How Causal Reasoning Can Bias Empirical Evidence

Momme von Sydow[1] (momme.von-sydow@bio.uni-goettingen.de)
York Hagmayer[1] (york.hagmayer@bio.uni-goettingen.de)
Björn Meder[1,2] (meder@mpib-berlin.mpg.de)
Michael R. Waldmann[1] (michael.waldmann@bio.uni-goettingen.de)
[1]Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany
[2]Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Abstract

Theories of causal reasoning and learning often implicitly assume that the structural implications of causal models and empirical evidence are consistent. However, for probabilistic causal relations this may not be the case. We propose a causal consistency hypothesis claiming that people tend to create consistency between the two types of knowledge. Mismatches between structural implications and empirical evidence may lead to distortions of empirical evidence. In the present research we used trial-by-trial learning tasks to study how people attempt to create consistency between structural assumptions and learning data. In Experiment 1 we show biasing of empirical evidence with causal chains even after repeated testing of direct and indirect relations. Experiment 2 investigates whether different causal models lead to different judgments, despite identical data patterns. Overall, the findings support the idea that people try to reconcile assumptions about causal structure with probabilistic data, but also suggest that this may depend on the type of causal structure under consideration.

Keywords: causal reasoning; induction; Markov condition; top-down effects; heuristics and biases

Causal Reasoning and Empirical Evidence in Covariation Assessment

Probability judgments about indirect causal relationships may be based on direct observations of covariations between events (empirical evidence) or they may be derived from top-down assumptions about the underlying causal structure (structural knowledge). The crucial advantage of causal model knowledge is that we can make inferences about relations which we have not directly observed. For example, we may first learn about a causal relation $A \rightarrow B$, and later about a causal relation $B \rightarrow C$. By combining the single links into a causal chain $A \rightarrow B \rightarrow C$ we can make inferences regarding the initial event A and the final event C. For example, deterministic causal relations warrant transitive inferences, that is, the occurrence of A allows us to infer that C is present, too (like in logical 'Modus Barbara'). However, most causal relationships tend to be probabilistic: a virus does not always cause a disease; a gene does not always cause a phenotypic trait. Crucially, in the case of probabilistic relations, transitivity relations do not necessarily hold (see Ahn & Dennis, 2000; von Sydow, Meder, & Hagmayer, 2009). However, causal models may nevertheless be used for assessing indirect relations from knowledge of direct relations, in a way that is inconsistent with direct empirical evidence.

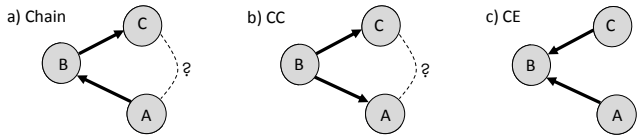

Figure 1: A causal chain, a common cause (CC) and a common effect (CE) model.

The representation of causal relationships in qualitative causal models (Gopnik et al., 2004; Rehder, 2003; Sloman, 2005; Waldmann, Hagmayer, & Blaisdell, 2006; Waldmann & Holyoak, 1992) and in causal Bayes nets (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993) suggests that people only represent direct causal relations and infer other relations from these causal models based on abstract assumptions about the structures. At the center of the Bayes net formalism is the *causal Markov condition*, which states that a variable in a causal network is conditionally independent of all other variables apart from its effects, given its direct causes. If the Markov condition holds, a causal chain (Fig. 1a) with positive direct relations, $A \rightarrow B$ and $B \rightarrow C$, entails a positive contingency between variables A and C. More specifically, the conditional probability of A given C, $P(C|A)$, is given by:

$$P(C|A) = P(B|A) \cdot P(C|B) + P(\neg B|A) \cdot P(C|\neg B) \quad (1)$$

Similarly, other indirect conditional probabilities can be derived from applying the Markov condition to the causal model. If we have a common cause model (CC, cf. Fig. 1b) $A \leftarrow B \rightarrow C$, the Bayes net formalism implies a positive relation between A and C. On the other hand, if the variables are linked in a common effect structure $A \rightarrow B \leftarrow C$ (CE, cf. Fig. 1c), no positive relation between A and C is entailed.

From a computational point of view, the Markov assumption is used as prerequisite for inducing causal structures from conditional dependency and independency relations, and as a basis for probabilistic inferences across complex causal networks (Spirtes et al., 1993; Pearl, 2000). On the other hand, the status of the Markov condition as a necessary and universal feature of causal representations has been criticized (Cartwright, 2001). However,, the status of the Markov condition in human causal reasoning is still under dispute (e.g., Rehder & Burnett, 2005; Mayrhofer, Goodman, Tenenbaum, & Waldmann, 2008).

A Causal Consistency Hypothesis

A number of studies in causal learning have shown that people tend use initial assumptions about causal models and do not tend to necessarily verify whether the assumptions underlying the model hold in the data. For example, Waldmann and Hagmayer (2001) showed that people make use of instructions regarding causal structures when assessing causal strengths, even when the data contradicted the initially suggested causal model. Waldmann, Meder, von Sydow and Hagmayer (2010) connected this research with categorization and demonstrated similar effects of category transfer with variable categorization schemes.

Similar phenomena may arise when participants are requested to make inferences about indirect relations within causal models. Previous research on inferences about indirect relations in causal chains has shown that people have a tendency to assume the Markov condition when making inferences from an initial event *A* to the final event *C*. Ahn and Dennis (2000) and Baetu and Baker (2009) have presented participants with data about direct relations between binary events. Learners' inferences about the indirect relations were consistent with the use of the Markov condition. However, they only investigated inferences in the absence of any evidence regarding the indirect relation.

Von Sydow, Meder and Hagmayer (2009) provided direct evidence about the indirect relation when learning causal chains. They showed that participants reasoned transitively (apparently assuming the Markov condition) even if the learning data provided evidence against transitivity. The present research continues in the wake of this work. Participants are again provided with data about the indirect relation. In addition, the influence of the amount of learning input, task features, and different causal structures are examined. We here particularly focus on the interplay between the implications of causal structures when the Markov condition is assumed and the observed data sample. Consider the data shown in Table 1. In these data, it holds that $P(B|A) = 0.75$ and $P(C|B) = 0.75$. Nevertheless, according to the data there is no contingency between *A* and *C*, since $P(C|A) = P(C|\neg A) = 0.5$. However, if we used these data to parameterize a causal chain $A \rightarrow B \rightarrow C$, and assumed the Markov condition, this causal model would imply that there is a positive contingency between the initial event *A* and the final effect *C* (i.e., $P(C|A) > P(C|\neg A)$, cf. Equation 1). Thus, depending on whether we assess the indirect relation between *A* and *C* directly from the data, or induce a causal model from the data and use the model to make inferences regarding indirect relations, we may arrive at very different conclusions. However, whether there is a potential tension between structural knowledge and data depends on the exact structure of the causal model. For example, a common effect model $A \rightarrow B \leftarrow C$ (Fig. 1c) does not entail a statistical dependency between *A* and *C*, as in this model the two events constitute independent causes of their common effect *B*.

Our causal consistency hypothesis suggests that when there is a mismatch between the causal model's structural implications and the observed data, people will create consistency by aligning the observed evidence with the causal model's implications. As a consequence, for an actually intransitive causal chain one should observe an overestimation of the statistical relations between the indirectly linked events *A* and *C*. For example, if learners assume the Markov condition when inducing a causal chain they should infer $P(C|A) > P(C|\neg A)$. This should also hold for common cause structures (but see von Sydow et al., 2009). By contrast, a common effect model implies no statistical dependency between *A* and *C*, as they represent independent cause of their common effect *B*. Thus, for this model there should be no conflict between the structural knowledge and the observed empirical probabilities.

Table 1: Sample of intransitive data from Experiment 1.

	A	B	C
1	present	present	present
2	present	present	present
3	present	present	absent
4	present	absent	absent
5	absent	present	present
6	absent	absent	present
7	absent	absent	absent
8	absent	absent	absent

Another potentially important factor which may affect how people deal with conflicts between structural implications and empirical evidence are the number and the focus of the test questions. In a previous study (von Sydow et al., 2009), participants were confronted with a causal chain and intransitive data, which did not show a positive statistical relation among the initial cause *A* and the final effect *C*, although the direct causal relations were positive. In these studies participants were first queried about the direct causal links before being asked about the indirect causal relation among *A* and *C*. Although participants had all relevant data available, they misjudged the relation between *A* and *C* to be positive. However, when participants are queried more often about the indirect relation, they may assess the relation directly, thereby arriving at estimates that correspond more closely to empirical probabilities.

Goals of Experiments and Hypotheses

The goal of the first experiment was to investigate how task features affect the integration of structural knowledge and empirical evidence. Participants were either asked frequently or only once about the indirect causal relation. We suspected that frequent queries would direct participants' attention to the empirical evidence regarding the indirect causal relation. Unlike in our previous studies (von Sydow et al., 2009) we presented subjects with trial-by-trial data instead of grouped data. Moreover, we used simpler dichotomous learning items, as opposed to variable category exemplars. Our goal was to find out whether these changes would make the empirical conditional probabilities more salient, thereby leading to judgments corresponding closer to the learning data. The main goal of Experiment 2 was to study other causal structures as well. While keeping the trial-by-trial

contingencies identical, we aimed to investigate whether the different possible causal structures modify the distortion of the empirical evidence. Participants were instructed about a causal chain, a common cause or a common effect model (Fig. 1). Their task was to investigate conditional probabilities between the direct and indirect causal relations. As outlined above, applying the Markov condition to these causal models leads only to a mismatch between data and model-based inferences in the chain model and in the common-cause model, but not in the common-effect model.

Experiment 1

Experiment 1 studied conditional probability judgments after successive trial-by-trial learning of two generative causal relations, $A \rightarrow B$ and $B \rightarrow C$, which were instructed to be part of a causal chain. Assuming the causal Markov condition, these relations imply a positive contingency between *A* and *C*. However, the learning data showed no statistical dependency between *A* and *C*, that is, $P(C|A) = P(C|\neg A) = 0.5$. We explored how the structure of the learning course, such as repeated queries about $P(C|A)$ might affect learners' estimates.

Methods

Design Experiment 1 had three conditions, each of which comprised eight learning phases and up to 12 test phases. Figure 2 depicts the succession of the phases. In all conditions, in the final test phase (P20) we requested estimates of the conditional probability $P(C|A)$ (Fig. 2). The state of all three events *A*, *B*, and *C* was presented simultaneously during learning, although the instructions focused participants on the direct relations of the causal chain, $A \rightarrow B$ and $B \rightarrow C$. Moreover, the directly linked pairs were circled to highlight their causal relation. After each learning phase, participants were requested to give probability estimates of the respective direct causal relation ($A \rightarrow B$ or $B \rightarrow C$). Because of the focus on direct relations we expected a substantial influence of structural knowledge.

In Condition 2 (C2), participants were also focused on the direct causal relations, but the conditional probability estimates of the indirect relation ($P(C|A)$) were additionally requested several times during learning (Fig. 2). This procedural change was intended to draw participants' attention to the indirect relation as well. We expected that repeated testing of the *A*-*C* relation would strengthen the influence of the empirical data.

Condition 3 (C3) served as control condition to ensure that participants used the scales correctly and were able to detect the zero contingency between *A* and *C*. In this condition participants only received the subset of information about the relation between *A* and *C* (cf. Fig. 2, Table 2).

Participants Sixty students from the University of Göttingen took part in the experiment for course credit or were paid 5€. They were randomly assigned to the conditions.

Procedure and Material Participants were instructed to take the role of a developmental biologist investigating newts that undergo a metamorphosis. The metamorphosis proceeded in three stages. In each stage a particular type of carotene (Alpha, Beta, and Gamma; henceforth denoted as *A*, *B*, and *C*) could occur or not occur. These carotenes may or may not affect the presence of other carotenes in a later stage.

Learning and Test Phases

		P1	P2	P3	P4	P5	P6	…	P20
Condition	C1	Learn $A \rightarrow B$, C	Test $P(B\|A)$	Learn A, $B \rightarrow C$	Test $P(C\|B)$	–	Learn $A \rightarrow$B, C	…	Test $P(C\|A)$
	C2	Learn $A \rightarrow B$, C	Test $P(B\|A)$	Learn A, $B \rightarrow C$	Test $P(C\|B)$	Test $P(C\|A)$	Learn $A \rightarrow B$, C	…	Test $P(C\|A)$
	C3	Learn $A \rightarrow C$	–	Learn $A \rightarrow C$	–	Test $P(C\|A)$	Learn $A \rightarrow C$	…	Test $P(C\|A)$

Figure 2: Design of Experiment 1

In the first condition (C1) participants were asked to assess one of the two causal relations after each learning phase (cf. Fig. 2), alternating between the first relation ($A \rightarrow B$) and the second ($B \rightarrow C$). Although in all learning phases all three events were shown, participants were only asked about the indirect relation between *A* and *C* after all eight learning phases. Condition 2 was similar, but here participants were asked to assess the indirect relation between *A* and *C* after every other learning phase (see Fig. 2). In the control condition (C3) participants only observed the relation between *A* and *C*. After every second learning phase learners had to assess $P(C|A)$.

In the two experimental conditions (C1 and C2) information about the state (present vs. absent) of all three types of carotene was presented in a trial-by-trial learning procedure. Table 2 shows the learning input. In each of the eight learning phases, 24 newts were presented in randomized order. In total, 192 newts were shown. The empirical conditional probabilities were: $P(B|A) = P(C|B) = 0.75$, $P(B|\neg A) = P(C|\neg B) = 0.25$, $P(C|A) = P(C|\neg A) = 0.5$. Thus, in the data there was a zero contingency between *C* and *A*. The probabilities entailed by the chain model assuming the Markov condition were $P(C|A) = 0.625$ and $P(C|\neg A) = 0.375$, that is, a positive contingency.

Table 2: Learning data in Experiments 1 and 2.

Pattern			Phase	Total
A	B	C	6	48
A	B	¬C	3	24
A	¬B	¬C	3	24
A	¬B	C	0	0
¬A	B	¬C	0	0
¬A	B	C	3	24
¬A	¬B	C	3	24
¬A	¬B	¬C	6	48
All			24	192

Based on the outlined design (cf. Fig. 2) three types of test phases were used, in which we assessed participants estimates of the relations between *A* and *B*, *B* and *C*, and *A* and *C*. For each judgment we used a rating scale ranging from -100 to +100. For instance, when accessing $P(C|A)$, participants were asked whether newts that had developed Alpha carotene (*A*) in the first stage rather tended to develop

Gamma carotene (*C*) or to develop no Gamma carotene (¬*C*) in the subsequent stage. The scale ranged from -100 ('newts with Alpha carotene never develop Gamma carotene') to +100 ('newts with Alpha carotene always develop Gamma carotene') in steps of 10. The middle point of the scale, 0, was labeled 'Alpha and Gamma carotene occurred together only by chance' (i.e., with $P(C|A) = 0.5$).

Results

Figure 3 shows the means of participants' ratings concerning the probability of *B* given *A*, of *C* given *B*, and of *C* given *A* over the course of learning (the different measurement points are denoted as t1 to t4).

Panels 1 and 2 reveal that participants detected the positive causal relation between the directly linked events quickly and rated the probabilities $P(B|A)$ and $P(C|B)$ roughly correctly ($P(B|A) = P(C|B) = .75$ or +50 on the used scale). Panel 3 shows learners' estimates of the indirect relation $P(C|A)$. The results suggest that in both experimental conditions (C1 and C2) the estimates were affected by structural knowledge. While in the control condition (C3) learners' estimates were around zero (corresponding to a probability of $P(C|A) = 0.5$), a very different pattern of judgments was obtained in the two experimental conditions. In both condition C1 and C2 participants gave judgments above zero; and the obtained estimates also differed from the control condition (C3). The results of C1 complement previous findings by showing that abstract causal knowledge guides learning and reasoning even when people are provided with almost 200 trials on the state of all three variables with a shown objective zero contingency. Nonetheless, the average estimate of $P(C|A)$ was actually about as high as if it were exclusively based on inference assuming the Markov condition (cf. Equation 1, $P(C|A) = .625$, or +25 on the used scale). The second condition (C2) illustrates the interplay between abstract causal knowledge and empirical evidence over the course of learning. From the first (t1) to the last measurement (t4) participants' estimates of the indirect relation *A*-*C* declined, showing the influence of the learning data. Nevertheless, even in the last test phase (t4) the judgments were above zero and higher than in the control condition. An analysis of variance of the final judgments with the three conditions as between subjects factor yielded significant results, $F(2, 56) = 6.95$, $p < .05$, $MSE = 594.0$. Additionally, we computed pair-wise comparisons, with a significant contrast between C1 and C2 ($M_{C1} = 24.5$, $M_{C2} = 8.5$) ($F(1, 56) = 17.21$, $p < .0001$), as well as between C2 and C3 ($M_{C3} = -7.5$; $F(1, 56) = 4.40$, $p < .05$) and between C1 and C2 ($F(1, 56) = 4.31$, $p < .05$).

In sum, Experiment 1 supports the idea that subjects' judgments for indirect causal relations were derived from causal model representations obeying the Markov condition, even when the available evidence indicated that this condition did not hold. C1 shows that even after a long period of learning of zero contingencies, a positive contingency between the initial and final event was inferred.The difference between C1 and C2 shows that the impact of evidence also depends on the attentional focus during learning: when the attention is directed more clearly to the indirect relation, the distortion of the learning by top-down inferences is reduced. But even after almost 200 trials the bias did not disappear completely.

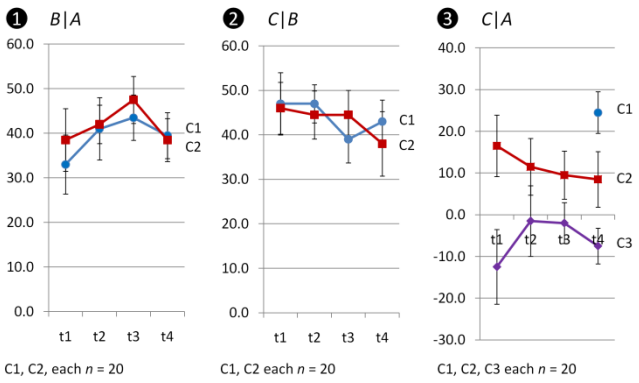

Figure 3: Mean judgments (±SE) in Experiment 1. t1 to t4 denote measurements at different points in time over the course of learning.

Experiment 2 – Causal Models

In Experiment 2 we investigated further causal structures. In addition to a causal chain we also used a common cause and a common effect model (Fig. 1). The learning data presented to participants were identical in all conditions and corresponded to Experiment 1 (Table 2). Although in the experiment participants were confronted with identical data about the three events the mapping of the events to their causal roles differed. In the chain condition *A* caused *B* and *B* caused *C*, in the common cause condition *B* was the common cause of *A* and *C*, and in the common effect condition *A* and *C* were independent causes of their common effect *B* (Fig. 1). If participants' mental causal models obeyed the Markov condition, increased values of $P(C|A)$ should be obtained in the chain and the common cause condition, but not in the common effect condition. Due to the lack of a mismatch between model and data in the CE model, participants should provide ratings corresponding to the empirical conditional probability of $P(C|A) = 0.5$.

Methods

Participants 150 students from the University of Göttingen participated for course credit or 5€. They were randomly assigned to one of the three causal model conditions.

Procedure and Material The procedure was almost identical to Condition 2 of Experiment 1 (Fig. 2), apart from the manipulations of the initial causal model assumptions. A different cover story was used, concerning the development of the metabolism of ravens. As causes and as effects we used three substances, which could be present or absent in different developmental stages of the ravens: Xantan, Yojan, and Zetosan (henceforth denoted as *A*, *B*, *C*). Participants in all condition were informed that they would have to answer questions about the potential direct causal relations (be-

tween *A* and *B*, and between *B* and *C*) as well as about the indirect relation between *A* and *C* after the learning phases. The causal links be present or absent.

The task investigated whether people assumed the Markov condition to hold when integrating single links into a larger causal structure. Although the instruction may well be interpreted to put a higher prior probability on the respective causal structures, the instructions were completely silent on whether one should assume the Markov condition. Hence, this provides a test for whether participants implicitly asserted the Markov condition and distorted the empirical probabilities accordingly.

Like in Condition 2 of Experiment 1 there were eight successive learning phases showing all three events *A*, *B*, and *C*. Again participants were focused on the respective direct causal relationship by the instructions and a circling of the directly related events. The data patterns were randomized within each learning phase; the learning data was identical in all conditions (Table 2).

In the test phases participants were again asked to assess conditional probabilities on a scale between +100 and -100 (cf. Experiment 1). When investigating the direct relations between *A* and *B* and *B* and *C* we assessed conditional probabilities in the causal direction (chain: $P(A|B)$, CC: $P(B|A)$). But note that in our learning data both conditional probabilities were identical ($P(B|A) = P(A|B)$). The wording of the question for $P(C|A)$ was identical, irrespective of condition.

Learning and Test Phases

P1	P2	P3	P4	P5	P6	…	P20
Learn $A \rightarrow B$, C	Test $P(B\|A)$	Learn A, $B \rightarrow C$	Test $P(C\|B)$	Test $P(C\|A)$	Learn $A \rightarrow B$, C	…	Test $P(C\|A)$

Figure 3: Design of Experiment 2.

Results

Figure 4 shows participants' mean estimates in the three conditions (Panel 1 – 3) across the four test phases (t1 – t4). Estimates of $P(B|A)$ and $P(C|B)$ were all positive, although they underestimated the correct value. With regard to the crucial estimate, $P(C|A)$, an inspection of the data reveals that the results of the chain condition replicate the results of Exp. 1, but that the expected effect for the CC model was not obtained. Consistent with our predictions, participants' estimates in the CE condition were close to zero. We conducted an ANOVA with the test phases (t1 to t4) as within-subject factor and causal structure (Chain, CC, CE) as between-subject factor. This resulted in a significant main effect of causal structure, $F(2, 147) = 4.34$, $p < .05$, $MSE = 2312$. No other effects proved significant. The pair-wise contrasts between the chain and the CE condition and between the chain and the CC condition yielded significant differences, $F(1, 147) = 5.31$, $p < .05$ and $F(1, 147) = 7.52$, $p < .01$. However, the contrast between the CC and CE condition was not significant: $F(1, 147) = 0.19$, $p = .66$. A test of the mean estimates of $P(C|A)$ against zero showed that only the chain condition consistently and significantly differed from zero, with no reduction over time. (Chain: t1, $t(50) = 2.49$, $p < .05$; t2, $t(50) = 2.04$, $p < .05$; t3, $t(50) = 2.98$, $p < .01$; t4, $t(50) = 3.27$, $p < .01$; CC: t1, $t(50) = -0.90$, $p = .37$; t2, $t(50) = -0.66$, $p = .51$; t3, $t(50) = .43$, $p = .66$; t4, $t(50) = -0.41$, $p = .68$; CE: t1, $t(50) = -0.19$, $p = .84$; t2, $t(50) = .99$, $p = .32$; t3, $t(50) = -0.49$, $p = .62$; t4, t1, $t(50) = .15$, $p = .88$).

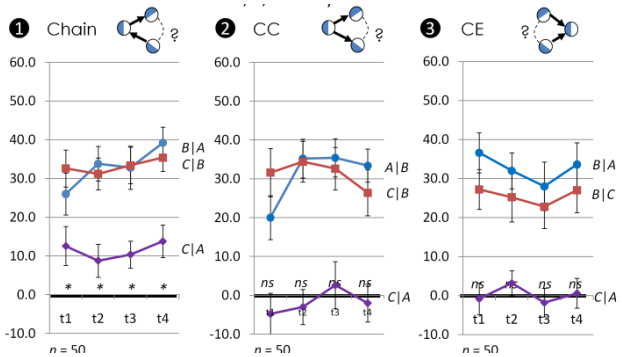

Figure 4: Means (±SE) of conditional probability estimates on a scale from -100 to +100 for the three causal structures (chain, common cause (CC), and common effect (CE)) across the four test phases (t1 to t4).

In sum, Experiment 2 replicated the biasing effect of structural knowledge with causal chains. As predicted by Bayes nets, no such effect was found for the common effect model for which top-down assumptions and empirical evidence were consistent with each other. Interestingly, no effect was obtained for the common cause model.

We can only speculate why we did not find an effect in the CC condition. Maybe the Markov condition is more intuitive in causal chains, in which the intermediate event can be easily represented as separating the initial from the final event. In contrast, screening-off relations may be harder to envision in common cause structures in which the intermediate event simultaneously causes several effects (see Cartwright, 2001). Actually, von Sydow et al. (2009) suggested that CC structures may often be interpreted to violate the Markov assumption, at least if one is concerned with the predication of attributes of a category (without representing alternative causes of the attributes). Attributes of categories are often represented as CC structures (Rehder, 2003). It has been argued that people may represent different kinds of noisy logical interaction patterns of such attributes (including XOR) (von Sydow, 2009). If such judgments correspond to a *causal* logic of CC structures, they would violate the assumption of conditional independence and unconditional positive correlation between effects (the Markov condition). However, further research is needed to connect models of noisy logical predication with theories of causal induction.

Another possibility may be that attentional factors caused the low ratings in the CC and CE condition, since we switched the direction of the question formats for the local causal links (e.g., $P(A|B)$ in the chain and $P(B|A)$ in the CC condition). Although, a predictive question format seemed to be most natural to elicit the causal representations that we

aimed to manipulate, this remains a factor that should be controlled for in future research.

General Discussion

The results of Experiment 1 corroborate our prediction that in a causal chain $A \rightarrow B \rightarrow C$ conditional probability judgments about the indirectly linked events A and C will be distorted by structural assumptions of the underlying causal model. We investigated the influence of causal inferences based on the Markov condition when learning such relations. Going beyond previous studies (Baetu & Baker, 2009; Ahn & Dennis, 2000), we provided data on the indirect relation, which showed a zero contingency. Hence, transitivity did not hold in the data (cf. von Sydow et al., 2009). In Experiment 1 we investigated this issue in a trial-by-trial learning scenario, assessing the role of repeated questions. The conditional probability estimates of $P(C|A)$ matched the values that would have been predicted if people estimated this probability based on their knowledge about the direct relations and structural assumptions about causal models (i.e., the Markov condition). This biasing effect was remarkably stable even if people obtained contradicting empirical evidence in several learning phases and were repeatedly queried about the indirect relation, which was intended to draw participants' attention to the indirect relation. With repeated queries the influence of causal reasoning became smaller, but did not disappear even after almost 200 trials.

Experiment 2 confirmed that chains and common effect structures ($A \rightarrow B \leftarrow C$) led to different judgments of $P(C|A)$ despite identical learning input. As predicted by causal Bayes nets, a biasing effect only occurred in the chain condition in which the data violated the structural constraints underlying chains. Consistent with this idea, no influence of structural knowledge was obtained for the common effect model. Interestingly, in the common cause structure ($A \leftarrow B \rightarrow C$) we did not found an influence of the causal model on participants' judgments. The reasons for this failure are unclear at present. One hypothesis may be that people find the Markov condition less plausible for these models (see also von Sydow et al., 2009). Alternatively, attentional effects during learning may have had an effect.

Taken together, the results provide further evidence for our claim that people try to create consistency between structural top-down knowledge and empirical evidence when making probabilistic causal inferences (von Sydow et al., 2009; cf. also Waldmann et al., 2010).

Acknowledgments

This research was supported by a grant 'Bayeslogik' by the *Deutsche Forschungsgemeinschaft* (DFG, Sy 111/1-2 [MvS]). We thank Johanna Frisch and Deborah Wolff for their help and assistance with the data collection.

References

Ahn, W., & Dennis, M. (2000). Induction of causal chain. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 19-24). Lawrence Erlbaum Associates, NJ: Mahwah.

Baetu, I., & Baker, A. G. (2009). Human judgments of positive and negative causal chains. *Journal of Experimental Psychology: Animal Behavior Processes. 35(2)*, 153-168.

Cartwright, N. (2001). What is wrong with Bayes nets? *The Monist, 84*, 242-264.

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.

Mayrhofer, R., Goodman, N. D., Waldmann, M. R., & Tenenbaum, J. B. (2008). Structured correlation from the causal background. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 303-308).

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, MA: 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.

Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*, 264-314.

Sloman, S. (2005). *Causal Models. How People Think about the World and Its Alternatives*. Cambridge, MA: Oxford University Press.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

von Sydow, M. (2009). On a general Bayesian pattern logic of frequency-based logical inclusion fallacies. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 248-253). Austin, TX: Cognitive Science Society.

von Sydow, M., Meder, B., & Hagmayer, Y. (2009). A transitivity heuristic of probabilistic causal reasoning. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 803-808). Austin, TX: Cognitive Science Society.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*, 27-58.

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science, 15*, 307-311.

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., Meder, B., von Sydow, M. & Hagmayer, Y. (2010). The Tight Coupling between Category and Causal Learning. *Cognitive Processing, 11*, 143-158.