EXPLANATION AND STRUCTURE-MAPPING IN THEORY-BASED CATEGORIZATION

Jason Jameson
j-jameson@u.northwestern.edu

Dedre Gentner
genner@northwestern.edu
Department of Psychology,
Northwestern University, 633 Clark Street
Evanston, IL, USA

ABSTRACT

According to the theory-based approach to categorization, explanatory knowledge plays a critical role in supporting categorization decisions. This paper presents and tests a version of this view—the Explanation-Mapping hypothesis—that integrates explanation with a model of the structure-sensitive comparison processes that may underlie theory-based categorization decisions. Two experiments test the hypothesis that processes underlying analogy—specifically, structural alignment—also underlie explanation-based categorization decisions.

INTRODUCTION

Categorization is the process by which objects or events are grouped into categories—collections of objects treated in some respect as equivalent—for the purpose of making sense of the world. As a reflection of its importance, many researchers have expended great effort in understanding the mechanisms that support categorization decisions. This paper presents an approach to categorization inspired by the explanatory function of category knowledge, and assumes that categorization is a type of inference to the best explanation (Murphy & Medin, 1985; Rips, 1989).

The importance of explanations for categorization has been proposed (e.g., Murphy & Medin, 1985; Rips, 1989), but its promise as a fully developed psychological model has yet to be realized. This paper tests an Explanation-Mapping hypothesis, based on research on the psychology of explanation and the processes that could support the use of explanatory information.

We claim that categorization decisions depend on the explanatory principle common to category members (Murphy & Medin, 1985), and further that structural alignment processes underlie co-categorization judgments. Structure-mapping has been implicated in many kinds of comparison-based judgments, particularly in analogy and similarity (Gentner, 1983; Markman & Gentner, 1993). We extend the scope of structure-mapping processes to theory-based categorization decisions that rely on the use of explanatory information.

The motivation for this hypothesis is based on several observations. First, theoretical knowledge typically comprises a rich system of interrelated beliefs—processes involved in comparing an instance to a category representation (in this case, taken to be an explanatory principle) must thus be sensitive to structural information. Second, because explanations differ in quality (Keil, 2006; Lombrozo & Carey, 2006; Thagard, 1989), the comparison process should incorporate some process for evaluating the quality of competing knowledge structures. Structure-mapping processes are sensitive to an important property of structured representations that may underlie explanatory quality, namely, systematicity: the degree to which a representation forms a system of interrelated relations, with relations among objects and object attributes, and higher-order relations among relations (Gentner, 1983; Keil, 2006).
Two experiments test the hypothesis that structure-mapping processes support theory-based categorization decisions. Moreover, because structure-sensitive comparison processes have been implicated in many kinds of conceptual tasks, we broaden the scope of our investigation of explanatory information to similarity and analogy tasks. The first experiment tests the prediction that theory-based categories, in a co-categorization task, will be formed around the best shared explanation between category members. Moreover, we also test this hypothesis with similarity judgments, since many empirical studies have shown that similarity judgments are biased towards systematic relational structure (e.g., Markman & Gentner, 1993). The first experiment employs categories whose members share both structural as well as surface information. The second experiment extends the first in two ways: (1) to categories whose members share little surface information (so-called relational categories; see Gentner & Kurtz, 2005); (2) to an analogy task, which is predicted to exhibit a focus on shared explanatory information, as in the categorization and similarity tasks.

### EXPERIMENT 1

The first study tests whether people (a) prefer to co-categorize, and (b) judge as most similar, stories that share a systematic explanatory structure, as compared to those that do not. The logic of the study is as follows. A standard with a rich explanatory structure is compared to two alternatives, and separate groups of participants are asked either (a) which should be co-categorized or (b) which is more similar. The process underlying is hypothesized to comprise an alignment and evaluation of competing shared relational structures.

To illustrate what we mean by an explanatory structure, consider the sample materials in Table 1. In the story, some statements provide good explanations for many events described in the story; other statements provided poorer explanations, if any at all.

<table>
<thead>
<tr>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter was the most faithful and devoted of all the younger monks in the monastery. In addition, he was as skilled at illuminating texts as he was at tending the gardens. His simple passion for God was admired by all. One day at dinner, he spilled his food onto the floor. Peter was dismayed as he viewed the wasting of food as a sin. He went to his cubicle and began to fast in silence. His elders became worried, but they decided to say nothing. When a week had passed, Peter felt that he had repented, and resumed his life as usual.</td>
</tr>
</tbody>
</table>

Table 1. Sample item from Experiment 1.

For example, Peter’s belief that wasting food is a sin supports an explanatory structure that accounts for why he was so upset, why he went to his room, and why he repented for days. However, the fact that Peter is skilled at illuminating texts and tending gardens provides relatively little explanatory support for the story.

The first experiment compares the relative focus on shared explanatory structure for similarity and categorization judgments, and tests the prediction that similarity can also pattern after categorization for a broader class of categories than that yet explored by causation-based approaches. This prediction is of broader interest to the current literature on theory-based categorization, as it runs counter to the current consensus (e.g., Ahn, et al. 2000; Rehder & Hastie, 2001) that treats the background knowledge and processes used in theory-based categorization tasks as distinct from those underlying similarity judgments.

The experiment tests the prediction that both categorization and similarity judgments will exhibit a preference to match the best shared explanatory structures. This prediction follows from the demonstrations that similarity judgments can be sensitive to relational structure (e.g., Markman & Gentner, 1993).
METHOD

Participants

Forty undergraduate students from Northwestern University participated for partial credit in an introductory psychology course.

Design, materials, and procedure

Two conditions were tested between-subjects (20 per condition), either a similarity or categorization judgment. For both conditions, the task a match-to-sample (paper-and-pencil) forced choice task: that is, a single target item was presented, and participants were instructed to pick the one of two alternatives that was either most similar to or that should be categorized with the standard. The materials comprised four triads of stories (a separate triad on each page), with a triad composed of a standard story and two alternatives. Both alternatives were modifications of the standard. The structure-preserving alternative shared with the standard a structure-preserving statement (to be defined below), but differed on a less important explanatory statement. The structure-altering alternative differed on the more important structure-preserving statement, but shared the less important explanatory statement.

The materials were based on Bowdle and Gentner (1997), who had independent judges rate the systematicity of the stories. The contribution of a statement to the systematicity of a story (i.e., the support a statement provided to the explanatory structure of the story) was estimated by computing the reduction in systematicity that resulted from modifying the statement so that it no longer played as important an explanatory role (i.e., changing “Peter was dismayed as he viewed the wasting of food as a sin” to “Luke was dismayed as the bean porridge they were serving that night was his favorite”). Specifically, participants in their study rated the coherence of a standard story and four variants, with each variant differing in some way from the standard (e.g., deleted key sentences, shuffled sentences, or modified key sentences). Bowdle and Gentner computed a systematicity imbalance score by subtracting the mean coherence ratings from each of the four variants from the mean coherence rating for the standard. We used the systematicity imbalance score for the variant in which a key sentence was modified (the ‘causal pivot’) as an estimate of the statement’s contribution to the systematicity of the standard story.

These estimates served as the basis for choosing a key explanatory statement. One other statement from each of the four stories was chosen that intuitively provided less explanatory support for the story. As a manipulation check, each story and the associated pair of statements were given to an independent set of judges (N = 19), who were instructed to pick the statement that provided a better explanation for the events in the story. The proportion of choices for the a priori more explanatory statement (91%) confirmed the manipulation, a proportion that differs significantly from chance (50%), t(18) = 11.89, p < .0001, d = 2.73.

Results & Discussion

The proportion of choices for matching on the basis of a better shared explanation (across participants) was submitted to an independent samples t-test. As predicted, the categorization group (M = .84, SD = .15) exhibited a stronger preference for the structure-preserving alternative compared to the similarity group (M = .69, SD = .29), t(38) = 2.36, p = .05, d = .65. Also as predicted, when tested against chance, the similarity group exhibited a reliable preference for the structure-preserving alternative, p < .05, d = .64. Testing the categorization results against chance yielded a much larger effect, p < .01, d = 2.3.

These findings are consistent with the claim that explanation-based categorization judgments depend on the alignment and evaluation of matching relational structures. Moreover, these results also implicate a pref-
EXPERIMENT 2

Experiment 1 provides evidence that co-categorization and similarity judgments are sensitive to explanatory structure embedded in stories, a prediction that follows from considering candidate structure-mapping processes that may support those judgments. The second experiment tests the follow-up prediction that the preference to match on the basis of explanatory information holds even for entities that share little surface similarity, as long as they share systematic relational structure. If we find this pattern, then this will be evidence that the Explanation-Mapping approach can extend to relational categories (Gentner & Kurtz, 2005; Markman & Stilwell, 2001; Rehder & Ross, 2001). For these categories, membership is based primarily on a shared relational schema. For example, the category barrier has members that share few surface features, but still form a comprehensible class (e.g., wall, ideas, socio-economic status), because they share a common relational schema (e.g., x PREVENTS y FROM ACHIEVING GOAL z) (Gentner & Kurtz, 2005). Although, comparatively under researched, relational categories are prevalent in adult discourse. As Asmuth and Gentner (2005) note, relational nouns (names for such categories) comprise nearly half of high-frequency nouns. The Explanation-Mapping hypothesis supports a unified approach to understanding these kinds of categories, and thus goes beyond the kinds of categories studied by previous approaches to theory-based categorization (e.g., Ahn, et al, 2000; Rehder & Hastie, 2000).

The goal of the second experiment is to replicate the first experiment using categories in which membership is based primarily on the shared relational structure, with little to no influence from shared surface properties. In addition to similarity and categorization judgments, the second study will also employ an analogical relatedness task. According to structure-mapping theory, analogies are based primarily on evaluating the systematicity of shared relational structure (e.g., solar system and the atom), and thus should pattern after the similarity and categorization tasks. Experiment 2 thus tests the prediction that all three judgment types will show a preference to match on the basis of shared explanatory structure.

METHOD

Participants

Sixty undergraduate students from Northwestern University participated for partial credit in an introductory psychology course.

Design, materials, and procedure

The design was nearly identical to Experiment 1, with an analogical relatedness task included, which was tested between-subjects. Specifically, participants were asked to pick the one alternative of two that was most analogous to the standard. Further, the standard stories were the same as in Experiment 1, but the alternatives in each triad were constructed to share a relational schema, but little in the way of surface properties.

Results

A between-subjects ANOVA revealed no significant differences across the conditions, $p > .05$. However, if we compare each condition separately to chance (.5), all three tasks showed clear preferences for matching on the basis of the best shared explanation: for similarity ($M = .65, SD = .30$), $t(19) = 2.26$, $p = .04, d = .51$; for categorization ($M = .74, SD = .24$), $t(19) = 4.50$, $p < .001, d = 1.01$; and for analogical relatedness judgments ($M = .74, SD = .19$), $t(19) = 5.60, p < .0001, d = 1.25$.

Discussion

As predicted, all three tasks showed definite preferences to match analogous stories on the basis of shared explanatory structure. The
larger effect sizes for the categorization and analogical relatedness tasks suggest that matching on the basis of relational structure was especially important. This possibility is not surprising if we consider that analogies depend primarily on shared relational structure, and that explanation-based categorization depends on shared explanatory structure. A natural conjecture is that shared relational structure and shared explanatory structure are one and the same. If so, this opens up an avenue of research into relational categories such as barrier and robbery: their members should be found to share relational structure, though not necessarily surface attributes. Further research should test the possibility analogy and explanation-based categorization are indeed so closely related.

**GENERAL DISCUSSION**

This paper presents an Explanation-Mapping approach to theory-based categorization. Two studies test three key predictions: (1) explanatory information underlies theory-based categorization; (2) categorization processes involve structure-sensitive comparison processes: specifically, alignment of explanatory structures; and (3) because structure-sensitive comparison processes underlie both categorization and judgments of similarity and analogy, these tasks should pattern together for explanation-based categories.

Because most research on theory-based categorization has focused on causal categories (e.g., Ahn, et al. 2001; Rehder & Hastie, 2001), our studies examine causal categories from an Explanation-Mapping perspective. The long-range goal of this research is to investigate whether causal categories can be viewed as a special case of explanation-based categories. Experiment 1 adopted story categories with rich knowledge structures, and showed that people prefer to co-categorize, and judge as most similar, stories that share a systematic explanatory structure. This evidence is consistent with the claim that structure-sensitive comparison processes underlie both similarity judgments and co-categorization judgments, particularly for explanation-rich categories.

Experiment 2 showed that the Explanation-Mapping approach extends to relational categories—categories whose membership is determined solely by shared relational structure among members. The materials for this study were the complex story categories used in Experiment 1, modified to remove surface commonalities. Thus co-categorization decisions could be made solely on the basis of shared relational structure, and shows that people prefer to match items that share systematic explanatory structure, in spite of differences in surface similarity. As before, common patterns were found for similarity and co-categorization; in addition, analogical relatedness judgments showed the same pattern. This confluence of similarity, analogy and co-categorization is not predicted by current leading approaches to theory-based categorization, but follows from the Explanation-Mapping approach.

The Explanation-Mapping approach provides a general account of theory-based categorization, covering a wider range of categories than is currently possible by causation-based approaches. It can address relational categories (such as barrier) (Gentner & Kurtz, 2005; Markman & Stilwell, 2001; Rehder & Ross, 2001) and non-causal categories (such as mathematical and logical categories). A focus on explanation may also provide a more satisfying account of causal categories. For example, Ahn’s causal status hypothesis (e.g., Ahn, 1998; Ahn, et al., 2000) may be viewed in some situations as a special case of the Explanation-Mapping hypothesis (e.g., the rock caused the window to break; the collision of the rock with the window explains the broken window).

An explanation-based approach can also help to explain previous results in theory-based categorization, such as why some properties (such as grey, white, and black) may seem more important than others for theory-based categorization decisions. For example, grey hair seems more similar to white hair than to black hair, because grey and white properties
play similar supporting roles in the same explanation (appealing to age-related changes). However, grey clouds seem more similar to black clouds than to white clouds, because of a shared explanation related to weather (example taken from Medin and Shoben, 1988). This account is consistent with Medin and Shoben’s (1988) claim that some properties of objects play important theoretical roles in the knowledge of a category. For example, curvedness plays a more important theoretical role for the category of boomerangs than for the category of bananas. What makes a theoretical role important can be understood in explanation-based terms: curvedness is explained by the function of boomerangs; and thus fits within a larger explanatory framework containing information about designers’ intentions, the object’s function, and its relationships to other properties (for example, how its curved shape causes its path to curve). For bananas, in contrast, curvedness bears no strong relation to other important functional properties.

The Explanation-Mapping hypothesis also bears on the relationship between an instance and a category. Specifically, the likelihood that an instance belongs to a category may depend on the fit between the representation of the instance and the representation of the category (i.e., how well the instances of the property are explained by the category explanatory principle). This proposal is consistent with well-known findings. For example, Wisniewski (1995) found that, for the category of ‘things that capture animals,’ more likely members possessed pairs of connected features, such as “caught an elephant” and “contains peanuts,” than members with unrelated features such as, “caught an elephant” and “contains acorns.” Clearly, the former feature pair is better explained by the knowledge of the category than the latter pair (i.e., elephants like peanuts, which could be used as bait). Moreover, the most likely members of explanation-based categories could be considered ‘theoretical ideals’ (Rehder, 2003; see also Barsalou, 1983), exemplars that “instantiate the full set of theoretical knowledge one should possess about a category” (p. 1155). These ideals may be those exemplars whose structure fits best with the category’s explanatory structure.

The Explanation-Mapping proposal also helps to illustrate how instances within the category may be related. Specifically, structure-mapping processes may serve as general learning mechanism for selectively abstracting explanatory information from instances, and thus supporting the development of theory-based categories (Catrambone & Holyoak, 1989; Gentner, Loewenstein & Thompson, 2003, 2004; Gentner & Medina, 1998; Gick & Holyoak, 1983; Kurtz, Miao & Gentner, 2001; Loewenstein & Gentner, 2001; Loewenstein, Thompson & Gentner, 2003; Thompson, Gentner & Loewenstein, 2000). What this research suggests is that over successive similarity comparisons, important explanatory background knowledge may be abstracted.

In summary, focusing on the explanatory function of background knowledge together with a consideration of potential processes that may use that information has the potential for revealing the intricate role that theoretical knowledge plays across a broad range of conceptual tasks.

REFERENCES


Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. Journal of Educational Psychology, 95, 393-408.


