

# Simulating Causal Models: The Way to Structural Sensitivity

**York Hagmayer** (york.hagmayer@bio.uni-goettingen.de)  
Department of Psychology, University of Göttingen,  
Gosslerstr. 14, 37073 Göttingen, Germany

**Michael R. Waldmann** (michael.waldmann@bio.uni-goettingen.de)  
Department of Psychology, University of Göttingen,  
Gosslerstr. 14, 37073 Göttingen, Germany

## Abstract

The majority of psychological studies on causality have focused on simple cause-effect relations. Little is known about how people approach more realistic, complex causal networks. Two experiments are presented that investigate how participants integrate causal knowledge that was acquired in separate learning tasks into a coherent causal model. To accomplish this task it is necessary to bring to bear knowledge about the structural implications of causal models. For example, whereas common-cause models imply a covariation among the different effects of a common cause, no such covariation between the different causes of a joint effect is implied by a common-effect model. The experiments show that participants have virtually no explicit knowledge of these relations, and therefore tend to misrepresent the structural implications of causal models in their explicit judgments. However, an implicit task that only required predictions of singular events showed surprisingly accurate sensitivity to the structural implications of causal models. This dissociation supports the view that people's sensitivity to structural implications is mediated by running simulations on mental analogs of the causal situations.

## Introduction

In everyday life as well as in scientific research we rarely observe the behavior of complex causal networks at once. A more typical scenario is that we learn about single causal relations separately, and later try to integrate the different observed relations into a more complex interconnected causal model. For example, we might first learn that aspirin relieves headache. Later we may observe that aspirin unfortunately also creates stomach problems. Now we are in the position of putting these two pieces of knowledge together. The question is how? How are different fragments of causal knowledge integrated into coherent complex structures?

## Bayesian Causal Models

One recent approach to this problem that has become increasingly popular in the past few years postulates *Bayesian network models* for representing causal knowledge (see Pearl, 1988, 2000; Glymour & Cooper, 1999). Bayesian network models provide compact, parsimonious representations of causal relations. For example, Figure 1 displays a causal model that connects five events,  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ ,  $X_5$ . One way to represent this domain is to list the

32 probabilities of the joint probability distribution,  $P(X_1, X_2, X_3, X_4, X_5)$ , by considering every combination of present and absent events. Another possible strategy is to encode the base rates and all covariations that can be computed between five events. However, even with modestly complex structures the number of covariations becomes very large, especially when more complex higher-order covariations between multiple events also are considered. Bayesian network models reduce the complexity of representing causal knowledge by distinguishing between direct causal relations (the arrows in Fig. 1), and covariations that can be derived by using information encoded in the *structure* of the causal models. The structure of causal models primarily expresses information about conditional independence between events. For example, in Figure 1 event  $X_4$  is coded as being independent of event  $X_5$  conditional upon event  $X_3$ . Conditional independence greatly simplifies computations by allowing the derivation of the indirect relations from products of the relevant components (see Pearl, 1988; Glymour & Cooper, 1999). In Figure 1 the joint probability distribution can be factorized into the product of a small number of unconditional and conditional probabilities,

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_5 | X_3) \cdot P(X_4 | X_2, X_3) \cdot P(X_3 | X_1) \cdot P(X_2) \cdot P(X_1).$$

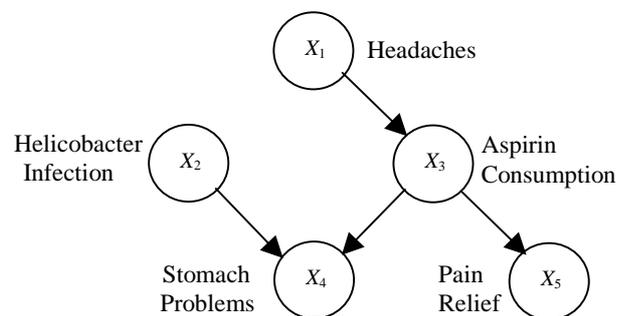


Figure 1: Example of a Bayesian Network

The distinction between direct causal relations and indirect relations can also be used for the *integration* of separate pieces of causal knowledge. Combining the information that aspirin relieves headache with the information that it additionally causes stomach problems yields a

common-cause model with aspirin playing the role of the common cause of two independent effects, relief of headache and stomach problems (see Fig. 2, left). By contrast, integrating the two causal relations “Aspirin causes stomach problems” with “Helicobacter pylorus causes stomach problems” would yield a different structure, a common-effect model, in which two independent causes converge on a joint effect (see Fig. 2, right). In both examples, two independent causal relations are being integrated. However, the outcome of the integration process is different. The two different causal models entail different implications for the indirect relations between events.

### Structural Implications of Causal Models

The basis for the possibility of integrating different causal links into coherent wholes are the structural implications of causal models. In our experiments we focused on two simple models, a common-cause and a common-effect model. Both models integrate two causal links but entail distinctly different implications for the non-causal relations.

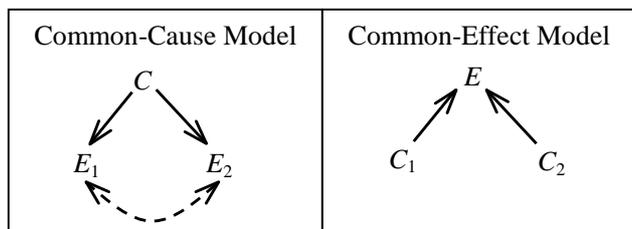


Figure 2: Implications of Different Causal Models

Figure 2 (left) depicts a common-cause model with a common cause  $C$  producing two independent effects  $E_1$  and  $E_2$ . Common-cause models of this kind entail a (spurious) covariation among the effects. Provided the common cause independently generates the two effects, the joint probability of the effects,  $P(E_1, E_2)$ , can be calculated by taking the product of the base rate of the cause,  $P(C)$ , and the two conditional probabilities,  $P(E_1|C)$  and  $P(E_2|C)$  (see also Appendix). Thus, although the two effects may never have been observed together, the causal model still allows it to derive a prediction for the patterns that should be expected. Common-cause models clearly differ from common-effect models. Figure 2 (right) shows an example in which two causes,  $C_1$  and  $C_2$ , are linked with a joint effect  $E$ . Common-effect models do not imply covariations among the different causes of the joint effect. The causes may covary in a specific learning situation but this covariation is not implied by the model, it is something that has to be explicitly encoded. This is the reason why in the example shown in Figure 1 common effects were conditionalized on patterns of its direct causes (e.g.,  $P(X_4|X_2, X_3)$ ). However, this is only possible when all the relevant events have been observed together, and when the number of relevant patterns is small enough not to surpass information processing limitations. In more complex cases and in situations in which causal knowledge has to be generated from different learning experiences, *causal schemas* have been postulated in the literature (Pearl,

1988). For common-effect models, the *noisy-or schema* has been proposed as a plausible integration schema (see also Waldmann & Martignon, 1998). According to this schema  $P(X_4|X_2, X_3)$  can be reduced to  $[1-(1-P(X_4|X_2)) \cdot (1-P(X_4|X_3))]$ , an expression that only contains probabilities referring to direct causal relations. The noisy-or schema assumes that different causes have independent and additive influences on the common effect. Given that common-effect models do not imply covariations among the causes a further reasonable default assumption is that they occur independently. A number of psychological experiments have shown that learners indeed tend to initially assume independence (see Waldmann, Holyoak, & Fratianne, 1995).

### Sensitivity to Structural Implications: Computation vs. Causal Simulation

Previous research has demonstrated sensitivity to structural implications of causal models in causal learning (Waldmann & Holyoak, 1992; Waldmann, 2000), causal reasoning (Waldmann & Hagmayer, 1998), and categorization (Waldmann et al., 1995). The processes underlying this sensitivity are unclear, however. The standard approach within the area of Bayesian modeling is to explicitly derive the predicted event patterns or covariations and test these predictions against the data at hand. It appears unlikely that this strategy could be followed in intuitive everyday reasoning. Despite the fact that Bayesian models provide a parsimonious way of representing domain knowledge it is also clear that the explicit derivation of indirect relations is often complex and computationally demanding (Glymour & Cooper, 1999). In fact, one reason for the increasing number of automated statistical tools that are currently offered to researchers lies with the fact that the task surpasses the capacity limitations of intuitive reasoning.

However, there is an alternative, more implicit strategy. Instead of explicitly computing covariations we may form mental representations of causal structures that are analogous to the graphical structures used in Bayesian network modeling (e.g., Fig. 1). Similar to toy models, these causal models can then be used to run mental simulations (see also Barsalou, 1999). For example, instead of calculating the probability of patterns within a common-cause model with one cause and two effects we could mentally imagine the presence or absence of the cause, and then generate predictions for each individual effect based on the observed covariations between the cause and either effect. Since these predictions are triggered by a common event within a mental common-cause model the predicted patterns should show the covariations that are implied by the structure of the mental model. These covariations are not the consequence of an explicit computation, they rather are a side effect of the structure of the causal model used to simulate the causal situation in the real world. Therefore it may well be that the predicted patterns exhibit covariations of which the learners are not aware. For the learner it is only necessary to focus on the direct causal relations. All the indirect relations are taken

care of by running simulations on mental analogs of the objective causal situation.

Two experiments will be presented in which participants acquired partial knowledge about separate fragments of common-cause or common-effect models. To test whether they were sensitive to the additional covariations implied by the different causal models, two types of measures were collected. *Explicit knowledge* was assessed by means of probability estimates in which participants were requested to estimate the strength of the indirect, not directly observed relation. Based on the assumption that explicit computations of the answers to these questions are hard we expected poor performance with this task. However, the second task was designed to tap into *implicit knowledge* generated by causal simulations. In this task, participants were requested to predict the pattern of events they expected to see. For example, in a common-cause condition (see Fig. 2) the experimenter instructed participants to imagine that the cause was present and to make a prediction about the two effects. A typical finding with this type of task is that participants tend to match the probabilities they have seen in the learning situation. Since in the present task the two effects never have been seen together, direct experience with the patterns is not available. However, it is possible that participants match the probabilities for each relation independently within a mental analog of a common-cause model. The model itself generates covariations that have never been observed directly. The crucial measure in this task is the covariation between the predicted effects that can be derived from participants' responses. The causal-simulation account predicts that these patterns should display the covariations implied by the causal models even when no explicit knowledge could be detected in the explicit task.

## Experiment 1

The goal of this experiment was to investigate whether learners who have acquired partial knowledge about fragments of causal models are sensitive to the structural implications of these models. Participants were given the task to learn about the causal relations between the mutation of a gene and the prevalence of two (fictitious) substances (enzyme BST and brasus protein). We used a trial-by-trial learning procedure in which participants worked through a stack of index cards with information on the front side about whether a mutation of the gene occurred or not. By turning around the individual cards participants received information about the presence or absence of either the enzyme BST or the brasus protein. To ensure that no covariation between the enzyme and the protein could be observed the cards were divided into two different stacks, one for each substance. Participants were instructed to alternate between the stacks in the course of the learning phase. In the initial instruction the separate stacks were characterized as displaying the raw data of two different research projects located at different universities. The task and the presentation of the data were identical for all participants. They first received information about the mutation of the gene on the front side of the cards, and then were shown information about the occur-

rence of either the enzyme or the protein on the backside. The learning phase consisted of 80 cards, 40 for each substance.

Two factors, type of causal model and degree of covariation, were manipulated yielding four experimental conditions. The first factor contrasted two different causal models. One group of participants read in the initial instructions that the researchers were interested in finding out whether the mutation causally influences the two substances (common-cause model)(see Fig. 2, left). In contrast, for the second group the two substances were described as potential causes of the mutation (common-effect model)(see Fig. 2, right). The second factor manipulated the strength of the relation between mutation and the two substances. The strength was always equal for both substances and either weak or strong. Table 1 displays the absolute frequencies used in this experiment. Thus, for example, participants in the condition with strong connections saw 16 cases for each substance in which the presence of a mutation of the gene was paired with the presence of the substance.

Table 1: Frequencies in Experiment 1

	Strong Condition		Weak Condition	
	Substance	No Substance	Substance	No Substance
Mutation	16	4	10	10
No Mutation	0	20	6	14

Apart from the different initial instructions about the underlying causal model the learning phases and the test phases were identical within the conditions with strong or weak relations. Regardless of whether the mutation of the gene was introduced as a cause or as an effect, information about its presence or absence was delivered before information about the substances was given.

The learning phase was followed by a test phase in which participants' assumptions about the covariation between the two substances was assessed. This covariation had to be inferred because the two substances had never been seen together. To test whether participants were sensitive to the different implications of the two causal models we compared an implicit with an explicit measure of knowledge. In the *implicit* test procedure participants received 20 new index cards in a random order, half of them indicating that in this particular case a mutation had occurred. The rest of the index cards described cases in which no mutation had occurred. Participants' task was to predict for each case individually whether either of the two substances was present or absent. No feedback about the substances was provided during this test phase. Since patterns of substances had to be predicted it was possible to analyze the amount of covariation between the substances in the responses of the participants. We used the *phi* correlation coefficient as a measure of the degree of the implicitly predicted covariation (see Appendix). In a second task that followed the

implicit task, we investigated participants' *explicit* expectations. In this task they had to estimate the probability that the second substance is present conditional upon the first being present ( $P(\text{substance}_2|\text{substance}_1)$ ) and being absent ( $P(\text{substance}_2|\sim\text{substance}_1)$ ). As with the implicit measures the explicit estimates were transformed into *phi* correlations that allowed us to directly compare the implicit with the explicit measure.

What are the normative Bayesian predictions for the presented data? When a common-cause model is assumed it is appropriate to encode the conditional probabilities directed from the cause (i.e., mutation) to its effects (i.e., substances). In this direction, the data display a conditional probability of either substance in the presence of the mutation (i.e.,  $P(\text{substance}|\text{mutation})$ ) of .80 in the strong and .50 in the weak condition. The corresponding values in the absence of a mutation ( $P(\text{substance}|\sim\text{mutation})$ ) are 0 in the strong versus .30 in the weak condition. Taking the difference of these numbers yields the widely used *contingency* ( $\Delta P$ ) measure of statistical strength (Eells, 1991). Accordingly, the contingency is  $\Delta P = .80$  in the strong and  $\Delta P = .20$  in the weak condition.

Within the framework of a common-effect model the same data should again be analyzed from causes to effects. In this condition the substances play the role of the causes. Thus, it is appropriate to compare  $P(\text{mutation}|\text{substance})$  with  $P(\text{mutation}|\sim\text{substance})$ . The data yield a probability of the mutation in the presence of the substance of 1 in the strong and of .63 in the weak condition. The corresponding values in the absence of the substance are .17 in the strong and .42 in the weak condition. These numbers imply almost the same contingencies as in the common-cause condition of  $\Delta P = 0.83$  (strong condition) and  $\Delta P = .21$  (weak condition).

On the basis of structural information from the causal models these numbers can be used to derive the predicted covariation between the substances. While the common-effect model does not imply a covariation, the common-cause model entails that the observed joint probability should correspond to the product of the base rate of the cause and the conditional probabilities observed for each causal link. These probabilities can be transformed into a *phi* coefficient of correlation (see Appendix). The data presented imply a *phi* correlation of  $r = .67$  between the substances in the strong condition and of  $r = .042$  in the weak condition.

## Results and Discussion

The results are based on 48 students from the University of Göttingen who were randomly assigned to one of the four learning conditions. Table 2 shows the means for both the explicit and the implicit measure obtained in the four conditions.

The correlations that the participants generated in the implicit prediction task resemble very closely the ones normatively implied by the causal models. Participants in the common-cause condition generated a high mean correlation of .62 between the substances when the causal connections were strong and a mean correlation of -.004 when they were weak. In contrast, in the common-effect

condition in which they received identical learning inputs as participants in the corresponding common-cause condition the prediction responses displayed generally low correlations in both conditions. An analysis of variance revealed a significant main effect for the factor causal model,  $F(1, 44) = 7.28$ ,  $p < .05$ ,  $MSE = .14$ , and a significant main effect for the factor strength of covariation,  $F(1, 44) = 18.4$ ,  $p < 0.01$ ,  $MSE = .14$ . The interaction failed to be significant,  $F(1, 44) = 2.33$ ,  $p = .13$ ,  $MSE = .14$ .

Table 2: Means of Implicit and Explicit Measures (Experiment 1)

	Implicit Measure: Generated Correlations		Explicit Measure: Estimated Correlations	
	Common-Cause Model	Common-Effect Model	Common-Cause Model	Common-Effect Model
Strong Relations	.622	.168	.286	.161
Weak Relations	-.004	-.130	-.109	.039

The explicitly estimated correlations clearly differed from the implicitly generated ones (see Table 2). There was no significant difference of the estimated correlations in the two contrasted causal models,  $F < 1$ . Only the difference between the conditions in which strength of covariation was manipulated proved significant,  $F(1, 44) = 8.05$ ,  $p < .01$ ,  $MSE = .10$ .

These results indicate that participants showed little sensitivity to the implications of causal models when the task required explicit estimates. They seemed to be aware of the fact that the inferred covariations somewhat depend on the strength of the causal links responsible for the covariations, but they did not explicitly grasp the structural difference between common-cause and common-effect models. By contrast, the implicit measure displayed surprisingly accurate predictions. In this task, participants clearly differentiated between common-cause and common-effect models despite identical learning inputs. In our view, this finding supports the prediction that sensitivity to structural implications can be achieved by running simulations on mental analogs of causal models.

## Experiment 2

In Experiment 1 participants first were informed about whether a mutation of the gene occurred or not, and then learned for each substance separately whether it was present or absent. This procedure served the goal of presenting identical learning inputs to participants in the different conditions. It raises the question, however, whether the observed asymmetries of sensitivity to implied covariations are due to the contrasted causal models or rather to differences in the direction of required inferences during learning. In the common-cause condition learning was directed from cause to effects (predictive learning), whereas in the common-effect conditions the very same

learning items implied that learning proceeded from effect to causes (diagnostic learning). Thus, it may be speculated that differences between predictive and diagnostic learning rather than differences in the underlying causal models may be the reason for the obtained results.

The goal of Experiment 2 was to replicate the results of Experiment 1 and to control for the direction of learning. Moreover, unlike in Experiment 1 the conditional probabilities and contingencies were equalized in the contrasted conditions. Material and procedure were taken from Experiment 1. All participants had the task to learn about the causal connection between mutation and the two substances. Again, as learning input they received index cards separated into two stacks which either provided information about the relation between the mutation and the enzyme BST or between the mutation and the brusus protein. Table 3 shows the frequencies of the different patterns that were presented during the learning phase.

Table 3: Frequencies in Experiment 2

	Substance	No Substance
Mutation	25	5
No Mutation	5	25

These frequencies implied conditional probabilities between the mutation and the substances that were completely symmetric ( $P(\text{mutation}|\text{substance}) = P(\text{substance}|\text{mutation}) = .8$ , and  $P(\text{mutation}|\sim\text{substance}) = P(\text{substance}|\sim\text{mutation}) = .2$ ). Thus, the contingencies were identical in both directions ( $\Delta P = .60$ ).

Two factors were manipulated in Experiment 2. The first factor manipulated the assumed causal model by means of differential initial instructions. As in Experiment 1, the mutation of the gene was either introduced as the cause of the two substances (common-cause model) or as their effect (common-effect model). The second factor manipulated the learning direction. Learning proceeded either from causes to effects (predictive learning) or from effects to causes (diagnostic learning). Thus, half of the participants received information about the mutation first before learning about the substances whereas the other half first read information about the presence or absence of one of the substances, and then received feedback about the mutation. In fact, the same index cards were used for all participants, the only difference was which side they saw first. Information about the mutation was shown first in the predictive version of the common-cause condition and in the diagnostic version of the common-effect condition. The reversed cards showing information about the substances first were given to participants in the predictive common-effect and the diagnostic common-cause conditions. Using the procedures described in the Appendix, a *phi* correlation of  $r = .37$  between the substances can be derived for the common-cause model in which they played the role of effects. This is about half the size of the implied covariation in the condition with strong relations of Experiment 1. Thus, a smaller effect

size is to be expected in the present experiment. In contrast to the common-cause model, the common-effect model does not imply any covariation between the causes. These different structural implications are, of course, independent of the direction of learning.

As in Experiment 1, sensitivity to implied covariations was assessed by means of implicit and explicit measures. Regardless of the learning direction the implicit test always presented information about the mutation of the gene as the cue for the predictions. Participants were shown 20 new cases, half of which describing mutations, and had to predict for each case individually whether either of the substances was present or not. The explicit task in which participants estimated conditional probabilities followed the implicit one (see Experiment 1).

## Results and Discussion

64 students from the University of Göttingen were randomly assigned to one of the four conditions. The means of the *phi* correlations that were either generated (implicit measure) or estimated (explicit measure) in the four different conditions are shown in Table 4.

Table 4: Means of Implicit and Explicit Measures (Experiment 2)

Learning Direction	Implicit Measure: Generated Correlations		Explicit Measure: Estimated Correlations	
	Common-Cause Model	Common-Effect Model	Common-Cause Model	Common-Effect Model
Predictive	.243	.013	.258	.239
Diagnostic	.186	-.001	.129	.210

As in Experiment 1, assumptions about the underlying causal model clearly influenced the implicit measure. The main effect for the factor causal model was significant for the generated correlations,  $F = 4.97$ ,  $p < .05$ ,  $MSE = .14$ . In general, participants generated higher correlations between the substances when they were viewed as effects (common-cause model) than when they had been characterized as causes of the mutation (common-effect model). In the common-effect condition the generated covariations between the two substances (i.e., the causes) were very close to 0 which supports our prediction that independence between causes is assumed in common-effect models. Neither the factor learning direction nor the interactions with this factor proved significant ( $F < 1$ ).

In contrast to the implicit measures, no sensitivity to the structural implications of causal models could be detected with the explicit measures. In general, participants tended towards correlations that clearly differed from 0 but showed no sensitivity to the assumed causal model. None of the effects approached significance in an analysis of variance in which type of causal model and learning direction entered as factors ( $F < 1$ ).

These results clearly support the conclusions of Ex-

periment 1 by demonstrating sensitivity to structural implications with an implicit but no sensitivity with an explicit measure. Consistent with the normative analysis, the implicit measures yielded higher covariations for the common-cause than for the common-effect model. The present experiment also shows that this pattern of results is not due to differences in the learning procedure (predictive vs. diagnostic) but rather is based on differences of the assumed causal models.

## Conclusions

Research on causality belongs to the truly interdisciplinary topics of cognitive science. There are differences in the research focus between disciplines, however. Whereas the majority of studies within cognitive psychology have focused on single cause-effect relations, researchers in the areas of computer science and philosophy have become increasingly interested in complex causal structures (e.g., Glymour & Cooper, 1999; Pearl, 2000). The goal of the present research is to bridge this gap without forgetting the inherent information processing limitations of humans. It is unlikely that untutored human learners are able to store and use the complex information embodied in even fairly simple causal structures. Therefore we have focused on a more realistic task in which participants learned about different fragments of a causal model separately, and later were confronted with the task to integrate the different pieces in order to predict unobserved covariations. To solve this task correctly, knowledge about structural implications of different causal models has to be activated. Research on Bayesian networks has shown that structural information greatly simplifies causal computations but it also has demonstrated that the task still remains complex. Consistent with this analysis both experiments have demonstrated that participants showed little explicit knowledge about differences between causal models, even when the models were extremely simple. Participants' explicit judgments did not distinguish between a condition in which the target events were two effects of a common cause and a condition in which these events represented two causes of a common effect. This result raises doubts as to humans' competence to correctly learn about causal structures in the world. However, a second, more implicit measure displayed surprisingly accurate inferences. When the task required predicting individual events, participants proved sensitive to the difference between common-cause and common-effect models. This dissociation between explicit and implicit measures is consistent with the view that mental simulations of causal models support the implicit task. Generating predictions by means of a mental simulation capitalizes on causal structure without requiring explicit knowledge. As long as the mental representation mirrors the causal features of the represented domain, simulations should display the same structural constraints. Therefore causal simulations allow us to generate correct predictions without requiring complex, explicit computational inferences.

## References

- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Glymour, C. N., & Cooper, G. F. (1999). *Computation, causation, and discovery*. Cambridge: MIT Press.
- Eells, E. E. (1991). *Probabilistic causality*. Cambridge: Cambridge University Press.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Francisco: Morgan Kaufmann.
- Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 53-76.
- Waldmann, M. R., & Hagmayer, Y. (1998). Estimating causal strength: The role of structural knowledge and processing effort. *Unpublished manuscript*.
- 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.
- Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, 124, 181-206.
- Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Derry, *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

## Appendix

The following derivation shows how joint probabilities and correlations can be derived for common-cause models. In the formulas,  $s_1$ ,  $s_2$  represent the two substances and  $m$  the mutation. “~” signifies the absence of an event. The joint probability of the two substances can be computed by

$$P(s_1.s_2) = P(s_1.s_2|m) \cdot P(m) + P(s_1.s_2|\sim m) \cdot P(\sim m) \quad (1)$$

Common-cause models assume that the effects are independent conditional upon the states of the common cause, that is:

$$P(s_1.s_2|m) = P(s_1|m) \cdot P(s_2|m)$$

Thus, Equation 1 can be simplified:

$$P(s_1.s_2) = P(s_1|m) \cdot P(s_2|m) \cdot P(m) + P(s_1|\sim m) \cdot P(s_2|\sim m) \cdot P(\sim m)$$

The joint probabilities for the other patterns (e.g.,  $P(s_1.\sim s_2)$ ) can be calculated in a similar fashion. These probabilities can be used to compute *phi* correlation coefficients based on the following formula:

$$r = \frac{P(s_1.s_2) \cdot P(\sim s_1.\sim s_2) - P(s_1.\sim s_2) \cdot P(\sim s_1.s_2)}{\sqrt{(P(s_1.s_2) + P(s_1.\sim s_2)) \cdot (P(s_1.s_2) + P(\sim s_1.s_2)) \cdot (P(s_1.\sim s_2) + P(\sim s_1.\sim s_2)) \cdot (P(\sim s_1.s_2) + P(\sim s_1.\sim s_2))}}$$

This procedure of computing *phi* correlations can be applied to the patterns predicted by the participants (implicit task) as well as to the estimated conditional probabilities (explicit task).