COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 2137–2150

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12318

Sufficiency and Necessity Assumptions in Causal Structure Induction

Ralf Mayrhofer, Michael R. Waldmann

Department of Psychology, University of Göttingen

Received 14 January 2013; received in revised form 14 September 2015; accepted 16 September 2015

Abstract

Research on human causal induction has shown that people have general prior assumptions about causal strength and about how causes interact with the background. We propose that these prior assumptions about the parameters of causal systems do not only manifest themselves in estimations of causal strength or the selection of causes but also when deciding between alternative causal structures. In three experiments, we requested subjects to choose which of two observable variables was the cause and which the effect. We found strong evidence that learners have interindividually variable but intraindividually stable priors about causal parameters that express a preference for causal determinism (sufficiency or necessity; Experiment 1). These priors predict which structure subjects preferentially select. The priors can be manipulated experimentally (Experiment 2) and appear to be domain-general (Experiment 3). Heuristic strategies of structure induction are suggested that can be viewed as simplified implementations of the priors.

Keywords: Bayes nets; Causal learning; Causal induction; Structure induction

1. Introduction

Causal learning and reasoning are ubiquitous. Causal knowledge enables us to predict future events, explain past events, and plan actions to achieve goals. Whereas early psychological theories of causal reasoning tried to reduce causal relations to non-causal associative, probabilistic, or logical links between cues and outcomes (for an overview, see Waldmann & Hagmayer, 2013), more recent theories assume causal model representations in which causes are distinguished from effects (e.g., Gopnik et al., 2004; for an overview, see Rottman & Hastie, 2014). The majority of work on causal learning has been devoted to the question of how people infer causal strength based on contingency

Correspondence should be sent to Ralf Mayrhofer, Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany. E-mail: rmayrho@uni-goettingen.de

data (e.g., Cheng, 1997; Griffiths & Tenenbaum, 2005; see Perales & Shanks, 2007; Hattori & Oaksford, 2007, for overviews) and how people select between competing candidate causes of an observed effect (i.e., causal attribution; see, e.g., Downing, Sternberg, & Ross, 1985; Hewstone & Jaspars, 1987; Kelley, 1967; Mackie, 1974). In both cases, however, the causal roles of the variables (i.e., causes vs. effects) were pre-specified by the task instructions. In contrast, our focus here is on how people determine the causal role of the involved variables, that is, which ones are causes and which ones are effects (i.e., causal structure induction).

There has been a debate in psychology about how people infer causal structure. One approach argues that learners combine non-statistical cues, such as interventions, temporal order, or domain-specific prior knowledge to specify which variables serve as causes and which ones as effects (see Fernbach & Sloman, 2009; Gopnik & Wellman, 2012; Lagnado, Waldmann, Hagmayer, & Sloman, 2007; Waldmann, 1996). According to this approach, these cues suggest hypothetical causal structures while covariation information is used to estimate the causal parameters, such as causal strength (e.g., the probability that a cause brings about its effect) and base rate (e.g., the probability that an effect is brought about by unobserved background causes).

In contrast, causal Bayes net theory (see Gopnik et al., 2004), originally developed as a normative theory of how computer systems and experts should make causal inferences, provides us with mechanisms for the induction of causal structures from covariation information alone (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000). Whereas researchers who claim that learners use non-statistical cues to causal structure are skeptical about the capability of learners to induce structure from covariation information alone, Gopnik et al. (2004) suggested that people are capable of inducing causal structure from statistical patterns, even when other cues (e.g., temporal order) are not available (see also Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; but see Gopnik & Wellman, 2012).

Gopnik et al. (2004) discussed two strategies of causal structure induction: In *constraint-based* learning, triples of variables are analyzed to assess patterns of conditional dependencies that are constrained by the underlying causal structure. For example, common-cause structures with three variables imply that each pair of variables is correlated but that the two effect variables are independent (i.e., uncorrelated) when conditionalized on the states of their common cause. In a common-effect structure, by contrast, the alternative causes are assumed to be independent (unless there are additional causes affecting both), but they become dependent conditional on their joint effect. An alternative to the constraint-based approach are Bayesian algorithms that calculate the posterior probability of the different candidate structures by combining the likelihood of the data given the alternative structures with their prior probability (see Steyvers et al., 2003; see also Meder, Mayrhofer, & Waldmann, 2014).

Both induction strategies use conditional and unconditional probability information in the data to assess the likelihood of the competing structure hypotheses. Whereas both approaches allow learners to decide whether a dataset is more likely to be generated by a common-cause or a common-effect structure, common-cause structures cannot be discriminated from causal chains, for example. These two causal structures entail the same

conditional dependencies (i.e., they are Markov equivalent). Thus, additional cues are required to discriminate between such Markov equivalent structures.

1.1. Empirical evidence for causal structure induction

In the past decade, causal structure induction has attracted more and more attention in cognitive psychology. Steyvers et al. (2003) introduced the mind-reading alien paradigm to test whether people are capable of inducing the causal structure underlying three variables based on covariation data only. They presented subjects with three aliens that had particular thoughts from a small dictionary (e.g., "POR," "TUS," etc.). Some of the aliens were mind readers and were therefore able to read the thoughts of other aliens (which led to sharing their thoughts). Subjects were requested to decide which out of several causal structures generated the presented thought configurations of the aliens. Overall, Steyvers et al. observed above-chance but poor performance when only covariation information was available.

Lagnado and Sloman (2004, 2006), who also studied structure induction, observed that people tend to prefer temporal and interventional cues over contradicting covariational information (see also McCormack, Frosch, Patrick, & Lagnado, 2015). Fernbach and Sloman (2009) consequently argued that structure learning is local and primarily driven by temporal cues. They concluded that people "do not rely on covariation when learning the structure of causal relations" (p. 678). White (2006) arrived at a similar conclusion after showing that people are not capable of inducing causal structure from patterns of co-occurrences. More recent research, however, has shown that humans can infer causal structure when the causal system is deterministic and spontaneous occurrences of effects are rare (Deverett & Kemp, 2012), or when the observed system shows dynamic regularities over time (Rottman & Keil, 2012; see also Rottman, Kominsky, & Keil, 2014, for evidence regarding children).

1.2. Goals of the present research: The role of abstract prior beliefs in causal structure induction

The presented theories of causal structure induction generally do not consider the possibility that learners might use abstract prior assumptions about the parameters of a causal system when inducing causal structures. However, more recent research about causal learning has shown that people have general assumptions about causal strength and about how causes potentially interact with background causes. In this research, the causal roles of the variables were pre-specified in the task instructions. For example, Lu, Yuille, Liljeholm, Cheng, and Holyoak (2008) argued that people's estimation of causal strength is best explained by a Bayesian model that incorporates the quantitative assumption that each effect is either generated by a strong observable cause or by the background cause (i.e., either high causal strength and low base rate of the effect, or low causal strength and high base rate of the effect; i.e., strong and sparse prior). In a similar vein, Yeung and Griffiths (2011, 2015) showed that people have a bias toward high causal strength

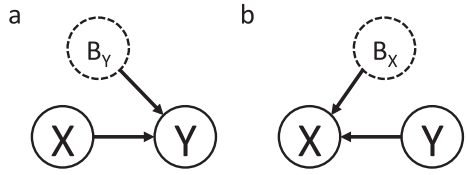

Fig. 1. The two possible causal structures between two observable variables X and Y along with background causes (B_X, B_Y) that explain occurrences of the effects in the absence of the observable cause.

(i.e., sufficiency). Thus, it appears that people tend to assume that causal relations are (quasi)-deterministic (see also Goldvarg & Johnson-Laird, 2001; Griffiths & Tenenbaum, 2009; Lu et al., 2008; Schulz & Sommerville, 2006).

Similarly, research in causal attribution has shown that people prefer to select sufficient causes rather than necessary causes when explaining an effect (Downing et al., 1985; see also Hewstone & Jaspars, 1987) and weigh evidence regarding violations of sufficiency stronger than evidence regarding violations of necessity (Schustack & Sternberg, 1981). In philosophy, it has also been discussed whether causal attributions should be based on sufficient conditions, necessary conditions, or more sophisticated variants, such as the INUS condition (i.e., a cause is an *I*nsufficient but *N*on-redundant part of a condition which is itself *U*nnecessary but *S*ufficient for the effect's occurrence; see Mackie, 1974).[1]

Our new idea is that abstract prior assumptions about the nature of causal relations should not only manifest themselves in strength estimates or in the selection of actual causes but also in the induction of causal structure. To test how different prior assumptions influence causal structure induction, we used the simplest possible causal network with two observable variables, X and Y (see Fig 1. for an illustration). It is well-known that the question whether X causes Y or Y causes X is not decidable when only contingency data are available. Both graphs are Markov equivalent. For each parameterization of Graph 1 (X→Y; see Fig. 1a), there exists a parameterization of Graph 2 (X←Y; see Fig. 1b) yielding the exact same likelihood for any given set of contingency data. Without any additional assumptions, a structure induction algorithm has to guess the underlying structure. The goal of our research, therefore, is to investigate whether learners show systematic stable preferences when selecting between alternative structures that may be traced back to their prior assumptions about the causal system's parameterization.

2. Experiment 1

The goal of Experiment 1 was to test whether learners employ abstract prior assumptions about a causal system's parameterization in a simple two-variable structure induction task and whether there are interindividual differences regarding these prior assumptions. We constructed different datasets in such a way that different priors make different predictions regarding the competing underlying causal structures. This method allowed us to cluster subjects who differed in their prior assumptions about the

relationship between cause and effect (see Section 2.1. below). We tested two determinism priors: a sufficiency prior (i.e., preference for high causal strength) and a necessity prior (i.e., preference for low base rate of effect).[2]

For the structure induction task, we adapted the mind-reading alien story introduced by Steyvers et al. (2003). Steyvers and colleagues presented subjects with a scenario in which three aliens thought of different distinct words from a small dictionary. Some of these aliens were capable of reading the thoughts of the other aliens. Subjects' task was to identify the aliens capable of mind reading on the basis of data showing how thoughts are distributed at a specific point in time across the three aliens. Since the thoughts of mind readers depend upon the thoughts of the aliens whose thoughts are read, the thoughts of the mind reader represent a causal effect of the thoughts of the alien whose mind is read.[3] Thus, this identification task is equivalent to structure induction.

In the present experiment, we presented subjects with two aliens, X and Y, either thinking of nothing or of the word POR. It was stated that either alien X was able to read the POR-thoughts of alien Y, or alien Y was able to read the POR-thoughts of alien X and that either alien also may think of POR on its own. The task was to identify the mind-reading alien on the basis of information about thought configurations (i.e., datasets).

2.1. Method

2.1.1. Participants

Fifty students from the University of Göttingen (37 female; 21.2 years old on average) participated as part of a series of unrelated computer experiments in exchange for course credit or €8/h.

2.1.2. Procedure and material

The experiment was conducted on desktop computers and consisted of an instruction, an instruction test, and 16 test phases. In the instruction phase, subjects were presented with a story about two aliens X and Y (called Gonz and Brxxx) that either thought of nothing or of POR (indicated by a bubble containing either nothing or POR). It was stated that one of the two aliens was capable of reading the POR thoughts of the other alien, that either alien could also think of POR on its own, and that the subject's task was to identify the mind reader. After reading the instruction, subjects were asked to answer a few multiple-choice questions about the instructions and had to re-read the instruction until they passed the comprehension test without any errors.

In each of the 16 test phases, we presented participants with 12 patterns (i.e., thought configurations) each showing either alien thinking of POR or nothing (see Fig. 2, for an example). The 12 patterns were separately presented on the computer screen in random order one by one (self-paced). Both aliens and their thoughts appeared simultaneously (i.e., no temporal cue was provided). After observing the 12 patterns of a set, subjects were requested to decide whether Alien X or Alien Y was the mind reader (i.e., forced choice). Then, subjects continued with the next test phase until all 16 pattern sets were presented. (The pattern sets were randomly assigned to the test phases.)

Fig. 2. An example of a thought configuration shown in Experiment 1 and 2.

Table 1 shows the frequencies of the eight distinct pattern sets used in all experiments (each was presented twice) along with the predictions of a Bayesian structure selection procedure employing the two types of determinism priors.[4] The sets were selected with the goal that the cause-and-effect roles of the variables were counterbalanced for each prior (subsets indicated by an "a" vs. "b"), ensuring that preferences for position (e.g., a preference to assign the cause role to the left-hand side variable) do not influence the assignment of subjects to clusters.

2.2. Results and discussion

Since sufficiency and necessity priors predict exactly opposite choices (see Table 1), we recoded the selections of the subjects with respect to the prediction of the sufficiency prior (i.e., 1 = predicted by sufficiency and, therefore, 0 = predicted by necessity). For

Table 1
Pattern sets used in all three experiments, predictions of the different priors, and the selection rates

							Data (% of "X→Y" Selections)					
	Pattern Frequency				Prediction			Exp 2 Condition		Exp 3 Domain		
Set	00	01	10	11	Suff	Ness	Exp 1	Suff	Ness	Biol	Chem	Phys
1a	7	4	0	1	X→Y	X←Y	74	98	55	56	54	54
1b	7	0	4	1	X←Y	X→Y	27	7	45	44	44	44
2a	6	5	0	1	X→Y	X←Y	78	95	62	54	54	60
2b	6	0	5	1	X←Y	X→Y	27	2	31	48	48	46
3a	1	5	0	6	X→Y	X←Y	63	93	52	54	52	56
3b	1	0	5	6	X←Y	X→Y	36	5	45	44	44	42
4a	1	4	0	7	X→Y	X←Y	64	93	55	56	46	58
4b	1	0	4	7	X←Y	X→Y	43	2	57	50	50	44

Note. The four "pattern frequency" columns show how often each pattern was shown within each of eight pattern sets of size 12 (rows; "01," for example, means that X = 0 and Y = 1, as in the example depicted in Fig. 1). The "prediction" columns show which structure should be selected according to the respective prior (Suff: sufficiency prior, Ness: necessity prior; for details see note 4). Each set was shown twice in Experiments 1 and 2 and shown once in each domain in Experiment 3. The "data" columns show the percentage of X→Y selections in the respective experiments.

each subject, an average score was calculated that indicated how many of the 16 choices corresponded to the predictions of a sufficiency prior. Then, each subject was assigned to the cluster that minimized the difference between the subject's average score and the prediction of the respective prior. Because a structure selection procedure based on maximum likelihood (i.e., without any prior assumptions) would predict random guessing, we additionally included a random-guesser cluster (i.e., with an expected average score of 0.5).[5] Note that with our procedure the probability of a random guesser accidentally being assigned to the sufficiency or necessity cluster, respectively, is only .011. Thus, in case of random guessing we would only expect about one of our 50 subjects being falsely assigned to the sufficiency or necessity cluster.

Using our clustering procedure, 27 of 50 subjects (54%) were assigned to the sufficiency cluster, 7 subjects (14%) to the necessity cluster, and 13 subjects (26%) to the random-guesser cluster. Three subjects (6%) could not uniquely be assigned by the procedure. For the sufficiency and the necessity clusters, 92.4% and 92.0% of participants' selections, respectively, were consistent with the predictions of a Bayesian selection procedure with the respective prior (see Table 1). These numbers indicate that participants' responses were highly consistent, which shows that people seem to use an intraindividually stable prior in the experiment.

In sum, the experiment demonstrates the role of priors in structure induction and shows that biases differ between individuals. This demonstration goes beyond previous studies that typically compare human data with Bayesian methods incorporating uninformative priors. However, the evidence for the strong influence of different prior assumptions is only correlational so far. To strengthen our case that differences in prior assumptions are an important factor in structure induction, we experimentally manipulated subjects' priors in Experiment 2.

3. Experiment 2

To manipulate prior assumptions, we instructed subjects about how the causal relations within the learning domain usually work. We used the same materials as in Experiment 1 but added a short description about the to-be-expected type of causal relation. We instructed either high causal strength (i.e., that a cause is usually sufficient to bring about its effect) or low base rate of effect (i.e., that a cause is usually necessary for the effect to occur).

3.1. Method

3.1.1. Participants

Forty students from the University of Göttingen (27 female; 22.5 years old on average) participated as in Experiment 1.

3.1.2. Procedure, materials, and design

The procedure, instruction, and pattern sets were identical to those used in Experiment 1 except for the manipulation of the prior assumptions in the instruction phase. In the *suf-*

ficiency condition, subjects were told that mind readers mostly succeed in reading the thoughts of the other alien (=high causal strength; i.e., sufficiency prior), whereas in the *necessity* condition we instructed participants that mind readers only rarely think of POR on their own (=low base rate of effects; i.e., necessity prior). The prior assumptions were manipulated between subjects (2 × 20).

3.2. Results and discussion

Subjects' choices were coded and averaged as in Experiment 1, such that the resulting average score indicated how many of the choices corresponded to the predictions of a sufficiency prior (see also Table 1). With respect to our experimental manipulation, we expected higher average scores in the sufficiency condition compared to the necessity condition.

In the *sufficiency* condition, 95.0% of the selections were consistent with a sufficiency prior. In the *necessity* condition, only 53.4% of the selections were predicted by sufficiency (hence 46.6% of the cases were consistent with a necessity prior). Thus, the manipulation of the prior through initial instructions made a substantial difference, $t(38) = 3.98$, $p < .001$. Nevertheless—although our manipulation proved successful—there was a general tendency to assume sufficiency even in the *necessity* condition.

4. Experiment 3

In the previous experiments, we have shown how assumptions about sufficiency and necessity guide the induction of causal structures in the rather artificial domain of mind-reading aliens. We used this domain to minimize the influence of domain-specific knowledge about causal mechanisms. However, it would be interesting to know whether our findings generalize to other domains, and how stable the individual priors are across domains. It may well be that the priors of people are domain-specific. For example, they may have a strong sufficiency intuition for physical domains, but not for biological ones (see Saito & Shimazaki, 2013, for evidence suggesting this possibility). Research in the categorization literature indicates that different domains may be associated with different assumptions about the underlying causal structure (see Wattenmaker, 1995). However, an alternative possibility is that people have intraindividually stable abstract intuitions about the nature of causality that are relatively stable across domains. People may, for instance, believe that causal relations generally tend to express sufficiency or necessity (Yeung & Griffiths, 2015). These priors may of course be overridden by specific knowledge about mechanisms but since we are going to present variables that are not associated with prior knowledge of directionality, only abstract intuitions can drive subjects' judgments.

In the present experiment, we used the general method of Experiment 1 again but varied the cover stories describing variables from the domains of biology, chemistry, and physics. In each of these domains, we chose variables that do not suggest a specific cau-

sal direction. Our goal was to explore whether priors vary across domains or whether we will see again intraindividual stability.

4.1. Method

4.1.1. Participants

Forty-eight students of the University of Göttingen (33 female; 23.5 years old on average) participated in the same manner as in the previous experiments.

4.1.2. Procedure, materials, and design

The initial instructions described the general setting and procedure of the experiment. Then subjects were presented with three consecutive experimental tasks that were essentially three shortened versions of Experiment 1 adapted to three different domains. We used the same eight pattern sets as in Experiment 1 (see Table 1), but now subjects only saw each set once (instead of twice as in Experiments 1 and 2), resulting in eight structure judgments per task. The three experimental tasks only differed in their cover story that either referred to the domain of biology, chemistry, or physics. The domains (i.e., tasks) were presented in random order.

In the *biological domain* task, subjects were asked to imagine being a biologist who investigates specific kinds of bugs in eight different regions of the world (i.e., eight pattern sets). It was mentioned that some bugs produce excrement that attracts other bugs. The task was to figure out which of two bug species produces the excrement (i.e., cause), and which bug species was attracted (i.e., effect), based on information about whether bugs from the two different species are present, whether both species are absent, or whether only one species is present in the region.

In the *chemical domain* task, subjects were asked to imagine being a chemist who investigates specific kinds of chemical substances in eight different chemical labs. It was pointed out that some substances initiate the synthesis and therefore the presence of other substances. Again, subjects were asked to identify which of the two substances initiates the synthesis of the other substance and, therefore, causes its presence. Again the learning data provided information about the frequencies of the paired presence/absence of the two substances.

In the *physical domain* task, subjects were told to imagine being a physicist who investigates specific kinds of subatomic particles in eight different particle physics laboratories. It was stated that there are kinds of particles that interact with neutrinos, thus creating other kinds of particles in the laboratory. The task was to identify which of the two kinds of particles interacts with neutrinos and, therefore, causes the presence of the other particle, based on contingency information about the presence and absence of the two kinds of particles.

4.2. Results and discussion

As in Experiment 1, we recoded subjects' responses with respect to the predictions of the sufficiency prior; then we averaged across choices for each domain separately (see

Table 1). Subjects were clustered according to their average scores, resulting in 22 of 48 subjects (45.8%) being assigned to a *stable-sufficiency* cluster (i.e., using a sufficiency prior across all domains), 16 (33.3%) to a *stable-necessity* cluster (i.e., using a necessity prior across all domains), and 9 (18.8%) to an *alternating* cluster (i.e., subjects who switched between necessity and sufficiency across domains). None of the subjects were assigned to the random-guesser cluster. One subject could not be uniquely assigned to a cluster using our procedure. Note that the probability of a random guesser accidentally being assigned to the *stable-sufficiency* or *stable-necessity* cluster, respectively, is only .00004 each; and the probability of accidentally being assigned to the *alternating* cluster is only .00026.

Experiment 3 extends the findings of Experiments 1 and 2 to more realistic domains and shows that the majority of subjects seems to use a stable determinism prior across domains (i.e., either a sufficiency or a necessity prior) that does not appear to be sensitive to abstract domain characteristics.

One important finding in Experiment 3 is that different domains do not seem to be associated with different priors. Although it cannot be conclusively ruled out that this finding is in part also due to subjects trying to be consistent across tasks, a more plausible explanation is that subjects have abstract intuitions about what it means to be causally related, and bring to bear domain knowledge when specific mechanism knowledge is available. Since the goal of our studies was to investigate the role of priors on structure induction, a methodological requirement was to choose variables for which prior knowledge was minimized.

There is one important difference between Experiment 3 and the other two studies, however. In comparison to Experiments 1 and 2, a higher proportion of subjects were associated with a necessity prior in Experiment 3 (see Table 1). Given that there were no differences between domains in Experiment 3, the most plausible explanation of this effect is that background assumptions about the causal mechanism underlying mind reading differed from the more abstract characterization in Experiment 3. In Experiment 3, we used a neutral description of the causal relation (e.g., "X causes the presence of Y") to equate the tasks as much as possible across domains. It may be that mentioning a cause of the presence of the effect in the three cover stories led subjects to think that there may be other causes of the presence of each variable. In contrast, in Experiments 1 and 2 we used a more concrete description of the causal mechanism which described the causal relation as arising from a disposition of one of the causal participants (e.g., "Y is able to read X's mind"; see also Mayrhofer & Waldmann, 2015). This cover story might have highlighted the sufficiency of the cause of the thoughts of the mind reader because alternative causes of the thoughts of Y seem less plausible.

5. General discussion

Our experiments suggest that people enter the task of causal structure induction with the strong bias that the underlying causal relations are deterministic and that causes are

either sufficient or necessary for their effects. This holds true despite the fact that the observable input is probabilistic (see also Goldvarg & Johnson-Laird, 2001; Griffiths & Tenenbaum, 2009; Lu et al., 2008; Schulz & Sommerville, 2006, for related views). The results of our experiments suggest that learners employ priors to select causal structures in elemental causal induction tasks with Markov equivalent structures. These priors can be altered through initial instructions (Experiment 2). A second important novel finding is that in the absence of biasing instructions different participants vary with respect to the priors they prefer (Experiments 1 and 3) and that the majority of subjects use these priors stably across different domains (Experiment 3). An important finding is that for variables that are not preferentially associated with the cause or effect role, biases seem primarily to be driven by the characterization of the causal relation and intraindividual preferences. We found a pre-dominance of sufficiency intuitions in the alien mind reader task, whereas the sufficiency prior was less prevalent when a more abstract characterization of the causal relation was chosen (see also Saito & Shimazaki, 2013, for converging evidence supporting this hypothesis).

An interesting direction for future research is to study the role of prior assumptions in the induction of more complex causal models. One possibility to implement such an a priori bias is to assign appropriate prior distributions to the parameters in a Bayesian structure induction procedure (see Deverett & Kemp, 2012; Lu et al., 2008). One problem with this approach is that Bayesian procedures pose strong demands upon the reasoners' statistical processing capacities. A simpler, more heuristic way to reconcile, for instance, a sufficiency bias with probabilistic data is to assume that the generating causal model contains deterministic causal relations that may occasionally be broken due to random disturbances, such as the presence of a hidden preventer or the absence of a necessary enabler (i.e., quasi-determinism). Patterns in which the cause is present and the effect is absent can be interpreted as largely inconsistent with the determinism assumption and should, therefore, count as evidence against the existence of a causal relation. Based on this assumption, Mayrhofer and Waldmann (2011) have proposed a "broken link" heuristic as the basis of the induction of more complex causal structures (e.g., common-cause vs. common-effect models). For example, if a case with one present and one absent event is presented, a hypothetical model in which the present event is assumed to be the cause would entail a broken link which should weaken this particular causal model hypothesis. According to the "broken link" heuristic, the causal structure is chosen for which the sum of the number of broken links is minimal. When applying the "broken link" heuristic, models are selected that are maximally consistent with a sufficiency bias. A corresponding heuristic for a necessity bias might be to preferentially select structures that minimize the occurrence of unexplained effects (i.e., choose the model hypothesis that minimizes the number of cause-absent/effect-present pairs).

Unlike Bayesian or constraint-based procedures, these heuristics only look at pairwise relations between variables even with more complex models and do not need to consider complex conditional dependency information (see Mayrhofer, Hagmayer, & Waldmann, 2010; Mayrhofer & Waldmann, 2015). Thus, an advantage of a heuristic implementation

is that it allows for relatively simple processing strategies to induce causal structure. Future research will have to test these hypotheses further—especially in more complex scenarios (see also Deverett & Kemp, 2012; Mayrhofer & Waldmann, 2011).

Acknowledgments

Portions of this research were presented at the Annual Meeting of the Cognitive Science Society in 2011 (see Mayrhofer & Waldmann, 2011). We wish to thank Anselm Rothe for his help in preparing the experiments. This research was supported by research grants of the Deutsche Forschungsgemeinschaft (Wa 621/22-1; Ma 6545/1-2) as part of the priority program "New Frameworks of Rationality" (SPP 1516).

Notes

1. The focus in causal attribution research, however, is token-level or actual causation; that is, the question which cause brought about the effect in a particular instance. Causal structure induction, by contrast, is concerned with type-level causal claims between generic variables (e.g., "smoking causes cancer").
2. In addition, we tested the *strength version* of the strong-and-sparse prior (proposed by Lu et al., 2008). Because this prior did not fit the data well, we do not present the results here (see Mayrhofer & Waldmann, 2011, for more details). In the present task, the *structure version* of the strong-and-sparse prior is dominated by its sufficiency component, which leads, at least for this study, essentially to the same predictions as a pure sufficiency prior.
3. One might argue that the task may be confusing because the mind reader, that is, the causal agent, is the causal effect. However, Mayrhofer and Waldmann (2015, Experiment 1) have shown that people can distinguish between agent–patient relations and causal dependency (which is of interest here). To simplify the task of reporting the causal structure further, we asked subjects to identify the mind-reading alien (i.e., the effect)—in contrast to Steyvers et al. (2003), who requested participants to draw a causal arrow between the variables.
4. We calculated the posterior distribution over both structures for each set given (a) a Beta(100, 1) prior over causal strength (i.e., a strong preference for high causal strength, that is, sufficiency) and (b) a Beta(0, 100) prior over base rate of effect (i.e., a strong preference for a low base rate of effect, that is, necessity) while the priors of the remaining parameters were set to Beta(1, 1) distributions (i.e., flat priors). In each case, the posterior probability of the predicted structure (see Table 1) was >.99.
5. Note that a subject who, for instance, always picked the right-hand-side alien as the mind reader would also be classified as a random guesser since the pattern sets were counterbalanced with respect to position.

References

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

Deverett, B., & Kemp, C. (2012). Learning deterministic causal networks from observational data. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 288–293). Austin, TX: Cognitive Science Society.

Downing, C. J., Sternberg, R. J., & Ross, B. H. (1985). Multicausal inference: Evaluation of evidence in causally complex situations. *Journal of Experimental Psychology: General*, *114*, 239–263.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 678–693.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610. doi:10.1016/S0364-0213(01)00046-5

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.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*, 661–716.

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, *31*, 765–814.

Hewstone, M., & Jaspars, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology*, *53*, 663–672.

Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*, *15*, 192–238.

Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 856–876.

Lagnado, D. A., & Sloman, S. A. (2006). Time as guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 451–460.

Lagnado, D. A., Waldmann, M. A., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation. Cues to causal structure. In A. Gopnik & L. E. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford, England: Oxford University Press.

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.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, England: Clarendon.

Mayrhofer, R., Hagmayer, Y., & Waldmann, M. R. (2010). Agents and causes: A Bayesian error attribution model of causal reasoning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 925–930). Austin, TX: Cognitive Science Society.

Mayrhofer, R., & Waldmann, M. R. (2011). Heuristics in covariation-based induction of causal models: Sufficiency and necessity priors. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3110–3115). Austin, TX: Cognitive Science Society.

Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. *Cognitive Science*, *39*, 65–95.

McCormack, T., Frosch, C., Patrick, F., & Lagnado, D. A. (2015). Temporal and statistical information in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 395–416.

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

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

Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin and Review*, *14*, 577–596.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*, 109–139.

Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: Observations and interventions. *Cognitive Psychology*, *64*, 93–125.

Rottman, B. M., Kominsky, J. F., & Keil, F. C. (2014). Children use temporal cues to learn causal directionality. *Cognitive Science*, *38*, 489–513.

Saito, M., & Shimazaki, T. (2013). Interpreting covariation in causal structure learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3348–3353). Austin, TX: Cognitive Science Society.

Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*, 427–442.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, *110*, 101–120.

Spirtes, P., Glymour, C., & Scheines, P. (2000). *Causation, prediction, and search* (2nd ed.). New York: Springer.

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

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 34: Causal learning* (pp. 47–88). San Diego: Academic Press.

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 733–752). New York: Oxford University Press.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology*, *28*, 274–328.

White, P. (2006). How well is causal structure inferred from co-occurrence information? *European Journal of Cognitive Psychology*, *18*, 454–480.

Yeung, S., & Griffiths, T. L. (2011). Estimating human priors on causal strength. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1709–1714). Austin, TX: Cognitive Science Society.

Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, *76*, 1–29.