Recognizing Estimation-Relevant Structural Differences in Predictive Causal Reasoning

Simon Stephan (simon.stephan@psych.uni-goettingen.de), Sarah Placì & Michael R. Waldmann
Department of Psychology, University of Göttingen

Abstract

Predictive causal reasoning requires estimating the probability of an effect given a cause, $P(e|c)$. This estimate is context-dependent: a change in the causal structure may render a previous estimate biased. We investigate whether reasoners can discern estimation-relevant from estimation-irrelevant structural changes of causal models. Using the Causal Bayes Net framework, we compare situations where adding a causal link is either estimation-relevant (e.g., introducing a confounder) or irrelevant. In an experiment ($N = 204$), participants successfully discerned these cases, adjusting estimates in the normative direction only when structural shifts warranted it. However, we also found that few participants provided normative point-estimates. We identified two dominant heuristic strategies: a "normative change direction" strategy, where reasoners adjusted estimates but *overemphasized* alternative causes, and a persistent "alternative cause neglect" strategy, where participants ignored alternative causes entirely. Our findings suggest that while people recognize whether structural shifts require re-estimation, they struggle with the numeric integration of multiple causal pathways.

Keywords: predictive reasoning, causal reasoning, causal structure, confounding, causal Bayes nets, neglect of alternative causes

Introduction

We often find ourselves in situations where we want to predict an outcome. One way to do this is to rely on what we know about the outcome's causes and how they are embedded in the causal model reflecting the target situation. For example, imagine a parent who wonders whether their child will get sick after playing with another child who turns out to be suffering from hand-foot-and-mouth disease. Or imagine a publishing house trying to predict the probability that book sales will increase by a certain amount after a famous author has won a prestigious prize. Formally, what a reasoner wants to estimate in these cases is $P(e|c)$, the predictive probability of an effect given knowledge about the presence of a known cause.

Such causal estimates are context-dependent. A causal estimate obtained in one situation may become biased when the underlying causal structure changes. Imagine, to give yet another example, policy makers seeking to improve academic success (E) in one country based on data from another. In Country 1, study habits (A) and parental support (C) are independent drivers of success. In Country 2, however, a collectivist culture links the two: increased study effort (A) causes an increase in parental support (C). In other words, in this target environment A and C are confounded. If policy makers fail to recognize that the $A \rightarrow C$ link exists in the novel context, their prediction of academic success given parental support, $P(e|c)$, will be wrong. As Cartwright and Hardie (2012) have put it: even solid evidence that "it worked here" (i.e., in one context) does not automatically warrant the conclusion that "it will work there" (i.e., at all or to the same extent in a different context).

The ability to rely on causal knowledge to make predictions is a hallmark of human reasoning, and has therefore been the subject of numerous studies (see Rottman & Hastie, 2014; Waldmann, 2017, for overviews). Most studies presented participants with only a single causal structure and tested how they respond to changes in the size of its parameters, or to changes in the status of observable variables (see, e.g., studies on the causal Markov condition, Rehder, 2014; Rehder & Burnett, 2005; Rehder & Waldmann, 2017). An aspect that previous studies did not address, however, is how reasoners adjust their estimates when moving from a situation governed by one structure to one governed by a different one. We look at how reasoners' predictive causal judgments (i.e., estimates of $P(e|c)$) are influenced by such structural causal changes. To succeed in causal transfer, a reasoner must analyze the differences in the underlying causal structures and decide whether they are estimation-relevant or estimation-irrelevant.

Estimating P(e|c) When Structures Change

As theoretical framework, we use causal Bayesian nets (CBN) (Glymour, 2001; Pearl, 2000; Spirtes et al., 1993; Spohn, 2001), which have proven to be a fruitful normative benchmark for human causal learning and reasoning (see Rottman, 2017; Rottman & Hastie, 2014; Sloman, 2005, for overviews).

Causal structures – and the differences between them – can be arbitrarily complex. This means we had to decide which structures should be compared. We decided to begin our investigation with the structures illustrated in Pair 1 in Fig. 1. Each structure of Pair 1 represents one situation. We model the task where a reasoner is confronted with a structural causal change between the two situations.

What differs between the two situations of Pair 1 is the presence vs. absence of a causal arrow between cause A and cause C. The left graph in Pair 1 is a simple common effect CBN in which causes C and A independently cause the target effect E. This causal model has been the focus of analysis in numerous theories, such as Cheng's (1997) power PC theory and Grif-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

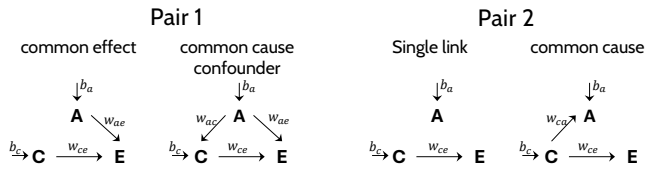

Figure 1: Pairs of causal Bayesian networks (CBNs) we compared in the experiment. The additional causal link in Pair 1 is estimation relevant. In Pair 2, it is not.

fiths and Tenenbaum's (2005) causal support model (see also Meder et al., 2014; Stephan & Waldmann, 2018). This basic common effect CBN has also been studied in numerous experiments that tested lay people's predictive reasoning abilities (Fernbach & Rehder, 2013; Fernbach et al., 2010, 2011a).

The right graph of Pair 1 differs in one structural aspect from the left one: it contains an additional arrow going from cause A to cause C, turning A into a common-cause of both C and E. Technically, this structure represents a so-called common cause confounder CBN. One reason why we included a common cause confounder CBN is that confounding, although a prominent topic in normative disciplines of causal inference (cf. Pearl, 2000; Pearl & Mackenzie, 2018), has so far received little attention in psychological studies on causal reasoning (but see Meder et al., 2009; Meder et al., 2006, 2008).

A normative reasoner should conclude that this structural causal difference is *estimation-relevant* with respect to $P(e|c)$. Assuming that all shared parameters of the two structures are identical, and assuming that the additional link between A and C under the common-cause confounder CBN indicates a positive causal influence of A on C, $P(e|c)$ must be higher under the common-cause confounder CBN than under the independent-causes CBN. Conceptually, the existence of a generative causal arrow from A to C implies that observing C provides *diagnostic* evidence for the presence of an additional cause of target effect E, A. This is not the case under the left CBN, where C and A are independent causes of E.

In all structures, the parameters, w_i, written on the endogenous causal arrows represent the strengths of these individual links, and the parameters, b_i, written on the exogenous causal arrows denote the base rates of the variables. For example, b_c represents the base rate of C occurring, and w_{ce} represents the individual causal strength with which C causes E. It corresponds to the probability of E given C in the absence of alternative causes (here A) of E (cf. Cheng, 1997; Griffiths & Tenenbaum, 2005). This means that if we were in a situation where A did not exist, w_{ce} would correspond to our target estimate $P(e|c)$. Yet, since A does exist in the present structures, a normative reasoner who aims to compute $P(e|c)$ must incorporate its influence on E. A occurs with probability b_a and causes E with an individual strength of w_{ae}.

In the case of binary variables, the typical assumption is that C and A combine their influences according to a "noisy-or" rule (or noisy-OR gate), i.e., that E can independently be caused by C or A (or by both). Under these assumptions, our target estimate for the simple common effect CBN with two independent causes, $P(e|c)_{icce}$[1], is given by the following equation see also Cheng, 1997:

$$P(e|c)_{icce} = w_{ce} + b_a \cdot w_{ae} - w_{ce} \cdot b_a \cdot w_{ae}. \quad (1)$$

The equation shows that $P(e|c)_{icce}$ increases as the strengths of the contributing causes (w_{ce} and w_a) and the base rate of the alternative cause (b_a) increase.

We now consider a reasoner who has estimated $P(e|c)_{icce}$ under the common effect CBN and then moves to a situation where the underlying structure is the common cause confounder CBN (right CBN of Pair 1 in Fig. 1). Normatively, the existence of the additional causal arrow between C and A requires the recalculation of the target estimate, now being $P(e|c)_{cccf}$[2]. On average, $P(e|c)_{cccf}$ will be higher than $P(e|c)_{icce}$ because the additional causal link between A and C implies that an observation of C provides "diagnostic" evidence for A's presence[3]. If A is a confounding variable of C and E, then C and A are positively correlated, which means that observing the presence of C increases the probability of A, i.e., $P(a|c) > P(a) > P(a|\neg c)$.

A useful first step in the calculation of $P(e|c)_{cccf}$ is the calculation of $P(a|c)_{cccf}$, the probability that A is present given that C is present. We assume that A and C's background (b_c) also combine according to a noisy-OR gate (cf. Meder et al., 2014, who also modeled diagnostic inferences this way). The equation for $P(a|c)_{cccf}$ is then given by:

$$\begin{aligned} P(a|c)_{cccf} &= \frac{P(c|a)_{cccf} \cdot P(a)_{cccf}}{P(c)_{cccf}} \\ &= \frac{(w_{ac} + b_c - w_{ac} b_c) \cdot b_a}{(w_{ac} + b_c - w_{ac} b_c) \cdot b_a + b_c (1 - b_a)} \quad (2) \\ &= \frac{w_{ac} b_a + b_c b_a - w_{ac} b_c b_a}{w_{ac} b_a + b_c - w_{ac} b_c b_a}. \end{aligned}$$

The first part simply re-writes $P(a|c)_{cccf}$ in accordance with Bayes theorem to highlight the relevant probabilities. The second step translates these probabilities into causal parameters. The last step provides a simplification. It can be seen that the equation formalizes the following logic: first, C provides more diagnostic evidence of A, the stronger the link between A and C is. In addition to the strength of the link, the amount of diagnostic evidence that C provides of A also depends on the causes' base rates: C becomes less diagnostic

[1] We here use *icce* as an abbreviation of independent causes of a common effect, referring to the common effect structure shown in Fig.1.

[2] We here use *cccf* as an abbreviation of common cause confounder.

[3] Boundary conditions are cases where either b_c or b_a, or both, are 1.0. In theses cases $P(e|c)_{cccf} = P(e|c)_{icce}$.

for A the more likely it is that C and A are present due to background causes. In the extreme case, for instance, where C or A are always present, observing C has no additional inferential value, implying $P(a|c) = P(a)$. Conversely, the diagnostic relevance of C increases the lower C's and A's base rates are. Eq. 2 also shows that C's diagnostic value is at maximum whenever its base rate is zero ($b_c = 0$; assuming that all other parameters take on values greater than zero); if the only existing cause of C is A, then observing the presence of C implies the presence of A because A must have caused the presence of C.

In a next step, $P(e|c)_{cccf}$ can be calculated by inserting $P(a|c)_{cccf}$ into the following equation:

$$P(e|c)_{cccf} = w_{ce} + P(a|c)_{cccf} \cdot w_{ae} - w_{ce} \cdot P(a|c)_{cccf} \cdot w_{ae}. \quad (3)$$

Substituting $P(a|c)_{cccf}$ and factoring the common terms yields the following equation:

$$P(e|c)_{\text{cccf}} = w_{ce} + (1 - w_{ce})\, w_{ae} \frac{w_{ac} b_a + b_c b_a - w_{ac} b_c b_a}{w_{ac} b_a + b_c - w_{ac} b_c b_a}. \quad (4)$$

What Equations 1 to 4 show is that a rational reasoner who shifts from a simple common effect CBN to a common cause confounder CBN should increase the estimates of $P(e|c)$. Conversely, a rational reasoner going in the opposite direction should respond with lower estimations of $P(e|c)$.

Heuristic Response Strategies

In addition to the normative solution, we consider three heuristic response strategies a reasoner may rely on, beginning with the least cognitively demanding (aside from guessing). One heuristic strategy would be to focus solely on the causal link between C and E, and to ignore the alternative cause A. Since the link between C and E remains the same, a reasoner adopting that strategy will give identical ratings in both situations. A series of studies by Fernbach and colleagues (Fernbach & Rehder, 2013; Fernbach et al., 2010, 2011a) on predictive reasoning under the simple common effect CBN (left in Pair 1) has shown that people widely apply this strategy. Fernbach and Rehder (2013) conclude that one reason for participants' failure to incorporate knowledge about unobserved alternative causes is the computational difficulty of having to integrate over their possible states, which would require the multiplication of their base rates and strengths. They conclude: "Situations in which variables or their states are unidentified invite reasoners to reduce effort by invoking shortcuts that can lead to error" (p. 84 Fernbach & Rehder, 2013). When estimating $P(e|c)_{icce}$, the applied cognitive shortcut can be the neglect of alternative causes.

The studies demonstrating the existence of this neglect of alternative causes did not model contexts where a reasoner encounters a structural causal change between two situations, however. We looked at a case where the change involves a novel link between A and C (i.e., confounding). We assume that this will make A's estimation-relevance more salient, which should result in an attenuated tendency to neglect it. Participants who adopt the alternative-cause neglect strategy and who solely focus on the target link between C and E should provide estimates that correspond to w_{ce}. We call this strategy "alternative cause neglect".

Participants who recognize A's relevance should give ratings higher than w_{ce}. What it takes to distinguish between both structures of Pair 1, however, is the realization that observing C has different implications for the probability of A under the simple common effect CBN and the common cause confounder CBN: $P(a|c)_{cccf} > P(a|c)_{icce}$. A reasoner who recognizes A's estimation-relevance but fails to understand that $P(a|c)$ differs between the structures would provide the same $P(e|c)$ ratings under both structures, but their estimates would be higher than those given by a reasoner who entirely ignores A. For example, if reasoners who adopt this strategy integrate C and A's influence according to a noisy-OR integration rule, then their estimates should correspond to $w_{ce} \cdot b_a \cdot w_{ae}$, the normative value of $P(e|c)$ under the simple common effect CBN. However, they should underestimate the link under the common cause confounder CBN. We call this strategy "simple alternative cause integration".

Another heuristic strategy might be to simply increase or decrease one's estimation of $P(e|c)$ whenever a causal arrow gets added or removed in a novel situation. In the case of Pair 1, the application of this strategy would result in an estimation difference in the normative direction. Yet, not every addition or removal of a causal arrow is estimation-relevant. Thus, one way to see if reasoners resort to this heuristic strategy is to look at conditions under which the addition or subtraction of a causal arrow should not lead to a change in $P(e|c)$. The two situations grouped into Pair 2 of Fig. 1 instantiate such a condition. Unlike in Pair 1, a normative reasoner should conclude that this structural causal difference is *estimation-irrelevant*. As the added causal arrow connects C and A, and as A is itself not causally linked to E, the value of $P(e|c)$ remains unaltered. A pattern where $P(e|c)$ estimations increase under structures containing an additional causal arrow would be evidence of the use of this strategy. We refer to it as "unspecific additional link increase".

A third heuristic strategy we consider is one in which a reasoner (1) reliably detects whether the presented structural causal difference has estimation relevance, (2) understands in which direction the estimate needs to change, but (3) fails to conduct the exact normative computation. Participants in this category should produce a rating pattern where estimates of $P(e|c)$ go into the normative direction – and display invariance in the case of Pair 2 – but should not converge on the exact estimate entailed by the equations. We refer to this strategy as "normative change direction".

We tested these possibilities in our Experiment. To foreshadow, we found that most participants were best described by the "normative change direction" strategy. They recognized estimation-relevant structural differences and appeared

Table 1: CBN parameters and resulting values for $P(e|c)$ and $P(a|c)$ used in the experiment.

	b_c	b_a	w_{ce}	w_{ca}/w_{ac}	w_{ae}	$P(a\|c)$	$P(e\|c)$
common effect	0.2	0.2	0.3	0.9	0.9	0.2	0.43
common cause confounder	0.2	0.2	0.3	0.9	0.9	0.53	0.64
single link	0.2	0.2	0.3	0.0	0.0	0.2	0.3
common cause	0.2	0.2	0.3	0.9	0.0	0.2	0.3

to understand in which direction $P(e|c)$ needs to change. Yet, they mostly did not succeed in calculating the precise value. In addition to this group of participants who responded in the normative direction, we also identified a substantial group of participants who seemed to have ignored A entirely, i.e., showed the "alternative cause neglect" strategy. This suggests that reasoners' tendency to ignore alternative causes is a prevalent heuristic, even when the relevance of alternative causes is particularly salient.

Experiment

The goal of the experiment was to see if lay people react normatively to the encountered structural causal differences illustrated in Fig. 1. The selected set of parameter values is listed in Table 1. With these parameter values, $P(e|c)$ clearly differs between the two structures. We kept the values of parameters that are shared by the different structures constant, to ensure that differences in $P(e|c)$ solely result from structural differences. The experimental materials and the data analysis files are available in a GitHub repository here: https://github.com/SimonStephan31/prediction_under_causal_changes

Methods

Participants Two hundred and four participants recruited via www.prolific.co participated in this online experiment (M_{age}=39.2 years, SD_{age}=13.6 years, age range 18 to 72 years). Participants had to be at least 18 years old, be native English speakers, have an approval rate of at least 90%, participate via a PC or laptop, not have taken part in pilot studies, and have at least some formal educational qualification (i.e., participants reporting "no formal qualifications" were excluded). We also excluded participants from the analyses who failed to respond correctly to two comprehension test questions administered at the end of the study.

Sample size rationale The sample size was determined a priori in R (R Core Team, 2025) with the pwrss package (Bulus, 2023), based on pilot data. We planned two sets of tests: a directed interaction contrast on mean $P(e \mid c)$ ratings, testing whether the difference between structures was larger (in the normative direction) for Pair 1 than for Pair 2. This was approximated as an ANOVA interaction with $\eta_p^2 = .07$, and powered at .90.

Second, we planned two one-sided one-sample proportion tests to assess whether a majority of participants favored the common-cause structure in Pair 1 and whether they gave equal ratings in Pair 2. We wanted to determine with 0.9 power whether an observed proportion of 64% differs from 50% (the smallest effect size we saw in a pilot study).

The analyses indicated a required sample size of $n = 51$ per condition (including a between-subjects counterbalancing factor that we will explain below), yielding a total of $N = 204$. Data collection ended once this sample size was reached.

Design, materials, and procedure Participants were assigned either to the condition in which we compared the CBNs shown in Pair 1 of Fig. 1 or to the condition in which we compared the CBNs shown in Pair 2 of Fig. 1. To test how participants respond to the causal differences between the structures of a pair, ratings of $P(e|c)$ for each structure within each pair were measured within subjects. Which structure of the presented pair was shown on the left side and which one on the right side was counterbalanced between subjects. This means that there were four different participant groups (Pair 1 vs. Pair 2 $\times$ Side-order 1 vs. Side-order 2).

Participants read about a fictitious biological scenario about the mating success of a New Guinean bird-of-paradise species (cf. Meder et al., 2006). We introduced two subspecies. Each of the subspecies was associated with one of the two causal structures of the respective pair of structures (cf. Fig. 2). Two causal factors were mentioned: the ability to produce a distinct sound (C), and the presence of a hormone, Amino-Regulin (A). The descriptions were accompanied by the corresponding causal graphs. An example from the Pair 1 condition is shown in Fig. 2. The description also included information on the biological mechanisms linking the variables mentioned to influence mating success. In the Pair 1 condition, the hormone (A) was described as also influencing the ability to produce the distinct sound (C) in one of the sub-populations but not in the other, due to a genetic mutation. In the Pair 2 condition, it was pointed out that the hormone (A) had no influence on mating success (E), contrary to what the biologists had initially assumed. The instructions also excluded any additional causes of mating success.

On a subsequent screen, participants learned about the meaning of the causal strength parameters shown in the graphs. For example, in the Pair 1 condition, the arrow from sound production (C) to mating success (E) was described as indicating a 30% chance of finding a mating partner for males who can produce the sound, even in the absence of the hormone.

Subsequently participants completed the test phase, which requested them to provide estimates of $P(e \mid c)$. Participants were shown two male birds (one from each subspecies, presented next to each other) and were asked to rate the likelihood that each would find a mating partner. Participants saw that both birds had the ability to produce the distinct sound ($C = 1$) and that the status of the hormone was unobserved (A =?). Ratings were provided on continuous numerical sliders. The causal structures remained visible on top of the screen during this phase to minimize memory demands.

Participants also answered a comprehension check question about the causal role of A to ensure that they understood

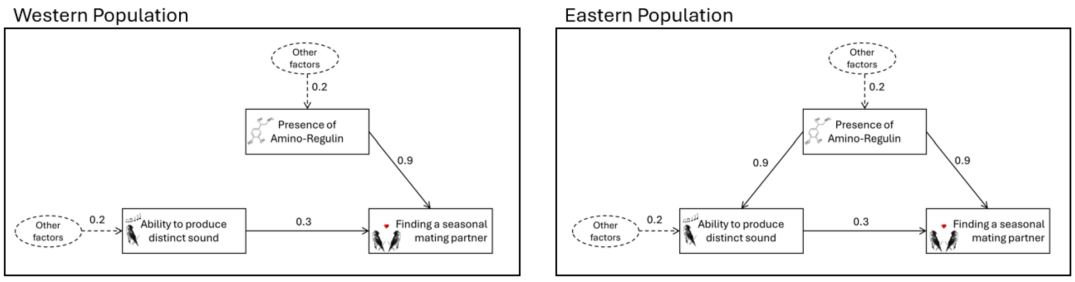

Figure 2: Illustration of how the structures of Pair 1 from Fig. 1 were shown to participants in the experiment.

the structural causal differences. They then provided demographic information and were debriefed.

Results and discussion

Fig. 3 shows participants' $P(e|c)$ ratings. The mean ratings were close to the normative values in both the Pair 1 (M_{S1} = 0.42, 95% CI [0.38, 0.45]; M_{S2} = 0.61, 95% CI [0.56, 0.65]) and the Pair 2 (M_{S1} = 0.36, 95% CI [0.33, 0.40]; M_{S2} = 0.40, 95% CI [0.36, 0.45]) condition. An interaction contrast testing whether the change from Structure 1 to Structure 2 was larger in the Pair 1 than the Pair 2 condition was significant, $t(202) = 4.93$, $p < .001$, $\eta_p^2 = .107$. This shows (1) that our participants were able to discern the estimation-relevant from estimation-irrelevant structural causal differences. (2) In the case of estimation-relevant differences (Pair 1), they were able to adjust their estimate in the normative direction when moving from the first to the second causal structure.

While participants' mean $P(e|c)$ ratings were close to the normative values, it can also be seen from the distributional information included in Fig. 3 that there were substantial interindividual differences, especially in the Pair 1 condition. The jittered dots in Fig. 3 show that many participants provided $P(e|c)$ ratings that corresponded to the causal strength C (w_{ce}). These participants seem to have expressed alternative-cause neglect (direct-link bias), as documented in the literature (Fernbach & Rehder, 2013; Fernbach et al., 2010, 2011a). While clusters of participants who neglected the alternative cause can be seen in both causal structures of Pair 1, this cluster became smaller under the cccf CBN. To better understand the differences, we clustered participants by rating change (decrease, increase, or no change; coded from the perspective of a participant moving from the icce CBN to the cccf CBN)[4]. The results are shown in Fig. 4. As can be seen there, while 37.3% gave identical ratings, 55% increased their $P(e|c)$ estimates. This is what we had anticipated because we assumed that a causal link between C and A would make it harder to ignore A. At the same time, in our Pair 2 control condition, only about a fifth (21.6%) of the participants gave higher ratings for the structure that included an additional causal arrow. A directed 2-sample test for equality of proportions with continuity correction indicated that the difference between Pair 1 and Pair 2 is significant, $\chi^2(1) = 23.88$, $p < .001$ (one-tailed). This finding corroborates the assumption that participants were able recognize estimation-relevant causal differences, and speaks against the alternative explanation that they simply relied on the "unspecific additional link increase" heuristic.

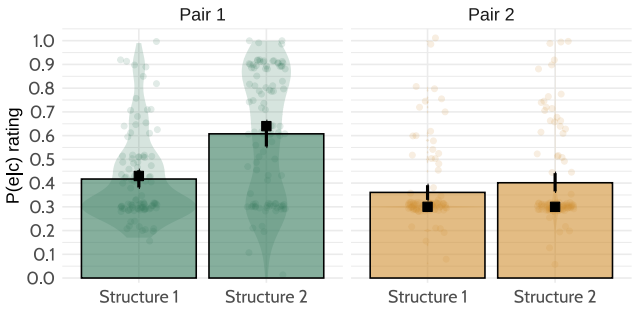

Figure 3: Results of the experiment. Bars represent means and error bars 95% CIs. Jittered dots show participants' individual ratings, violin plots their distribution. Squares denote normative values. For Pair 1, Structures 1 and 2 refer to the icce and the cccf CBN, respectively. For Pair 2, Structures 1 and 2 refer to the sl and the ce CBN, respectively.

Although we observed in the Pair 1 condition a tendency to adjust estimates in the normative direction, and although the means were close to the normative values, many participants failed to estimate the normative value. There was substantial interindividual variability, especially for the cccf CBN. As can be seen by the violin plots in Fig. 3, the distribution of the cccf CBN ratings shows two clusters of ratings, one around 0.3 and another 0.9. At the same time a relatively small number of values fell into the region of the normative value (of 0.64). This shows that most participants who changed their estimate in the normative direction still struggled to arrive at the normative solution (see Rottman & Hastie, 2014, for a list of factors that may explain deviations). In sum, the results suggest that participants are best described by the strategy we call "normative change direction".

[4]We tolerated a margin of noise of three points on our continuous rating scale, which means that participants with a rating difference smaller than four points were assigned to the no-change cluster.

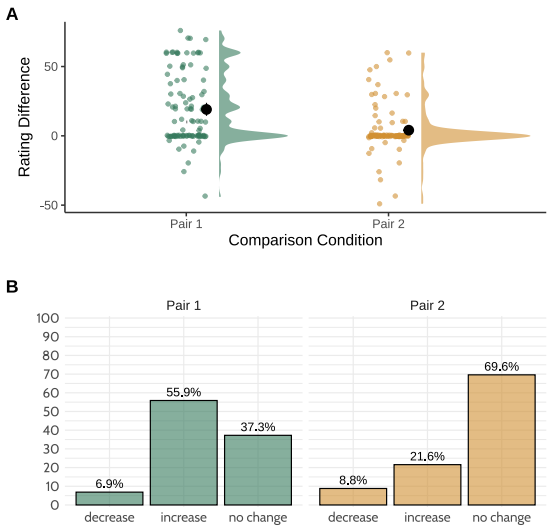

Figure 4: P(e|c) rating changes in the experiment (A) and the corresponding rating change clusters (B).Category labels ("decrease" and "increase") are coded from the perspective of a participant who first rated Structure 1 and then Structure 2.

General Discussion

A major challenge in human causal cognition is the ability to distinguish causally relevant from irrelevant changes between causal contexts. This discrimination is a prerequisite for successful causal transfer: the ability to apply knowledge from a source environment to a different, yet causally similar target (Bareinboim & Pearl, 2012; Cartwright & Hardie, 2012; Pearl & Bareinboim, 2011). We here began to investigate reasoners' responses to such changes by asking participants to make predictive probability judgments ($P[e|c]$) in the context of structural causal changes. We found that reasoners have this capacity, as our participants were able to adjust their predictions in the normative direction – or keep their original estimate in the case of estimation-irrelevant changes.

Despite an overall normative tendency, we also observed that many individual estimates deviated from the normatively correct values. A substantial number of our participants neglected the alternative cause A completely, a tendency well documented in the literature (Fernbach & Rehder, 2013; Fernbach et al., 2010, 2011a). At the same time, a noteworthy novel finding is that the inclusion of an additional link between the alternative cause and the target cause (i.e., confounding) helps reasoners overcome the alternative cause neglect. However, many of our participants who considered alternative causes still failed to integrate both relevant causal paths normatively. An observed tendency in participants who recognized the relevance of alternative causes was to *overemphasize* their contribution. Hence, a key complication in people's predictive reasoning seems to be the correct integration of multiple causal pathways that lead to a target effect (cf. Rottman & Hastie, 2014, 2016). As Rottman and Hastie (2014) have pointed out: "Performing the full Bayesian calculations requires reasoning about many nodes simultaneously, understanding how causes combine in complex ways (e.g., noisy-OR rule), and understanding how to use multiple parameters for a single inference" (p. 27). This is a cognitively demanding task.

The study employed a "learning from description" type of paradigm (Hertwig & Erev, 2009; Hertwig et al., 2018), for three main reasons: (1) Learning by description is an important task in its own right. We often use text books or receive verbal instructions to learn about causal relationships. (2) We wanted to rule out uncertainty about the underlying CBNs and their parameter sizes (see Griffiths & Tenenbaum, 2005; Lu et al., 2008; Meder et al., 2014, for studies on the role of structure and strength uncertainty in causal learning). (3) We wanted to stay close to previous studies on the neglect of alternative causes (Fernbach & Rehder, 2013; Fernbach et al., 2010, 2011a, 2011b), which also applied a learning from description paradigm. Our paradigm resembled the one used by Fernbach and Rehder, 2013, who also gave their participants graphical illustrations of the instructed common effect CBN and its parameters.

An aspect about our experiment that may be considered problematic is that the additional link we added to the second causal model in our Pair 2 control condition goes from C to A and not from A to C, like in the Pair 1 condition. Our reason for doing so was that we wanted to avoid the presentation of a causal chain model ($A \rightarrow C \rightarrow E$), for which test cases like the one we used in our experiment have been found to elicit lower than normative ratings due to a violation of the causal Markov condition (see, e.g., Rehder, 2014; Rehder & Waldmann, 2017). An alternative concern might be that our decision to add a link that goes from C to A might risk the elicitation of the so-called strength dilution effect (Stephan et al., 2023). However, in this study we showed that the dilution effect requires continuous target causes and does not occur with binary ones, which our cause variable C was in the present experiment. In our experiment, dilution was not observed, which is consistent with this theory.

We investigated only four causal structures, grouped into two pairs. We plan to extend the set of target causal structures in future studies. Moreover, we here began to investigate reasoners' sensitivity to estimation-relevant causal changes by focusing on *structural changes*, i.e., cases where causal links are added or removed. In future studies, we also plan to look at the estimation-relevance of *parametric changes*. To give just one example, changing the magnitude of b_c constitutes an estimation-irrelevant change under the simple common effect CBN, but an estimation-relevant change under the common cause confounder CBN. It is an open question whether reasoners are sensitive to this more subtle manipulation.

Acknowledgments

We thank Tom Wysocki for helpful comments. This work was supported by a Reinhart Koselleck project (WA 621/25-1), funded by the Deutsche Forschungsgemeinschaft (DFG).

References

Bareinboim, E., & Pearl, J. (2012). Transportability of causal effects: Completeness results. *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 698–704.

Bulus, M. (2023). *pwrss: Statistical power and sample size calculation tools* [R package version 0.3.1]. https://cran.r-project.org/package=pwrss

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press. https://doi.org/https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405.

Fernbach, P. M., Darlow, A., & Sloman, S. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, *21*, 329–336.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011a). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, *140*, 168–185.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011b). When good evidence goes bad: The weak evidence effect in judgment and decision-making. *Cognition*, *119*(3), 459–467.

Fernbach, P. M., & Rehder, B. (2013). Cognitive shortcuts in causal inference. *Argument & Computation*, *4*, 64–88.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523.

Hertwig, R., Hogarth, R. M., & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science*, *27*(2), 123–128. https://doi.org/10.1177/0963721417740645

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*, 955–982.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & cognition*, *37*, 249–264. https://doi.org/https://doi.org/10.3758/MC.37.3.249

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2006). Understanding the causal logic of confounds. In R. Sun (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 579–584). Mahwah, NJ: Erlbaum.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review*, *15*, 75–80. https://doi.org/https://doi.org/10.3758/PBR.15.1.75

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*(3), 277–301.

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

Pearl, J., & Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 247–254.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, *72*, 54–107.

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

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition*, *45*(2), 245–260. https://doi.org/https://doi.org/10.3758/s13421-016-0662-3

Rottman, B. M. (2017). The acquisition and use of causal structure knowledge. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 85–114). New York: Oxford University Press.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*, 109–139. https://doi.org/https://psycnet.apa.org/doi/10.1037/a0031903

Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? markov violations, weak inferences, and failures of explaining away. *Cognitive Psychology*, *87*, 88–134. https://doi.org/https://doi.org/10.1016/j.cogpsych.2016.05.002

Sloman, S. A. (2005). *Causal models: How we think about the world and its alternatives*. Oxford University Press.

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

Spohn, W. (2001). Bayesian nets are all there is to causal dependence. In M. C. Galavotti, P. Suppes, & D. Costantini (Eds.), *Stochastic causality* (pp. 157–172). Chicago: University of Chicago Press.

Stephan, S., Engelmann, N., & Waldmann, M. R. (2023). The perceived dilution of causal strength. *Cognitive Psychology*, *140*, 101540.

Stephan, S., & Waldmann, M. R. (2018). Preemption in singular causation judgments: A computational model. *Topics in Cognitive Science*, *10*(1), 242–257.

Waldmann, M. R. (Ed.). (2017). *The Oxford handbook of causal reasoning*. New York: Oxford University Press. https://doi.org/https://doi.org/10.1093/oxfordhb/9780199399550.001.0001