

# Causal Information as a Constraint on Similarity

Jessica M. Choplin (choplin@psych.ucla.edu)

Patricia W. Cheng (cheng@psych.ucla.edu)

Keith J. Holyoak (holyoak@psych.ucla.edu)

Department of Psychology  
University of California, Los Angeles  
405 Hilgard Ave.  
Los Angeles, CA 90095-1563

## Abstract

Considerable evidence indicates that causal information provides a vital constraint on conceptual representation and coherence. We investigated the role of causal information as a constraint on similarity, exploiting an asymmetry between predictive causal reasoning (given the cause, predict the effect) and diagnostic causal reasoning (given the effect, diagnose the cause). This asymmetry allowed us to isolate the effects of causal understanding from the effects of sharing non-causal features. We found that judged similarity between two objects that are identical except for one feature was affected by whether that feature was a competing potential cause of an effect or an effect of a common cause.

## Causation Constrains Similarity

Any two objects have indefinitely many features in common. For example, Murphy and Medin (1985) pointed out that the number of features that plums and lawn mowers have in common is, in principle, infinite. Both weigh less than 1000 kg, and both are found on earth, in the solar system. Both cannot hear well, have an odor, are used by people, not by elephants, and so on. But despite these shared features people do not generally consider plums and lawnmowers to be similar. Intuitively, the features plums and lawnmowers have in common are not considered important. But what features are important? Why are some features important and not others? There must be criteria for constraining the sheer number of these features (Goodman, 1972; Medin, Goldstone, & Gentner, 1993). Previous researchers (e.g., Murphy & Medin, 1985) have suggested that features will be considered more important when they are diagnostic of causal function and part of a larger explanatory framework. Whether or not a feature is causal serves as a criterion by which people select important features and separate them from unimportant ones (Sloman, Love, & Ahn, 1998).

We propose a new paradigm to investigate the influence of causal knowledge on similarity and categorization. This paradigm provides an optimal way of equating the number of common and distinctive features between two objects across conditions:

Equality is ensured because the stimuli were in fact identical across conditions. We use this paradigm to investigate the influence of causal knowledge as a constraint on similarity.

## Experiment

To avoid confounding the effects of causal understanding with the effects of shared non-causal features, we utilized an asymmetry tested by Waldmann and Holyoak (1992). When a feature is identified as a cause, other potential causes are discounted (Morris & Larrick, 1995). Other potential causes (presented at the same time) are redundant to explain the effect and are assumed not to be causes after all. For example, if one notices a bowling ball moving, knowing that it was kicked is enough to explain the motion. There is no need to postulate another cause. By contrast, when a feature is identified as an effect, other potential effects are not discounted. For example, if one were to kick a bowling ball, the person might move the bowling ball and hurt his or her foot at the same time. The cause (the kick) would have two effects (the moving of the bowling ball and the pain in the foot). One would not assume the ball did not move simply because of the pain in the foot. The significance of this asymmetry for our current project is that it can be used to distinguish the influence of causal understanding from that of mere associations.

Two tasks were used in this experiment: first a training task and subsequently a similarity rating task. The purpose of the training task was to allow participants to learn a competing cause of an effect or an effect of a common cause. It used a two-phase blocking paradigm similar to that employed by Waldmann and Holyoak (1992). We trained participants by showing them the keys presented in Table 1. Half of the participants (predictive condition: identify features that predict an effect) were told that some of the keys could open any safe (target event in predictive condition), and they were asked to identify which features were responsible for this special ability. The other half of the participants (diagnostic condition: diagnose whether features were caused by the carving process) were told that some of the keys were carved by a special carving process (target event in diagnostic condition), and they

were asked to identify whether each feature was caused by the special carving process.

Participants saw two series of keys. In the first series, the keys were all missing a section. In the second series, this section was restored. The restoration provided an enabling condition (provided a location on which a feature could be located) for the existence of a redundant new feature that could be causally related to the target event. In the predictive condition, because old features were sufficient to explain the opening of safes, we predicted that this new, redundant feature would not be thought responsible for opening safes. By contrast, in the diagnostic condition, because a cause can have multiple effects, the new feature could very well be the result of the special carving process. We predicted that the carving process would be assumed to produce the redundant feature. Non-causal features would not produce this asymmetry.

We predicted that this causal knowledge would affect whether features are considered important in a similarity judgment. Consider a similarity judgment between a key that had the restored new feature and one that did not. If the old feature was sufficient to open safes, the key without the restored feature could open safes just as well as the key with this feature, the causal relations would not be altered, and the restored feature would not be important. By contrast, when the restored feature is the result of the carving process, the key with the restored feature received different treatment from the key without it, the causal relations would be altered, and the feature would be important.

## Method

### Participants

The participants were 298 students enrolled in an introductory psychology class at the University of California, Los Angeles, and 106 travelers waiting to board airplanes at Los Angeles International Airport (LAX). The students participated as part of a class exercise. The travelers participated voluntarily. Of these 404 participants, 5 failed to answer critical questions, leaving 399 responses for analysis.

### Materials

As shown in Table 1, we created pictures of keys with features attached. To parallel Waldmann and Holyoak's (1992) disease symptoms, these keys had four features that we manipulated during the training phase as follows. One feature was constantly present (the t-shape on the left, called Cue C for *Constant*) and was not correlated with the target event. Another feature was present half of the time (the b-shape on the middle left, called Cue U for *Uncorrelated*) and also was not correlated with the target event. The L-shape on the middle right (called Cue P for *Predictor*) was a perfect deterministic predictor of the target event. The k-shape on the right (called Cue R for *Redundant*) was not presented at all during the first series and was only

presented in the presence of Cue P during the second series. Within the second series, Cue R was a perfect predictor of the target event but only provided redundant information after Cue P.











1 <sup>st</sup> Series (missing section)	2 <sup>nd</sup> Series (restored section)	Target Event Occurs?
		
		No
		Yes
		No
		Yes

Table 1: Key stimuli. The top key is a blank used as an initial illustration for participants. Cues C (t-shape on the left) and U (b-shape on the middle left) were not correlated with the target event. Cues P (L-shape on the middle right) and R (k-shape on the right) were correlated with the target event, but Cue R only provided redundant information.

We prepared two three-page booklets, one booklet for the predictive condition and another for the diagnostic condition. The first page was used to present the first series of keys and to elicit the causal inferences that participants first made. The second page was used to present the restored feature in the second series of keys and to elicit the causal inferences participants made after this item was restored. The third page was used to collect similarity judgments.

At the top of the first page, participants were told that some of the keys they would see either could open any safe (predictive condition) or were caused by a special carving process (diagnostic condition). It was necessary to explain why Cue R was missing from this first page. To do so, participants in both conditions were told that the keys for the first page were carved from blanks with a missing notch that looked like the first key in the left column of Table 1. In the middle of the booklets' first page a two-column table was used to present the keys. The left-hand column was labeled "Keys" and presented the four keys in the bottom four rows of the left column in Table 1. The right-hand column was used to indicate whether the target event occurred for each key.

In the predictive condition, the right-hand column asked, "Will this key open any safe?" In the diagnostic condition, it asked, "Is this key carved by the special process?" The word "Yes" was used to indicate that the target event occurred and the word "No" was used to

indicate that the target event did not occur. At the bottom of the first page, participants were asked questions on their understanding of the causal relations. In the predictive condition, participants indicated whether they thought each of the three features, Cues C, U, and P respectively, caused safes to open by circling the word “yes” or the word “no.” In the diagnostic condition, they likewise indicated whether they thought the carving process caused each of the three features.

The second page resembled the first, except that the keys on the middle column of Table 1 replaced those on the left. The first key in the middle column of Table 1 replaced the first key in the left-hand column as the blank. The keys in the bottom four rows on the right replaced the keys in the bottom four rows on the left as the keys causally related to the target event. Participants were asked to indicate whether they thought each of the four features, Cues C, U, P, and R, respectively, was causally related to the target event.

The final page was used to collect similarity judgments. Seven-point scales were provided, with “1” labeled “not at all similar,” and “7” labeled “very similar.” Participants compared a key containing all four features to a key that lacked Cue R. In the predictive condition, since Cue P would be sufficient to open safes, judged similarity between the two keys would be high—both keys could open safes. By contrast, in the diagnostic condition, since Cue R was caused by the carving process, judged similarity between the two keys would be low—the keys received different treatments. Participants likewise judged the similarity between a key containing all four features to one that lacked Cue P. This similarity serves as a baseline for evaluating the importance of Cue R, to reduce noise in the analyses.

## Procedure

The introductory psychology students received our materials within a larger packet of materials. The packets were randomly assigned and completed within the classroom. The travelers were approached by the experimenter and asked to participate. After agreeing to participate, participants were randomly given a booklet. Participants were given pencils and instructed to follow the instructions inside the booklet. After participants had finished, the experimenter returned to collect the booklets.

## Results and Discussion

The results of the training task mirrored the basic findings of Waldmann and Holyoak (1992). We divided participants into those who showed a coherent pattern of responses that indicates blocking (judged Cues C, U and R non-causal and Cue P causal) and those who showed other patterns. A blocking pattern of responses was observed for 12.5% of the participants in the predictive condition and only 4.7% of the participants in the diagnostic condition. This difference was

significant,  $X^2(1, n=399) = 6.13, p < .025$ . Only a minority of the participants in both conditions, however, showed this pattern. Questioning participants afterwards revealed that many based their judgments solely on the information provided on the second page, the page on which P and R both predicted the target event. Many thought the second page represented a separate updated scenario. This misunderstanding likely explains why the majority of participants in both conditions judged Cue R to be causal.

To measure the importance of Cue R (relative to Cue P) in similarity judgments, we subtracted judged similarity between a key containing all four features and one that lacked Cue R from judged similarity between a key containing all four features and one that lacked Cue P. Positive scores indicated that Cue P was more important than Cue R and vice versa. As predicted, participants ( $n = 198$ ) in the predictive condition judged the relative importance of Cue P to Cue R ( $M = 1.45$ ) higher than participants ( $n = 201$ ) in the diagnostic condition ( $M = 0.99$ ),  $t(397) = 2.27, p < .025$ .

Because the size of this effect was relatively small and large numbers of participants misunderstood the task, we analyzed the relative importance of Cue P and Cue R conditionalized on causal response patterns. We divided participants into 3 groups: those who showed a coherent pattern of responses that indicates blocking (judged Cues C, U, and R non-causal and Cue P causal), those who showed a coherent pattern of responses that indicates “no blocking” (judged Cues C and U non-causal and Cues P and R causal), and those who showed other patterns. Of the participants in the predictive condition, those who showed blocking ( $n = 22$ ) judged the relative importance of Cue P to Cue R ( $M = 2.64$ ) greater than those who showed “no blocking” ( $n = 118, M = 1.28$ ),  $t(138) = 2.71, p < .01$ . [Note: Of the participants in the diagnostic condition, the very few ( $n = 9$ ) who showed blocking also judged the relative importance of Cue P to Cue R ( $M = 2.33$ ) greater than those who showed “no blocking” ( $n = 134, M = 0.80$ ),  $t(141) = 2.81, p < .01$ .] Most important for our hypothesis, those participants in the predictive condition who showed the blocking pattern judged the relative importance of Cue P to Cue R ( $M = 2.64$ ) much higher than participants in the diagnostic condition ( $M = 0.99$ ),  $t(221) = 3.89, p < .001$ , and even higher yet than participants in the diagnostic condition who did not show blocking ( $M = 0.8$ ),  $t(154) = 4.83, p < .0001$ . We also replicated these results in another experiment (not reported here). It is notable that we obtained these results from a simple paper-pencil task with participants who were probably unmotivated or unable to pay close attention to the instructions or answer the questions carefully. It is likely that the differences will be enhanced under more controlled conditions, providing a means for future

studies to isolate the effects of causal understanding from the effects of mere associations.

### Conclusions

The importance of causal information in category formation and conceptual coherence has been emphasized by a number of researchers (e.g., Lien & Cheng, 2000; Murphy & Medin, 1985). The results reported here extend this research to similarity judgments among novel stimuli and provide a paradigm to study this effect. A two-phase blocking paradigm such as the one used by Waldmann and Holyoak (1992) can be used to isolate the effects of causal understanding from the effects of mere association. We found that different causal schemas applied to the same correlational data affected rated similarity of novel stimuli. Of course, causality is not the only mechanism that constrains similarity. Analogical processes and higher-order relations are also responsible for providing constraints (Medin, Goldstone, & Gentner, 1993).

Why should causal information be so important? People need to discover causal relations because only such relations allow the prediction of the consequences of interventions. This constraint makes discovery of causal features vital for understanding the world and for predicting future events. Having a notion of similarity that reflects the causal relations in the world is critical for inductive inference.

### Acknowledgments

We thank John Hummel, Michael Waldmann and Douglas Medin for helpful conversations and Askim Okten for running participants. This research was supported by NSF grants SBR-9729726 and SBR-9729023.

### References

- Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects*. New York: Bobbs-Merrill.
- Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, 40, 87-137.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- Morris, W. M. & Larrick, R. (1995). When one cause casts doubts on another: A normative analysis of discounting in causal attribution. *Psychological Review*, 102, 331-355.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.

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

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.