Heuristics in Covariation-based Induction of Causal Models: Sufficiency and Necessity Priors

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

Abstract

Our main goal in the present set of studies was to re-visit the question whether people are capable of inducing causal models from covariation data alone without further cues, such as temporal order. In the literature there has been a debate between top-down and bottom-up learning theories in causal learning. Whereas top-down theorists claim that in structure induction, covariation information plays none or only a secondary role, bottom-up theories, such as causal Bayes net theory, assert that people are capable of inducing structure from conditional dependence and independence information alone. Our three experiments suggest that both positions are wrong. In simple three-variable domains people are indeed often capable of reliably picking the right model. However, this can be achieved by simple heuristics that do not require complex statistics.

Keywords: causal induction; causal Bayes nets; heuristics

Introduction

How are the causal regularities in the world learned? One popular answer can be traced back to the philosopher Hume (1748/1974), who famously argued that a temporal order (causes precede their effects) along with covariation information are the basis of the inference about the existence of a causal relation. Hume’s analysis creates a puzzle, though, when we consider more complex causal models (see also Waldmann & Hagmayer, in press). Our everyday knowledge is not neatly organized in single cause-effect relations but is interrelated in complex models with multiple causes of common effects (common effect model), common causes of multiple effects (common cause model), or causal chains. Just looking at pairwise covariations will often not help us to recover the underlying causal model even when the temporal order cue is available. Lagnado et al. (2007) have therefore proposed the view that people use multiple cues, such as temporal order, interventions, or prior knowledge, to form hypothetical models (see also Waldmann, 1996). Covariation information may be used to validate these hypotheses but it plays a subordinate role entering the induction process after the other cues have been applied in a top-down fashion. Fernbach and Sloman (2009) have even argued that people “do not rely on covariation when learning the structure of causal relations” (p. 678).

Our main goal in the present research is to re-visit the question whether people are really incapable of inducing causal models from covariation alone in situations in which no other cues are available. To anticipate the results we have found an amazing ability to select the right causal model based on covariation data alone, which surpasses previous
use conditional and unconditional probability information in
the data to assess the likelihood of the competing model.
In sum, whereas Gopnik and colleagues argue that people
are capable of recovering causal models from covariation
alone using conditional dependence and independence in-
f ormation, other researchers (Fernbach & Sloman, 2009;
Lagnado et al., 2007) doubt that people have such sophist i-
cated statistical competencies.

Empirical Evidence
Steyvers et al. (2003) introduced the alien mind-reading
paradigm to test whether people are capable of inducing the
causal structure underlying three events based on infor-
mation about conditional dependence and independence.
They presented subjects with three mind readers who, based
on mind reading, had particular thoughts or not. Overall
Steyvers et al. observed above-chance but poor performance
 when only covariation information was available. Although
Steyvers et al. (2003) claimed to have studied purely bot-
tom-up learning, in all experiments participants were pro-
vided with graphs showing alternative models (e.g., com-
mon cause, common effect model) between which subjects
had to choose. Thus, subjects could use these graphs as
potential top-down hypotheses. Moreover, although lear-
ners’ performance is partly consistent with the proposed
Bayesian learning models, Steyvers et al. (2003) acknowled-
ged that people might have used simple heuristics that
approximate rational inference. We will propose a simple
heuristic that poses far fewer demands on statistical pro-
cessing capacity.

Broken link heuristic
The general idea motivating the proposed heuristic is that
people enter the task with the bias that causal relations are
deterministic and causes sufficient for their effects, despite
the fact that the observable input is typically probabilistic
(Goldvarg & Johnson-Laird, 2001; Grif-
fiths & Tenenbaum 2009; Lu et al.,
2008; Schulz & Sommerville, 2006).
One way to reconcile a sufficiency bias with probabilistic data is to assume that
the generating model contains determin-
istic causal relations, which may occa-
sionally be broken due to random dis-
turbances, such as the presence of a hidden preventer or the absence of a
necessary enabler. However, these cases
should be rare. Thus, relations in which
the cause is present and the effect is
absent can be interpreted as largely in-
consistent with the determinism assump-
tion, and should therefore count as evi-
dence against the existence of a causal
relation. For example, if a case with one
present and two absent events is present-
ed, a hypothetical common cause model
in which the present event is assumed to
be the cause would entail two broken links, which should weaken this particular causal model hypothesis. Applying
the broken link heuristic is simple: (1) Learners observe
individual learning patterns with three events that can be
present or absent, (2) based on the hypothetical assumption
of each of the alternative causal models under consideration,
the number of broken links (i.e., cause-present, effect-absent
pairs) is counted across all learning patterns, (3) at the end
of learning, the causal model is chosen for which the sum of
the number of broken links proved minimal. This way mod-
els are chosen that are maximally consistent with the deter-
minism bias. Unlike Bayesian models, the heuristic only
looks at pairwise relations between events, and does not
need to consider complex conditional dependency informa-
tion.
Although the broken link heuristic typically approximates
the normative inference of specific causal Bayes net strat e-
gies, it is possible to design patterns in which the predi-
cions of these models and of the heuristic diverge. In Expe-
riment 1 we designed such patterns to provide a more speci-
fic test of the heuristic. Of course, it is not possible to test a
heuristic against all possible future Bayesian models of
structure induction. Our test is therefore restricted to a com-
parison between the broken link heuristic and the model that
has thus far been proposed in the literature as underlying
structure induction (Steyvers et al., 2003; see also General
Discussion).

Experiment 1
To test our heuristic we presented subjects with sets of 12
patterns each containing three binary variables being present
or absent (here: aliens thinking of “POR” or nothing; see
Fig. 2). The sets were randomly generated so that in each
case the heuristic predicts exactly the opposite causal struc-
ture as a Bayesian structure selection procedure with unin-
formative priors (which is similar to a maximum likelihood
models to simplify the task. Pilot research had shown that six target structures. We used this task to identify causal configurations of mind readers that corresponded to the six patterns each showing aliens thinking of “POR” or nothing (see Fig. 2 for an example). The patterns within were capable of reading the “POR”-thoughts of the other aliens and that participants will have to identify these mind readers on the basis of information about thought configurations. Since the thoughts of such a reader of thoughts therefore depend upon the thoughts of the non-mind reader(s), the mind readers constitute effects and the other aliens represent causes within the causal model (see also Mayrhofer et al., 2010). It was stated that either one or two of the aliens were capable of reading the “POR”-thoughts of the other aliens and that participants will have to identify these mind readers on the basis of information about thought configurations.

Procedure and Material In the instruction phase we presented subjects with an instruction about three aliens: Gonz, Brxxx, and Zoohn, who either thought of nothing or of “POR” (indicated by a bubble containing nothing or “POR”). It was stated that either one or two of the aliens were capable of reading the “POR”-thoughts of the other aliens and that participants will have to identify these mind readers on the basis of information about thought configurations. Since the thoughts of such a reader of thoughts therefore depend upon the thoughts of the non-mind reader(s), the mind readers constitute effects and the other aliens represent causes within the causal model (see also Mayrhofer et al., 2010). Participants were requested to choose one out of six configurations of mind readers that corresponded to the six target structures. We used this task to identify causal models to simplify the task. Pilot research had shown that subjects are often confused when asked about causes and effects in the mind reading alien task, which may have contributed to the low performance in Steyvers et al. (2003).

In the test phase, participants were presented with 48 sets of 12 patterns each showing aliens thinking of “POR” or nothing (see Fig. 2 for an example). The patterns within each set were presented in random order one by one. The aliens and their thoughts appeared simultaneously (i.e., no temporal cue was provided). After observing a set, subjects were requested to choose the causal structure that presumably generated that set.

To generate the pattern sets, we randomly drew five million sets of size 12 from a multinomial distribution with equal probabilities and preserved all unique sets for which the broken link heuristic uniquely predicted a common cause structure (CC pattern sets) or a common effect structure (CE pattern sets) while the Bayesian structure selection procedure predicted the opposite structure (i.e., reversed causal links). From this pool, for each subject 24 CC pattern sets and 24 CE sets were randomly selected (yielding 48 pattern sets in total per subject), and then presented in random order.

Results and Discussion The responses were aggregated within subjects for the CC pattern sets and the CE patterns sets separately (left vs. right hand side in Fig. 3). For analyzing purposes, all data sets and the participants’ corresponding responses were rotated so that the structures’ common elements were in the upright position.

Overall, the data demonstrate impressively high performance given that only covariation information was available. With respect to the CE pattern sets subjects substantially preferred to select the structures predicted by the heuristic compared to the structures predicted by the Bayesian structure selection procedure (42.7% vs. 12.8%, t[59]=9.26, p < .001; see Fig. 3a, right hand side). For the CC pattern sets, the results are less clear. Subjects generally showed a preference for the structures with the correct common element, but there seems to be no systematic preference for the structures predicted by the heuristic vs. those predicted by the Bayesian structure selection procedure (24.9% vs. 23.1%, t[59]=0.42, p=.34; see Fig. 3a, left hand side).

A more detailed analysis of the data on the subject level revealed that there are at least two groups of subjects using different strategies in solving the task. Based on each subject’s average response to the CC pattern sets with respect to the structure predicted by the broken link heuristic, we divided participants into two groups (Group 1: average response above chance level; Group 2: average response below chance level). The groupwise aggregated data are shown in Figs. 4b and 4c.

Group 1 (37 subjects) responded as predicted by the broken link heuristic: 36.2% (CC pattern sets, left panel) and 45.8% (CE pattern sets, right panel) of the selections were made consistent with the heuristic and only 15.8% (CC sets) and 16.0% (CE sets) consistent with the Bayesian procedure.

Group 2 (23 subjects) responded very differently: Whereas this group seemed to have adopted the heuristic for the CC pattern sets (right panel (37.6% vs. 7.5%), for the CC pattern sets (left panel) the structures seemed to be chosen according to the Bayesian procedure (6.7% vs. 34.9%).
In sum, Group 1 behaved largely consistent with the broken link heuristic, whereas Group 2 deviated from this heuristic. One possible interpretation of these differences is that the two groups interpret their preference for deterministic structures differently. Determinism may be associated with a bias for sufficiency, which leads to the maximization of causal strength. The broken link heuristic used by Group 1 is consistent with such a sufficiency bias. However, the preference for deterministic structures may also be associated with a bias for necessity. This bias leads to a preference for effects with low base rates because effects are expected to be accompanied by observable causes. A third possibility is the strong and sparse (SS) prior by Lu et al. (2008) which qualitatively entails that people should expect that either the observable cause or the unobserved background cause are necessary and sufficient for the effect, respectively.

Experiment 2
To test the idea that different prior assumptions about sufficiency or necessity in the causal system underlie different strategies in solving the induction task, we used networks with two variables, A and B, only. It is well known that the question whether A causes B or B causes A is not decidable with covariation data alone. Both graphs are Markov equivalent (see also Fig. 1). For each parameterization for graph 1 (A → B) there exists a parameterization for graph 2 (A ← B) yielding the exact same likelihood. A potential preference for graph 1 or graph 2 is therefore necessarily due to prior assumptions about the causal system’s parameterization. The goal of Experiment 2 was to test whether we can identify different classes of people that differ in their prior assumptions in this simple task (sufficiency, necessity, or SS prior).

Method
Participants 50 students from the University of Göttingen participated in exchange for course credit, or were paid 8 € per hour.

Procedure and Material We used the same cover story and instructions as in Experiment 1. The only difference was that there were only two aliens. Subjects were instructed that only one of the two aliens was able to read the “POR”-thoughts of the other alien and that they had to find out which one had this capacity.

In the test phase, participants were presented with 16 sets of 12
patterns (see Table 1) with each of the two aliens thinking of “POR” or nothing. After observing each set, subjects were asked to decide whether Alien A or B could read the mind of the other alien.

Using the same method as in Experiment 1, we chose sets of patterns that yielded distinctive predictions for the three assumed priors. Table 1 shows the frequencies of the eight pattern sets (each was presented twice) along with the predictions of the three priors.

Table 1: Used pattern sets in Experiment 2

<table>
<thead>
<tr>
<th>Data</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
</tr>
<tr>
<td>00</td>
<td>A</td>
</tr>
<tr>
<td>01</td>
<td>A</td>
</tr>
<tr>
<td>02</td>
<td>A</td>
</tr>
<tr>
<td>03</td>
<td>A</td>
</tr>
<tr>
<td>04</td>
<td>A</td>
</tr>
<tr>
<td>05</td>
<td>A</td>
</tr>
<tr>
<td>06</td>
<td>A</td>
</tr>
<tr>
<td>07</td>
<td>A</td>
</tr>
<tr>
<td>08</td>
<td>A</td>
</tr>
</tbody>
</table>

Notes. The four “data” columns (left side) show how often each pattern was shown within each of eight pattern sets (e.g. “01” means that A=0 and B=1). The “prior” columns show the predictions of the three different priors. The letters indicate which variable should be chosen as cause according to the respective prior (P1: high causal strength, P2: low base rate of effect, P3: strong and sparse). Each set was shown twice.

Results and Discussion

For analyzing purposes, we coded participants’ selections of cause A and B with respect to the different prior profiles (1: as predicted, 0: not as predicted) and assigned each subject to the profile cluster that minimized the mean squared distance. Additionally, we included a “random guesser” cluster.

Using this procedure, 28 out of 50 subjects (56%) were assigned to the sufficiency (i.e., high causal strength) cluster, 7 subjects (14%) to the necessity (i.e., low base rate) cluster, 2 subjects (4%) to the strong and sparse prior cluster, and 10 subjects (20%) to the “random guesser” cluster. Three subjects (6%) could not be assigned by the procedure. Within the sufficiency cluster, 91.7% of participants’ selections were consistent with the sufficiency bias. The corresponding numbers were 92.0% for the necessity cluster, and 77.1% for the strong and sparse prior cluster.

We assume that the broken link heuristic may underlie inferences of subjects with a sufficiency bias. But what about the necessity oriented subjects? A corresponding heuristic for the necessity bias might be to preferentially select structures that minimize the occurrence of unexplained effects.

In sum, the results show that different prior assumptions may play a role in causal structure induction. In Experiments 1 and 2 we have shown that different groups of learners either are biased in the direction of a sufficiency or a necessity bias. However, the evidence for interindividual differences is only correlational. To strengthen our case that differences in prior assumptions are the relevant causal factor, in Experiment 3 we experimentally manipulated subjects’ biases.

Experiment 3

To test whether different prior assumptions play a role in strategy selection, we used the materials of Experiment 2. Through instructions we manipulated whether learners expect high causal strength (i.e., sufficiency prior) or low base rate of effect (i.e., necessity prior). Based on the results of the previous experiment we did not test the sparse and strong prior again.

Method

Participants 40 students from the University of Göttingen participated in exchange for course credit, or were paid 8 € per hour.

Procedure, Materials, and Design The procedure, instruction and pattern sets were identical to those used in Experiment 2, except for the manipulation of the priors. In one condition, subjects were told that mind readers mostly succeed in reading the mind of the other alien (= high causal strength), whereas in the other condition we instructed participants that mind readers only rarely think of “POR” on their own (= low base rate of effect). The priors were manipulated between subjects (2 × 20).

Results and Discussion

We recoded subjects’ answers so that the selection of the variable predicted by a sufficiency prior was coded as 1 and the selection of the other variable, which is predicted by a necessity prior, as 0. For each subject, an average score was calculated. The results are shown in Fig. 4; higher ratings indicate the use of a sufficiency prior.

![Figure 4: Results of Experiment 3 for the high-causal-strength vs. low-base-rate-of-effect conditions.](image-url)

In the high causal strength condition, 95.0% of the selections corresponded to the sufficiency prior (i.e., broken link...
heuristic), in the low base rate condition only 53.4% of the selections were predicted by this prior (hence 46.6% of the cases were consistent with a necessity prior). Thus, the manipulation of prior information made a substantial difference, t(38)=3.98, p<.001. Although our manipulation proved successful, there was a general tendency toward the sufficiency bias.

General Discussion
Our main goal in the present set of studies was to re-visit the question whether people may be capable of inducing causal models from covariation data alone without further cues, such as temporal order or intervention. In the literature, there has been a debate between bottom-up and top-down learning theories of causal learning: Whereas top-down theorists claim that covariation information plays none or only a secondary role after cues have been used to select potential hypothetical models, bottom-up theories, such as causal Bayes net theory, assert that people are capable of inducing structure from conditional dependence and independence information alone. Our three experiments show that both positions may be wrong. In simple three-variable domains with clear instructions and a (relatively large) set of alternative models people were indeed often capable of reliably picking the right model. However, in Experiment 1 we also showed that learners can solve the task using a simple heuristic that does not require conditional dependence and independence information. Another novel discovery was that subjects may differ with respect to their preferred bias, and consequently their preferred heuristic. We have shown that the determinism bias can come in two variants, a sufficiency or high causal strength bias, or a necessity or low base rate of effect bias.

We have focused on the Bayesian model without specific priors, proposed by Steyvers et al. (2003), because it is the only one that has so far been tested as underlying structure induction. It is certainly possible to adapt more complex models incorporating various biases (e.g., Lu et al., 2008) to the present task. Some of these models incorporate sufficiency and necessity biases, so that it is likely that they will fare equally well as our simple heuristics. However, thus far there is no unambiguous empirical evidence that people in fact can make the elaborate, complex statistical computations required by these models. Moreover, our much simpler heuristics represent an existence proof that causal induction may be equally successful with much simpler procedures motivated by intuitive biases, such as the intuition that causes should be typically accompanied by their effects).

Acknowledgments
We wish to thank Anselm Rothe for help in preparing the experiments. This research was supported by a research grant of the Deutsche Forschungsgemeinschaft (DFG Wa 621/20).

References

Acknowledgments
We wish to thank Anselm Rothe for help in preparing the experiments. This research was supported by a research grant of the Deutsche Forschungsgemeinschaft (DFG Wa 621/20).

References

Acknowledgments
We wish to thank Anselm Rothe for help in preparing the experiments. This research was supported by a research grant of the Deutsche Forschungsgemeinschaft (DFG Wa 621/20).