Lee, H. S., & Holyoak, K. J. (2007). Causal models guide analogical inference. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the Twenty-ninth Annual Conference of the Cognitive Science Society* (pp. 1205-1210). Austin, TX: Cognitive Science Society.

Causal Models Guide Analogical Inference

Hee Seung Lee (heeseung@ucla.edu)
Keith J. Holyoak (holyoak@lifesci.ucla.edu)
Department of Psychology, University of California, Los Angeles
Los Angeles, CA 90095 USA

Abstract

Computational models of analogical inference have assumed that inferences about the target are generated solely by "copying with substitution and generation" from the source, guided by a mapping based on similarity and parallel structure. In contrast, work in philosophy of science has stressed that analogical inference is based on causal models of the source and target. In two experiments, we showed that reducing analogical overlap by eliminating a matching higher-order relation (a preventive cause) from the target *increased* inductive strength even though it *decreased* similarity of the analogs. Analogical inference appears to be mediated by building and then "running" a causal model.

Keywords: analogical inference; causal model; similarity; inductive strength; preventive cause; generative cause.

Introduction

In everyday life, people often face uncertainty, which permeates such situations as meeting a new person, solving a novel math problem, or resolving a foreign policy crisis. To reduce uncertainty about a novel *target* situation, people frequently make analogical inferences based on similar *source* situations they experienced in the past. When a source includes properties that the target initially lacks, any of these missing properties in the target can become a candidate for analogical inference. Yet people do not draw all possible inferences. How, then, do people make analogical inferences *selectively*?

Similarity and structure as inductive constraints

People are more likely to draw strong analogical inferences when they perceive a source and a target to be similar. According to the contrast model of similarity (Tversky, 1977), common properties tend to increase perceived similarity of two concepts whereas differences tend to reduce the perceived similarity. By manipulating the number of shared properties of the source and target, Lassaline (1996) demonstrated that both similarity and inductive strength increased with addition of shared properties.

Although overlap of properties influences analogical inference, formal models have placed major emphasis on the role of *structural* parallels between relations in the source and target. The importance of relations in analogical inference provided the basis for Gentner's (1983) structure mapping theory, which has been implemented in a computational model, SME (Structure Mapping Engine; Falkenhainer, Forbus, & Gentner, 1989). Gentner distinguished between first-order relations, which take objects as arguments (e.g., "The dog is bigger than the cat") and higher-order relations, which include propositions as arguments (e.g., "Because the dog is bigger than the cat, the cat ran away from the dog"). Gentner argued that higher-order relations, such as "cause", are more important for analogical inference than first-order relations. The priority of higher-order relations is due to what she termed the *systematicity* principle, which hypothesizes a preference for inferences based on predicates having many mutually interconnecting relationships. An entity embedded within corresponding higher-order relations has more matched structures, and hence will receive a higher "match" score in SME, than would an entity embedded in first-order relations, or an isolated entity. SME therefore predicts that deeply hierarchical information is especially likely to be transferred to the target.

Lassaline (1996) demonstrated that when a causal relation in the source is unmapped, and the cause property is shared by the source and target, then people are likely to infer the corresponding effect in the target. For example:

Animal A has properties X, W, and Z.
For Animal A, X causes Z.
Animal B has X, W, and Y.
Therefore, Animal B also has Z.

Here property X is the cause property shared by Animal A and Animal B, leading to the inference that effect Z found in Animal A will also be present in the target, Animal B.

Using a similar paradigm, Rehder (2006) showed that category-based generalizations are preferentially guided by causal relations, such that standard effects of typicality, diversity, and of similarity itself are eliminated when a causal relation is present. However, in Rehder's experiments the single causal relation, when present, was also the sole higher-order relation. Given this confounding, his findings are consistent with Gentner's (1983) systematicity principle.

In addition to SME, other computational models of analogical inference, such as ACME (Analogical Constraint Mapping Engine; Holyoak & Thagard, 1989) and LISA (Learning and Inference with Schemas and Analogies; Hummel & Holyoak, 1997, 2003) incorporate similar relation-based constraints. All these models generate candidate inferences using variants of a procedure termed "copy with substitution and generation", or CWSG (Holyoak, Novick, & Melz, 1994), in which inferences about the target are constructed directly from the mapping between source and target relations.

Causal models as inference engines

The systematicity principle explicitly eschews any role for the *meaning* of relations in guiding analogical reasoning: "the processing mechanism that selects the initial candidate set of predicates to map attends only to the structure of the knowledge representations for the two analogs, and not to the content" (Gentner, 1983, p. 165). Nonetheless, it is striking that virtually all examples of "higher-order relations" mentioned in the psychological literature involve the relation "cause". In philosophy of science, Hesse (1966) was the first to emphasize the role of causal relations in analogical inference. Lassaline (1996) found that people make stronger analogical inferences based on the higher-order relation, "cause" than based on a non-causal relation, "temporally prior to." In general, the same syntactic "order" of relations may not always yield the same degree of inductive strength about the target property to be inferred. The ultimate goal of analogical inference is to know whether or not the outcome will be present in the target, not to just blindly copy the presence of the outcome in the source domain based on semantic or structural correspondences between the source and target. According to Holyoak (1985), "the goal is a *reason* for the solution plan; the resources *enable* it; the constraints *prevent* alternative plans; and the outcome is the *result* of executing the solution plan" (p. 70). Thus the most important consideration in analogical inference is *how* each factor influences the outcome in the source domain; hence causal relations will play a central role.

Although some computational models of analogical inference postulate a special role for causal relations (Hummel & Holyoak, 2003), models of analogy have not been closely connected to models of human causal reasoning. In the present paper we explore the possibility that people may use *causal models* to guide analogical inference. Graphical representations of causal links have been used extensively in work on causal reasoning in philosophy (Reichenbach, 1956; Salmon, 1984), artificial intelligence (Pearl, 1988), and psychology (Waldmann & Holyoak, 1992). Causal models postulate that causes can be either *generative* (making the effect happen) or *preventive* (stopping the effect from happening; see Cheng, 1997; Griffiths & Tenenbaum, 2005). A generative cause increases the probability of an outcome whereas a preventive cause decreases the probability of the outcome. Because generative and preventive causes exert their power in opposite directions, the distinction between generative and preventive causes is crucial in predicting the outcome. Previous studies of category-based induction (e.g., Rehder, 2006) have not employed preventive causes.

In philosophy of science, Bartha (in press) has recently extended Hesse's (1966) work on the role of causal models in analogy. Bartha distinguished between *contributing* causes (generative) and *counteracting* causes (preventive) in assessing the normative strength of arguments by analogy. He pointed out that the *absence* of a correspondence in the target for a counteracting cause might actually strengthen an argument from analogy. In contrast, previous computational accounts based solely on CWSG algorithms predict that any causal relation shared by the source and target, regardless of causal direction, can only increase the strength of an analogical inference.

Figure 1 shows how people might reach different inductive conclusions about the probability of a possible target property based on the presence or absence of a preventive cause in the target. The source has four properties: G_1, G_2, P_1, and E. Properties, G_1 and G_2 are generative causes that increase the probability of outcome E occurring, whereas property P_1 is a preventive cause that decreases the probability of outcome E occurring. Target 1 has three properties, G_1, G_2, and P_1, whereas Target 2 has only two properties, G_1 and G_2. Given the same source, which of Target 1 and Target 2 will yield a stronger analogical inference about the presence of effect E?

All extant computational models of analogical inference predict that people will draw a stronger analogical inference about Target 1 than Target 2, because Target 1 shares more properties with the source than does Target 2. Moreover, Target 1 shares three higher-order relations with the source, whereas Target 2 shares only two higher-order relations. Because both similarity and structural approaches focus solely on correspondences of properties and relations between the source and target, they predict that Target 1will yield a stronger analogical inference than will Target 2.

However, if people use causal models in analogical inference, as suggested by Bartha (in press), then Target 2 will actually yield *greater* inductive strength than Target 1. Target 1 is more similar to the source than is Target 2; however, Target 1 includes the preventive cause, P_1, and this preventive cause will *decrease* the probability of outcome E. In contrast, even though Target 2 is less similar to the source than is Target 1, because it includes only generative causes, G_1 and G_2, and not the preventive cause, P_1, the probability of outcome E will be increased. We performed two experiments to test these competing predictions.

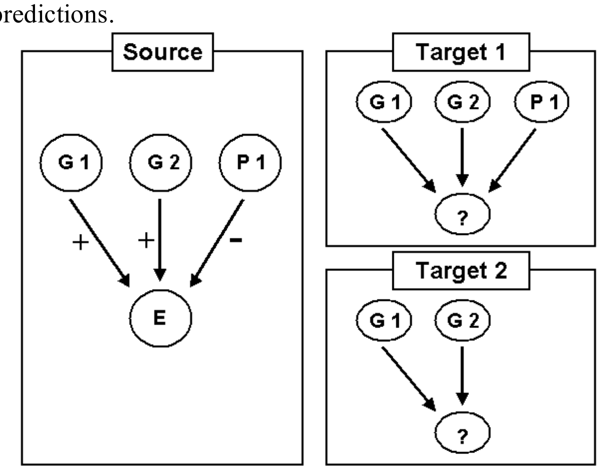

Figure 1: Example of use of causal models in analogical inference. G, P, and E represent a generative cause, preventive cause, and effect, respectively.

Experiment 1

Experiment 1 investigated the effect of a preventive causal relation on analogical inference, using a paradigm adapted from that of Lassaline (1996).

Method

Participants Forty-two undergraduate students at the University of California, Los Angeles (UCLA) participated in the experiment for course credit. Half of the participants provided inductive strength judgments, and the other half provided similarity ratings.

Design and Materials Participants read a description of two imaginary animals referred to as *Animal A* and *Animal B,* and then they evaluated either inductive strength of an analogical inference or similarity of the two animals. Across the arguments, the number of shared properties, and presence or absence of a preventive relation connecting one shared property to a non-shared property, were manipulated. Three types of arguments were created: N_1N_2P, N_1N_2, and N_1P, where N and P represent neutral and preventive properties, respectively. The N_1N_2P argument includes two shared neutral properties and one shared preventive property; the N_1N_2 argument includes two shared neutral properties without the preventive property; and the N_1P argument includes one shared neutral property and one shared preventive property. All three types of arguments involved the same source analog, which had four properties, one of which was stated to prevent another. An example of argument type N_1N_2P is the following:

> Animal A has dry flaky skin, muscular forearms, a weak immune system, and blocked oil glands.
> For animal A, a weak immune system PREVENTS blocked oil glands.
> Animal B has dry flaky skin, muscular forearms, and a weak immune system.
> Therefore, animal B also has blocked oil glands.

For similarity ratings, the same argument lists were used, but the argument included only the premises without the conclusion sentence.

Nine property lists were created and three argument types were created for each property list, resulting in 27 items altogether. Of the total of 27 items created, nine items were used to create a booklet for each participant, three of each argument type (N_1N_2P, N_1N_2, and N_1P). The larger pool of items served to avoid repeated use of the same property lists for an individual participant, as only one type of argument per list was selected for each participant, generating three sets. Within each set, the order of items was randomized for each participant.

Procedure Participants were tested individually in a small testing room. Instructions and experimental trials were self-paced and there was no time limit. Both groups of participants (similarity rating and inductive strength judgment groups) were given a booklet that included instructions and nine arguments. Participants were instructed that they were to assume all the information given in the descriptions is true. Each participant judged either how likely a conclusion would be true, or how similar the pairs of animals were, based on the information given in the description. For the group making inductive strength judgments, the task after reading descriptions of Animal A and Animal B (the premise statements) was to judge how likely Animal B has a certain property (the conclusion statement). In making their judgments, they were asked to imagine there were 100 examples of Animal B, and to estimate how many out of these 100 cases would have the property stated as the conclusion, assigning a number between 0 and 100 for each item. For the group making similarity ratings, participants were given only premise statements with descriptions of the two animals, not a conclusion statement. They evaluated how similar Animal A and Animal B were based on the descriptions they read. For each description of two animals, a similarity rating scale from 0 to 10 was provided. Under the numbers 0 and 10, the words *totally different* and *identical* were written, respectively. Participants were asked to try to use the entire scale, but to feel free to use any number as often as they felt it was appropriate.

Results and Discussion

Similarity ratings and inductive strength judgments were analyzed separately. For each dependent measure, a two-way analysis of variance (ANOVA) was performed with the three argument types (N_1N_2P, N_1N_2, and N_1P) as a within-subjects variable and the three sets as a between-subjects counterbalancing variable. The results for both similarity ratings and inductive strength judgments are shown in Figure 2.

The ANOVA on similarity ratings revealed a main effect of argument type, $F(2, 36) = 16.79$, $p < .001$. N_1N_2P arguments were evaluated as having the highest similarity of the three argument types. N_1N_2P arguments were rated as having higher similarity than either N_1N_2 arguments, $F(1, 18) = 21.38$, $p <.001$, or N_1P arguments, $F(1, 18) = 28.65$, p

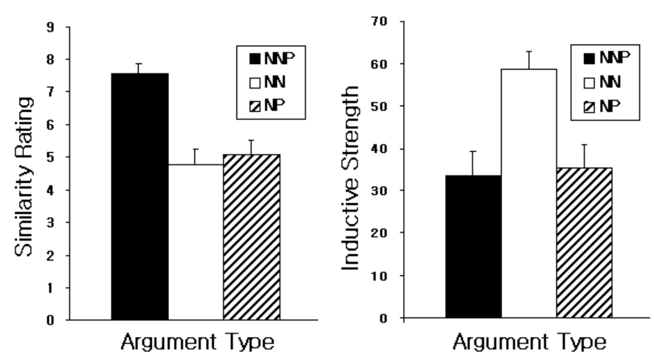

Figure 2: Mean similarity ratings (left) and mean inductive strength judgments (right) for each argument type in Experiment 1.

< .001. The difference between N_1N_2 and N_1P arguments, however, was not significant, $F < 1$. The mean similarity ratings of N_1N_2P, N_1N_2, and N_1P arguments were 7.58, 4.77, and 5.09, respectively. There was no significant effect of set, nor was there an interaction between argument type and set.

The ANOVA on inductive strength judgments also revealed a main effect of argument type, $F(2, 36) = 8.44$, $p < .01$, but showed a different pattern from that obtained for similarity ratings. As predicted, N_1N_2 arguments were rated as having higher inductive strength than either N_1N_2P arguments, $F(1, 18) = 9.19$, $p < .01$, or N_1P arguments, $F(1, 18) = 11.08$, $p < .01$. The difference between N_1N_2P and N_1P arguments was not significant, $F < 1$. Mean inductive strength judgments of N_1N_2P, N_1N_2, and N_1P arguments were 33.60, 58.57, and 35.32, respectively. There was no effect of item set, nor was there an interaction between argument type and set.

To summarize, similarity ratings increased with the number of shared properties, but inductive strength judgments were reduced by the presence of a shared preventive property. These results suggest that people use causal models to guide analogical inferences.

Experiment 2

A possible problem with the design of Experiment 1 is that the preventive cause may have been interpreted as deterministic, making the source seem implausible. In fact, some of the participants pointed out that since the premise stated that one property prevents another, to have both attributes (preventive cause and the effect it purports to prevent) co-occur was not reasonable (e.g., for Animal A, property P was said to prevent property E, but Animal A in fact had both properties, P and E). Although people may interpret the term "prevent" as deterministic by default, causal models are designed to represent probabilistic causes. Accordingly, in Experiment 2 the phrase "tends to" was introduced in order to make all causes appear probabilistic. Also, generative causes were introduced as well as preventive causes to allow direct comparison between the two types of causal relations.

Method

Participants Sixty undergraduate UCLA students received course credit for participating in the experiment. Half of the participants provided inductive strength judgments, and the other half provided similarity ratings.

Design There were two independent variables. The first independent variable was presence versus absence of the preventive property. In the generative-only condition, the source did not have the preventive property, but instead had three generative properties. In the generative + preventive condition, the source had two generative properties and one preventive property. All the source properties were causally related to the effect property, E. Each generative property tended to produce E whereas the preventive property tended to prevent E. The second independent variable was argument type. The generative + preventive condition included three argument types: G_1G_2P, G_1G_2, and G_1P. In the generative-only condition, because there was no preventive property, only two argument types were possible: $G_1G_2G_3$ and G_1G_2 (counterbalanced with G_1G_3). As in Experiment 1, participants provided either similarity ratings or inductive strength judgments.

A causal-model analysis predicts that in the generative-only condition, similarity ratings and inductive strength judgments will follow the same pattern: the $G_1G_2G_3$ argument will have higher perceived similarity and higher inductive strength than the G_1G_2 argument. However, in the generative + preventive condition, similarity ratings and inductive strength judgments will follow different patterns. The G_1G_2P argument will have higher perceived similarity than the G_1G_2 and G_1P arguments, but the G_1G_2 argument will have higher inductive strength than the G_1G_2P or G_1P arguments. In addition, the G_1G_2P argument will have higher inductive strength than the G_1P argument. In contrast, all extant computational models of analogy predict that similarity and inductive strength will be positively correlated regardless of the content of the causal relations.

Materials and Procedure Each participant was given a booklet consisting of six descriptions of animal pairs, referred to as Animal A and Animal B. Three of the six items were $G_1G_2G_3$, G_1G_2, G_1G_3 arguments (generative-only condition), and the other three items were G_1G_2P, G_1G_2 and G_1P arguments (generative + preventive condition). Six property lists were created and six sets were constructed by counterbalancing which property list was assigned to each condition and argument type. Within each set, the order of items was randomized for each participant. The procedure was the same as that of Experiment 1.

Results and Discussion

Similarity ratings and inductive strength judgments were analyzed separately. For the generative-only condition, argument types G_1G_2 and G_1G_3 were literally the same (differing only by counterbalancing), so these data were collapsed together for both similarity ratings and inductive strength judgments. The results of similarity ratings are shown in Figure 3 (top). In the generative-only condition, the mean similarity ratings for arguments $G_1G_2G_3$ and G_1G_2 were reliably different, $t(29) = 4.44$, $p < .001$, such that perceived similarity increased by 1.55 points from two shared attributes to three shared attributes. The mean similarity ratings of arguments $G_1G_2G_3$ and G_1G_2 were 8.17 and 6.62, respectively. In the generative + preventive condition, the mean similarity ratings showed a similar pattern to that observed in the generative-only condition. A one-way ANOVA was performed to examine the differences among the three argument types, G_1G_2P, G_1G_2, and G_1P. This ANOVA revealed a significant effect of argument type, $F(2, 58) = 29.79$, $p < .001$, such that perceived similarity ratings increased from two shared properties to three shared properties. Also, even though arguments G_1G_2, and G_1P have the same number of shared properties (two), G_1G_2 arguments were rated as having higher similarity than G_1P arguments, $t(29) = 3.92$, $p < .001$. The mean similarity ratings of argument types G_1G_2P, G_1G_2, and G_1P were 7.87, 5.73, and 3.90, respectively.

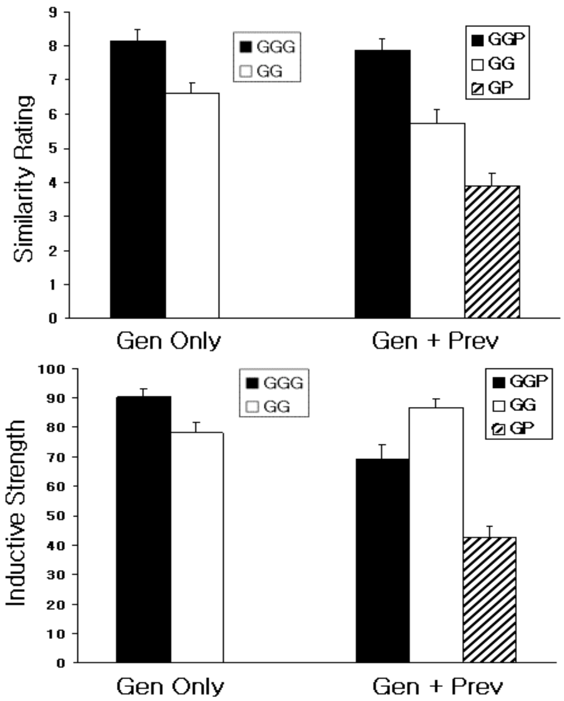

Figure 3: Mean similarity ratings (top) and mean inductive strength judgments (bottom) for each argument type in the generative-only and generative + preventive conditions of Experiment 2.

The results for inductive strength judgments are shown in Figure 3 (bottom). For inductive strength judgments, the generative-only condition and the generative + preventive condition showed different patterns. In the generative-only condition, the mean inductive strength judgments for arguments $G_1G_2G_3$ and G_1G_2 differed, $t(29) = 3.29$, $p < .01$, such that inductive strength increased by 12 points from two shared attributes to three shared attributes. The mean inductive strength judgments of arguments $G_1G_2G_3$ and G_1G_2 were 90.27 and 78.25, respectively. However, in the generative + preventive condition, the results for inductive strength judgments showed a different pattern. A one-way ANOVA was performed to examine the differences among the three argument types, G_1G_2P, G_1G_2, and G_1P. This analysis revealed a significant effect of argument type, $F(2, 58) = 37.03$, $p < .001$. Unlike the pattern for similarity ratings, argument G_1G_2 was rated as having higher inductive strength than either argument G_1G_2P, $t(29) = 4.87$, $p < .001$, or G_1P, $t(29) = 8.13$, $p < .001$. Also, argument G_1G_2P was rated as having higher inductive strength than argument G_1P, $t(29) = 3.84$, $p < .01$. The mean inductive strength judgments for arguments G_1G_2P, G_1G_2, and G_1P were 69.4, 86.6, and 42.5, respectively.

As in Experiment 1, similarity ratings increased with the addition of shared attributes between the source and target in both the generative-only and generative + preventive condition. However, in the generative + preventive condition, argument G_1G_2 was rated to have higher similarity than argument G_1P even though the number of shared attributes was the same. One possible explanation of this difference is that people may have sometimes made use of a causal model in making similarity comparisons. In argument G_1G_2, because there are only generative factors, people may have considered effect E to be probable, and therefore inferred that the target would actually share three properties with the source: G_1, G_2, and inferred outcome E. In contrast, because argument G_1P includes a preventive property, people may have considered the probability of effect E to be low, thus inferring that the target would have only two shared attributes: G_1 and P.

In the generative-only condition, the argument $G_1G_2G_3$ had higher inductive strength than did the argument G_1G_2, the same pattern as for similarity ratings. This result confirms that people consider the number of generative causes when inferring the presence of the outcome in the target. However, in the generative + preventive condition, the argument G_1G_2P had *lower* inductive strength than the argument G_1G_2. This result suggests that people do not simply map the source and target and then make inferences by a CWSG procedure. If people just focused on the number of correspondences between the source and the target, as the similarity and structural views assume, the argument G_1G_2P should have yielded higher inductive strength than the argument G_1G_2.

Not surprisingly, the argument G_1G_2P had higher inductive strength than the argument G_1P. Thus even when a preventive cause is present, people also consider the number of generative causes.

General Discussion

In both of the experiments reported here, inductive strength of analogical arguments increased with number of shared generative causes and *decreased* with the presence of a shared preventive cause. Yet when the source included both a preventive cause and the outcome, presence of the preventive cause in the target necessarily increases the overall correspondence between the source and target (and indeed, yielded higher rated similarity of the analogs in both of our experiments). These findings cannot be explained by the systematicity principle (Gentner, 1983), nor by any computational model of analogical inference that relies solely on a CWSG procedure (e.g., SME, Falkenhainer et al., 1989; ACME, Holyoak & Thagard, 1989; or LISA, Hummel & Holyoak, 2003).

In accord with the recent proposal of Bartha (in press), the present experimental findings suggest that people use causal models when they draw analogical inferences. People are likely to first evaluate whether the causal relations in the source are generative or preventive. When mapped to the target, the resulting causal model then provides the basis for inferring the likelihood of a corresponding effect in the target. Presence of a generative cause in the target increases the probability of occurrence of the effect, whereas presence of a preventive cause decreases the probability of the effect. As a consequence, presence or absence of a preventive cause in the target has different effects on perceived similarity versus inductive strength: when the target includes the preventive cause, perceived similarity increases, but inductive strength decreases. The absence of a preventive cause in the target increases net positive causal power and yields a stronger analogical inference.

Researchers in the area of analogy have long acknowledged the importance of causal relations in analogical reasoning (Winston, 1980), and some work has focused directly on the role of pragmatic factors in guiding analogical mapping and inference (e.g., Holyoak, 1985; Spellman & Holyoak, 1996). Nonetheless, little effort has been devoted to building models of analogy that actually incorporate the basic elements of causal models (e.g., Cheng, 1997; Pearl, 1988; Griffiths & Tenenbaum, 2005; Waldmann & Holyoak, 1992). What is most surprising about the present demonstration that people treat generative and preventive causes differently in evaluating analogical inferences is not the finding itself (arguably little more than common sense), but the fact that it challenges all extant computational models of analogical inference.

If theories of analogy can benefit from work on causal models, the field of analogy may have much to offer in return. In particular, analogical inference provides a possible answer to a central question that looms in current work on causal models, namely, how are causal hypotheses first formed? One possible answer is a kind of causal grammar (Griffiths & Tenenbaum, in press); another is the CWSG procedure for building relational structures by analogy. Analogy and causal inference are intricately related (Holland, Holyoak, Nisbett & Thagard, 1986), and a full theory of human induction will need to provide a unified account of both.

Acknowledgments

Preparation of this paper was supported by NSF grant SES-0350920 to KH. We thank Patricia Cheng for helpful discussions.

References

Bartha, P. (in press). *By parallel reasoning: The construction and evaluation of analogical arguments*. Oxford, UK: Oxford University Press.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367-405.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1-63.

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

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334-384.

Griffiths, T. L., & Tenenbaum, J. B. (in press). Two proposals for causal grammars. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford, UK: Oxford University Press.

Hesse, M. (1966). *Models and analogies in science*. Notre Dame, IN: University of Notre Dame Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation (Vol. 19).* New York: Academic Press.

Holyoak, K. J., Novick, L. R., & Melz, E. R. (1994). Component processes in analogical transfer: Mapping, pattern completion, and adaptation. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Analogical connections* (pp. 113-180). Norwood, N.J.: Ablex.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466.

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

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 754-770.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition, 34*, 3-16.

Reichenbach, H. (1956). *The direction of time.* Berkeley: University of California Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology*, *31*, 307-346.

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

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222-236.

Winston, P. (1980). Learning and reasoning by analogy. *Communications of the ACM*, *23*, 689-703.