The Effects of Relational Structure on Analogical Learning

Daniel Corral & Matt Jones

University of Colorado Boulder

In press, *Cognition*

Abstract

Relational structure is important for various cognitive tasks, such as analogical transfer, but its role in learning of new relational concepts is poorly understood. This article reports two experiments testing people's ability to learn new relational categories as a function of their relational structure. In Experiment 1, each stimulus consisted of 4 objects varying on 2 dimensions. Each category was defined by two binary relations between pairs of objects. The manner in which the relations were linked (i.e., by operating on shared objects) varied between subjects, producing 3 logically different conditions. In Experiment 2, each stimulus consisted of 4 objects varying on 3 dimensions. Categories were defined by three binary relations, leading to six logically different conditions. Various learning models were compared to the behavioral data, based on the theory of schema refinement. The results highlight several shortcomings of schema refinement as a model of relational learning: (1) it can make unreasonable demands on working memory, (2) it does not allow schemas to grow in complexity, and (3) it incorrectly predicts learning is insensitive to relational structure. We propose schema elaboration as an additional mechanism that provides a more complete account, and we relate this mechanism to previous proposals regarding interactions between analogy and representation construction. The current findings may advance understanding of the cognitive mechanisms involved in learning and representing relational concepts.

Keywords: relational structure; schema refinement; schema elaboration; structured representation; analogy; relational category learning.

1. Introduction

The ability to generalize and transfer knowledge from a given problem to an analogous task has been of great interest to cognitive scientists. Over the last 30 years, one of the most influential ideas to come out of the large body of work on analogical learning is that transfer is driven by discovering the relational structure shared between two analogous scenarios (Gentner, 1983; Gick & Holyoak, 1983; Hummel & Holyoak, 1997). Penn, Holyoak, and Povinelli (2008) posit that recognition and processing of relational information are critical for many other higher cognitive processes as well, including inference, causal reasoning, and theory of mind. Recognizing a problem's relational structure can aid in filtering out superfluous information that can inhibit transfer (Cooper & Sweller, 1987; Gick & Holyoak, 1983; Sweller, Mawer, & Ward, 1983). Furthermore, representing problems in terms of their relational structure plays a critical role in expert problem solving (Chase & Simon, 1973; Chi, Feltovich, & Glazer, 1981). Taken together, these findings suggest that cognitive functions such as concept learning, reasoning, problem solving, and decision-making are greatly dependent upon and can be improved by learning to recognize a scenario's underlying relational structure.

How people learn structured relational concepts is thus a fundamental question for the study of higher-level cognition. However, there is reason to believe the cognitive processes that subserve relational learning are more complex than simply aligning examples to discover their shared structure. For scenarios of even moderate complexity, there is an exponential explosion of possible relations, which moreover can be represented at various levels of abstraction. Finding the right relational representations of both scenarios in an analogy is crucial for analogical reasoning to succeed (Chalmers, French, & Hofstadter, 1992; French, 1997; Mitchell & Hofstadter, 1990). Furthermore, some relational concepts seem intuitively more coherent than others, in that they are easier to learn and use. Conceptual coherence has been argued to arise from higher-order relational constraints that exist both within and between concepts (Murphy &

Medin, 1985), but there is still a lack of understanding of what those constraints are or of how they confer coherence.

The primary aim of this article is to provide a rigorous definition of relational structure that is grounded in the literature on analogical reasoning, and to conduct a systematic empirical investigation of whether and how relational structure influences learnability of relational concepts. We define *relational structure* as the manner in which a system of relations is linked together by shared role-fillers (e.g., the object filling role 1 of relation 1 also fills role 1 of relation 2). We elaborate on this definition in Section 1.1 and explain how it is closely related to structure mapping, the dominant psychological theory of analogy (Gentner, 1983). In short, structure mapping is based on the principle of parallel connectivity, according to which two systems are alignable only if their constituent objects and relations are connected in the same way. Thus structure-mapping theory holds that analogy is precisely the process of identifying the shared relational structure between two scenarios (as defined here).

The experiments reported here bridge research on category learning and analogy, using a categorization task to test subjects' ability to learn relationally defined categories (for precedents on this approach, see Kittur, Hummel, & Holyoak, 2004; Rehder & Ross, 2001; Tomlinson & Love, 2010). Each subject learned a category defined by a particular relational structure, by learning to distinguish category members from non-members. Each stimulus was an arrangement of four objects, varying on simple perceptual dimensions that defined binary comparative dimensions between objects (e.g., LARGER and BRIGHTER). Critically, the categories in all experimental conditions were defined by the same two (Experiment 1) or three binary relations (Experiment 2). The conditions differed only in relational structure, that is, in how those relations were linked by operating on common objects. Thus, comparing performance across conditions provides a pure test of whether relational structure affects learnability of a relational concept. In the spirit of Shepard, Hovland, and Jenkins' (1961) classic study on learning of feature-based categories, we conducted an exhaustive comparison of the three (Experiment 1)

and six (Experiment 2) logically different categories within this design (i.e., the different possible relational structures). The goals were to determine how well people can learn these relational categories, and whether learning performance differs according to relational structure.

A second focus of this article is how the representation of a relational concept evolves over the course of learning, upon encounters with multiple instances of the concept. The predominant theory of how people represent a structured relational concept (i.e., an abstract relational system) is as a *schema*, which can be defined as a set of (usually abstract) objects together with relations operating on those objects (Gentner, 1983; Hummel & Holyoak, 2003; Rumelhart & Norman, 1978).[1] One common proposal for how people learn schemas for relational concepts is through *schema refinement*, whereby analogical comparison of a schema to successive examples strips away superfluous details, leaving only the necessary information that defines the concept. Although schema refinement plays a critical role in several influential models of relational learning (Doumas, Hummel, & Sandhofer, 2008; Kuehne, Forbus, Gentner, & Quinn, 2000), the present results highlight some critical shortcomings of this mechanism. In particular, schema refinement alone predicts no learning differences across different relational structures (whereas the data show learning is structure-sensitive), and it cannot explain the overall level of human performance in these experiments. We introduce *schema elaboration* as an additional mechanism that leads schema-based models to better match human behavior. We consider four variants of schema elaboration, motivated by different theoretical perspectives, and compare their ability to predict the relative learnability of different relational structures.

The primary theoretical message of this article is that, although the framework of structure mapping and schema refinement has been valuable and successful, it does not capture the full richness of the cognitive processes underlying analogy. As French (1997) has argued, the

[1] We maintain a distinction between the terms *schema* and *relational structure* so as to separate the subject's representation from the objective concept defined by the experimenter.

mapping-and-refinement framework oversimplifies the problem, by ignoring the interactive dynamics between analogy and representation construction. These points are paralleled in Section 1.2, where we argue against schema refinement as too simplistic to explain how people develop representations of new relational concepts. We address these shortcomings through testing various relational learning models (presented in Section 3), which focus on the development of representations for relational concepts. A strength of these models is that the combination of schema refinement and elaboration produces an interplay between bottom-up and top-down processes that enables the models to dynamically converge on the correct representation (cf. Mitchell & Hofstadter, 1990). Although this mechanism is by no means a complete answer, it contributes a significant step toward understanding the power and flexibility of human analogical learning.

1.1. Structure-Mapping Theory

Many current models of analogy have been strongly influenced by structure-mapping theory. Since its initial proposal (Gentner, 1983), structure-mapping theory has provided a great deal of insight into the cognitive processes of analogical learning and transfer. Structure-mapping theory posits that forming an analogy between two scenarios involves an alignment process that puts their constituent objects and relations into one-to-one correspondence. The primary goal of the alignment process is *parallel connectivity*, the property that if two relations are aligned then their corresponding role-fillers are aligned as well. Consider the classic solar-system–atom analogy (Gentner, 1983), a subset of which is shown in Figure 1: Planets revolve around the sun, and planets are smaller than the sun; electrons revolve around the nucleus, and electrons are smaller than the nucleus. The analogy works not only because the same first-order relations are present in both scenarios (i.e., SMALLER THAN and REVOLVES AROUND), but also because these relations share objects in the same way. Specifically, in both scenarios the object that is smaller is also the object doing the revolving, and the object that is larger is also the object being revolved around.

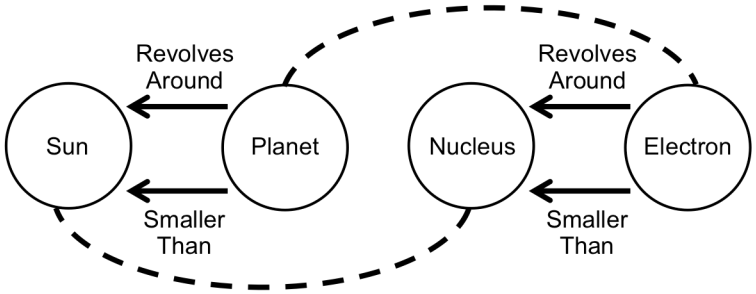


Figure 1. Diagram of simplified solar-system–atom analogy. Dashed lines indicate the alignment between the two scenarios. Relations are also aligned (SMALLER THAN to SMALLER THAN and REVOLVES AROUND to REVOLVES AROUND), but these lines are omitted for clarity.

It is easy to see that structure-mapping theory's account of analogy is equivalent to recognition that two scenarios have the same relational structure, under the definition given here. Formally, we define the relational structure in (a person's representation of) a scenario to comprise (1) the set of relations that are present and (2) how those relations are linked together by shared role-fillers. In the example above, there are two linkages between the SMALLER THAN and REVOLVES AROUND relations: one object fills both of their agent roles, and one object fills both of their patient roles. Thus two scenarios have the same relational structure if and only if (a) their relations can be put into correspondence and (b) the role-fillers of those relations can be put into correspondence in a way that preserves role binding (i.e., which object fills each role of each relation). The latter condition is exactly the parallel connectivity constraint. Hence relational structure appears to be a driving force for analogy, and the ability to learn and recognize different relational structures likely plays an important role in the effectiveness of analogical reasoning.

1.2. Schema Refinement

Evidence that analogy facilitates learning of relational concepts comes from studies showing that comparison between two analogous scenarios leads to improved reasoning about

a new scenario sharing relational structure with the first two (e.g., Gick & Holyoak, 1983). These findings have been explained by the proposal that analogy leads to induction of a schema that embodies the relational structure common to both analogues (Hummel & Holyoak, 2003; Kuehne et al., 2000). The schema can then be mapped to future scenarios, allowing for inferences about scenarios that are recognized as analogous to the schema. For example, the schema for the solar-system–atom analogy (or the simplification given in Figure 1) would consist of two generic object tokens linked by the relations SMALLER THAN and REVOLVES AROUND. This schema could be used to predict outcomes of new scenarios, such as the behavior of a man-made satellite projected beyond the earth's atmosphere (i.e., the smaller object, the man-made satellite, will revolve around the larger object, the earth).

A schema derived in this way from comparison of two scenarios is likely to include idiosyncratic information that happens to be present in both scenarios but would not apply to other analogous scenarios. For example, a schema derived from the solar-system–atom analogy might include the property that both objects are naturally occurring. To develop a truly abstract concept that will generalize over superficial properties, many models of analogical learning propose an additional process of *schema refinement* (Doumas et al., 2008; Hummel & Holyoak, 2003; Kuehne et al., 2000). In schema refinement, analogical mapping between a schema and a new scenario results in modification (or replacement) of the schema, so that it contains only the structure that is common to the original schema and the new scenario. Continuing the example, comparison to the man-made satellite scenario would lead the NATURALLY OCCURRING property to be removed from the schema. Through comparison to successive instances of an abstract concept (as they are encountered), a schema can be refined so that it eventually includes only the information that is common to all instances.

Schema refinement thus offers a natural explanation for how people might learn a relational category in classification tasks like the ones reported here. Nevertheless, we argue that schema refinement has important shortcomings as a theory of relational learning. These shortcomings

follow from the assumption that schemas are only modified by simplification (i.e., removal of objects or relations). First, contrary to the results presented in the experiments below, learning by refinement alone leads to a prediction of no false alarms during learning of a relational category. A false alarm occurs when a stimulus that does not fit the category is mistaken for a category member, because it contains all of the structure present in the subject's current schema. This can only happen if the schema lacks relational constraints for what constitutes category membership. This is impossible under an idealized model of pure schema refinement, because relations that are present in all category members will never be removed from the schema.

Second, refinement cannot add new information to a schema. Consequently, upon its first encounter with a member of a relational category, a pure refinement model would have to retain all information contained in that instance, because any of that information may turn out to be necessary in defining the category. For example, after being presented with the first instance of the solar-system scenario, a subject would have to retain the exact sizes of both objects, the speed at which the small object revolves around the large object, the fact that the smaller object hosts living creatures and the larger one does not, and so on. Given the processing constraints of working memory (Baddeley, 2003), such an assumption seems psychologically implausible. Instead, subjects should be expected to quickly forget a large amount of the information that was initially processed.

On the other hand, this argument from memory limitations relies on an assumption that schema information is stored propositionally, or in some other capacity-limited form. If relational information can be stored in a way not subject to strong capacity limits, then learning by pure refinement may be viable. For example, the relations used in the present experiments could in principle be maintained spatially, in a system like a visuospatial sketchpad (Baddeley, 1986). To preview the results reported below, we find significant limitations in subjects' relational learning,

especially with more complex structures (Experiment 2), thus arguing against a model of relational learning by refinement alone.

1.3. Schema Elaboration

The arguments in Section 1.2 suggest schema refinement is not a complete account of how relational concepts are acquired. Thus, we propose that a more complete model of relational learning must incorporate forgetting, and, consequently, the ability to add information to the schema rather than only simplifying it. Specifically, we propose that when a schema is insufficiently complex (i.e., is missing appropriate relational constraints), the subject can update it by adding new relations. We refer to this process as *schema elaboration*.

One paradigm that highlights the differences between schema refinement and elaboration is category learning. Consider a task like the ones reported in the experiments below, in which the subject is presented with a series of stimuli, is asked to classify each as a category member or nonmember, and learns from feedback following each response. A simple schema-based model assumes the subject maintains a schema representing his or her current hypothesis about the category structure (this is essentially a prototype model; we consider exemplar-based approaches at the conclusion of this article). Each new stimulus is compared to the schema, and the subject classifies the stimulus as in the category if and only if it satisfies all relations in the schema (i.e., if the schema can be fully mapped into the stimulus). Under this framework, opportunities for refinement occur precisely whenever the subject commits a miss (i.e., misclassifies a category member as a nonmember). On these trials, the stimulus lacks one or more relational constraints in the current schema, and refinement involves removing those constraints by intersecting the schema with the stimulus. Consequently, the stimulus would be correctly recognized as a category member if it were to be presented again.

If the subject learns by refinement alone, the relational constraints that make up the category rule will never be lost, because they are present in every member of the category. It follows that a pure refinement model will never commit a false alarm (i.e., misclassify a

nonmember as a category member). However, if a subject forgets one of these true relational constraints, false alarms become possible. A false alarm then affords an opportunity for schema elaboration, because it indicates the schema is incomplete. Elaboration following a false alarm involves augmenting the schema with relational constraints that the current stimulus violates, so that the stimulus would be recognized as a nonmember if it were to be presented again.

Thus, according to this framework, schema refinement involves removing extraneous information from the schema following a miss, and schema elaboration involves adding information following a false alarm. A further important difference between refinement and elaboration is that the subject can logically deduce what constraints need to be removed following a miss (i.e., all constraints in the schema that the current stimulus violates), whereas following a false alarm it is not logically determined which constraints need to be added (all that is known is that the stimulus violates some aspect of the category). One important consequence of this difference, as demonstrated by the models described in Section 3, is that schema refinement alone cannot predict learning differences between different relational structures, whereas schema elaboration can.

2. Experiment 1

The first experiment investigated people's ability to learn arbitrary new relational categories. The study used a standard category-learning paradigm, with an A/not-A design in which the subject was presented with a series of stimuli and was asked to decide whether each stimulus did or did not belong to the category. The category to be learned was manipulated between subjects. The category structures all contained the same number and types of first-order relations, but differed in how those relations were connected (i.e., in their relational structure). We aimed to test models of relational concept acquisition by assessing how this manipulation of relational structure affects learning.

2.1. Method

2.1.1. Participants

Ninety-four undergraduates participated for course credit in an introductory psychology course. Subjects were randomly assigned to three conditions (*N*s = 34, 33, 27 for Conditions 1-3, respectively), which differed in the type of category structure to be learned.

2.1.2. Materials and Design

Each stimulus comprised four objects (filled circles), varying along two perceptually separable dimensions: brightness and size (e.g., Smith & Kemler, 1978). Each dimension had four levels, assigned without replacement to the four circles in each stimulus. The brightness values of the circles were 15, 45, 120, and 250, as defined on a 0-255 gray scale on a standard LCD monitor. The sizes of the circles (in diameter) were .85, 1.36, 2.30, and 3.18 cm. All adjacent pairs of size and brightness values were easily discriminable.

Each dimension defines a comparative binary relation among the objects (i.e., BRIGHTER and LARGER). The category to be learned by each subject was defined by two such relations, one on each dimension. A stimulus was a member of the category if and only if it satisfied both of these relations. Figure 2A illustrates an example category structure, in which the upper-right object must be larger than the upper-left object, and the lower-right object must be brighter than the lower-left object. Figure 2B shows an example stimulus that is a member of this category.

The category structures varied in how the relations were connected to each other, in terms of the objects on which they were defined, as shown schematically in Figure 3A. Specifically, the two relations could operate on disjoint objects (Condition 1), on one shared object with one unique object for each relation (Condition 2), or on the same pair of objects (Condition 3). It is easy to see that, topologically, these are the only three possibilities. That is, these are the only three relational structures that can be formed by two binary relations operating on four objects (ignoring the directionality of the relations).

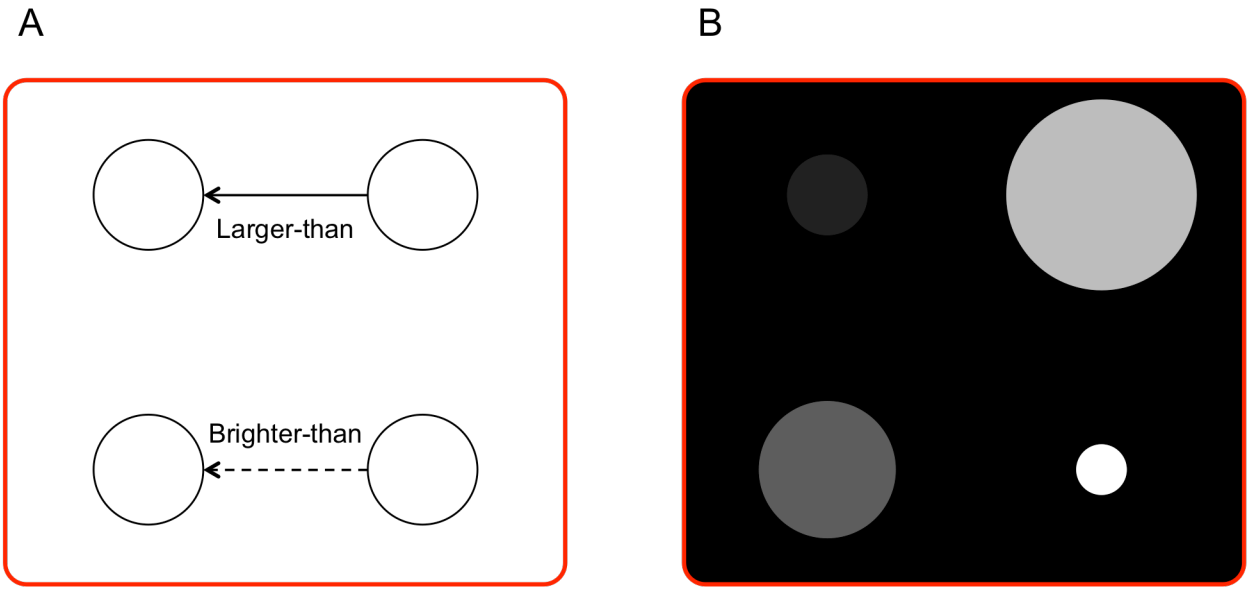


Figure 2. A: Example category structure from the main task of Experiment 1 (Condition 1). The arrows indicate that for a stimulus to be in the category, the upper-right object must be larger than the upper-left object, and the lower-right object must be brighter than the lower-left object. This diagram is for illustrative purposes only; subjects were presented only with actual stimuli, as in Figure 2B. B: Example stimulus that would be a member of the category in Figure 2A.

Figure 3B shows different spatial instantiations of each condition. Although these instantiations differ in the spatial arrangement of relations, the relational structure in each condition is preserved. For example, both instantiations of Condition 1 consist of two relations operating on disjoint pairs of objects. The subconditions shown in Figure 3B constitute all possible spatial instantiations of each condition, up to rotation and reflection. Subconditions were counterbalanced across subjects within each condition. To determine the specific category structure for each subject, that subject's subcondition was mapped to the physical arrangement of the objects by random rotation and reflection (see Figure 3C). The two relations were then randomly assigned to the two dimensions, and a binary direction was randomly selected for each relation (i.e., which of the two objects was brighter or larger).

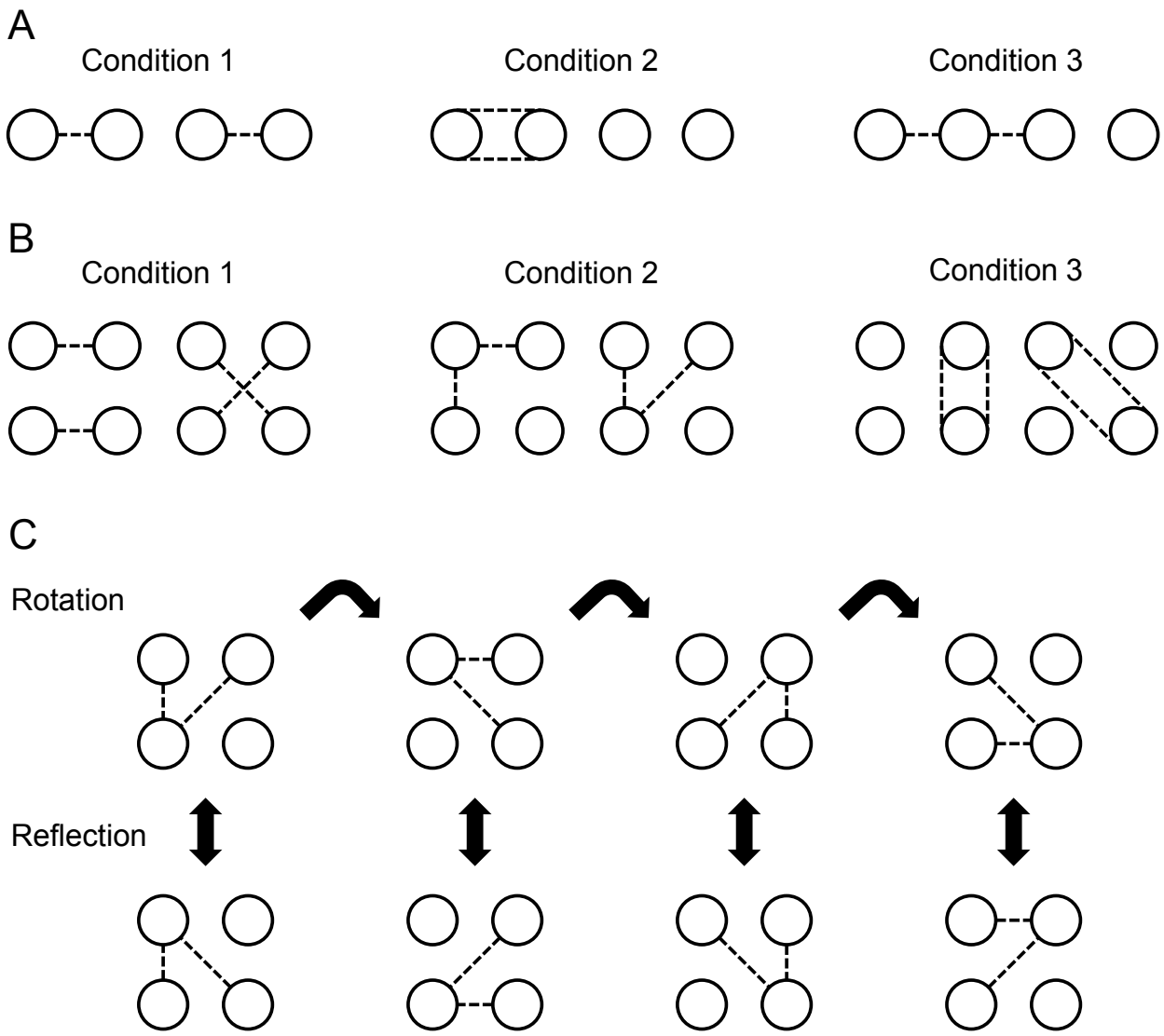


Figure 3. Design of category structures in Experiment 1. A: Abstract relational structure of each condition, showing how the relations are connected by operating on common objects. Circles indicate objects and lines indicate relations (one being BRIGHTER and the other LARGER). These conditions constitute all topologically unique ways of defining two binary relations over four objects. B: Subconditions for each condition. These subconditions constitute all possible spatial arrangements of each condition, up to rotation and reflection. C: The possible physical instantiations of the second subcondition of Condition 2.

2.1.3. Procedure

To familiarize subjects with both of the relations, they were given two training tasks prior to the main task, one for the BRIGHTER relation and one for the LARGER relation. The training tasks were the same as the main task, except that each stimulus contained only two objects (see Figure 4) instead of four, and each category was defined by only one relation (e.g., the right object must be darker than the left object). Each training task ended once the subject gave eight consecutive correct responses. This criterion was designed to ensure that the subject learned or attended to the relation defining that task. The expectation was that this training would facilitate learning of the main task, by encouraging subjects to look for pairwise relations of size and brightness between the objects. The order of training tasks was randomized for each subject.

All three tasks (i.e., training and main tasks) followed the same procedure. On each trial, a stimulus was sampled randomly, subject to equal probabilities of choosing a stimulus in or out of the category. The subject responded by pressing Y or N, indicating the stimulus is or is not in the category. The correct answer was subsequently displayed in the center of the screen directly below the stimulus for 800 ms. The screen was then cleared for 400 ms before the stimulus for the following trial was presented.

Subjects were given a cover story in which the stimuli were optical key cards for a building, and their goal in each task was to learn which key cards would open a particular door. The instructions for each task indicated the category was different from previous tasks (i.e., each task was about a different door of the building) and included a random, positive example (i.e., a

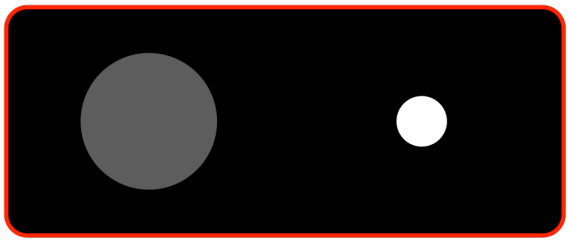


Figure 4. Example stimulus for training tasks of Experiment 1.

key card that opens the current door). On the main task, after every 20 trials the subject was given a self-paced rest break and was told his or her performance on those trials. The full experiment (i.e., training and main tasks combined) was programmed to end after 55 minutes.

2.2. Results and Discussion

To ensure subjects learned the appropriate first-order relations, the analysis excluded subjects who took over 125 trials to meet the learning criterion on either of the training tasks, leaving 77 subjects. Importantly, this subject exclusion does not bias the comparison among conditions in the main task, because it is based on events that occurred before the experimental manipulation (i.e., the training procedure was identical in all conditions).

Because the duration of the experiment was determined by time, there was significant variability in the number of trials that subjects completed in the main task, ranging from 201 to 1493 (M = 863, SD = 233). The statistical requirement to analyze the same number of trials for all subjects forces a tradeoff between including a maximal number of subjects and analyzing enough trials to allow differences among conditions to emerge. By inspecting the distribution of number of trials across subjects—aggregated across all conditions—we settled on a minimum of 600 trials. This criterion eliminates seven additional subjects who were outliers in this distribution: two each in Conditions 1 and 2, and three in Condition 3. Thus we were left with 70 subjects: 27 in Condition 1, 23 in Condition 2, and 20 in Condition 3. More stringent criteria (i.e., more trials and fewer subjects) led to the same pattern of significant results as reported next.

Figure 5 displays average learning curves for the three conditions, in blocks of 50 trials. The plot shows moderate learning in all three conditions, with performance best in Condition 3 and worst in Condition 1. This difference is marginally significant when tested over all trials, M_1 = .574, M_2 = .593, M_3 = .635, MSE = .008, $F(2,67) = 2.61$, p = .081. The effect becomes more reliable if we restrict to later trials, when differences among conditions have had time to emerge. Restricting to the second half (trials 301-600) yields M_1 = .578, M_2 = .616, M_3 = .663, MSE = .013, $F(2,67) = 3.27$, p = .044.

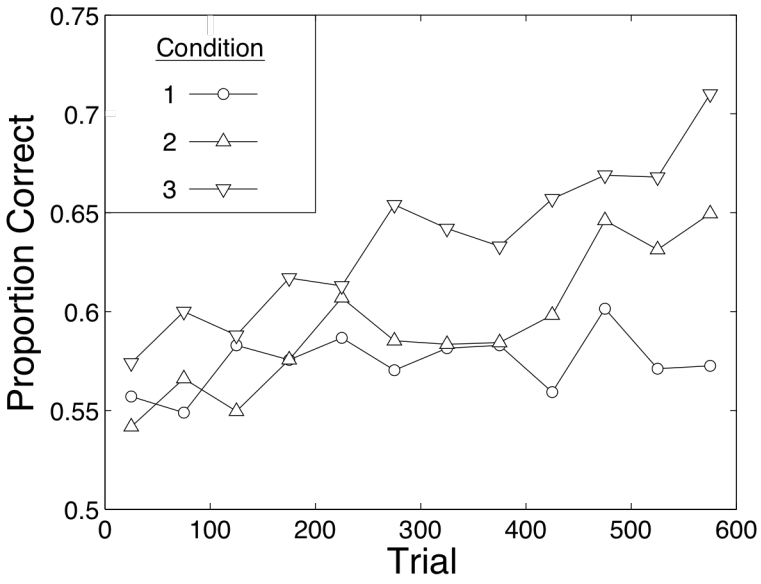


Figure 5. Learning curves for Experiment 1.

Because of model predictions reported in Section 4, we also decomposed each subject's errors into misses and false alarms. On average, 56.6% of subjects' errors were false alarms, with no significant difference in this proportion across conditions ($p > .1$).

In conclusion, subjects were able to learn this task, but only imperfectly. Even after 600 trials, performance was 57%-71% for the three types of category structure. In addition, performance differed reliably across conditions, especially in later trials. Best performance was observed in Condition 3, in which both category-defining relations operated on the same pair of objects. Worst performance was observed on Condition 1, in which the two relations operated on disjoint pairs of objects. Condition 2, in which the relations shared one common object, exhibited intermediate performance. Thus learning of the categories was sensitive to relational structure. We turn next to implications of these findings for schema-based models of relational learning.

3. Models of Schema Learning

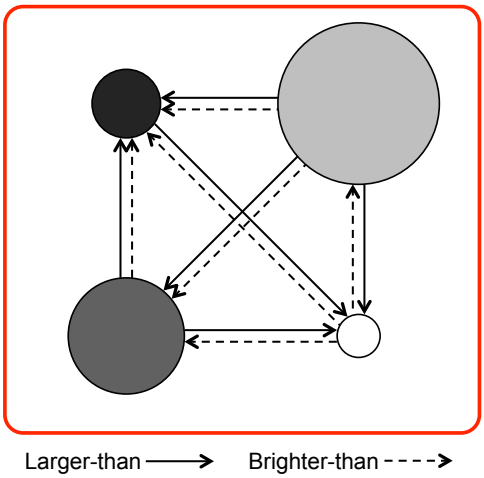
The models considered here operate by maintaining a schema from trial to trial that contains some set of relations among the objects within the stimuli. The schema represents the learner's current hypothesis about the category structure. The model classifies the stimulus on each trial as in the category if it satisfies all relations currently in the schema. Thus, the models implicitly assume a conjunctive category rule, in which all relations must be satisfied for category membership (as was true for the categories in this study). This assumption is in line with prior research showing that, in contrast to feature-based categories, people learn relational categories defined by conjunctions much better than relational categories defined by family resemblance (Jung & Hummel, 2009; Kittur et al., 2004).

Figure 6 illustrates the operation of the models as applied to Experiment 1. Each stimulus embodies 12 relations, one for each dimension on each of the six possible object pairs, as shown by the example in Figure 6A. The schema on any trial can contain any subset of these relations (each in either direction). We assume the schema is initialized as a complete representation (i.e., all 12 binary relations) of the example stimulus provided in the instructions for the main task. The schema then evolves over time by comparison to the stimulus presented on each trial.[2] We consider three mechanisms for schema change: refinement, forgetting, and elaboration. After defining these mechanisms, we combine them into a hierarchy of models that allow assessment of each mechanism's contribution to relational category learning.

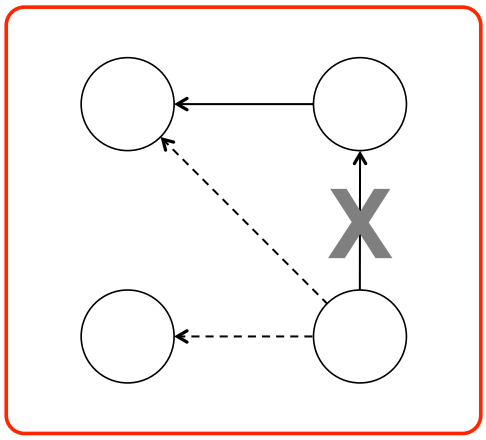
---

[2] The mapping between schema and stimulus is assumed to preserve spatial location (i.e., the upper-left object is always mapped to the upper-left object, etc.). This assumption is necessary because, if arbitrary mappings were allowed, then the task would be unlearnable. The assumption can be accommodated within the structure-mapping framework by assuming all objects have highly salient features (or unary relations) indicating their spatial positions.

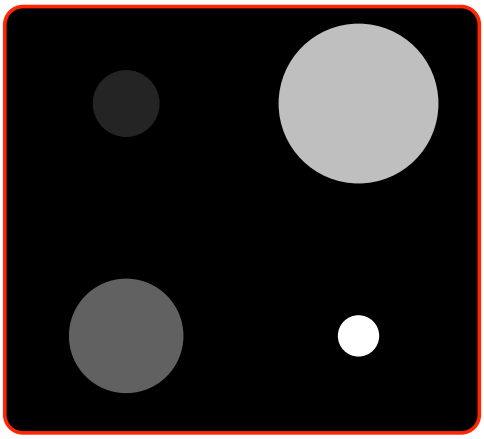
A



B

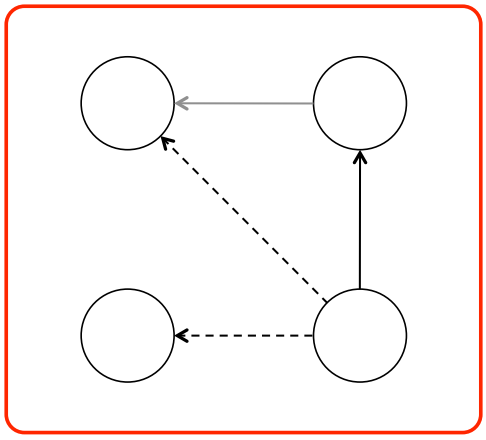


Response: “No”

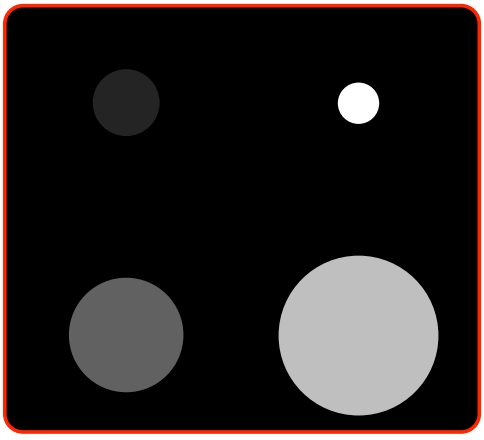


Wrong. This DOES open the door.

C



Response: “Yes”



Wrong. This does NOT open the door.

Figure 6. Illustration of how the models operate in Experiment 1. A: The 12 binary relations present in a stimulus. B: Example of schema refinement, following a miss. The model begins the trial with the schema on the left (including all four relations shown). It classifies the stimulus on the right as not in the category, because the stimulus violates the relation in the schema that the lower-right object is larger than the upper-right object. After feedback is given, the offending relation is removed from the schema (indicated by the X). The new schema is the intersection of the old schema and the relational representation of the stimulus (as shown in Figure 6A). C: Example of schema elaboration, following a false alarm. The model begins the trial with the schema on the left, including only the three relations shown in black. It classifies the stimulus on the right as in the category, because the stimulus satisfies all relations in the schema. After feedback is given, the schema is augmented with a relation the stimulus violates. In this case, the model chooses the relation shown in grey.

3.1. Refinement

Schema refinement involves removing relations that are absent in a newly encountered category member, as illustrated in Figure 6B. As noted in Section 1.2, refinement can occur only following a miss, that is, when feedback indicates the stimulus is in the category but it was classified as a nonmember because it violates some relation(s) in the current schema. When the model commits a miss, any relations the stimulus violates logically cannot be part of the category rule. Therefore these relations are discarded from the schema, with all other relations in the schema retained. Thus the new schema is the intersection of the old schema and the stimulus.

3.2. Forgetting

Forgetting of information from the schema can occur between trials. For instance, even though we assume the schema is initialized as a complete representation of the example category member given during the instructions, much of that information might be forgotten before the first learning trial. We model forgetting by assuming that prior to each trial, every relation in the schema has an independent probability *p* of being forgotten, which depends on the total number of relations currently in the schema (*r*):

$$p = 1 - \frac{L}{r}\left(1 - e^{-r/L}\right), \tag{1}$$

where *L* can be interpreted as a memory-capacity parameter. This formulation has the property that the expected number of retained relations equals *L* * (1 – exp(-*r*/*L*)). This function uniquely derives from two simple assumptions: (1) as the number of elements to be remembered approaches zero, the probability of retaining each element approaches unity, and (2) the expected number of elements that will be retained after one step converges exponentially (as a function of the initial schema size) on some limiting capacity *L*.

Although the form of forgetting in Equation 1 is fairly specific, it embodies a simple assumption that information is forgotten probabilistically, at a rate that is positively dependent on the current memory load (i.e., more forgetting when there is more to remember). This assumption has an established history in traditional models of memory (e.g., Shiffrin & Cook, 1978). Moreover, we do not take the form in Equation 1 as a strong theoretical commitment. Any form of forgetting or interference, by which relations can be lost from the schema, would suffice for the present purpose. We have taken two steps to ensure that the details of Equation 1 do not affect the pattern of model predictions. First, we verified that the qualitative patterns of predicted performance across conditions are effectively unchanged across different values of the *L* parameter. Thus, the forgetting parameter serves only to calibrate overall performance; the model's critical predictions concerning relative learnability of different relational structures

are essentially invariant with respect to this parameter. Second, we simulated a new family of models in which the forgetting probability was held constant (i.e., independent of schema size), using a new free parameter for the constant forgetting rate that replaced the *L* parameter. The qualitative predictions of these constant-forgetting models are the same as those of the capacity-dependent models reported here. Thus the models' essential predictions do not depend on the particular form of forgetting assumed.

3.3. Elaboration

Schema elaboration involves adding one or more relations to make the schema more constrained. As noted in Section 1.3, the need for elaboration is indicated by a false alarm, that is, when a stimulus is classified as a category member because it satisfies all relations in the current schema, but feedback indicates it is not in the category. This situation is illustrated in Figure 6C. Following feedback for a false alarm, the schema can be elaborated by adding one or more relations that the stimulus violates. As with schema refinement following a miss, this learning mechanism ensures the correct answer would be given if the same stimulus were immediately presented again.

There is thus a symmetry between schema refinement and elaboration in a binary classification task, each interpretable as a learning mechanism for correcting the two types of errors that can occur (misses and false alarms, respectively). An important difference is that, after receiving feedback on a false alarm, the subject cannot logically infer how the schema needs to be changed. That is, it is ambiguous which relation(s) in the stimulus constitute a violation of the category and hence which relational constraints must be added to the schema. Thus schema elaboration is a potentially more flexible mechanism than is schema refinement. This flexibility enables sensitivity to relational structure that is not predicted by schema refinement alone.

To implement schema elaboration, we assume that following a false alarm the model identifies all relations that the stimulus violates but that are absent from the schema, and treats

each as a candidate to be added. For simplicity, we assume exactly one relation is added to the schema following any false alarm. This relation is chosen probabilistically among the candidates, by assigning an *elaboration score*, *s*, to each candidate, and applying a standard softmax or Luce choice rule:

$$p(i) = \frac{e^{\phi s_i}}{\sum_j e^{\phi s_j}} \quad (2)$$

Here $p(i)$ is the probability of selecting candidate i, and ϕ is a free parameter governing the degree of determinism in the selection process (see Mitchell & Hofstadter, 1990, for a more sophisticated implementation of temperature-dependent stochastic representation construction).

Flexibility in the elaboration process lies in how elaboration scores are assigned to candidate relations. Different assumptions lead to different predictions about the relative learnability of the category structures in the current experiments. We consider five possibilities, described next (see Table 1 for summary). The first uses uniform scores (i.e., random selection), whereas the other four are sensitive to relational structure, in ways motivated by different theoretical perspectives in the literature. These were not intended as an exhaustive set of possibilities, but rather as an exploratory set of reasonable options, to test whether structure-sensitive schema elaboration could in principle explain differences in learning as a function of relational structure.

3.3.1. Random Elaboration

The simplest version of schema elaboration chooses a candidate relation at random for addition to the schema. This assumption can be implemented in the framework of Equation 2 by assuming uniform elaboration scores, for example $s = 1$ for all candidates (note that ϕ is inconsequential for this model).

Table 1: Elaboration scores under different models of schema elaboration

Model	Elaboration Score
Random Elaboration	$s = 1$
Object Centrality	$s = R_1 + R_2$
Economy of Objects	$s = \sum_k (R_k > 0)$
Plurality of Objects	$s = \sum_k (R_k = 0)$
Relational Chaining	$s = \frac{1}{R_1 + 2} + \frac{\left(\frac{1}{R_1 + 1} - \frac{1}{R_1 + 2}\right)}{R_2 - R_1 + 1}$ or $s = 0$ if $R_1 = R_2 = 0$

Note: The given formulas determine the elaboration score (s) for each candidate relation, for use in determining which candidate to add to the schema following a false alarm (see Equation 2). R_1 and R_2 are the numbers of relations already in the schema that operate on the candidate relation's two objects, with $R_1 \leq R_2$.

3.3.2. Object Centrality

An alternative possibility is that schema elaboration is guided by a preference for more coherent relational concepts. In their classic work on conceptual coherence, Murphy and Medin (1985) proposed that coherence is determined by people's lay theories about the world, which impart relational constraints within and between concepts. Related research has shown that features central to the relational network of a concept more strongly influence the concept's conceptual coherence (Sloman, Love, & Ahn, 1998). Taken together, these ideas suggest a relational category will be more psychologically coherent if it contains a central object participating in multiple relations. Thus, performance in Experiment 1 should be best for Conditions 2 and 3. This principle is formalized by defining the elaboration score for each candidate relation as the sum, over the two objects that relation operates on, of how many

relations already in the schema each object participates in. For example, if the current schema contains 2 relations including object A and 1 relation including object B, then a new relation linking A and B would have a score of 3. This assumption leads the model to prefer relations built on objects that are already more central, thus favoring categories with a centralized structure.

3.3.3. Economy of Objects

Due to the processing constraints of working memory (Baddeley, 2003), it may be easier to maintain a schema involving fewer objects or to discover an analogy that requires mapping fewer objects between scenarios. Therefore, when elaborating a schema, people may be inclined to select relations that minimize the total number of objects involved. This hypothesis predicts that learning will be superior for category structures involving a smaller number of objects. Thus performance in Experiment 1 should be best for Condition 3 and worst for Condition 1. This principle is formalized by defining the elaboration score for each candidate relation as the number of its objects (0, 1, or 2) that participate in other relations already in the schema. This assumption leads the model to avoid adding relations that require tracking additional objects.

3.3.4. Plurality of Objects

Cognitive load theory (van Merriënboer & Sweller, 2005) suggests that objects that do not participate in any of the category's relations will act as extraneous distractors. Subjects' attention may be drawn to such objects, making them more likely to add relations on new objects when elaborating a schema. Similarly, because people often struggle to recognize surface features as irrelevant information (Cooper & Sweller, 1987), irrelevant objects may place additional strain on working memory, while obscuring the category's underlying relational structure. Consequently, category structures that contain the greatest number of irrelevant objects, such as Condition 3, would be most difficult to learn. This principle is formalized by defining the elaboration score for each candidate relation as the number of its objects (0, 1, or

2) that do not participate in any relations currently in the schema. This scoring rule is exactly opposite the Economy of Objects rule, and the two are equivalent under a substitution $\phi \rightarrow -\phi$.

3.3.5. Relational Chaining

Lastly, learning may be better for category structures composed of relations that are chained together (e.g., Condition 2), because people may be inclined to link known relational structure to new objects. Such a preference might arise from a causal learning perspective, in which subjects seek to discover causal chains among the objects. For example, upon learning that object A must be bigger than object B, people may be inclined to test whether object B must be brighter than object C. Thus, structures composed of relations that are chained together may be easier to acquire.

This principle can be formalized in many ways, but we chose the following approach. For a candidate relation between objects A and B, let R_1 be the number of relations already in the schema that operate on A, and let R_2 be the number of relations already in the schema that operate on B. Assume A is the object with fewer relations, so that $R_1 \leq R_2$. In other words, the candidate relation operates on two objects, one of which already participates in R_1 relations in the schema, and the other of which participates in R_2 relations. The candidate's elaboration score is given by the equation in Table 1. Although the equation appears complicated, it simply implements a lexicographic preference for small values of R_1 followed by small values of $R_2 - R_1$. Table 2 shows elaboration scores for different values of R_1 and R_2. As a special case, the elaboration score when $R_1 = R_2 = 0$ was set to zero, to implement a preference not to add isolated relations. Thus, the ideal candidate is one that extends an existing chain: $R_1 = 0$, $R_2 = 1$.

3.4. Models tested

The mechanisms of schema refinement, forgetting, and schema elaboration as described in Sections 3.1-3.3 were combined into seven different models, which collectively enable assessment of each mechanism's contribution to relational category learning. The models are

Table 2: Elaboration scores for Relational Chaining model

	R_2				
R_1	0	1	2	3	4
0	0	.75	.667	.625	.6
1		.5	.417	.389	.375
2			.333	.292	.278
3				.25	.225
4					.2

Note: Shown is the elaboration score (*s*) under the Relational Chaining model for any candidate relation, as used in Equation 2. These values are determined by the equation in Table 1. R_1 and R_2 are defined as in Table 1. Elaboration scores are shown only for values of R_1 and R_2 up to 4, although larger values of these variables are possible.

summarized in Table 3. The models are organized in a semi-nested hierarchy and are explained from simplest to most complex.

The *pure refinement (PR)* model assumes schema refinement is the only mechanism of schema updating. Under the PR model, this learning process will continue until all incorrect relational constraints are removed, at which point the schema will necessarily coincide with the true category rule.

The *refinement-with-forgetting (RF)* model extends the PR model by assuming relations can be forgotten from the schema between trials. This forgetting can be interpreted as due to soft capacity limitations of working memory, under the assumption that the schema is actively maintained and rehearsed or refreshed from trial to trial, or it can be interpreted as due to encoding and retrieval failures to and from long-term memory. Refinement ensures all incorrect relations will eventually be removed (as in the PR model), but forgetting can lead to correct relations being removed as well.

Table 3. Assumptions of tested models, and their strengths (+) and weaknesses (–) across both experiments.

	Model						
Assumption	PR	RF	RE	OC	EO	PO	RC
Schema refinement	Y	Y	Y	Y	Y	Y	Y
Forgetting	N	Y	Y	Y	Y	Y	Y
Schema elaboration	N	N	Y	Y	Y	Y	Y
Structure-sensitive elaboration	N	N	N	Y	Y	Y	Y
Free parameters	None	*L*	*L*	*L*, ϕ	*L*, ϕ	*L*, ϕ	*L*, ϕ
Match to data							
Performance in the range of subjects'	–	–	+	+	+	+	+
Predicts differences among conditions	–	–	–	+	+	+	+
Predicted differences agree with data							
Experiment 1	–	–	–	+	+	–	–
Experiment 2	–	–	–	–	–	–	–

Note: Rows 1-4 indicate which models do (Y) and do not (N) incorporate each assumption. Parameters *L* and ϕ respectively determine rate of forgetting and strength of preferential schema elaboration. Each row in the lower half of the table indicates which models have (+) or lack (–) the indicated property. For example, the "Performance in the range of subjects'" row indicates which models are able to produce the same overall level of performance as was observed from the experimental subjects, and which models make dramatically wrong predictions (much too high for PR and too low for RF). PR = pure refinement; RF = refinement with forgetting; RE = random elaboration; OC = object centrality; EO = economy of objects; PO = plurality of objects; RC = relational chaining.

The *random elaboration (RE)* model extends the RF model by introducing schema elaboration, with random selection among all candidate relations to be added following a false alarm. The PR model cannot produce a false alarm, because its schema always contains all true relational constraints, and hence adding elaboration to that model would not change it. However, with forgetting included, false alarms are possible and there is thus opportunity for elaboration. This is why the models tested progress from refinement only (PR) to refinement

with forgetting (RE) to refinement, forgetting, and elaboration (RE). The interplay among these three mechanisms in the RE model leads to its continually adding and removing constraints in an attempt to converge on the true category structure.

The final four models are *object centrality (OC)*, *economy of objects (EO)*, *plurality of objects (PO)*, and *relational chaining (RC)*. These models extend the RE model by assuming schema elaboration is sensitive to relational structure, according to the four schemes described in Sections 3.3.2-3.3.5.

All models operate on the main task of Experiment 1 as follows. At the beginning of the task, the model is presented with a random category member, mirroring the positive example given to subjects in the instructions, and the schema is initialized to include all 12 relations present in this stimulus. The model then proceeds through a series of learning trials mirroring those given to subjects. Before each trial, relations are randomly forgotten from the schema according to Equation 1, except in the PR model (which does not assume forgetting). During each trial, a stimulus is presented, and the model provides a categorization response according to whether the stimulus satisfies all relations currently in the schema. Then feedback is given. If the trial is a miss, then the schema is updated by refinement (i.e., intersection with the stimulus). If the trial is a false alarm, then except in the PR and RF models, the schema is updated by elaboration (i.e., adding a relation the stimulus violates, chosen by Equation 2). This cycle is repeated throughout the course of the task.

Several predictions can readily be made regarding the models' performance in this paradigm. The first set of predictions concerns overall learning performance across all experimental conditions. The PR model is an ideal observer for this task, and hence performance was expected to be high for all conditions. In the RF model, once a true relational constraint is forgotten, it has no way of being reincorporated into the schema; hence all relations will eventually be lost and the model should asymptote at chance performance. In the other five models, elaboration and forgetting can combine to produce intermediate levels of performance

(depending on the capacity parameter *L*, with lower levels leading to more forgetting and lower performance). Although these models can forget true category-defining relations (as in the RF model), elaboration enables those relations to be reincorporated into the schema.

A second set of predictions concerns differences in learning performance across conditions. In the PR, RF, and RE models, all relations are treated independently. Schema refinement considers each relation individually, according to whether it is present in a given positive stimulus. Forgetting probabilities are also independent across relations (notwithstanding a global dependence on schema size). Therefore the manner in which relations are linked through operating on shared objects should have no effect on learning, and these models should exhibit equal performance in all conditions. In contrast, the models with structure-sensitive elaboration (OC, EO, PO, and RC) should predict condition differences. For example, the EO model should exhibit better performance in Condition 3 (in which both relations operate on the same pair of objects), because its elaboration mechanism tends to consider relational structures that are defined on fewer objects. In general, the performance of the structure-sensitive models should be determined by how well the category structure agrees with the preferences implicit in the models' elaboration scores.

4. Simulation of Experiment 1

A simulation study was conducted to evaluate the models' predictions on the main task of Experiment 1. Each model was simulated 10000 times in each condition, for 600 trials per simulated subject. To compare model predictions to subject performance, mean proportion correct was computed for trials 551-600 for each simulated subject. These trials represent the final block of the subject data, where differences across conditions were strongest (see Figure 5). Because we were primarily interested in the models' qualitative predictions, parameters were fit by hand. The *L* parameter (which is relevant to all but the PR model) was set to 12, a value that leads all five elaboration models to approximately match subjects' average performance.

The same value of *L* was used for the RF model (see following paragraph for further discussion). The ϕ parameter for the structure-sensitive elaboration models was set to 10, which is a large value that accentuates predicted differences among conditions. Different parameter choices merely change the models' overall performance (greater for larger values of *L*) or the magnitude of differences among conditions for the structure-sensitive models (weaker for smaller values of ϕ, under the constraint $\phi > 0$); the qualitative patterns described next remain unchanged.

Figure 7 shows model performance by condition, compared to the subject data. Table 3 (bottom half) summarizes the critical comparisons between model and subject performance. As predicted, the PR model far outperforms the subjects, in fact producing perfect performance in the range of trials analyzed. The model makes only a few errors early in learning, and all of these are misses, in contrast to the subjects, who made a large proportion of false alarms. These results imply pure schema refinement is not representative of how people acquire relational categories. Adding forgetting lowers performance, but the RF model eventually performs at chance, because it forgets all relations and has no way of reincorporating true relational constraints into its schema. The RF model can match subjects' mean performance on any given range of trials if *L* is set large enough that relations are forgotten very slowly (e.g., *L* = 265 matches the subjects' grand mean of 64.4% performance over trials 551-600), but this value of *L* is arguably implausible given its intended psychological interpretation. More importantly, the RF model under this parameterization produces a nonmonotonic learning curve, reaching 90% correct on the second block (trials 51-100) and slowly declining toward 50% thereafter, in clear disagreement with the increasing learning curves produced by subjects (Figure 5).

The RE model can match subjects' average performance level (with the right choice of *L*), demonstrating that refinement, forgetting, and elaboration can together produce intermediate

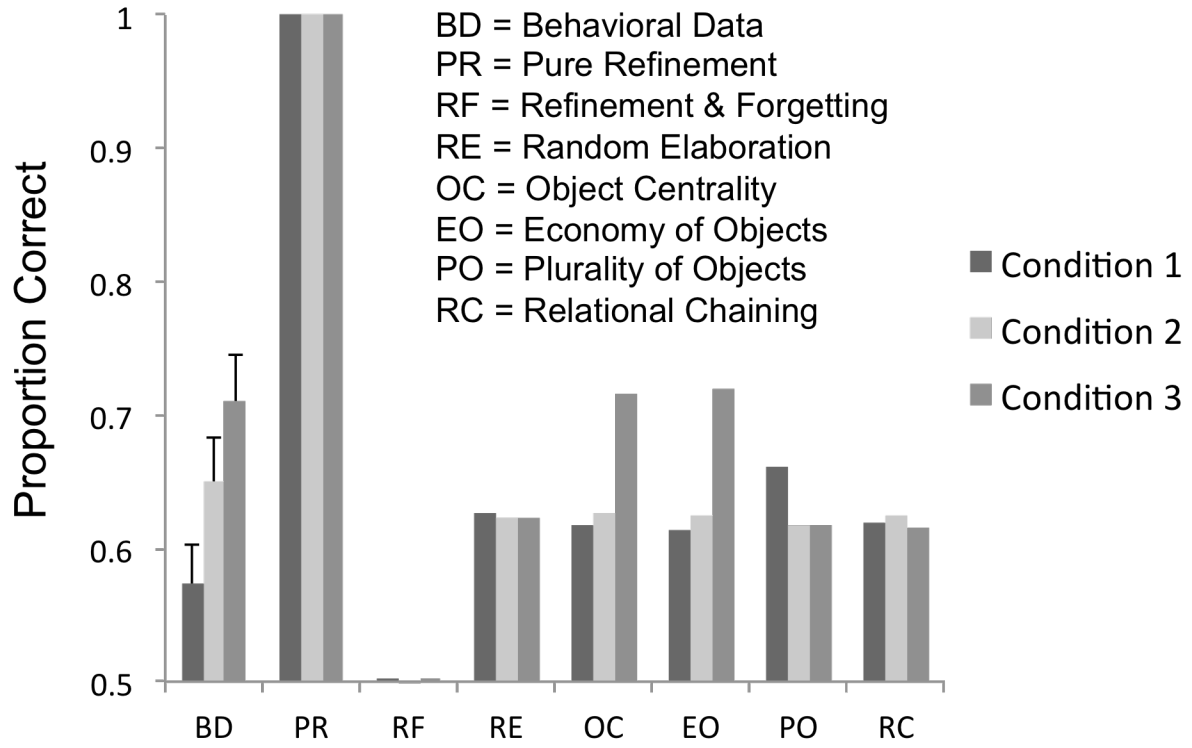


Figure 7. Mean performance by condition in Experiment 1 for subjects and models on final block (trials 551-600). The RF model performs at chance. Error bars for behavioral data indicate standard error of the mean.

performance. However, the RE, RF, and PR models all fail to predict learning differences among conditions. Because subjects' learning differed reliably across conditions, this suggests a successful model of relational learning must be sensitive to relational structure. As expected, the models that include structure-sensitive elaboration do predict condition differences. In particular, the predictions from the OC and EO models qualitatively match the pattern in the behavioral data, with performance being best for Condition 3, followed by Conditions 2 and 1.

In summary, the simulation results support the need for both forgetting and schema elaboration, in addition to schema refinement, to explain relational category learning. Moreover, they show how structure-sensitive elaboration can explain learning differences across different

relational structures. The pattern of differences among conditions is in line with the predictions of the OC and EO models, which both seek compact or centralized relational structures.

One shortcoming of Experiment 1 is that the stimuli and category structures are not sufficiently complex for the OC and EO models to be differentiated. Moreover, patterns of learning may be different with more complex relational concepts. For instance, people might initially hold a preference for fewer objects, to establish a conceptual base to further explore the problem space, but upon establishing that conceptual base, their preferences for seeking out additional relational constraints may change. Therefore, a second experiment was conducted to investigate relational learning in conditions of greater task complexity.

5. Experiment 2

Experiment 2 was identical in design to Experiment 1, differing only in the complexity of the stimuli and category structures. The four objects in each stimulus varied along three dimensions instead of two, thus defining three binary comparative relations (see Figure 8). The category learned by each subject was determined by three instances of these relations, one on each dimension. The manner in which these relations were connected to each other was varied across conditions, allowing for six topologically unique relational structures, shown schematically in Figure 9A. As in Experiment 1, these conditions are exhaustive and are the only logical possibilities for three binary relations shared among four objects. As with Experiment 1, the primary questions were how well subjects could learn these categories and how learning would depend on relational structure (i.e., on experimental condition). We also aimed to test whether the data would support models of schema elaboration, and in particular whether either the OC or EO model would provide a good account of learning differences across conditions, as they were found to do in Experiment 1.

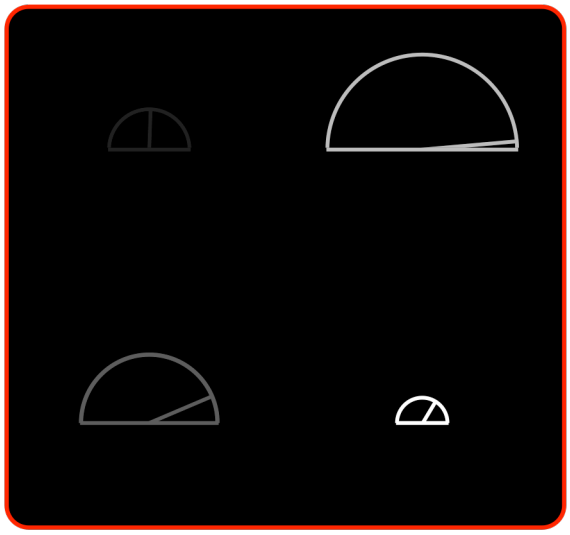


Figure 8. Example stimulus from main task of Experiment 2. The four objects differ in brightness, size, and tilt of the radius.

5.1. Method

5.1.1. Participants

One hundred thirty-seven undergraduates participated for course credit in an introductory psychology course or were paid to participate. Subjects were randomly assigned to six conditions (*N*s = 21, 23, 21, 28, 21, and 23 for Conditions 1-6, respectively), which differed in the type of category structure to be learned.

5.1.2. Materials and Design

Each stimulus contained four objects, each a semicircle with a radius (a variant of what are known in the perceptual categorization literature as *Shepard circles*; Shepard, 1964). The objects varied along three perceptually separable dimensions: brightness, size, and tilt of the radius. Each dimension had four levels, assigned without replacement to the four objects in each stimulus. The size and brightness levels were identical to those used in Experiment 1. The radius tilts were 5°, 23°, 59°, and 88°, counterclockwise from horizontal. Adjacent pairs of values on all three dimensions were easily discriminable.

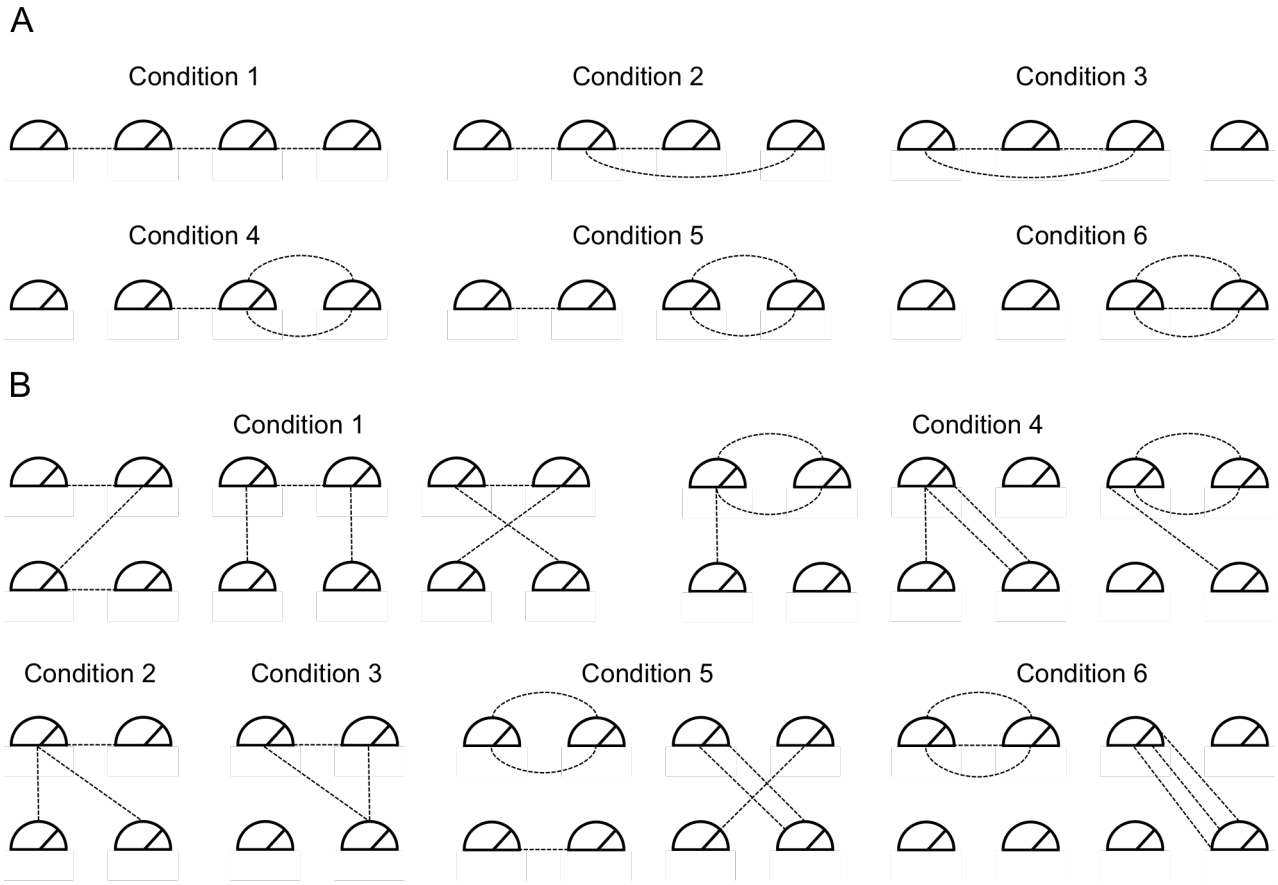


Figure 9. Design of category structures in Experiment 2. A: Relational structure of each condition, showing how the relations are connected by operating on common objects. These conditions constitute all topologically unique ways of defining three binary relations over four objects. B: Subconditions for each condition. These subconditions constitute all possible spatial arrangements of each condition, up to rotation and reflection.

The category to be learned by each subject was defined by three binary relations, one on each of the dimensions (i.e., BRIGHTER, LARGER, and STEEPER). A stimulus was a member of the category if and only if it satisfied all three of these relations (e.g., upper-left object must be brighter than upper-right, upper-right must be larger than lower-right, and lower-right must be steeper than lower-left). The experimental conditions differed in the relational structure of the category, that is, in how the three category-defining relations were connected (see Figure 9A).

Each condition was divided into subconditions as shown in Figure 9B, which were counterbalanced across subjects. The subconditions for each condition constitute all of its possible spatial instantiations, up to rotation and reflection. Once a subject's subcondition was determined, the specific category structure was defined by choosing a random rotation and reflection of that subcondition, randomly assigning the three relations to the three dimensions, and choosing a random direction for each relation.

5.1.3. Procedure

The procedure was identical to that of Experiment 1 with one exception. Participants were required to complete three training tasks instead of two, one for each first-order relation (i.e., BRIGHTER, LARGER, and STEEPER), in randomized order.

5.2. Behavioral Results

As in Experiment 1, subjects were excluded from analysis if they took more than 125 trials to complete any of the training tasks. This criterion left 64 subjects out of the original 137. This degree of exclusion is not ideal, but subjects taking longer to learn a training task could not be expected to be aware of the corresponding relation when doing the main task. The number of subjects failing this criterion was roughly uniform across training tasks: 37 for brightness, 35 for size, and 41 for tilt (with overlap among these groups). The relatively low performance on these training tasks is not surprising given the number of hypotheses subjects had to consider and is consistent with previous experiments using unidimensional category rules in high-dimensional stimulus sets (e.g., Pashler & Mozer, 2013).

Because the task was more difficult than in Experiment 1, and because there were three training tasks and the entire experiment was still limited to 55 min, subjects tended to complete fewer trials on the main task than did subjects in Experiment 1. Of the 64 subjects taken to have learned the training tasks, the number of trials completed in the main task ranged from 37 to 1253 (M = 623, SD = 201). By inspecting this distribution, we chose a minimum of 400 trials, as a tradeoff between retaining a maximal number of subjects and analyzing a maximal number of

trials per subject. As in Experiment 1, this cutoff was determined only from the distribution of number of trials completed, aggregated over all conditions. This requirement eliminates 6 additional subjects who were outliers in this distribution: one each in Conditions 1 and 2 and two each in Conditions 4 and 5.

Figure 10 displays average learning curves by condition for the remaining 58 subjects, in blocks of 50 trials. An ANOVA comparing proportion correct on all trials reveals a significant effect of condition, $F(5, 52) = 3.41$, $p < .01$, $MSE = .0049$. An ANOVA restricted to the second half (trials 201-400) also shows a significant effect, $F(5, 52) = 2.52$, $p = .04$, $MSE = .011$. Learning performance was best in Condition 1, where all three relations were connected in a chain, followed by Condition 4 (two relations operating on the same pair of objects and the third forming a chain with them) and Condition 6 (all three relations on the same pair of objects).

We interpret the learning differences across conditions as being due to the different relational structures formed by each category. However, another possible reason for the learning differences is the spatial arrangement of the relations. For example, the relatively high average performance in Condition 1 might be due to a particular advantage of one of its sub-

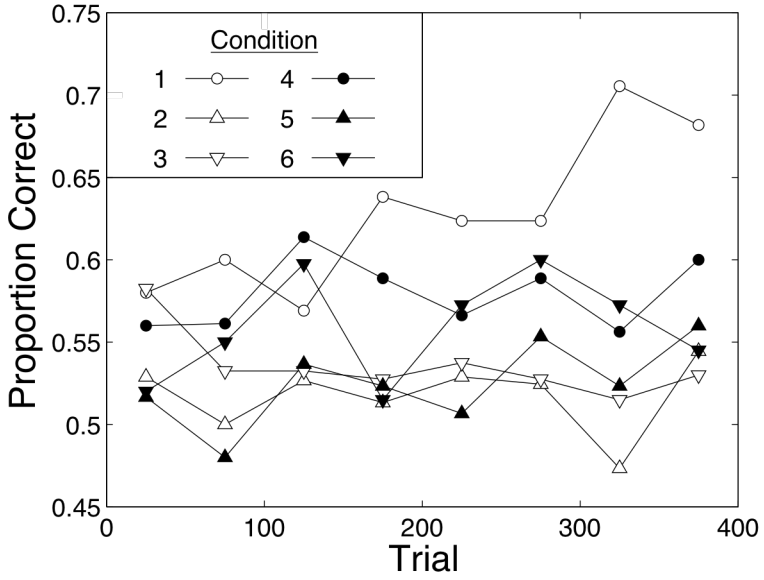


Figure 10. Learning curves for Experiment 2.

conditions. If spatial arrangement were a factor, we would expect differences among sub-conditions within each condition. To test this possibility, we ran additional ANOVAs with subcondition nested within condition, based on trials 1-400 and based on trials 201-400. The effect of subcondition was nonsignificant in both analyses (*F*s < 1). The fact that spatial arrangement appears to have had no effect on learning supports the interpretation that the observed learning differences among conditions reflect the manipulation of relational structure.

5.3. Model Simulations

The models were simulated on the main task of Experiment 2, following the same procedures as in simulating Experiment 1. All models were simulated 10000 times per condition, for 400 trials per simulated subject. Mean performance was computed for trials 201-400, for comparison to the subject data. Model parameters were set similarly as in Experiment 1. The *L* parameter was set to 9, which leads the five elaboration models to approximately match subjects' overall average performance. The ϕ parameter for the structure-sensitive elaboration models was set to 10, to accentuate predicted differences among the conditions.

Figure 11A shows the model predictions. The predictions from the PR, RF, and RE models are the same as in Experiment 1: PR predicts perfect performance in the range of trials analyzed, RF predicts chance performance, and RE predicts intermediate performance at a level dependent on *L*. All three of these models fail to predict differences among conditions. The models that assume structure-sensitive elaboration predict condition differences, but unlike in Experiment 1 none of these models even qualitatively matches the empirical ordering among conditions.

Another discrepancy between the structure-sensitive elaboration models and the subject data is the models under-predict the magnitude of condition differences. Even correcting for sampling error, the variance across the six conditions in subjects' mean performance (i.e., the

variance of the 6 condition means) was .0015.[3] The greatest variance across conditions predicted by any of the models (OC) is .00025. One factor limiting the models' predicted condition differences is that in this low range of performance, each model's schema tends to contain very few relations at the start of any given trial (mean 1.03 in the present simulations). The models exhibit less structure-sensitivity when the schema is small, because elaboration scores are based on how each candidate relation connects to relations already in the schema. In the extreme when the schema is empty, elaboration must be random. Thus the predicted condition differences should be greater when *L* is increased, to allow more complex schemas. This change would also increase performance beyond that exhibited by subjects, but one possibility is that people maintain richer schemas and make errors for other reasons not captured by models (e.g., noise in mapping the schema to the stimulus). Therefore, the structure-sensitive elaboration models were simulated again, with *L* set to 40, to test whether they better match the empirical pattern of condition differences. As shown in Figure 11B, the predicted condition differences are increased, but the models all still fail to reproduce the correct ordering among conditions from the behavioral data.

5.4. Discussion

The empirical results of Experiment 2 are largely consistent with the conclusions of Experiment 1. Subjects exhibited partial learning of the task, but performance with these complex stimuli was quite low. Nevertheless, there were reliable differences across conditions, indicating that subjects were sensitive to relational structure. The model simulations again show that schema refinement alone achieves near-perfect performance that is well above that of subjects, adding forgetting leads to eventual chance performance, and adding elaboration

[3] The corrected variance is equal to $\mathrm{var}(M) - MSE \cdot \frac{1}{6}\sum_i \frac{1}{N_i}$, where var(*M*) is the uncorrected variance of the six condition means, *MSE* is the mean squared error (i.e., residual variance) reported in the ANOVA in Section 5.2, and N_i is the sample size of Condition *i*.

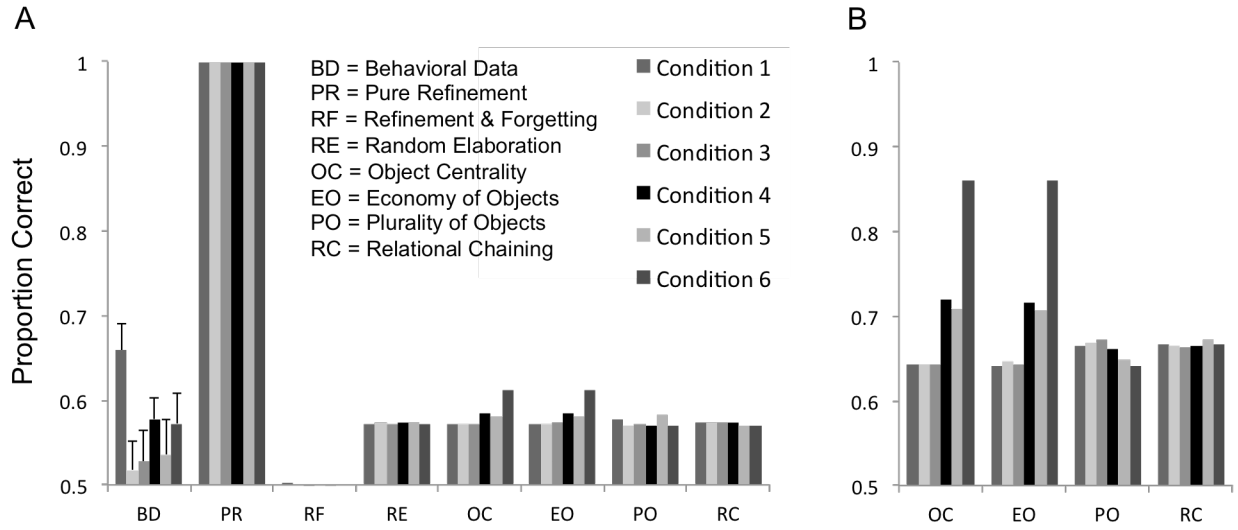


Figure 11. A: Mean performance for subjects and models on trials 201-400 of Experiment 2. Error bars for behavioral data indicate standard error of the mean. The RF model performs at chance. B: Predictions of structure-sensitive elaboration models at higher levels of performance (using a different value of the forgetting parameter), where condition differences are greater.

allows for intermediate performance. Additionally, only the models that assume structure-sensitive elaboration predict learning differences among conditions. However, unlike in Experiment 1, none of these models reproduces the pattern of differences exhibited by subjects.

The pattern of condition differences in the behavioral data was also somewhat different from that in Experiment 1. In Experiment 1, subjects' performance was best when the two relations operated in parallel on the same pair of objects (Condition 3) and second-best when they formed a chain (Condition 2). In Experiment 2, subjects' performance was best when all three relations formed a chain (Condition 1), and the conditions with the next best performance (4 and 6) both involve parallel relations. Performance was also better in Condition 4, where the third relation forms a chain with the two parallel relations, than in Condition 5, where the third relation is disconnected from the other two. Therefore, chaining and parallel relations both

appear to facilitate learning, although their relative importance seems to differ between the two experiments.

One reason none of the structure-sensitive models matches the empirical pattern in Experiment 2 is that each embodies only one type of preference for relational structure. In contrast, the behavioral data suggest that human learning is sensitive to multiple aspects of relational structure. Furthermore, people's expectations may shift as a function of which relations are already in the schema. For instance, if the schema contains two relations forming a chain, people may have an expectation that the chain will be extended. If instead the schema contains two parallel relations, people may have an expectation that additional relations will be defined on the same objects.

The simulations also reveal some unanticipated predictions of the models. First, the RC model shows almost no learning advantage in Condition 1, even though the category structure in that condition perfectly matches the model's elaboration preferences. More generally, the RC model shows little variation in performance across conditions, and likewise the PO model shows less variation than do OC and EO. The reason for these differences in structure-sensitivity is that the models' elaboration preferences lead to different numbers of preferred relations when selecting a candidate to add to the schema. That is, there are many more ways to extend a chain of relations (RC) or to add a relation on one or two new objects (PO) than there are to add a relation on existing objects (OC, EO). Therefore, even when the RC and PO models' expectations match the category structure (in Condition 1 for RC, and in Conditions 1, 2, and 5 for PO), those expectations do not provide strong guidance and hence do not significantly improve performance. In contrast, the OC and EO models have a large advantage in conditions with parallel relations (Conditions 4-6), because once one of these relations is learned, there is a small search space for discovering the other(s).

Second, the models' predicted order of performance across conditions does not fully align with their motivating theoretical principles. For example, the EO model performs better in

Condition 5 than in Condition 3, even though the latter involves fewer objects in the category rule. This result is due to partial matches between the category structures and the elaboration preferences of the models. More specifically, a model's performance is dependent not only on its likelihood of converging on the exact category structure (i.e., all 3 correct relations), but also on its likelihood of acquiring a partial schema (e.g., 2 of the 3 correct relations). Once the model has one correct relation in its schema, performance will depend on its probability of adding a second correct relation. In the case of Condition 5, whenever one of the two parallel relations is in the schema, the EO model will have a strong tendency to add the other. Therefore the model has a higher probability of learning a partial schema in Condition 5 than in Condition 3, even though it is less likely to acquire the third relation.

6. General Discussion

Learning and processing of structured relational representations is critical for analogical transfer (Gick & Holyoak, 1983), development of expert knowledge (Chi et al., 1981), and many other high-level cognitive processes (Penn et al., 2008). Previous research on learning of relational concepts has focused on the relative learnability of feature- versus relation-based categories (Love & Markman, 2003; Tomlinson & Love, 2010) or of deterministic versus probabilistic relational categories (Jung & Hummel, 2009; Kittur, et al., 2004), and on the manner in which comparison facilitates relational learning (Andrews, Livingston, & Kurtz, 2011; Gentner & Namy, 1999; Kurtz, Boukrina, & Gentner, 2013). However, to our knowledge, prior work has not explicitly examined how learning of relational concepts is affected by their internal relational structure.

In the present study, we define relational structure as a pattern of connections among the relations that make up a relational concept or system, in terms of how those relations operate on shared objects (i.e., role fillers). Under the structure-mapping theory of analogy (Gentner, 1983), an analogy between two scenarios amounts to recognition that they have the same

relational structure. Thus an understanding of the impact of relational structure on the relative learnability of different relational concepts might lead to new insights regarding the mechanisms underlying analogical reasoning, as well as the conditions under which such reasoning will be more or less successful.

Both experiments reported here constitute pure tests of the effects of relational structure on learning, in that all stimuli contained the same objects, features, and relations. The categories in the different experimental conditions were defined by the same set of relations, only configured differently in terms of how they were linked by shared objects. The significant performance differences among conditions in both experiments suggest that acquisition of relational categories is indeed affected by their relational structure. In particular, it appears that a relational category easier to learn when multiple constituent relations operate on the same objects (to a greater extent in Experiment 1) and when relations are linked together in a chain (to a greater extent in Experiment 2). These results are far from providing a complete theory, but we believe they contribute an important first step in understanding the effects of relational structure on learning.

One direction for future research would be to investigate people's ability to learn the abstract relational structures that defined the experimental conditions (Figures 3A & 9A), as opposed to the particular instantiations learned by the present subjects (e.g., Figure 2A). The latter included information about the locations of the objects on which the relations operate, as well as the type (e.g., BRIGHTER) and directionality of each relation. This specificity was required in defining the categories because otherwise they would be unlearnable. For example, every possible stimulus in Experiment 1 contains an object that is bigger than a second object and brighter than a third object, and thus a purely abstract version of Condition 2 would yield a trivial (universal) category. Likewise, this was why the models assumed location-preserving mappings between the schema and each stimulus (or equivalently that location attributes were part of the schema). Nevertheless, it might be informative to examine how readily people transfer their

knowledge between different instantiations of these experimental conditions. If people trained on two relational categories in succession showed better transfer when the categories instantiated the same relational structure, this would be evidence for a more abstract form of relational learning than was tested in the present experiments.

6.1. Implications for Models of Relational Concept Learning

We have focused here on the implications of these results for schema-based models of relational learning. It is commonly proposed that relational concepts are represented as schemas (Rumelhart & Norman, 1978; Shank & Abelson, 1977) and that these representations are learned by a process of schema refinement via intersection with successive instances (Doumas et al., 2008; Kuehne et al., 2000). As demonstrated here, schema refinement alone cannot predict sensitivity to relational structure, because the intersection process treats each element of a schema independently. Moreover, schema refinement taken alone incorrectly predicts a lack of false alarms in relational category learning, and it requires unreasonable assumptions about working memory because all potentially relevant information must be maintained in the initial schema.

As a complement to schema refinement, we have proposed a schema elaboration mechanism that adds information to a schema when a false alarm indicates the schema is underconstrained. Because refinement alone can never lead to a false alarm, elaboration is only consequential in a model that also allows forgetting of true category-defining relations. An important property of schema elaboration is that it can be sensitive to relational structure. Unlike with refinement following a miss, there is ambiguity in how to elaborate the schema following a false alarm, and thus the choice can be guided by the other relations already present in the schema.[4] Thus, expectations about relational structure might influence the learnability of a

[4] Forgetting might also be structure-sensitive, if a relation's likelihood of being maintained were to depend on its connections to other relations in the schema. We tested a series of structure-sensitive forgetting models (also

relational category. A model combining refinement, forgetting, and elaboration also addresses the other shortcomings of pure refinement mentioned above: Rather than starting very complex and converging downward to the true category, a schema can be relatively small, growing and shrinking following false alarms and misses.

The formal models tested here define a hierarchy that, when evaluated in sequence, provides support for each of our theoretical proposals. First, schema refinement alone (PR model) is psychologically implausible for relational learning, predicting performance far beyond the range of subjects. Second, refinement plus forgetting (RF) produces declining learning curves and eventual chance performance, suggesting that an elaboration mechanism that can grow a schema's complexity is necessary to account for people's intermediate performance. Third, in conjunction with refinement and forgetting, elaboration offers a plausible mechanism for how new or forgotten information can be (re)incorporated into a schema (RE). Fourth, schema refinement cannot explain differential learning of categories defined by different relational structures, but structure-sensitive elaboration can (OC, EO, PO, RC).

Beyond these conclusions, there are aspects of the data that the models do not fully explain. In Experiment 1, both the OC and EO models qualitatively matched the empirical pattern of differences across conditions. However, when more complex category structures were introduced in Experiment 2, none of the models' predictions matched the empirical pattern of condition differences. Moreover, although the different forms of structure-sensitivity in these models were theoretically motivated, they were built in as scoring rules. Further work should investigate the learning or decision mechanisms that may underlie these tendencies, as well as whether they arise from explicit expectations about category structures or implicit learning

---

including refinement and random elaboration) and found their predictions to be very similar to those of the structure-sensitive elaboration models reported here. Importantly, these models still require schema elaboration, or else they eventually perform at chance, just as with the RF model reported here.

mechanisms (cf. Ashby, Alfonso-Reese, Turken, & Waldron, 1998). In short, although it is clear that relational structure affects relational learning, the specific mechanisms behind this effect remain an open question.

Another shortcoming of the present structure-sensitive models is that they produced weaker condition differences than were observed empirically. To match the subjects' overall performance level, the models had to assume a large forgetting rate (small *L*), leading to typically small schemas. Because the models' structure-sensitivity derives from the information already in the schema during elaboration, the result was small condition differences. One possible solution would be to substitute the present all-or-none approach (in which a relation is either in the schema or it is not) with a continuous learning model, in which continuous-valued strengths are maintained for all possible relations in the schema. Mutual excitation or inhibition between relations could account for sensitivity to relational structure, following principles similar to the ones tested here. For example, the model might assume mutual excitation between any two relations sharing a pair of objects. Such a continuous learning model might generate low performance when its association strengths are weak but still retain some knowledge of the global category structure, thus showing greater structure-sensitivity.

One aspect of the data not analyzed here is order effects. For example, encountering the appropriate example at the "right" time (e.g., early in learning) may play a critical role in concept acquisition. Much of our previous work has made detailed analyses of sequence effects as a means to uncover learning mechanisms and concept representations (Foster, Cañas, & Jones, 2012; Jones, Curran, Mozer, & Wilder, 2013; Jones, Love, & Maddox, 2006; Jones & Sieck, 2003). The present models also predict sequence effects, and thus a useful angle to explore in follow-up work could be to examine the role of such effects in the learning of structured relational concepts.

One reason for the modeling approach used here, as well as the type of relational categories tested, is that previous evidence suggests that both agree with the type of relational

categories people are predisposed to learning. The assumption of a single schema containing discrete (all-or-none) relations corresponds to a conjunctive category representation, in which a stimulus is a category member if and only if it satisfies all relations in the schema. The categories subjects learned were also of this nature (e.g., upper right object is bigger than lower right object AND upper left object is brighter than lower right object). Relational categories could also be defined by disjunctions (e.g., upper-right object is bigger than lower-right object OR upper-left object is brighter than lower-right object), but previous work indicates that people are very poor at learning such categories (Jung & Hummel, 2009, 2011; Kittur et al., 2004). For example, Kittur et al. (2004) contrasted learning of four different types of categories, defined by (1) the conjunction of three features, (2) the conjunction of three relations, (3) a family resemblance structure in which any three of four features were sufficient for membership, and (4) the same family resemblance structure defined over four relations. The first three categories were learned equally quickly, but the family resemblance relational category took more than twice as long. Kittur et al. explained this result using the same basic assumption made here: People represent relational categories by a single schema containing the relations present in all category members, which in the family resemblance case is the empty set. This finding supports the present modeling approach, and it offers a challenge to other models of relational category learning based on exemplar representations (Tomlinson & Love, 2006), which would predict good learning of a relational family resemblance category.

The models considered here can also be contrasted with models built on more complex architectures, such as DORA (Discovery Of Relations by Analogy; Doumas et al., 2008) and LISA (Learning and Inference with Schemas and Analogies; Hummel & Holyoak, 2003). These models learn relational concepts through schema induction and refinement, but they also include more fine-grained mechanisms that might lead to different predictions than the idealized PR model tested here. Evaluating these models on the present experiments could be a useful undertaking for future research, but they would require further specification before they could be

used to derive specific predictions. The various learning mechanisms in DORA (e.g., retrieval, structure mapping, schema induction, schema refinement, analogical transfer, predication) could each apply at multiple stages of a category-learning task, and in its current form DORA does not commit to when each mechanism will be triggered. Moreover, fits of DORA or LISA would arguably be less informative about the mechanisms involved in relational category learning, because of the challenge in isolating individual mechanisms or theoretical principles in such complex models. The simple architecture used in the present models allows for a clear interpretation of the mechanisms that drive relational learning in each model, enabling focused tests that isolate the empirical support for each proposed mechanism. Thus, despite the sophistication of DORA and LISA, we consider the present models more useful as diagnostic tools for theoretical development.

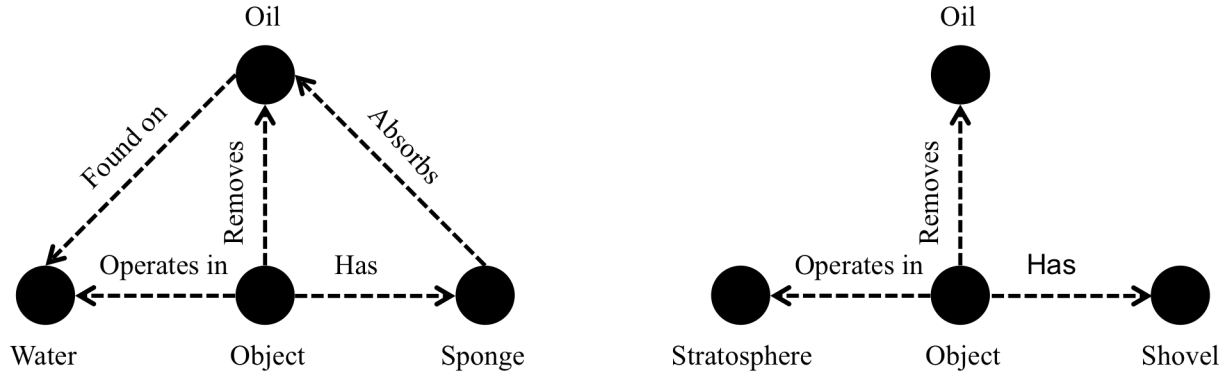
Notwithstanding the advantages of this analytic approach, the broader message of this article is that theories of analogy need to consider more complex mechanisms. Researchers in machine learning and related fields have been working on these kinds of problems as well and have developed computationally sophisticated proposals for how relational concepts can be learned, but unfortunately there has been little cross-pollination between psychology and machine learning in this domain. However, some models have made progress in bridging this gap, such as Copycat (Mitchell & Hofstadter, 1990) and AMBR (Associative Memory-Based Reasoning; Kokinov, 1988, 1994). Both of these models were designed with the goal of constructing flexible representations from scratch. The interactive dynamics between top-down and bottom-up learning mechanisms in these models may lead them to make differential predictions about the learnability of relational structures in the present experiments. Other extant modeling frameworks offer additional theoretical principles that may yield further insights into the present paradigm, including LISA (Hummel & Holyoak, 2003) and DORA (Doumas et al., 2008), which tie relational representations to subsymbolic (connectionist) representations; ReBel (relational Bellman operator; van Otterlo, 2009), which uses principles of reinforcement

learning to acquire structured relational representations equivalent to schemas; MAC/FAC (Forbus, Gentner, & Law, 1995), which separates the processes of analogical retrieval and analogical mapping; BART (Bayesian Analogy with Relational Transformations; Lu, Chen, & Holyoak, 2012), which induces new relations from unstructured input using probabilistic inference; and Companions (Forbus, Klenk, & Hinrichs, 2009), which links structure-mapping principles to inferential logic operating on a large conceptual knowledge base. Examining each of these models' predictions on the present experiments may provide important insights into additional mechanisms that are critical to the acquisition of relational concepts. Hopefully one contribution of this paper will be to spur the psychological community to think about more complex mechanisms that must be involved in more complete models of analogical learning.

6.2. Learning of Relational Representations

A primary aim of this article has been to investigate the factors that affect a relational concept's coherence. Although differences in learnability were observed as a function of relational structure, performance was relatively low in all conditions. Therefore the simple manipulations of relational structure used here appear to have affected coherence, but not very strongly. An important question for further research is to explore other variations of relational concepts that might have stronger effects on their coherence.

One possibility is that strong conceptual coherence requires relational structures that are more richly interconnected than were the concepts used here. This idea is supported by Rehder and Ross's (2001) study of what they termed abstract coherent categories. Rehder and Ross argued that relational categories can acquire psychological coherence when a category member's first-order relations fit together in a way that makes sense based on prior knowledge. For example, the predicates "operates in water", "removes spilled oil", and "coated with spongy material" are psychologically coherent because of additional relations (presumably known by the subject) that oil slicks are found on water and a sponge can absorb oil. This explanation naturally fits into the present framework, as illustrated in Figure 12. The coherent item forms a



Coherent Item **Incoherent Item**

Figure 12. Diagrams of the relational structures formed by items in the coherent and incoherent categories from Rehder and Ross (2001).

rich relational structure, in which the objects tie the relations together in multiple ways. In contrast, the incoherent item forms an impoverished structure, with no additional relations to "close the loops." This interpretation suggests that categories in our paradigm might be more coherent if they consisted of richer relational structures. It also suggests that the coherence of these categories might increase if their objects were tied together by prior knowledge (cf. van Overschelde & Healy, 2001).

There may also be other factors that are important for producing psychological coherence of a relational concept, beyond the sheer number of relations and their interconnections. One possibility is that psychological coherence is greater in schemas that exhibit certain types of self-similarity, such as mirror graph symmetry or fractal-like symmetries (i.e., similar patterns of connections at different levels of granularity). It is also possible that psychological coherence arises when a relational system as a whole exhibits emergent behavior that does not follow transparently from the sum of its parts, leading to a more compact higher-level representation (e.g., Holland, 1999). To take a classic example, the massively complex system of interactions among a cloud of gas molecules can be concisely characterized by the emergent variables of

temperature and pressure, creating a strong conceptual coherence in that thermodynamic level of description.

Alternatively, psychological coherence may be driven by the perception of causal connections or other second-order relations within a relational system, which take one or more first-order relations as their arguments (e.g., gravity *causes* the earth to revolve around the sun). Although such second-order relations may contribute to psychological coherence, this idea may also beg the question. First-order relations cannot be arbitrarily strung together into a coherent system simply by introducing causality. The reason a system like *cause*(*gravity*,*revolve*(*planet*,*sun*)) is coherent is likely that it makes sense for gravity to have this effect, in terms of the alignment of gravitational force with the vectorwise difference between the sun's and planet's locations, the relationship between force and acceleration of the planet, and the geometry of circular (or elliptical) orbits. Thus, the reason the concept is coherent (and the second-order CAUSE relation makes sense) is that it comprises a system of first-order relations that all "fit together" (based on previous knowledge) to form a coherent relational structure.

Another open question regarding representation in analogical learning is how people find representations of two scenarios that will support a successful analogy between them. Chalmers et al. (1992) argue that, precisely because of the great representational flexibility on which analogical reasoning capitalizes, finding or constructing the right representations of the separate scenarios is one of the most computationally challenging aspects of analogy. It is not enough to assume a separate “representation module” (French, 1997) that determines the representation and passes it to the analogy process, because different analogies involving the same item can require entirely different representations of that item. Nevertheless, many models of analogy bypass this issue by using hand-coded representations that are well-suited for the analogies the models are intended to discover. Furthermore, such models often lack conceptual flexibility and are unable to discover representations that have not been pre-coded

(Kokinov, 1994, 1998). Mitchell and Hofstadter (1990) argue that a complete model of analogy must incorporate interactive dynamics between bottom-up and top-down learning mechanisms, specifically via a tight coupling between the processes of representation construction and analogical mapping. Their Copycat model implements these ideas by building up representations of the base and target scenarios through the discovery of regular structures in each. The analogical mapping between the scenarios evolves simultaneously with the building of these representations. The representations within each scenario compete with each other, with only those that participate in the developing analogy surviving and growing into larger structures. Thus the model solves the representation and mapping problems concurrently, through an interactive process that enables discovery of representations that will support a good analogy.

The proposals in our models closely parallel these previous ideas. Schema refinement can be likened to bottom-up learning, as it is driven purely by local comparisons of individual relations between the base and target scenarios. Schema elaboration can be likened to top-down learning, as it is dependent on the global structure of the schema. Furthermore, schema elaboration involves competition among relations to be added to the schema, similar to the competition among representations in Copycat. Although the models formulated here operate on pre-coded representations of primitive relations, those representations contain a large amount of superfluous information, and thus there is a lot of work for the model to do in order to converge on the appropriate schema. This convergence is based on interactive dynamics between the pruning process of schema refinement and the constructive process of schema elaboration. One important difference is that Chalmers et al.'s (1992) critique of structure-mapping theory was cast primarily at a within-trial level, whereas the learning mechanisms in our models operate across trials. However, the issues raised by Chalmers et al. seem equally applicable to across-trial learning.

In addition to these parallels between within-trial and across-trial dynamics of representation construction, the ideas advanced here might directly inform the question of representation construction during the course of a single analogy. First, our investigation of the effects of relational structure could illuminate the question of representation selection, in that people might be guided toward representations that are more structurally coherent. Second, preferences for relational structure may also reduce the number of mappings that are considered, ameliorating the issue of combinatorial explosion. Third, the process of refining and elaborating a schema to zero in on the right concept over the course of learning parallels the process of zeroing in on the right level of abstraction for each scenario (by unpacking or chunking its substructure) during the course of a single analogy. This parallel may or may not turn out to reflect shared theoretical principles, but the connection is suggestive. In these ways, our work on schema-learning dynamics and effects of relational structure with relatively simple materials may ultimately serve as a stepping stone toward understanding how people construct representations at the optimal level of abstraction in more complex analogies.

6.3. Conclusions

The present findings indicate that relational category learning is sensitive to higher-order structure, such that some relational structures are harder to learn than others. The models developed here offer some constraints on the mechanisms of relational learning, but many questions remain. Understanding what drives the differences among the experimental conditions may provide important theoretical insight into the cognitive processes by which people acquire abstract, higher-order concepts. Such work may also have practical applicability for areas where the recognition of relational structure is important, such as education, problem solving, and decision-making.

Author Note

Daniel Corral and Matt Jones, Department of Psychology and Neuroscience, University of Colorado. This research was supported by AFOSR grant FA9550-10-1-0177. Portions of this work were presented at the 34th Annual Meeting of the Cognitive Science Society. Correspondence regarding this article may be addressed to Daniel.Corral@Colorado.edu or mcj@colorado.edu.

References

Andrews, J. K., Livingston, K., & Kurtz, K. (2011). Category learning in the context of co-presented items. *Cognitive Processing, 12*, 161-175.

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.

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4,* 829-839.

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.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

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

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347-362.

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.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141-205.

Forbus, K, Klenk, M., & Hinrichs, T. (2009). Companion cognitive systems: Design goals and lessons learned so far. *IEEE Intelligent Systems*, *24*, 36-46.

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*.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.

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

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development*, 14, 487-513.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Holland, J. H. (1999). *Emergence: From chaos to order*. Cambridge, MA: Perseus Books.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 220-264.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220-264.

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review, 120*, 628-666.

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

Jones, M. & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 626-640.

Jung, W., & Hummel, J. E. (2009). Probabilistic relational categories are learnable as long as you don’t know you’re learning probabilistic relational categories. In *Proceedings of The 31st Annual Conference of the Cognitive Science Society* (pp. 1042–1047).

Jung, W., & Hummel, J. E. (2011). Progressive alignment facilitates learning of deterministic but not probabilistic relational categories. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2643–2648).

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2004). Feature- vs. relation-defined categories: Probab(alistical)ly not the same. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 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.

Kokinov, B. (1998). Analogy is like cognition: dynamic, emergent, and context-sensitive. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 96-105). Sofia: NBU Press.

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* (pp. 770–775).

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.

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

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*, 617-648.

Mitchell, M. & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy-making. *Physica D, 42*, 322-334.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Pashler, H., & Mozer, M. C. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1162-1173.

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.

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

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: Three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition* (pp. 37-53). Hillsdale, NJ: Erlbaum.

Shank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

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

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*, Whole No. 517.

Shiffrin, R. M., & Cook, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior, 17*, 189-218.

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

Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology, 10,* 502–532.

Sweller, J., Mawer, R., & Ward, M. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, *112*, 639-661.

Tomlinson, M.T., & Love, B. C. (2006). From pigeons to humans: Grounding relational learning in concrete examples. *Twenty-First National Conference on Artificial Intelligence* (AAAI-2006), *17*, 136-141.

Tomlinson, M. T., & Love, B. C. (2010). When learning to classify by relations is easier than by features. *Thinking & Reasoning, 16*, 372-401.

van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147-177.

van Otterlo, M. (2009). *The logic of adaptive behavior*: *Knowledge representation and algorithms for adaptive sequential decision making under uncertainty in first-order and relational domains*. Fairfax, VA: IOS Press, Inc.

van Overschelde, J. P., & Healy, A. F. (2001). Learning of nondomain facts in high- and low-knowledge domains. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 1160–1171.