories were highly similar, with little irrelevant variation. In both cases, all stimuli followed a very specific pattern (a pair of semicircles with inscribed radii, or four regular geometric figures arranged in two concentric pairs), with only a few dimensions of variation (size, brightness, and angle, or color and shape). In contrast, the stimuli in Experiments 2 and 4 varied widely. In Experiment 2, the stimuli in the *containment* and *support* categories respectively ranged from {*bank*, *teller*} to {*forest*, *trees*} and from {*foundation*, *building*} to {*tires*, *vehicle*}. The same was true for the stimuli in Experiment 4, as exemplified in Figure 11. Despite these large differences across studies in the amount of irrelevant stimulus variation, the contrast advantage was observed in all four experiments.

Second, even if increasing the irrelevant stimulus variation did produce a match advantage in our paradigm, such an effect would support feature processing, not structural alignment. According to Carvalho and Goldstone (2014a), learning from between-category comparisons encourages identification of diagnostic differences, which are more salient when there are a small number of irrelevant features among the stimuli, whereas learning from within-category comparisons encourages subjects to seek commonalities between items, which are more salient under greater amounts of random stimulus variation. This is fundamentally an attentional explanation. Moreover, research on analogy has consistently demonstrated that analogical learning does not follow this pattern. Instead, decreasing surface similarity (i.e., increasing irrelevant variation) makes discovering the common relational structure between two scenarios more difficult (Gick & Holyoak, 1983; Holyoak & Koh, 1987; Reed, 1989; Ross, 1987, 1989). Therefore, finding that Carvalho and Goldstone's results regarding similarity structure extend from feature-based categories to relational ones (and from a blocked/interleaved manipulation to our match/contrast paradigm) would contribute further support for our conclusion that the effects of comparison on relational category learning support accounts based on feature processing over accounts based on analogy.

Unitary versus Compositional Representations of Relational Concepts

At the heart of the distinction between feature-based and analogical models is the information on which they operate. Feature-based models encode a stimulus as either a vector of dimension values (Estes, 1986) or a set of attributes (Tversky, 1977). Although vector and set approaches are often viewed as competitors, both imply a "flat" type of comparison process that can identify only the shared and unique attributes of two stimuli, without regard for their structure. In contrast, structure-mapping and related theories represent stimuli as relational systems defined by role binding, which captures how constituent objects and relations are composed to form the whole stimulus (Gentner, 1983). This structural information supports a richer and more sensitive form of comparison, embodied by structural alignment. In particular, structure-mapping theory's principle of parallel connectivity entails that having the same elements is insufficient for two stimuli to be aligned, and that instead what matters is that they are bound into isomorphic structures (Corral & Jones, 2014).

This difference in information and in corresponding comparison processes clarifies the conditions under which feature-based processing is in principle sufficient for a task like categorization. Specifically, if two categories differ in the individual objects, features, or relations their members contain—without regard for the bindings among them—then the sort of flat comparison assumed by feature-based models should be able to distinguish them. For example, a subject in Experiment 2 presented with the stimuli {*cup*, *water*} and {*chair*, *person*} might not have to construct a structured representation for each stimulus like *contains*(*cup*, *water*) and *supports*(*chair*, *person*) and then attempt to align these structures. Instead the subject may simply recognize that there is a containment relation implied by one stimulus and a support relation in the other—just as directly as she could recognize that one item contains something liquid or that the other contains something animate. The stimuli could hence be represented as lists or sets of elements, such as {*cup*, *water*, *contain*, …} and {*chair*, *person*, *support*, …}. The difference between the categories should be easily learnable under this representation, particularly after many trials in which the A and B items always include *contain* and *support*, respectively.

If two categories are defined by the same elements, differing only in their role-binding structure, then the difference will be invisible to feature processing and something like structural alignment will be necessary for successful learning. However, this is a psychological question, not a logical one. Logically, any relational system (i.e., set of relations linked through role-binding to shared objects) is equivalent to an atomic relation operating directly on all the objects jointly. Likewise, as noted by Gentner (1983, Footnote 4), an atomic relation among a set of objects is logically interchangeable with an attribute of the system as a whole. Thus, to use the spaceships example above from Foster et al. (2012), a relational system such as *beats*(*A*,*B*), *beats*(*B*,*C*), *beats*(*A*,*C*) might also be represented as an atomic relation *well-ordered*(*A*,*B*,*C*) or as an attribute *transitive*(*tournament*). To our knowledge these observations have never been pursued as a possible distinction among cognitive representations, with psychological consequences.

We thus propose a distinction between *unitary* and *compositional* representations of relational concepts. The key observation is that the same concept (i.e., the same semantics) can be represented in psychologically distinct ways. A unitary representation is either an atomic relation among objects, such as *bigger-than*(*x*,*y*), or an attribute of a whole system that expresses a relation among its components, such as *lopsided*(*z*) to express a size relation among the parts of *z*. These two possibilities have the same implications for the present investigation, because both can be treated as elements of a stimulus under flat comparison. Following the reasoning given earlier, unitary representations should lead to a contrast advantage in our paradigm, because between-category comparison highlights the elements that are diagnostic of category membership. In contrast, a compositional representation encodes a relational concept as a structure or system of (atomic) relations connected by role-binding to shared objects. The meaning of the concept as a whole is not explicitly represented but emerges only from this pattern of interconnections. Thus, learning or recognizing concepts represented compositionally requires structural alignment, which in turn predicts a match advantage in our paradigm.

It seems intuitively evident that many concepts can be represented both unitarily and compositionally. Indeed, most everyday objects that are thought of by default as atomic entities have substructure. What one typically conceives as simply a *dog* can also be understood as a complicated configuration of organs and systems (e.g., circulatory, respiratory, immune, digestive). The same flexibility arguably applies for abstract concepts that one might consider to be truly relational. Consider the concept *tradeoff* (see Loewenstein, Thompson, & Gentner, 1999). A tradeoff is a rich relational system involving an agent, two or more goal dimensions, two or more options, and a particular pattern of relations between the options' values on those dimensions (A is better on dimension X, but B is better on Y). Nevertheless, it seems likely that an average person can conceive of a tradeoff more directly as an atomic relation operating on the options and preferences of the agent, just as one can comprehend the sentence "I have a dog" without needing to actively represent how the dog digests food and delivers oxygen to its tissues. The availability of these unitary representations is consistent with superior learning from between-category comparisons: Just as one does not need to align the anatomies of two animals to recognize that one is a dog and the other is a cat, one might not need to align the structures of two decision scenarios to recognize that one involves a tradeoff whereas the other scenario is a win-win.

When people have both unitary and compositional representations available to solve a task, cognitive economy suggests they will prefer the former. Behavioral research on analogy has demonstrated that the cognitive processes of structural alignment load heavily on working memory (Forbus at al., 1995; Morrison, Holyoak, & Truong, 2001; Waltz et al., 2000), and neurophysiological studies have linked relational processing to dorsolateral prefrontal cortex (Christoff et al., 2001; Kroger et al., 2002; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997; Waltz et al., 1999). In contrast, flat comparison using feature vectors is generally assumed to be rapid and automatic, as in models of parallel retrieval from long-term memory (Forbus et al., 1995; Shiffrin & Steyvers, 1997). This difference in cognitive demand might lead people to seek unitary representations, especially in a categorization task that requires processing of hundreds of stimuli or comparisons.

The present findings can thus be interpreted as evidence of how adept people are at finding or inferring unitary properties that embody relational concepts, enabling them to solve the categorization tasks without the need for structural alignment. The category-distinguishing unitary relations in Experiment 2 are self-evident (*support* vs. *contain*, or perhaps more simply, *on* vs. *in*). In Experiment 4, the abstract relational coherence of Morkel stimuli could be captured by *found-in* and *can-remove* relations among their components (see Figure 10), or more generally by the property that the components (or any two of them) "work" or "make sense" together. Although these properties fit within structured role-binding representations of the stimuli, they would also be accessible to flat comparison on unstructured representations. In Experiments 1 and 3, the materials were designed with the aim of matching the atomic relations present in members of each category (*larger*, *same-color*, etc.), but it is not difficult to speculate on more sophisticated relations that subjects may have discovered. Although a category in Experiment 1 can be represented compositionally as, for example, *brighter*(*left object*, *right object*), it could be represented more compactly (and unitarily) as a Gestalt property of a combined object having more brightness in its left half (i.e., a left-right brightness gradient) that one might write as *left-brighter*. In Experiment 3, subjects may have recognized conjunctions of the primitive relations we assumed in designing the stimuli, such as *same-shape-and-color* or *same-except-for-size*, which was present only in Category A (see Figure 8). Alternatively, subjects may have recognized the patterns of spatial symmetry present in the stimuli: For a stimulus in Category B, the arrangements of colors on the left and right are the same, whereas for a stimulus in Category A the arrangements of color on the two sides are reversed. These perceptual patterns may have been directly apparent in each stimulus, that is, as global, unitary properties.

These particular possibilities are certainly post-hoc, and more generally the unitary/compositional distinction does not constitute a predictive theory without independent means to predict or assess what type of representation is operating in any given situation. As previous researchers in analogy have argued, people have enormous flexibility in how they represent concepts, and the real challenge of explaining abstract reasoning lies in how they select among these representations (Chalmers, French, & Hofstadter, 1992; French, 1997; Mitchell & Hofstadter, 1990). Although we do not aim to answer this question here, we believe the representational proposal sketched in this section offers a framework for formalizing these issues of flexibility. Moreover, it offers a way to bridge the literatures on feature-based learning models and structure-mapping theory, and to show concretely how relational categories might be learned in a feature-based manner. These theoretical considerations plus the present empirical results will hopefully provide the foundation for a systematic study of when people use feature representations and flat comparison processes, versus role-binding structure and structural alignment.

Recent work building on the present investigation has begun to address this question. One viable approach might be to ask subjects directly what representations they used. Self-reports of cognitive processes must of course be interpreted cautiously, because they often reflect post hoc inferences based on observing one's own behavior, rather than introspective access (Nisbett & Wilson, 1977). Within categorization, much of behavior is governed by processes and representations that are not verbalizable (Ashby et al., 1998), and subjects who believe they are following a particular rule exhibit influences of other processes in their classification responses (Allen & Brooks, 1991). On the other hand, the all-or-none character of learning in the present experiments suggests that subjects' category knowledge might be particularly explicit. With this in mind, Corral (2017) conducted a series of studies (with various materials, including those of Experiment 4) wherein subjects were asked after learning to select which of two options best described how they were thinking about the categories: a unitary description (e.g., the machine functioned) and a compositional one (e.g., what the machine worked to remove could be found where the machine operates and could be removed with the machine's tool). Subjects overwhelmingly selected the unitary description in 4 out of 5 of experiments, with no differences found in the fifth. In another recent study, using the materials from Experiment 4, Corral and Jones (2017) gave each subject a hint about the category rule at the outset of learning. The hint encouraged either a unitary representation ("the machine is intuitive") or a compositional one ("think about how each of the machine's components are related to one another"). Subjects who were encouraged to represent the category unitarily outperformed subjects who were encouraged to represent it compositionally. Together with the present results, these findings bolster the hypothesis that people prefer to represent relational concepts unitarily, at least in the context of repeated classification.

Finally, we note that the distinction proposed here may also inform research on metaphor comprehension. The debate between theories of metaphor based on structural alignment (Gentner & Bowdle, 2008; Gentner & Wolff, 1997) and based on categorization (Glucksberg, 2003; Glucksberg, McGlone, & Manfredi, 1997) might be seen to come down to a question of compositional versus unitary representations. If the concept conveyed by the metaphor can be represented directly as an attribute of the base, then it seems likely that the comprehender can evaluate whether that attribute applies to the target, in a process akin to categorization. If instead the concept is represented only as a system of substituent relations within the base, then some manner of structural alignment to the target should be necessary. We leave this idea as conjecture, but we find it intriguing that the mechanisms of metaphor processing might ultimately depend on the issues of representation considered here.

Implications for Structure-Mapping Theory

The central principle of structure-mapping theory is that analogical learning and reasoning are driven by discovery of shared relational structure between two scenarios (Gentner, 1983). The experimental results reported here challenge that principle, at least in the context of category learning, by showing that people learn relational categories better by comparing items instantiating contrasting relational concepts. The results suggest instead that people do not make use of compositional role-binding representations or of structural alignment processes, at least not when simpler, feature-based mechanisms are adequate for the task at hand. Although the representational apparatus of structure-mapping theory implicitly incorporates the distinction delineated above between unitary (global attributes and atomic relations) and compositional (systems of interconnected relations) representations of relational concepts, this distinction has not been previously appreciated for its psychological implications. More importantly, the core tenet of the theory is that relational concepts are learned in a compositional, structure-based manner (Markman, 1999, Markman & Gentner, 2000), and the present findings suggest significant limitations to the applicability of this idea.

One possibility that proponents of structure-mapping theory might pursue is that, in addition to learning from successful alignment, people can also learn from the ways in which alignment fails, by identifying relations that prevent a partial alignment between two scenarios from being extended further. These impediments to alignment could highlight the structural differences between opposing relational categories and thus drive learning from between-category comparisons. For example, a subject in Experiment 4 who compares the items in Figure 10 might align *sponge* with *shovel* and *oil* with *gas* and then recognize that the mapping cannot extend to include *can collect*. Indeed, recent work by Smith and Gentner (2014) suggests that encouraging people to partially align instances of contrasting relational categories makes them better able to identify the structural differences between them (see also Gick & Paterson, 1992). This idea can be seen as an extension of the work on alignable differences, which has shown that successful alignment between two members of the same relational category can highlight differences in their corresponding features (Gentner & Markman, 1994, 1997; Markman & Gentner, 1993a, 1993b). It also relates to Corral and Jones' (2014) proposal of schema elaboration, a mechanism for schema learning from between-category comparisons whereby the schema for a category is augmented with structure that an opposing item violates. Although learning from impediments to alignment seems plausible with the framework of structure-mapping theory, it begs the question of why or when learning from failed alignment should be more effective than from successful alignment. Thus it threatens to make the theory unfalsifiable, in that it could equally predict a match advantage through successful alignment or a contrast advantage through impediments to alignment, depending on which is assumed to be more effective for learning.

Perhaps a more interesting way forward would be to consider what aspects of the present methods might have encouraged subjects to embark on feature-based processing, and conversely to try to delineate the conditions under which people are more likely to engage in structural alignment. One factor that has already been discussed is the number of comparisons a person must make: structure-mapping theory is perhaps best suited for explaining cases of one-shot learning, whereas an extended categorization task might motivate subjects to rely on less cognitively demanding strategies. In support of this interpretation, Higgins and Ross (2011) found better learning of mathematical combinatorics concepts (permutation vs. combination) from within-category comparison, in a design involving only a single comparison between training items.

A second possible factor is the nature of the categorization task and the types of comparison it encourages. The present experiments used classification, by far the most common task used to study category learning (e.g., Ashby & Maddox, 2005), and they used neutral instructions designed to give subjects minimal directive on how to compare the stimuli on two-item trials (see Appendix C). In this regard, we take the present results to reflect default behavior in human category learning. However, different results might be obtained under conditions encouraging subjects to engage in more effortful comparison (see Alfieri et al., 2013). Previous research has used more elaborate comparison tasks, such as listing commonalities and differences or mapping corresponding elements, as a way to focus subjects on the substructure of stimuli and thereby encourage structural alignment (Doumas & Hummel, 2004; Gentner & Gunn, 2001; Kurtz, Miao, & Gentner, 2001). Additionally, inference learning—a categorization task wherein the category labels are given and the subject has to infer hidden features of the stimuli—has been shown to enhance learning of relations among stimulus features relative to classification learning (Markman & Ross, 2003; Yamauchi & Markman, 1998). As noted earlier, Higgins (2012) found a match advantage on a classification task using half of the stimuli we used in Experiment 4, after subjects were provided with an inference training task and a task of listing similarities and differences between items, providing evidence that requiring more effortful comparison can lead people to engage in structural alignment. This finding together with the present results suggests that systematic manipulation of classification tasks and comparison instructions could help to pin down the conditions under which people might use structural alignment in category learning.

A third potential factor influencing the use of structural alignment is the learner's prior knowledge. Observe that a subject in Experiment 3 is likely not learning anew the concepts *contain* and *support*, but rather determining that these already-known concepts provide the solution to the task. More generally, a person should have a unitary representation for a relational concept only if he or she has previously learned it (see also Doumas et al., 2008). Clark (2006) ascribes a critical role in this process to language, with verbal labels for relational concepts providing the explicit representations that enable them to be targets for explicit thought, much as we propose for feature-based processing of unitary representations. If such a representation is not available, then we would predict learners to be more likely to engage in structural alignment. To use Higgins and Ross's (2011) combinatorics categories as an example, advanced mathematics students might be more likely to exhibit a contrast advantage in our paradigm than high school students learning these concepts for the first time. As with language, the power of compositional representations lies in their productivity and systematicity, the fact that novel concepts can be built and understood by arranging known (unitary) relations in ways a person has not previously encountered (Fodor & Pylyshyn, 1988). Thus, although our results suggest that people engage in alignment of explicitly structured representations only when less-demanding approaches are inadequate, this does not diminish the importance of structural alignment for the more impressive feats of human cognition, such as problem solving, invention, and scientific discovery (French, 2002; Gentner & Forbus, 2011; Gentner & Markman, 1997).

Context of the Research

This project was motivated by two factors: (1) the connection between relational category learning and analogical reasoning and (2) the theoretical distinction between feature processing and structural alignment. Based on these ideas, we initially expected to find an interaction between category type and comparison type, with a contrast advantage for feature-based categories and a match advantage for relational categories. Our goal was to demonstrate that feature-based and relational categories are learned in fundamentally different ways. However, despite an extensive effort to find conditions that would produce a relational match advantage, we found a contrast advantage in every case that we tested, including one feature-based category structure and four relational ones. Therefore, we consider the present results especially strong evidence that relational category learning is more efficient from between-category comparisons. These findings have shifted our theoretical outlook, such that we have grown more skeptical towards the applicability of structure-mapping theory to relational category learning, at least in its most direct form. Our research program has shifted as well, such that the first author is now closely investigating how compositional versus unitary representations affect relational concept learning, as well as the factors that lead people to represent relational concepts compositionally and unitarily (Corral, 2017).

Conclusions

Although there has been much recent interest in potential connections between analogical reasoning and relational category learning, the present findings suggest important differences. Structure-mapping theory makes a clear prediction of an advantage of within-category comparisons, which is refuted by the experiments reported here. Instead, the results are consistent with feature-based processing. Rather than viewing feature-based and analogical models as strict competitors, we see more promise in integrated accounts like the unitary/compositional framework described here. Hopefully the present results and theoretical interpretation will lead to a better understanding of the interplay between representations of individual features and representations of relational structure, as well as the conditions under which people do and do not engage in structural alignment.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*, 87-113. https://doi.org/10.1080/00461520.2013.775712

Allen, S. W., Brooks, L. R., 1991. Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120*,

3–19.
Andrews, J. K., Livingston, K., & Kurtz, K. (2011). Category learning in the context of co-presented items. *Cognitive Processing, 12*, 161-175. https://doi.org/10.1007/s10339-010-0377-5
Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442-481.
Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. Journal of Experimental Psychology: General, 120, 150-172.
Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-78.
Barbella, D., & Forbus, K. D. (2013). Analogical word sense disambiguation. *Advances in Cognitive Systems*, *2*, 297–315.
Bowdle, B., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology*, 34, 244-286.
Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61-100. https://doi.org/10.2307/1167267
Carvalho, P. F., & Goldstone, R. L. (2014a). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*, 481-495. https://doi.org/10.3758/s13421-013-0371-0
Carvalho, P. F. & Goldstone, R. L. (2014b). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology*, *5*, 1-10. https://doi.org/10.3389/fpsyg.2014.00936
Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*, 1-12. https://doi.org/10.3389/fpsyg.2015.00505
Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical and Artificial Intelligence, 4*, 185-211. https://doi.org/10.1080/09528139208953747
Chang, M. D., & Forbus, K. D. 2013. Clustering hand-drawn sketches via analogical generalization. *Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence (IAAI-13),* 1507-1512.
Chang, M. D., & Forbus, K. D. (2014). Using analogy to cluster hand-drawn sketches for sketch-based educational software. *AI Magazine*, *35*, 76–84. https://doi.org/10.1609/aimag.v35i1.2505
Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage, 14*, 1136–1149.
Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*, 370-374.
Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15, 89-132.
Corral, D. (2017). A dual model of relational concept representation. *Doctoral Dissertation*, *University of Colorado Boulder*.
Corral, D., & Jones, M. (2017). Learning relational concepts through unitary- versus compositional-based representations. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 1830-1835).
Corral, D., & Jones, M. (2014). The effects of relational structure on analogical learning. *Cognition, 132*, 280-300. https://doi.org/10.1016/j.cognition.2014.04.007
DeCoster, J., Iselin, A. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods, 14*, 349-366. https://doi.org/10.1037/a0016956
Dietrich, E. S. (2010). Analogical insight: Toward unifying categorization and analogy. *Cognitive Processing*, *11*, 331-345. https://doi.org/10.1007/s10339-010-0367-7
Doumas, L. A. A., & Hummel, J. E. (2004). Structure mapping and the predication of novel higher-order relations. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 333-338).
Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1-43. https://doi.org/10.1037/0033-295X.115.1.1
Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18,* 500-549. https://doi.org/10.1016/0010-0285(86)90008-3
Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: algorithm and examples. *Artificial Intelligence, 41*, 1-63. https://doi.org/10.1016/0004-3702(89)90077-5
Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71. https://doi.org/10.1016/0010-0277(88)90031-5
Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141-205. https://doi.org/10.1207/s15516709cog1902_1
Foster, J. M., Cañas, F., & Jones, M. (2012). Learning conceptual hierarchies by iterated relational consolidation. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 324–329).
French, R. M. (1997). When coffee cups are like old elephants or why representation modules don't make sense. In A. Riegler & M. Peschl (Eds.), *Proceedings of the International Conference on New Trends in Cognitive Science* (pp. 158-163). Austrian Society for Cognitive Science.
French, R. M. (2002). The computational modeling of analogy-making. *Trends in Cognitive Science, 6*, 200-205. https://doi.org/10.1016/S1364-6613(02)01882-X
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155-170. https://doi.org/10.1207/s15516709cog0702_3
Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp.195-235). Cambridge, MA: MIT Press.
Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*, 752-775.
Gentner, D., & Bowdle, B. (2008). Metaphor as structure-mapping. In R. Gibbs (Ed.) *The Cambridge Handbook of Metaphor and Thought* (pp. 109-128). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.008
Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 266-276. https://doi.org/10.1002/wcs.105
Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory and Cognition, 29*, 565-577.
Genter, D., & Kurtz, K. J. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the laboratory* (pp. 151-175). American Psychological Association. https://doi.org/10.1037/11156-009
Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, *5*, 152-158. https://doi.org/10.1111/j.1467-9280.1994.tb00652.x

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56. https://doi.org/10.1037/0003-066X.52.1.45
Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487-513. https://doi.org/10.1016/S0885-2014(99)00016-7
Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25, 524-575.
Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, *37*, 331-355. https://doi.org/10.1006/jmla.1997.2527
Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38. https://doi.org/10.1016/0010-0285(83)90002-6
Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology*, *46*, 539-550. https://doi.org/10.1037/h0084333
Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227-247.
Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences, 7*, 92-96. https://doi.org/10.1016/S1364-6613(02)00040-2
Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language*, *36*, 50-67. https://doi.org/10.1006/jmla.1996.2479
Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 130*, 116–139.
Goldwater, M., Don, H., Krusche, M., & Livesey, E. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General, 147*, 1-35.
Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137-153. https://doi.org/10.1016/j.cognition.2014.12.001
Goldwater, M., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*, 729-757. https://doi.org/10.1037/bul0000043
Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory and Cognition,* *24*, 608–628. https://doi.org/10.3758/BF03201087
Hammer, R., Hertz, T., Hochstein, S., & Weinshall, D. (2009). Category learning from equivalence constraints. *Cognitive Processing, 10*, 211-232. https://doi.org/10.1007/s10339-008-0243-x
Hendrickson, A. T., Perfors, A., Navarro, D. J., & Ransom, K. (in press). Sample size, number of categories and sampling assumptions: Exploring some differences between categorization and generalization. *Cognitive Psychology.*
Higgins, E. J. (2012). Comparing comparisons in category learning. *Doctoral Dissertation, University of Illinois*.
Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 1388-1393.
Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332-340. https://doi.org/10.3758/BF03197035
Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.
Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American Psychologist, 52*, 35-44. https://doi.org/10.1037/0003-066X.52.1.35
Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 220-264. https://doi.org/10.1037/0033-295X.104.3.427
Hummel J. E., & Holyoak K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220-264. https://doi.org/10.1037/0033-295X.110.2.220
Jones, M., Maddox, W. T., & Love, B. C. (2005). Stimulus generalization in category learning. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, 1066-1071.
Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology, 55*, 196-231.
Jung, W., & Hummel, J. E. (2011). Progressive alignment facilitates learning of deterministic but not probabilistic relational categories. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 2643-2648.
Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97-103. https://doi.org/10.1002/acp.1801
Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol*, *17*, 249-262. https://doi.org/10.1207/S15327868MS1704_1
Kittur, A., Hummel, J. E., & Holyoak, K. J. (2004). Feature- vs. relation-defined categories: Probab(alistic)ly not the same. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 696-701.
Kokinov, B. (1988). Associative memory-based reasoning: How to represent and retrieve cases. In T. O'Shea & V. Sgurev (Eds.), *Artificial intelligence III: Methodology, systems, applications* (pp. 51-58). Amsterdam: Elsevier Science Publ.
Kokinov, B. (1994). Flexibility versus efficiency: The dual answer. In P. Jorrand & V. Sgurev (Eds.), *Artificial intelligence: Methodology, systems, applications*. Singapore: World Scientific Publ.
Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, *19*, 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x
Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*, 498–503. https://doi.org/10.1037/a0017807
Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797-2822. https://doi.org/10.2307/1131753
Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 12*, 477–485.
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44. https://doi.org/10.1037/0033-295X.99.1.22
Kuehne, S., Forbus, K., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, 770-775.
Kurtz, K. J. (2015). Human category learning: Toward a broader explanatory account. *Psychology of Learning and Motivation*, *63*, 77–114.
Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison

promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1303-1310. https://doi.org/10.1037/a0031847

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

Krawczyk, D. C., Holyoak, K. J., & Hummel, J. E. (2005). The one-to-one constraint in analogical mapping and inference. *Cognitive Science*, *29*, 29-38.

Larkey, L. B. & Love, B. C. 2003. CAB: Connectionist analogy builder. *Cognitive Science*, *27*, 781-794.

Lassaline, M. E., & Murphy, G. L. (1998). Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 144-160. https://doi.org/10.1037/0278-7393.24.1.144

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & Cognition*, *43*, 266-282.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, *6*, 586-597. https://doi.org/10.3758/BF03212967

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Maddox, W.T. (2002). Learning and attention in multidimensional identification, and categorization: Separating low-level perceptual processes and high level decisional processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 99-115.

Markman, A. B. (1999). *Knowledge representation*. Hillsdale, NJ: Erlbaum.

Markman, A. B., & Gentner, D. (1993a). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431-467. https://doi.org/10.1006/cogp.1993.1011

Markman, A. B., & Gentner, D. (1993b). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, *32*, 517-535. https://doi.org/10.1006/jmla.1993.1027

Markman, A. B., & Gentner, D. (2000). Structure-mapping in the comparison process. *American Journal of Psychology*, *113*, 501-538. https://doi.org/10.2307/1423470

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592-615. https://doi.org/10.1037/0033-2909.129.4.592

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence, 13*, 329-358. https://doi.org/10.1080/09528130110100252

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 128-148.

McLure, M., Friedman, S. E., & Forbus, K. D. (2010). Learning concepts from sketches via analogical generalization and near-misses. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1726–1731).

McLure, M., Friedman, S., Lovett, A., & Forbus, K. (2011). Edge-cycles: A qualitative sketch representation to support recognition. *Proceedings of the 25th International Workshop on Qualitative Reasoning*.

Mitchell, M. & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy-making. *Physica D, 42*, 322-334. https://doi.org/10.1016/0167-2789(90)90086-5

Morrison, R. G., Holyoak, K. J., & Truong, B. (2001). Working memory modularity in analogical reasoning. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 663-668).

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131*, 5-15. https://doi.org/10.1037/0096-3445.131.1.5

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review,84,* 231-259.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics, 45*, 279-290.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, 17(1), 3-27.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*, 266–300.

Patterson, J. D., & Kurtz, K. J. (2015). Learning mode and comparison in relational category learning. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 1841-1846.

Patterson, J. D., & Kurtz, K. J. (2016). Performance pressure and comparison in relational category learning. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 2333-2338.

Penn, D. C., Holyoak, K J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109-178. https://doi.org/10.1017/s0140525x08003543

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology, 33*, 43–63.

Ramscar, M., & Pain, H. (1996). Can a real distinction be made between cognitive theories of analogy and categorisation? *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, 346-351.

Reed, S. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, *81*, 532-540. https://doi.org/10.1037//0022-0663.81.4.532

Rehder, B. & Hoffman, A.B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*, 1-41.

Rehder, B. & Ross, B.H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1261-1275. https://doi.org/10.1037/0278-7393.27.5.1261

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 629-639. https://doi.org/10.1037/0278-7393.13.4.629

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468. https://doi.org/10.1037/0278-7393.15.3.456

Schwartz, D. L. & Bransford, J. D. (1998). A time for telling.

Cognition and Instruction, *16*, 475-522. https://doi.org/10.1207/s1532690xci1604_4

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*, 54-87. https://doi.org/10.1016/0022-2496(64)90017-3

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145-166.

Skorstad, J., Gentner, D., & Medin, D. (1988). Abstraction processes during concept learning: a structural view. *Proceedings of the 10th Annual Conference of the Cognitive Science Society*, 419-424.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22,* 189–228.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65*, 167–196.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1411–1436.

Smith, L. A., & Gentner, D. (2014). The role of difference-detection in learning contrastive categories. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, 1473-1478.

Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62,* 913-933.

Taylor, J. L. M., Friedman, S. E., Forbus, K. D., Goldwater, M., & Gentner, D. (2011). Modeling structural priming in sentence production via analogical processes. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629-640.

Tomlinson, M., & Love, B. C. (2006). Learning abstract relations through analogy to concrete exemplars. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2269-2274.

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

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*, 750-763. https://doi.org/10.3758/s13421-010-0063-y

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119–125.

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.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1-39. https://doi.org/10.1207/s1532690xci0701_1

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39*, 124-148.

Appendix A: Stimuli for Experiment 2

Category A			
1.	Bathtub, Water	41.	Pot, Coffee
2.	Bottle, Perfume	42.	Tube, Toothpaste
3.	Cup, Juice	43.	Aquarium, Fish
4.	Spoon, Medicine	44.	Farm, Pigs
5.	Reservoir, Water	45.	Stable, Horse
6.	Automobile, Gas	46.	Engine, Oil
7.	House, People	47.	Safe, Valuables
8.	School, Students	48.	Mall, Stores
9.	Bank, Teller	49.	Pitcher, Beer
10.	Vehicle, Passengers	50.	Classroom, Desk
11.	Ocean, Fish	51.	Volcano, Magma
12.	Solar System, Planets	52.	Spaceship, Astronauts
13.	Forest, Trees	53.	Airplane, Pilot
14.	Courtroom, Lawyers	54.	Wastebasket, Trash
15.	Swamp, Alligator	55.	Movie, Actors
16.	Jungle, Leopard	56.	Concert, Musicians
17.	Zoo, Animals	57.	Wallet, Money
18.	Antarctica, Penguins	58.	Mouth, Teeth
19.	Shoe, Foot	59.	Restaurant, Menu
20.	Glove, Hand	60.	Galaxy, Stars
21.	Building, Furniture	61.	Forest, Plants
22.	Living Room, Sofa	62.	Canyon, Rocks
23.	Library, Books	63.	Bedroom, Bed
24.	Frame, Picture	64.	Bathroom, Toothbrush
25.	Dishwasher, Utensils	65.	Hamper, Clothes
26.	Refrigerator, Food	66.	Kitchen, Sink
27.	Register, Money	67.	Birdcage, Bird
28.	Book, Words	68.	Firearm, Ammunition
29.	Glass, Water	69.	Military, Soldiers
30.	Toaster, Bread	70.	Race, Runners
31.	Jar, Cookies	71.	Pool, Swimmers
32.	Gym, Treadmill	72.	Beach, Sand
33.	Carafe, Lemonade	73.	Cabinet, Dishes
34.	Envelope, Letter	74.	Park, Swings
35.	Vault, Money	75.	Shower, Faucet
36.	Sky, Clouds	76.	Sky, Birds
37.	Hangar, Airplane	77.	Mirror, Reflection
38.	Harbor, Boat	78.	Printer, Paper
39.	Garage, Car	79.	Theater, Chairs
40.	Closet, Clothes	80.	Head, Brain

Category B			
1.	Foundation, Building	41.	Suspensions, Bridge
2.	Legs, Table	42.	Rope, Acrobat
3.	Stand, Television	43.	Rope, Piñata
4.	Shelf, Books	44.	Suspenders, Pants
5.	Tires, Vehicle	45.	Crane, Piano
6.	Parachute, Skydiver	46.	Scaffolding, Workers
7.	Hanger, Clothes	47.	Tree, Swing
8.	Road, Vehicle	48.	Hot Air Balloon, Basket
9.	Bench, People	49.	Tripod, Telescope
10.	Backboard, Basketball Rim	50.	Railing, Ballerina
11.	Walls, Ceiling	51.	Ceiling, Chandelier
12.	Tracks, Train	52.	Post, Sign
13.	Rod, Curtain	53.	Straps, Backpack
14.	String, Yoyo	54.	Wind, Kite
15.	Rope, Climber	55.	Life Preserver, Person
16.	Balance beam, Gymnast	56.	Stretcher, Person
17.	Bicycle, Bicyclist	57.	Pull-Up Bar, Person
18.	Nail, Clock	58.	Fishing Line, Fishing Hook
19.	Belt, Pants	59.	Branch, Nest
20.	Barbell, Weights	60.	Crane, Cable
21.	Bridge, Cars	61.	Branch, Leaves
22.	Neck, Tie	62.	Trunk, Branches
23.	Cable, Elevator	63.	Balloon, String
24.	Horse, Jockey	64.	Pole, Clothesline
25.	Pole, Flag	65.	Toothbrush, Toothpaste
26.	Ladder, Firefighter	66.	Face, Beard
27.	Crutch, Person	67.	Brackets, Shelf
28.	Rack, Towel	68.	Post, Fence
29.	Ear, Earring	69.	Building, Gargoyle
30.	Bed, People	70.	Tray, Food
31.	Closet Rod, Hangers	71.	Strap, Purse
32.	Hook, Picture	72.	Table, Plates
33.	Tree, Tree House	73.	Dolly, Equipment
34.	Stand, Aquarium	74.	Ski Lift, Skiers
35.	Nose, Glasses	75.	Pedestal, Statue
36.	Tree, Hammock	76.	Jetpack, Person
37.	Hinges, Door	77.	Ironing Board, Iron
38.	Stake, Scarecrow	78.	Basketball Rim, Net
39.	Perch, Bird	79.	Handle, Mug
40.	Columns, Coliseum	80.	Walker, Toddler

Appendix B: Stimuli for Experiment 4

Morkels	Non-Morkels
1. Operates on land Works to gather harmful solids Has a shovel	1. Operates on land Works to clean spilled oil Has an electrostatic filter
2. Operates on the surface of the water Works to clean spilled oil Has a spongy material	2. Operates on the surface of the water Works to collect dangerous gaseous ions Has a shovel
3. Operates in the stratosphere Works to collect dangerous gaseous ions Has an electrostatic filter	3. Operates in the stratosphere Works to gather harmful solids Has a spongy material
4. Operates in highway tunnels Works to remove carbon dioxide Has a large intake tank	4. Operates in highway tunnels Works to remove lost fishing nets Has a sifter
5. Operates in swamps Works to remove malaria-ridden mosquitos Has a finely woven net	5. Operates in swamps Works to remove broken glass Has a metal pole with sharpened end
6. Operates in warzones Works to gather shards of metal Has a large magnet	6. Operates in warzones Works to gather discarded paper Has a finely woven net
7. Operates in parks Works to gather discarded paper Has a metal pole with a sharpened end	7. Operates in parks Works to gather shards of metal Has a hook
8. Operates on the seafloor Works to remove lost fishing nets Has a hook	8. Operates on the seafloor Works to remove malaria-ridden mosquitoes Has a large intake fan
9. Operates on the beach Works to remove broken glass Has a sifter	9. Operates on the beach Works to remove carbon dioxide Has a large magnet
10. Operates on wood floors Works to remove stains Has absorbent cloth	10. Operates on wood floors Works to collect ocean sediments Has an intake port
11. Operates on solid surfaces Works to remove debris Has a dense set of bristles	11. Operates on solid surfaces Works to remove large rocks Has absorbent cloth
12. Operates on glass Works to remove liquids Has a rubber edge	12. Operates on glass Works to collect small particles Has a dense set of bristles
13. Operates on thick woven fabric Works to collect small particles Has an intake port	13. Operates on thick woven fabric Works to remove brush Has a large shovel
14. Operates in harbors Works to collect ocean sediments Has a large shovel	14. Operates in harbors Works to remove leaves Has metal teeth and a sieve
15. Operates in the jungle Works to remove brush Has sharp blades	15. Operates in the jungle Works to remove debris Has a rubber edge
16. Operates in farmland Works to remove rocks Has metal teeth and a sieve	16. Operates in farmland Works to smooth rough spots Has a rough metal surface
17. Operates on fine wood Works to smooth rough spots Has a rough metal surface	17. Operates on fine wood Works to remove liquids Has a nozzle and a motor
18. Operates in gardens Works to remove leaves Has a nozzle and a motor	18. Operates in gardens Works to remove stains Has sharp blades

Note: Morkels were coherent items and non-Morkels were incoherent items. Items1-3 in each category were taken from Rehder and Ross (2001), Items 4-9 were taken from Higgins (2012), and Items 10-18 were created by the present authors.

Appendix C: Experiment Instructions

Experiment 1

Match Condition

"Recently, a large number of unexplained crop circles have been found throughout the United States. A task force has been appointed to investigate the matter. The task force has identified two types of crop circles that are produced by two different alien races: the Alkins and the Bafsters. You have been selected to join the task force, but before you can begin you must learn to identify the crop circles that are being produced by each alien race. To help you in this task, we are going to show you pairs of crop circles, each produced by the same aliens. For each pair, your job is to decide whether they were both produced by Alkins or Bafsters. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Contrast Condition

"Recently, a large number of unexplained crop circles have been found throughout the United States. A task force has been appointed to investigate the matter. The task force has identified two types of crop circles that are produced by two different alien races: the Alkins and the Bafsters. You have been selected to join the task force, but before you can begin you must learn to identify the crop circles that are being produced by each alien race. To help you in this task, we are going to show you pairs of crop circles, one produced by Alkins and the other by Bafsters. For each pair, your job is to decide which is which. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Experiment 2

Match Condition

"Government intelligence has recently discovered that rogue agents are sending each other secret messages by using paired words. A task force has been appointed to investigate the matter. The task force has identified two types of secret codes: Code A and Code B. You have been selected to join the task force, but before you can begin you must learn to identify the different codes. To help you in this task, we are going to show you two pairs of words that the rogue agents have been using, both from the same code. For each pair, your job is to decide whether they are both from Code A or both from Code B. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Contrast Condition

"Government intelligence has recently discovered that rogue agents are sending each other secret messages by using paired words. A task force has been appointed to investigate the matter. The task force has identified two types of secret codes: Code A and Code B. You have been selected to join the task force, but before you can begin you must learn to identify the different codes. To help you in this task, we are going to show you two pairs of words that the rogue agents have been using, one from Code A and one from Code B. For each pair, your job is to decide which is from Code A and which is from Code B. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Experiment 3

Match Condition

"Recently, mysterious alien writing has been found throughout the United States. A task force has been appointed to investigate the matter. The task force has identified two types of alien symbols that are produced by two different alien races: the Alkins and the Bafsters. You have been selected to join the task force, but before you can begin you must learn to identify the symbols that are being produced by each alien race. To help you in this task, we are going to show you pairs of alien symbols, each produced by the same aliens. For each pair, your job is to decide whether they were both produced by Alkins or Bafsters. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Contrast Condition

"Recently, mysterious alien writing has been found throughout the United States. A task force has been appointed to investigate the matter. The task force has identified two types of alien symbols that are produced by two different alien races: the Alkins and the Bafsters. You have been selected to join the task force, but before you can begin you must learn to identify the symbols that are being produced by each alien race. To help you in this task, we are going to show you pairs of alien symbols, one produced by Alkins and the other produced by Bafsters. For each pair, your job is to decide which is which. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Experiment 4

Match Condition

"In this experiment, you will learn about a type of machine called a Morkel. Your job will be to learn which machines are Morkels and which are not. On each trial, you will read descriptions about two machines, and you'll need to decide whether they are both Morkels or not. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."

Contrast Condition

"In this experiment, you will learn about a type of machine called a Morkel. Your job will be to learn which machines are Morkels and which are not. On each trial, you will read descriptions about two machines, and you'll need to decide which is a Morkel and which is not. You will be told the right answer after you respond, so that you can learn. Press the spacebar to begin."