Causal Structure in Categorization

Woo-kyoung Ahn
Department of Psychology
University of Louisville
Louisville, KY 40292
wkahn001@homer.louisville.edu

Mary E. Lassaline
Department of Psychology
University of Illinois
Champaign, IL 61820
mlassali@psych.uiuc.edu

Abstract

What role does causal knowledge play in categorization? The current study tested the hypothesis that weight given to features is determined by the specific role they play within a causal structure. After learning typical symptoms of a disease, participants were asked to judge the likelihood that new patients had that disease. Half of the patients were missing one of the typical symptoms, and the other half had an extra symptom (a symptom typical of an alternative disease). For patients with a missing symptom, likelihood ratings were lower if the missing symptom was a cause of other symptoms than if it was an effect. However, for patients with an extra symptom, there was no difference between likelihood ratings when the extra symptom was a cause or an effect. These results suggest one mechanism underlying differences between experts and novices in categorization, and suggest an explanation for why different kinds of features (e. g., molecular or functional) are important for different kinds of categories (e.g., natural kinds or artifacts).

EBL and SBL

Categorization research falls roughly into two camps: similarity-based learning (SBL) and explanation-based learning (EBL). SBL theories are based on the assumption that concepts are formed by extracting similarity across multiple examples. EBL theories emphasize the role of prior knowledge in learning new concepts. Categorization research within cognitive science had initially been dominated by SBL models. Recently, researchers have pointed out a number of difficulties faced by SBL theories, and proposed a knowledge-intensive EBL view.

Importance of background knowledge

Many psychological experiments supporting SBL models have involved categorization in controlled settings using artificially constructed stimuli. Therefore, SBL theories tend not to take into account the role played by learners' domain knowledge in categorization. In contrast, many studies have shown that categorization processes are strongly affected by the learners' background knowledge (see Murphy & Medin, 1985, for a discussion). For example, category learning occurred faster with background knowledge (Ahn, Brewer, & Mooney, 1992; Pazzani, 1991); different features were abstracted for rule induction depending on subjects' domain theories (Wisniewski & Medin, 1991); and typicality judgments varied with subjects' perspectives (Barsalou, 1983; Roth & Shoben, 1983).

Importance of relational structure

In most SBL theories, objects are represented as sets of independent features (Tversky, 1977). However, features in natural objects are structured in many complex ways. For example, in the category bird, "having wings" and "building nests in trees" are not independent features because having wings enables birds to fly, which in turn allows birds to build nests in trees. Recent approaches to similarity, induction and categorization have emphasized the importance of structural relations, which are predicates taking two or more arguments (e.g., "is larger than" or "causes"), over simple attriburtes taking only one argument, such as "is red" (Gentner, 1989; Goldstone, 1994; Lassaline, 1995; Markman & Gentner, 1993; Medin, Goldstone, & Gentner, 1993). For example, it has been shown that the soundness of an analogy is determined by matching on structural relations rather than on simple attributes.

Importance of causal explanations

The idea that features connected to a relational structure are more important in categorization than unrelated features is consistent with many EBL systems developed in Artificial Intelligence (DeJong & Mooney, 1986; Mitchell, Keller & Kedar-Cabelli, 1986). In EBL systems, an instance is explained by using background knowledge, and then this explanatory structure is generalized to identify constraints and variables associated with a category. In making generalizations, causally connected features are incorporated into a schema as constraints that must be satisfied in categorization, and unconnected features are allowed to take variable values.

Feature Role within Relational Structure

So far, we have reviewed many studies showing that background knowledge is crucial in concept learning because it specifies which predicates are more important. The most prevalent theme in these studies is that features that belong to dependency structures (e.g., relational features or explanatory structures) are more important. However, previous studies have not looked at how arguments within the same relational structure vary in importance. That is, given the relational structure "X causes Y," which feature is more crucial in categorization, X or Y?

To understand the importance of this question, consider a clinical graduate student who is forming the concept "borderline personality disorder" by learning its typical symptoms, which include frantic efforts to avoid abandonment, identity disturbance, and recurrent suicidal behavior. When diagnosing a patient with borderline personality disorder, the student might simply set up a threshold for the number of typical symptoms that a patient must display in order to make a positive diagnosis (e.g., if the patient has 5 out of 9 critical symptoms, classify as borderline personality, as described in the DSM IV, APA, 1994). Note that if this is the manner in which the student makes the diagnosis, it does not matter which of the typical features a patient does not have (e.g., suicidal behavior or fear of abandonment). Furthermore, suppose the patient has an extra feature. If the student is simply matching the patient against a set of typical symptoms, it would not matter whether the extra feature is "chronic substance use" or "has six toes".

Alternatively, the student might learn the causal structure of the symptoms. For example, she might theorize that recurrent suicide attempts are caused by fear of abandonment. Consider the impact of both missing and extra features in this case. First, if a patient that fears abandonment does not currently display suicidal behavior, the student may be no less likely to make a positive diagnosis than if the patient did display suicidal behavior, because this symptom may show up later. In contrast, if the patient makes suicide attempts in the absence of fear of abandonment, the clinician might be more likely to make an alternative diagnose (e.g., major depression rather than borderline personality). Second, it seems that having some extra features would have more of an effect on the diagnosis than others. For example, chronic substance use may matter more as an extra feature than having six toes, because chronic substance use is causally connected to an alternative diagnosis (e.g., major depression) whereas having six toes is not.

Existing knowledge-based systems cannot account for such differences in the important of features when the features both belong to the same relational structure (as is the case in the example of the clinical diagnosis). The presence or absence of relational predicates is important for existing systems. We suggest that the differential position of arguments within the predicate also plays a role in categorization. The goal of the current experiment is to examine how the causal status of features affects their importance in categorization.

In this experiment, we manipulated the position of symptoms in a causal chain (e.g., X causes Y, and Y causes Z), such that a symptom was either a cause or an effect of other symptoms in the causal chain. We also manipulated whether the symptoms in the causal chain were associated with a target disease or an alternative disease. We refer to absent symptoms associated with a target disease as missing features, and existing symptoms associated with an alternative disease as extra features.

We examined the effect of a missing and extra symptom on diagnosis of a target disease as a function of the causal status (i.e., whether it was a cause or an effect of other symptoms) of that missing or extra symptom. We predicted that a symptom that was a cause of other symptoms would have greater impact on categorization of new patients than a symptom that was an effect in a causal chain, both when the cause was a missing feature and when it was an extra feature.

Method

Participants

Sixty undergraduate students at the University of Louisville participated in this study either in partial fulfillment of requirements of an Introductory Psychology course or for payment of $1.00. Students were randomly assigned to each experimental condition. Participation involved approximately 10 minutes.

Procedure

Each participant was given six problems requiring them to judge the likelihood that a patient with some set of symptoms has a target disease. Each problem consisted of a description of the symptoms associated with the target disease and a list of a patient's symptoms. Specifically, for each of the six problems, subjects were first told that there were three symptoms associated with the target disease 75% of the time. For some of the problems, subjects also received information about causal relationships between symptoms associated with the target disease or symptoms

Table 1. Likelihood judgment as a function of problem type.

		Likelihood Judgment	
		Information about Missing Symptom	
Problem Type	Background Knowledge	Unavailable	Explicit
Missing-Effect	A caused B, B caused C	P(X | A, B, (C))	P(X | A, B, no-C)
Missing-Cause	C caused B, B caused A	P(X | A, B, (C))	P(X | A, B, no-C)
Missing-None	none	P(X | A, B, (C))	P(X | A, B, no-C)
Extra-Effect	F caused E, E caused D	P(X | A, B, C, D)	P(X | A, B, C, D)
Extra-Cause	D caused E, E caused F	P(X | A, B, C, D)	P(X | A, B, C, D)
Extra-None	none	P(X | A, B, C, D)	P(X | A, B, C, D)

associated with an alternative disease. Subjects were then presented with a description of a particular patient's symptoms and asked to judge the likelihood, using a 0-100 scale, that the patient had the target disease. For example, one of the problems was the following:

> Scientists have found that symptoms A, B and C are associated with Disease Xeno 75% of the time. In addition, scientists have found that symptom A causes symptom B, and symptom B causes symptom C. Suppose Susan has symptoms A and B, but we do not have any information about whether she has symptom C or not. How likely is it that Susan has Xeno?

Design

Three variables were manipulated in this experiment: Missing Information Condition, which refers to the manner in which information about missing symptoms was given (Unavailable or Explicit); Symptom Type, which refers to the manner in which a patient description differed from the abstract description of the disease (Missing or Extra); and Knowledge Type, which refers to the manner in which missing or extra symptoms were causally connected to other features (Cause, Effect, or None).

Missing Information Condition varied between subjects. In the Unavailable condition, subjects were told that there was no available information about the missing symptom (as in the example above). In the Explicit condition, subjects were told that the patient definitely did not have a particular symptom known to be associated with the target disease.

Both Symptom Type (Missing or Extra) and Knowledge Type (Cause, Effect or None) varied within-subject to form a set of six problems. There were three problems in which the description of the patient was missing one of the symptoms associated with the target disease: (1) Missing-Cause, in which the missing symptom was the cause in a chain of symptoms associated with the target disease; (2) Missing-Effect, in which the missing symptom was the effect in a chain of symptoms associated with the target disease; and (3) Missing-None, in which no background knowledge about causal relationships between symptoms associated with the target disease was provided. The example in the procedure section is a Missing-Effect problem. There were three problems in which the description of the patient contained a symptom that was known to be associated with an alternative disease: (1) Extra-Cause, in which the extra symptom was the cause in a chain of symptoms associated with the alternative disease; (2) Extra-Effect, in which the extra symptom was the effect in a chain of symptoms associated with the alternative disease; and (3) Extra-None, in which no background knowledge about causal relations between symptoms associated with the alternative disease was provided. In addition, all participants were told that a person cannot have the target disease and the alternative disease at the same time.

Materials

Six problems were created for each of the Missing Information conditions (Unavailable and Explicit) by crossing the two levels of Symptom Type with the three levels of Knowledge Type. Likelihood judgments required for each of the six problem types are given in Table 1.

In Table 1, the three symptoms associated with the target disease are referred to as A, B and C, and the three symptoms associated with the alternative disease are referred to as D, E and F. The missing symptom, which is associated with the target disease but is not contained in the description of the patient, is C. The extra feature, which is associated with an alternative disease and is possessed by the patient, is D. The

Unavailable condition is denoted "(C)", to indicate that because no knowledge about C is available, the patient may or may not have symptom C. The Explicit condition is denoted "no-C", to indicate that it is known that C is not present. The likelihood judgment required for each question type is indicated by a probability statement (e.g., "P(X | A, B, C, D)" should be read "What is the likelihood that the patient has Disease X, given that the patient has symptoms A, B, C and D?"). The six problems given to subjects in the Unavailable and the Explicit conditions are listed under the heading "Information about Missing Symptom". Of the six problems given to each subject, three involved a missing feature and three involved an extra feature, one for each of the background knowledge conditions (Cause, Effect, and None). Although the letters A-F were used in Table 1, in the actual stimulus materials, different sets of letters were used for each problem to label the symptoms and different fictitious names were used for the diseases. Subjects were told that the names of the symptoms were difficult to pronounce, so letters of the alphabet were used to label the symptoms. Consecutive letters of the alphabet were not used to label the three symptoms associated with each disease.

Results

Average likelihood ratings for each of the six problem types in the Explicit condition are given in Figure 1, and average likelihood ratings for the six problems in the Unavailable condition are given in Figure 2. The impact of missing and extra features varied as a function of their status in the causal chains. We will first discuss results from the missing features followed by results from the extra features.

Analysis of categorization with missing features

As shown in Figure 1, subjects in the Explicit condition were least likely to believe that a patient had a disease if the patient was missing a feature that was a cause in the chain of symptoms associated with the target disease. Subjects rated the likelihood that a patient had the target disease 16.4% lower for the Missing-Cause problem than for the Missing-Effect problem, $t(29) = 3.17$, $p < .01$; and 5.2% lower for the Missing-Cause than for the Missing-None, $p > .10$. The missing cause led to a lower likelihood of diagnosis as the target disease than the missing effect. In addition, subjects rated the likelihood that a patient had the target disease 11.2% higher for the Missing-Effect problem than for the Missing-None problem, $t(29) = 3.17$, $p < .01$. The missing effect was even less important for target disease categorization than the symptom that did not have any causal relationship with other symptoms.

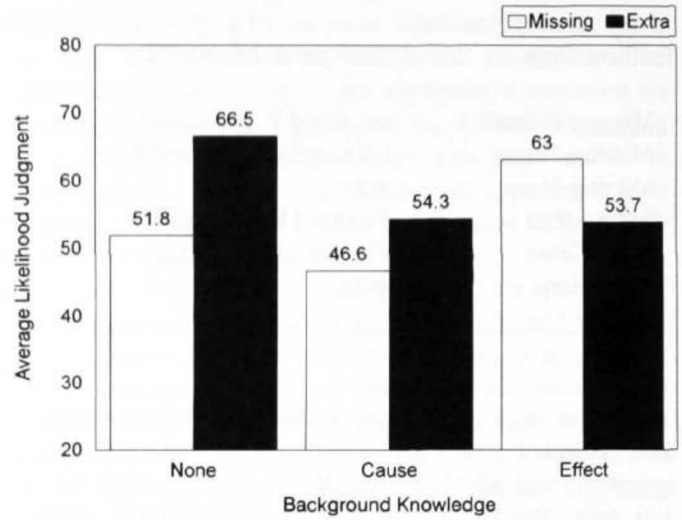

Figure 1: Average likelihood judgment as a function of background knowledge and symptom type for Explicit condition.

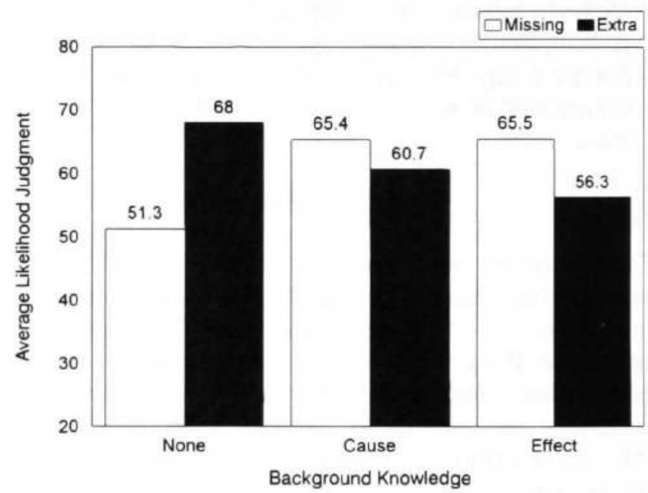

Figure 2. Average likelihood judgment as a function of background knowledge and symptom type for Unavailable condition.

Interestingly, this significant decrease in belief that a patient had the target disease as a result of a missing causal feature did not occur in the Unavailable condition, in which subjects were told that no information was available about the presence of symptom C for the patient. As shown in Figure 2, for the Unavailable condition, the mean likelihood judgment for the Missing-Cause problem was 14.1% greater than for the Missing-None problem, $t(29) = 4.54$, $p < .01$.

Judgments for the Missing-Cause problem did not differ from judgments for the Missing-Effect problem, $p > .10$. Again, as in the Explicit condition, judgments for the Missing-Effect problem were still 14.2% greater than for the Missing-None, $t(29) = 4.76$, $p < .01$.

To summarize results with a missing symptom, when the presence of the symptom was not explicitly denied (i.e., in the Unavailable condition), and therefore subjects could make inferences about the presence or absence of the symptom, causal background knowledge led to an increase in target disease categorization. People could infer the presence of the missing symptom if it was causally connected with other symptoms, either as a cause or an effect. However, if the presence of the symptom was explicitly denied, causal background knowledge had differential impact depending on whether the symptom was a cause or an effect. Missing a cause led to a lower likelihood of being diagnosed as the target disease than missing an effect.

Analysis of categorization with extra features

As with missing features, the importance of extra features was affected by causal background knowledge about the extra feature. Unlike with missing features, however, the particular causal status of an extra feature in an alternative disease did not matter. That is, in both the Explicit and the Unavailable conditions, the cause of an alternative disease and the effect of an alternative disease were equally likely to lower the subjects' ratings, $p > 0.10$. However, both the cause and the effect of an alternative disease mattered more than an extra feature that had no causal relationship with features in the alternative disease, $p < .05$.

Discussion

We investigated the impact of the absence of characteristic features and the presence of extra features on categorization as a function of the causal status of the feature. We found that weight given to extra and missing features is influenced by the specific role they play within the causal structure. Within a single causal structure, causes are believed to be more important in categorization than effects. People's categorizations reflect their belief in the existence of fundamental features shared exclusively by members of a category (Medin & Ortony, 1988). Indeed, many researchers in judgment and reasoning have found that inference proceeds from cause to effect rather than effect to cause (Einhorn & Hogarth, 1986; Tversky & Kahneman, 1982). Our results suggest that this is not always the case: the presence of a cause is more crucial when confirming target category membership, but the causal status of a feature in an alternative category was not important when disconfirming target category membership. When a cause known to be associated with a target category is missing, other associated features are given less weight as evidence that the example is a member than when that same feature is an effect of other features. In contrast, the presence of a cause associated with an alternative category is given just as much weight as counter evidence for target category membership as the presence of an effect.

A follow-up study suggests an explanation for the interaction between types of categories and types of features observed in previous categorization research (Barton & Komatsu, 1989; Gelman, 1988; Keil & Batterman, 1984). These studies have shown that changing a feature of an object has a different effect on belief about the identity of the object depending on both the kind of category to which the object belongs (i.e., natural kinds or artifacts) and the type of feature which was changed (i.e., "molecular", or having to do with what the object is made of, or "functional", having to do with what the object is used for). In these studies, molecular features (e.g., made of wood) are more important for identification as a member of a natural kind category than as a member of an artifact category. In contrast, functional features (e.g., used for writing, gives milk) are more important for artifact category membership than natural kind category membership.

These results can be explained within the current framework. Molecular features in natural kinds serve as causes for other features, including functional properties (e.g., goats' DNA structure allows them to give milk). It is the causal status of molecular features of natural kinds, rather than the fact that they are molecular features per se, which is responsible for their importance to natural kind category membership. In artifacts, molecular features do not necessarily determine the function of categories, whereas functional properties typically do determine the molecular features of the artifact (e.g., what it is made of). In our follow-up study, we pitted causal status of features against type of category (natural kind v. artifact) and type of feature (molecular v. functional) and confirmed our hypothesis. That is, when a molecular structure was described as an effect of a functional property in natural kinds, the functional property was considered to be more important for categorization than the molecular structure, reversing the previous findings. Similarly, when a functional property was described as an effect of a molecular property in artifacts, the effect found in previous studies was again reversed.

Finally, we speculate that this mechanism may be responsible for differences between experts and novices in categorization. Experts and novices may know the same features, but these features may reside in different causal structures. By changing the causal status of features, their importance for categorization also changes. In conclusion, the results of the current study suggest a mechanism for determining the importance of features in categorization.

References

Ahn, W., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example, *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 391-412.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders*. Fourth edition,Washington, E. C.: American Psychiatric Association.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*, 211-227.

Barton, M. E., & Komatsu, L. K. (1989) Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research, 18*, 433- 447.

DeJong, G. F., & Mooney, R. J. (1986). Explanation-based learning: An alternative view. *Machine Learning, 1*, 145-176.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 3-19.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning*. Cambridge: Cambridge University Press.

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

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

Lassaline, M. E. (1995). Structural alignment in induction and similarity. Manuscript submitted for publication.

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

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Medin, D. L., & Ortony, A. (1988). Psychological essentialism, In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-196), NY: Cambridge University Press.

Mitchell, T. M., Keller, R. M., & Kedar-Cabelli, S. T. (1986). Explanation-based generalization: A unifying view. *Machine Learning, 1*, 47-80.

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

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 416-432.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15*, 346-378.

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

Tversky, A., & Kahneman, D. (1982). In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Wisniewski, E. J., & Medin, D. L. (1991). Harpoons and Long Sticks: The interaction of theory and similarity in rule induction. In D. Fisher & P. Langley (Eds.) *Concept formation: Knowledge and experience in unsupervised learning*, p. 237-278, Morgan-Kaufman.