Causal Models and Cognitive Representations in Multiple Cue Judgment

Tommy Enkvist (tommy.enkvist@psyk.uu.se)
Department of Psychology, Uppsala University, SE-751 42, Uppsala, Sweden

Peter Juslin (peter.juslin@psyk.uu.se)
Department of Psychology, Uppsala University, SE-751 42, Uppsala, Sweden

Abstract

Recent studies suggest that humans can infer the underlying causal model from observing the distribution of variables. In a multiple-cue experiment we investigated if people can infer the causal structure from mere observation, and if different causal models invite different cognitive processes. Participants performed 220 training trials in two judgment tasks with different underlying causal structure. The result shows a poor ability to discriminate between causal models, and poor manipulation insight, but a correlation between causal models and cognitive processes. This study suggests that people do not represent multiple-cue judgment tasks in terms of causal models, but that common effect causal models invite reliance on processes of explicit cue abstraction.

Keywords: Multiple Cue Judgment; Causal Models; Cue Abstraction; Exemplar Memory.

Introduction

In everyday life, we make both judgments about common and rare events. From these judgments we might make important decisions, and act accordingly. When we make these judgments, how aware are we of the information upon which we base them? Are humans in general somewhat like amateur statisticians, calculating data and acting rationally according to the relevant information? And how much do we know about causality in our everyday life? Are human's also like amateur private investigators? Imagine, for example, yourself as a researcher investigating if different hormones in an exotic poisonous frog affect the toxicity of the frog, or if it is the toxicity that affects the level of the hormones. This is the intriguing task that our participants faced.

Advances in the formal modelling of causal relations (Pearl, 2000) has stimulated renewed interest for causal reasoning and its role in learning (Gopnik et al., 2004; Rehder, 2003; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). Beliefs about causality are increasingly used to explain how people reasons and make judgments (Sloman, 2005). Causal models have also emerged in the field of categorization (Rehder, 2003), but within a multiple cue learning paradigm causal models have been relatively absent (but see Schoppek, 2002).

The purpose of this study is to investigate if people can detect causal structure in a multiple-cue judgment task. When people are alerted to think about causality, can they then infer the causal structures merely from observing the cues and the criterion? Do they spontaneously represent the judgment tasks in terms of causal structure, or do they primarily reason in terms of exemplar memory or functional relations (Enkvist, Newell, Juslin, & Olsson, 2006; Juslin, Jones, Olsson, & Winman, 2003), as has been the belief in judgment research for a long time? These questions will be addressed in this study, and in a questionnaire that examines people's insights about the causal structure of the task.

Judgment Task and Cognitive Models

The two processes that are perhaps most often discussed in categorization learning and multiple cue judgment are rule-based and exemplar-based processes (Juslin, Jones et al., 2003; Smith, Patalano, & Jonides, 1998). Ruled-based models, like the *cue abstraction model*, implements the idea that people use controlled processes in working memory to mentally integrate cues according to a linear additive rule. In training participants abstract cue weights that are used to compute an estimate of the criterion when a new probe is presented (Juslin, Jones et al., 2003; Juslin, Karlsson, & Olsson; In press). In contrast, *exemplar models* assume that people make judgments by retrieving similar stored exemplars from memory (Medin & Schaffer, 1978; Nosofsky & Johansen, 2000), a process that involves rapid similarity-based processes. The exemplars retrieved from memory are representations of holistic concrete experienced instances encountered in training.

We rely on an experimental paradigm designed to help distinguish between cue abstraction and exemplar memory in a multiple-cue judgment task (Juslin, Jones et al., 2003). The judgment task involves a probe defined by four continuous cues and requires a judgment of a continuous criterion. Judgments are initially made in a training phase where feedback about the correct criterion is provided after every judgment. The cues $C1$, $C2$, $C3$ and $C4$ take on 11 discrete values between 0 and 10 and the toxicity c of a subspecies is a linear additive function of the cues:

$$c = 4 \cdot C1 + 3 \cdot C2 + 2 \cdot C3 + 1 \cdot C4 \qquad (1)$$

The criterion c is thus computed by assigning cue number one, C_1, most importance and therefore the largest weight and cue number four, C_4, the least importance.

When the participants make judgments of the continuous criterion the cue abstraction model suggests that they perform a mental analogue of linear multiple regression. For each cue, the weight ω_i (i=1…4) is retrieved and the estimate of c is adjusted accordingly:

$$\hat{c}_R = k + \sum_{i=1}^{4} \omega_i \cdot C_i \,, \qquad (2)$$

where $k = .5 \cdot (100 - 10 \cdot \Sigma \omega_i)$. The intercept k constrains the regression to be around the midpoint of the interval [0,100] (Juslin, Olsson, & Olsson, 2003). If $\omega_1 = 4$, $\omega_2 = 3$, $\omega_3 = 2$, and $\omega_4 = 1$, Equations 1 and 2 are identical and the cue abstraction model affords perfect judgments in this task.

The exemplar model implies that the participants make judgments by retrieving similar exemplars from memory (Medin & Schaffer, 1978). When the exemplar model is applied to judgments of a continuous criterion, the estimate $\hat{c}_E$ of the criterion c is a weighted average of the criteria c_j stored for the J exemplars, where the probe-exemplar similarities $S(p,x_j)$ are the weights:

$$\hat{c}_E = \frac{\sum_{j=1}^{J} S(p, x_j) \cdot c_j}{\sum_{j=1}^{J} S(p, x_j)} \,, \qquad (3)$$

where p is the probe to be judged, x_j is stored exemplar j (j=1…J), $S(p,x_j)$ is the similarity between probe p and exemplar x_j. The similarity between the probe p and exemplar x_j is computed according to the generalized context model (GCM:Nosofsky, 1986), a generalization of the original context model . The similarity $S(p,x_j)$ between a probe p and an exemplars x_j is found by transforming the distance between them.

According to GCM, the distance between a probe p and an exemplar j is,

$$d_{pj} = h\left[\sum_{M=1}^{M} w_m \left|x_{pm} - x_{jm}\right|\right], \qquad (4)$$

where x_{pm} are the value of the probe and x_{jm} are the values of an exemplar on the cue dimension m, the parameters w_m are the attention weight associated with cue dimension m, and h is a sensitivity parameter that reflects the overall property of discrimination in the psychological space. Attention weights can vary between 0 and 1 and are constrained to sum to 1. The similarity $S(p,x_j)$ between a probe p and an exemplar j is assumed to be a nonlinear decreasing function of the distance (d_{pj}) between them,

$$S(p, x_j) = e^{-d_{pj}} \,, \qquad (5)$$

Causal Models

The causal model network is a rather “new” framework based on Bayesian networks, a mathematical theory for representing probability (Sloman, 2005). In the causal modelling network, one of the core ideas is that an underlying causal network structure generates stable probabilistic relations of a system’s observed variables (Sloman, 2005). A particular kind of causal structure will generate a particular pattern of probability in the form of dependence and independence. Direct causal relations illustrated by arrows in a graph correspond to these relations between dependence and independence. Two variables are unrelated, or probabilistically independent, if there is no route from one variable to another on the direction of arrows in a causal graph.

Two causal models are used in this experiment, a common cause model and a common effect model, see Figure 1. In common cause the cues are affected by the criterion and thus have a high intercorrelation. In common effect the criterion is an effect of the cues and the cues have a low intercorrelation. The common effect model is in mathematical terms identical with functional models used in previous multiple cue judgment tasks (Enkvist et al., 2006; Juslin, Jones et al., 2003; Olsson, Enkvist, & Juslin, 2006), where the cues independent of each other affects a criterion.

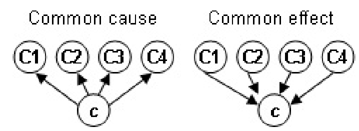

Figure 1. The Common cause and the Common effect models. *c* is the criterion, C1 to C4 are the cues.

The Experiment

The aim of the Experiment is threefold. First to investigate the ability to identify the underlying causal structure in a multiple-cue judgment task from merely observing the system. Recent studies suggests that humans and animals may have the ability to observe and infer underlying causal structure (Blaisdell, Sawa, Leising, & Waldmann, 2006; Gopnik et al., 2004; Steyvers et al., 2003).

Second, it investigates if different causal models tend to induce different cognitive processes in a multiple cue judgment task, like the exemplar based model and the cue abstraction model. One possibility is that cue abstraction is more prevalent in the common effect condition because it might be easier to estimate the weight of each independent cue. In the common cause condition cues are highly correlated and the weight of one individual cue could be more difficult to estimate. Therefore, more exemplar memory is expected in the common cause condition.

Third, the experiment highlights the effect of learning instructions about causal models on the ability to infer the underlying causal structure. Information about different causal models has been common in causal learning experiment (Lagnado & Sloman, 2004; Steyvers et al., 2003), and learning instructions can be used to boost performance in multiple-cue judgment tasks (Olsson et al., 2006). A questionnaire at the end of the experiment will try to capture participants’ insight about the underlying causal structures, by asking for model identification, manipulation of the variables, and the intercorrelation between cues.

Method

Participants. Forty-four students from Uppsala University participated. 19 males and 25 females with a mean age of

23.8 years (Range: 19-32, *SD*=2.9). The participants were rewarded with a cinema ticket or course credits.

Material and procedure. The experiment involved a 2∗2 between-subjects factorial design, with learning condition (neutral instructions vs. model instructions) and causal model condition (common effect vs. common cause) as independent variables. Participants were randomly assigned to one of the four group conditions.

Subspecies of fictitious poisonous frogs characterized by values on four continuous cues and a continuous criterion were used as stimuli. Each cue could take on 11 different values, represented in the experiment by a number ranging from 0 to 10 (Enkvist et al., 2006). There were two different causal environments with the poisonous frog. A common cause and a common effect environment. In the common cause environment the criterion is causing the levels of the cues and in the common effect environment the cues are causing the level of the criterion, see Figure 1. A random error was added to change the multiple correlation between the cues and the criterion to equal R=.9 in both causal learning conditions to reflect realistic and probabilistic learning environments.

The model instruction learning condition. Participants in the model instruction learning condition, regardless of what underlying model they were assigned to, were instructed to imagine themselves as doing research on a poisonous frog in South America, and that findings had suggested that they could guess the amount of poison contained in the frog by observing the amount of four different hormones; alpha, beta, phi and rho, in its blood. The participants were also informed that there were two different theories (the causal models) about the causal relations between the amount of poison and the value on the different hormones. Their task was to try to predict the amount of poison in the frog by observing the different values on the hormones, and also to try to decide which of the two theories that was correct. One theory was that the toxicity was caused by the different amounts of hormone in the blood, so that changing the hormone would also change the poison. The other theory was that the toxicity was instead causing the levels of hormones in the blood, so that changing the amount of poison would also change the level of hormones. They were informed that the first part of the experiment would involve a learning phase, where they would be provided with feedback on every trial, about the correct amount of poison that the frog was carrying. Also, the instructions said that every now and then, they were going to be asked to choose between the two theories. The underlying theory would remain the same the whole time, but they were asked this to see if their understanding of what theory was correct would change over time. They were also informed about the fact that they could not possible make perfect predictions every time, because of a random error.

The neutral learning condition. In this condition, participants were not informed about the possibility of alternative underlying structures like the common effect and the common cause models. The participants were only told that they would try to predict the amount of poison in the frog by observing the values of the hormones alpha, beta, phi and rho, and that a random error would make it impossible to make perfect predictions all the time.

The Experiment consisted of three parts. First a training phase where participants made outcome judgments of toxicity based on four cues and received outcome feedback about the correct criterion. 220 unique variants of poisonous frogs where presented in training. Participants in the model learning condition also made 14 judgments of the hypothetical model, common cause or common effect, that best described the relation between cues and criteria. The test phase consisted of 60 trials with no outcome feedback: 20 exemplars from the training phase and 10 new exemplars presented twice in random order.

After training and test all participants filled in a questionnaire, to find out if the participants had gained any insight about the system under study. The questionnaire was in three parts. First participants were asked 8 questions about the degree to which one variable can be used to manipulate another (first for how each cue affected the criterion and then how the criterion affected each cue). On each question participants estimated on a scale from 1 to 7, where 1 mean no effect and seven means high effect. Second, 6 questions were asked about cue correlations and estimated on a scale from 1 to 7 (1 indicate low correlation and 7 indicate high correlation). Finally participants made a model choice of which of the causal models (common cause or common effect) that best described the judgment task.

Dependent Measures. Judgment data is analyzed at three levels: *Performance and learning, Representation* and *Model Fit*. In the questionnaire, three different dependent variables were measured, *model insight*, *manipulation insight* and *intercorrelation insight*.

Performance is measure by Root Mean Square Error (RMSE), between judgment and criterion in the test phase. Lower RMSE suggests better performance. Learning is also measured by the correlation between judgments of old exemplars in test (exemplars encountered during training that also are included in the test phase) and the criterion of the old exemplars.

Representation is measured by Extrapolation index, the ability to extrapolate judgments beyond the previous learned criterion range. Extrapolation Index is the signed deviation from the prediction by a linear regression model with judgment as dependent variable and criterion as the independent variable. If the judgments for the extreme exemplars are as extreme as expected from linear extrapolation from the training exemplars the Extrapolation index is positive, otherwise Extrapolation index is negative.

Model fit is measured by Root Mean Square Deviation (RMSD), the absolute deviation between model prediction for each model (the cue abstraction model and the exemplar based model) and judgment data.

Results

Performance and learning. There where no differences in performance at test between the learning conditions or between the causal models. Figure 2 shows the judgment data from the test phase for both causal conditions.

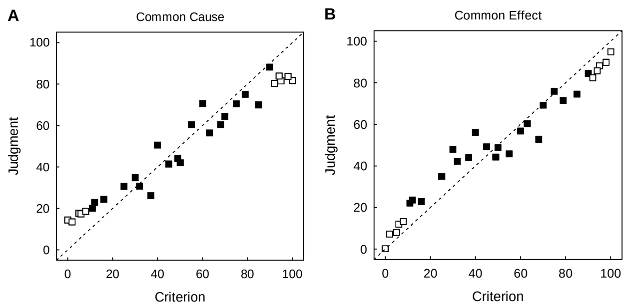

Figure 2. Judgment data from the test. Filled squares denote old exemplars seen both in training and test. Open squares denote new exemplars introduced at test. Panel A: The common cause condition. Panel B: The common effect condition.

Mann-Whitney U-tests on the correlations between the criterion and the judgments for old exemplars shows no differences between the learning conditions and causal models (causal models: $U_{1,\ 22}$=174, p=.11; learning conditions: $U_{1,\ 22}$=198, p=.31), suggesting that learning is roughly equal in all conditions. A two-way ANOVA with learning conditions (neutral and model based) and causal models (common cause and common effect) as independent variables and RMSE as dependent variable shows that there are no differences in RMSE between the learning conditions and the causal models in the last block (last 20 trials) of the training phase, p>.44, suggesting that learning is fairly similar at the end of training.

Representation and Model fit. A one-way ANOVA shows that the Extrapolation index is significantly higher for the common effect model ($F_{1,\ 42}$=74.85, p<.001). Extrapolation index is positive and significantly separated from zero for the common effect condition, suggesting the use of cue abstraction. For the common cause condition Extrapolation index is negative and significantly separated from zero, suggesting the use of exemplar memory, see Figure 3 left panel.

Model fit was analyzed to see which of the cognitive models, the exemplar model or the cue abstraction model, that best explained the judgment data. A two-way ANOVA with learning conditions (neutral and model based) and causal models (common cause and common effect) as between-subjects independent factors and the cognitive models (exemplar model and cue abstraction model) as within-subjects factor shows significant difference between the causal models ($F_{1,\ 40}$=16.95, p<.001), no difference between learning conditions ($F_{1,\ 40}$=.18, p=.67), but a significant interaction ($F_{1,\ 40}$=5.14, p=.029), see Figure 3. In the common cause condition there is no difference between the cue abstraction model and the exemplar model, but in the common effect condition cue abstraction has a significantly better fit than the exemplar model.

Different causal structures do not affect learning in a multiple-cue judgment task, but different causal structures thus seem to be related to different cognitive processes. Participants in the common cause condition were unable to fully extrapolate their judgment on new exemplars in test suggesting the use of exemplar memory. The Model fit shows that both the exemplar based model and the cue abstraction model fits data in the common cause condition suggesting that both models are in use in the common cause condition. Positive Extrapolation index that is significantly separated from zero and a significantly better fit for the cue abstraction model suggests that cue abstraction is the dominating process in the common effect condition.

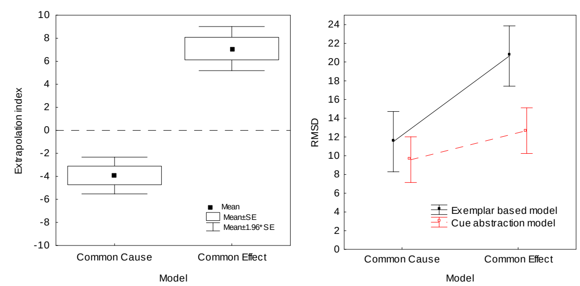

Figure 3. Left panel: Extrapolation index for the common cause model and the common effect model. Right panel: Model fit in terms of RMSD for each causal model condition and cognitive process.

Analysis of the Insight Questionnaire

Model insight. Participants were asked to assess the probability for each of the two causal models (common cause and common effect) after viewing two illustrations of the models like those in Figure 1. In a one-way ANOVA the probability for "common cause" was analyzed, and if the participants would have grasped the underlying causal structure, the participants in the common cause condition should judge a higher probability for common cause. This was not the case. There were no differences between the causal models (p=.55), see Figure 4A. In the model learning conditions participants where asked about the correct model 14 times during training. The results from the model choice in training shows a significantly better accuracy in model choice for the common effect condition, $F_{1,\ 20}$=16.9, p<.001, (accuracy for common cause .28 and for common effect .76). This, however, does not seem to indicate better performance in the common effect condition, but a general bias for the common effect model because the assessments in the model insight measure fall systematically below .5 in both conditions. This is also supported by the poor accuracy in model choice during training for participants in the common cause condition.

Manipulation insight. If participants have gained an insight about the causal models, they should also know how to manipulate the criterion or the cues. For each participant, a measure of the cue criterion relationship was calculated based on the causal model used in their task. A significant effect was found for causal model, $F_{1,\ 40}=8.4$, $p<.006$, indicating a difference between participants in the two models when it comes to insight about how to manipulate the criterion or the cues, see Figure 4B. A positive value suggests that participants have identified the correct manipulation pattern for their task. A negative value suggests that participants manipulate the task in the wrong way. Participants in the common cause condition seem to be completely clueless. The question seems to be if the participants in the common effect condition have acquired deeper insight, or if all participants are inclined to think in terms of the common effect model. One answer is that the poor insight in both model conditions suggests that there is a response bias towards the common effect model. No significant effects were found for learning condition, $F_{1,\ 40}=.004$, $p=.95$, or interaction, $F_{1,\ 40}=1.83$, $p=.18$.

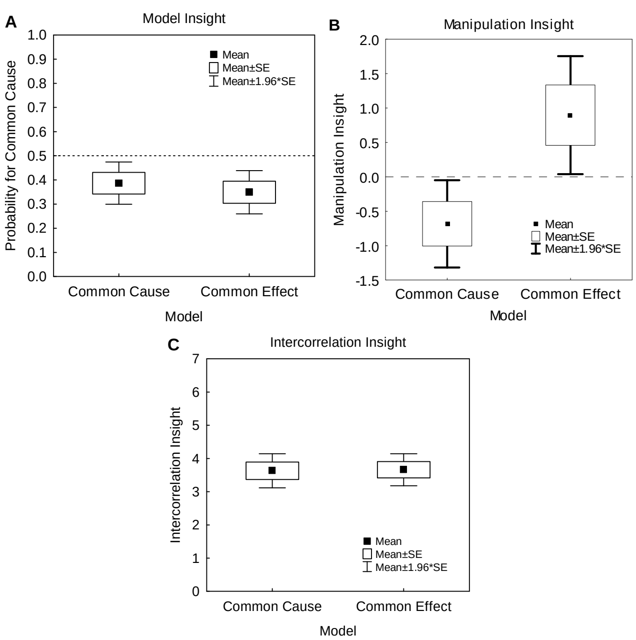

Figure 4. Panel A: Model Insight, probability for common cause. Panel B: Manipulation Insight. Panel C: Intercorrelation Insight.

Figure 4A and B are together strong evidence that there is a general bias towards thinking in terms of common effect. A probability judgment for the common cause model that are significantly below .5 and a significantly negative value on manipulation insight both suggest that participants in the common cause condition choose the common effect model and wrongly believes that changing values on the cues will affect the criterion in their task.

Intercorrelation insight is measuring participant's insight about the intercorrelation between the cues. Remember that in the common cause condition the cues have a tendency to be intercorrelated because of the underlying cause, while in the common effect condition the cues are independent from each other and not intercorrelated. A mean was calculated for the judgments of six assessments that expressed how much of the correlation between cues that was noticed by the participants. There were no significant effects of causal model, learning condition or interaction, see Figure 4C.

In summary the insight questionnaire suggests that participants have poor knowledge about the causal system that they are exposed to and that there seem to be a bias towards the common effect model. A surprising result is that there is no improvement in insight with learning instructions with information about causal models. Despite that the cover story contained written and visual information about the common cause model and the common effect model, participants were unable to gain any insight about the causality.

Discussion

The purpose of this study was to investigate if it is possible from observation to identify the underlying causal structure in a multiple-cue judgment task and if different cognitive processes could be associated to different causal models. In the experiment we also investigated the effects of learning instructions about causal models on the ability to infer the underlying causal structure in the judgment task.

Inferring Causal Structure in Multiple Cue Judgment. The question raised in this study was if it is possible to identify the underlying causal models in multiple-cue judgment tasks. This question was asked with reference to earlier studies, and those made by Steyvers et al. (2003) in particular, that humans seem to have a reasonably good ability to detect causal models, especially when they have been informed about possible causal explanations as producers of the patterns in the task.

The insight measures from the questionnaire in this experiment shows a poor understanding about the underlying causal structure. An explanation for the poor insight in the common cause learning condition is that there seems to be a bias towards thinking in terms of common effect (see Lagnado & Sloman, 2004 for similar findings). Figure 4A and B shows that participants in the common cause condition systematically make judgments about cue-criterion relationships and model choice that are consistent with common effect. A second explanation for the poor insight is the complexity of the judgment task compared to most causal learning tasks (Steyvers et al., 2003). Steyvers et al. for example used three binary cues compared to the four continuous cues and continuous criterion in the present study. It might be that it is more difficult to infer underlying causal structures in this multiple-cue judgment task because the four continuous cues were maybe at least one too many.

Causal Models and Cognitive Representation. The results show that there are no differences between the causal learning conditions in learning and performance. However when investigating representation and model fit, differences between the causal learning conditions occur. Participants in the common effect learning condition are able to extrapolate beyond the previous learned criterion range, suggesting the use of cue abstraction, but participants in the common cause learning condition are less able to extrapolate, as predicted by exemplar memory. The common cause learning condition shows no differences between cue abstraction model and exemplar based model in model fit calculations, but the common effect learning condition show a significantly better support for the cue abstraction model. The common effect condition which in mathematical terms equals Equation 1, is in line with previous studies, suggesting the use of cue abstraction in multiple cue judgment tasks with continuous cues and continuous criterion (Enkvist et al., 2006). The exemplar effects in the common cause condition could be an effect of the difficulty to abstract cue weights when cues are highly intercorrelated. When cue abstraction fails, participants shifts to use exemplar memory to succeed with the task (Juslin et al., in press; Olsson et al., 2006).

Learning Instructions as a Performance Booster? No benefits could be detected in the conditions that received learning instruction with model presentations. In Olsson et al. (2006) participants receiving information about the judgment task performed significantly better than participants with neutral instructions. In this experiment we found no benefits with instructions about the two causal models over neutral instructions with no model information.

Conclusions

The main conclusions from this experiment are; 1) there was no improvement in performance with initial instructions about causal models, 2) no differences in performance between the two causal learning conditions, 3) causal models seems to invite different cognitive processes. Common cause is relatively more associated with exemplar memory and common effect is associated with cue abstraction, 4) insight measures shows a poor understanding about the underlying causal structure and a strong bias towards the common effect model.

Acknowledgments

This research was supported by the Swedish Research Council. We are indebted to Linda Lindberget for helping with collecting the data, and Maria Henriksson for valuable comments. Bulletin

References

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*, 1020-1022.

Enkvist, T., Newell, B., Juslin, P., & Olsson, H. (2006). On the role of causal intervention in multiple cue judgment: Positive and negative effects on learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 163-179.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review, 111*, 3-32.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 924-941.

Juslin, P., Karlsson, L., & Olsson, H. (In press). Additive integration of information in multiple cue judgment: A division of labor hypothesis. *Cognition*.

Juslin, P., Olsson, H., & Olsson, A. C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133-156.

Lagnado, D. A., & Sloman, S. A. (2004). The Advantage of Timely Intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 856-876.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-61.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin Review, 7*, 375-402.

Olsson, A.-C., Enkvist, T., & Juslin, P. (2006). Go with the flow: How to master a nonlinear multiple-cue judgment task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1371-1384.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, England: Cambridge University Press.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1141-1159.

Schoppek, W. (2002). *Stochastic independence between recognition and completion of spatial patterns as a function of causal interpretation.* Paper presented at the Proceedings of the 24th Annual Conference of the Cognitive Science Society.

Sloman, S. A. (2005). *Causal Models - How People Think About the World and its Alternatives*. New York: Oxford University Press.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65*, 167-196.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453-489.