Beyond the Information Given

Causal Models in Learning and Reasoning

Michael R. Waldmann,[1] York Hagmayer,[1] and Aaron P. Blaisdell[2]

[1]University of Göttingen, Göttingen, Germany, and [2]University of California, Los Angeles

ABSTRACT—*The philosopher David Hume's conclusion that causal induction is solely based on observed associations still presents a puzzle to psychology. If we only acquired knowledge about statistical covariations between observed events without accessing deeper information about causality, we would be unable to understand the differences between causal and spurious relations, between prediction and diagnosis, and between observational and interventional inferences. All these distinctions require a deep understanding of causality that goes beyond the information given. We report a number of recent studies that demonstrate that people and rats do not stick to the superficial level of event covariations but reason and learn on the basis of deeper causal representations. Causal-model theory provides a unified account of this remarkable competence.*

KEYWORDS—*causality; learning; reasoning*

People's ability to predict future events, explain past events, and choose appropriate actions to achieve goals belongs to the most central cognitive competencies necessary for being a successful agent in the world.

How are the regularities in the world learned, stored, and accessed? An intuitively plausible story that has been told in philosophy for many centuries assumes that the world contains causes that have the power to generate effects and that people learn about these causal systems. The philosopher David Hume questioned this view in his seminal writings. He analyzed situations in which people learn about causal relations and did not detect any empirical input that might correspond to evidence for causal powers of events. What he found instead was temporally ordered successions of event pairs, but nothing beyond that. He concluded from this that causality is a cognitive illusion triggered by associations. It is a view that found famous followers, such as the philosopher Bertrand Russell, who considered causality a concept that has no place in modern science.

The psychology of learning has adopted Hume's heritage (see Allan, 2005). According to many learning theories, causal predictions are driven by associative relations that have been learned on the basis of observed covariations between events. Similar to Pavlov's dog who has learned to predict food when it hears a tone (i.e., classical conditioning), or to a rat's learning that a lever press produces food (i.e., instrumental conditioning), we learn about causal relations. There is no need for the concept of causality in this view.

CAUSAL-MODEL THEORY: BEYOND COVARIATIONS

Hume's analysis and subsequent associative theories present a puzzle. They seem to have correctly observed that causal learning's inputs largely consist of covariation information. It can be shown, however, that mental representations that merely mirror such inputs cannot explain the competencies people have in dealing with causal situations (see also Buehner & Cheng, 2005). If people had no causal knowledge, they could not represent the difference between causal and spurious statistical relations, such as the relation between barometers and the weather. Covariational knowledge also fails to differentiate between causes and effects, which is a central distinction for planning actions. Finally, causal models entail statistical relations between events that are helpful in learning. For example, multiple effects of a common cause are correlated in a predictable way; the same is not true of multiple causes of a common effect.

Figure 1 shows a simplified causal model of people's knowledge of the flu. Such diagrams are commonly used to represent *causal models*, which in psychology were first proposed by Waldmann and Holyoak (1992). The arrows represent asymmetric causal relations directed from causes to effects. People can reason correctly about causality if they have causal-model representations (as in Fig. 1), but how do people generate causal models out of covariational information? The answer is that people have a natural tendency to assume the existence of causal relations, which leads them to align covariational input with

Address correspondence to M.R. Waldmann, Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany; e-mail: michael.waldmann@bio.uni-goettingen.de.

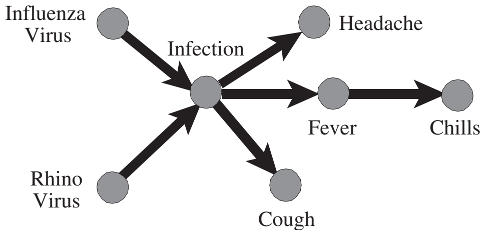

Fig. 1. Example of a causal model for the flu. The arrows represent causal relations directed from causes to effects.

causal-model representations (see also Buehner & Cheng, 2005). Similar to using sparse language input to induce a grammar that goes beyond the learning input, people have a tendency to represent some events as causes with the power to generate or prevent effects, and they build causal networks that can be used for inferences and planning. This core capacity to interpret the world in terms of causal relations can be applied to various domains (e.g., physical, biological, psychological) and merely requires the distinction between causes and effects. However, once acquired, knowledge may affect how further causal knowledge is learned.

Learning involves acquiring knowledge about the *structure* of the causal model and its *parameters*—such as the power of the causes to generate their effects (i.e., causal strengths) and the frequencies of the causes (i.e., base rates). It is obvious that parameters are estimated on the basis of the statistical properties of observations; but how do people learn about the structure of causal models? Although some researchers assume that causal structures are often learned through covariation information alone (Gopnik et al., 2004), our research indicates that *cues*—such as that causes typically temporally precede their effects—are used that suggest causal structures (see Lagnado, Waldmann, Hagmayer, & Sloman, in press). The main goal of our research has been to investigate how people enrich the covariational input to arrive at causal-model representations and how they use these representations for inferences and planning.

SENSITIVITY TO CAUSAL DIRECTIONALITY

The distinction between causes and effects is a central feature of causal representations. Thus, one line of research focused on whether there is evidence that people assign these distinct causal roles to the learning input (see Waldmann, 1996, for a summary of early work).

Fenker, Waldmann, and Holyoak (2005) have investigated this question in a semantic-memory task. The question of interest was whether causal relations are represented and accessed differently from associative relations in memory. In the experiments, participants were presented with pairs of words, one after another, either describing events that referred to a cause (e.g., “spark”) or an effect (e.g., “fire”). Both the temporal order of word presentation and the question to which participants had to respond were manipulated. Questions referring to the existence of a causal relation were answered faster when the first word referred to a cause and the second word to its effect than vice versa. No such asymmetry was observed, however, with questions referring to the associative relation—that is, whether the words were related in some meaningful way. People appear to distinguish the roles of cause and effect when queried specifically about a causal relation but not when the same information is evaluated for the presence of an associative relation (see also Satpute et al., 2005, for brain-imaging research supporting these conclusions).

Whereas semantic-memory tasks target the results of learning, there is also evidence that people go beyond the information given in trial-by-trial learning tasks (see also Waldmann, 2000). The general idea was to present participants in different conditions with identical covarying events. If Hume was right, and learning simply consists of picking up these observed covariations, the outcome of learning should be the same in the different conditions. In one study, Waldmann (2001) presented learners first with cues that represented substances in hypothetical patients’ blood and then gave feedback about fictitious blood diseases. According to associative-learning theories, learners should have learned to predict the disease from information about the presence of the substances. However, two conditions manipulated—through initial instructions—whether learners interpreted the substances (i.e., cues) as effects of the diseases (common-cause model) or as causes (common-effect model). The results showed that causal models guided how the learning input was processed. Learners treated the substances as potentially competing explanations of the disease in the common-effect condition, whereas the substances were treated as collateral effects of a common cause in the contrasting condition. Thus, despite the fact that all learners observed the same sequence of events, they assigned different causal roles to these events, and consequently they made different inferences. These inferences were based on statistical implications entailed by the different causal models; they didn’t solely reflect the observed statistical patterns.

Causal structures and their parameters are not independent entities but are deeply intertwined. The causal strength between a cause and an effect, for example, needs to be estimated differently depending on whether or not there is a confounding alternative cause. For example, if one learns the causal strength between rhino virus and infection, one needs to control for the possible confound, influenza virus, but not for effects of infection (e.g., fever; see Fig. 1). Waldmann and Hagmayer (2001) have shown that learners are indeed sensitive to the causal roles of events when estimating causal strength.

SEEING VERSUS DOING: TWO TYPES OF PREDICTIONS FROM OBSERVATIONAL DATA

Predicting and diagnosing on the basis of observed events are both examples of observational inferences (“seeing”). Causal knowledge also underlies interventional inferences (“doing”). Sometimes these two types of predictions coincide, but very often they do not. Manipulating barometers does not change the weather. Thus, the observed covariations alone do not allow for making correct inferences; the learner needs to go beyond the information given and assign causal roles to the observed events.

An associationist might respond that human and nonhuman animals could distinguish between seeing and doing on the basis of observational (i.e., classical) and instrumental conditioning. One may, for example, have learned that the barometer predicts the weather in an observational-learning setting and in parallel may have tried to fiddle with the barometer, which showed no effect on the weather. This solution only works, however, if learners are provided with both kinds of learning experiences, not if they only passively observe covarying events and then are requested to make both observational and interventional predictions.

We (Waldmann & Hagmayer, 2005) tested people’s competence to derive predictions for hypothetical observations and hypothetical interventions from causal models that had been learned purely through observation. In a fictitious medical scenario, participants were told that scientists hypothesized that three hormones, “pixin,” “sonin,” and “xanthan,” are related through either a common-cause or a causal-chain model (see Fig. 2, left). All participants in the two conditions received identical observational data indicating that the three hormones were connected by probabilistic causal relations. In the subsequent test phase, learners were asked to imagine a new group of test animals and to make predictions about hypothetical observations of sonin and about hypothetical interventions that increased sonin in the blood by means of inoculations. The observational inference could be modeled on the basis of the two presented causal models. Since the three hormones were statistically related in both causal models, the observation of the presence of sonin allowed participants to reason that pixin and consequently xanthan were also very likely to be present.

Interventional predictions often require modifications of causal models. In the common-cause model, an intervention that adds sonin to the blood leads to the consequence that the levels of sonin are now determined by this intervention and no longer by its usual cause (pixin), whose causal influence is preempted by the novel intervention. One way to model this intervention is to remove the arrow from pixin, sonin’s normal cause that is being explained away by the new intervention. The removal expresses that pixin is no longer a cause of sonin (see Fig. 2, right). Due to the removal of the arrow in the common-cause model, sonin becomes independent of xanthan, so that regardless of whether sonin is increased or decreased by an intervention, the level of xanthan should remain at the same level.

The chain condition served as a control that showed that seeing and doing don’t always lead to different predictions (see Fig. 2). Since there are no causes of sonin that are being preempted, an intervention does not necessitate a modification of the causal model. As a consequence, participants should make identical predictions for the observational and interventional questions. In our experiments, participants’ responses corre-

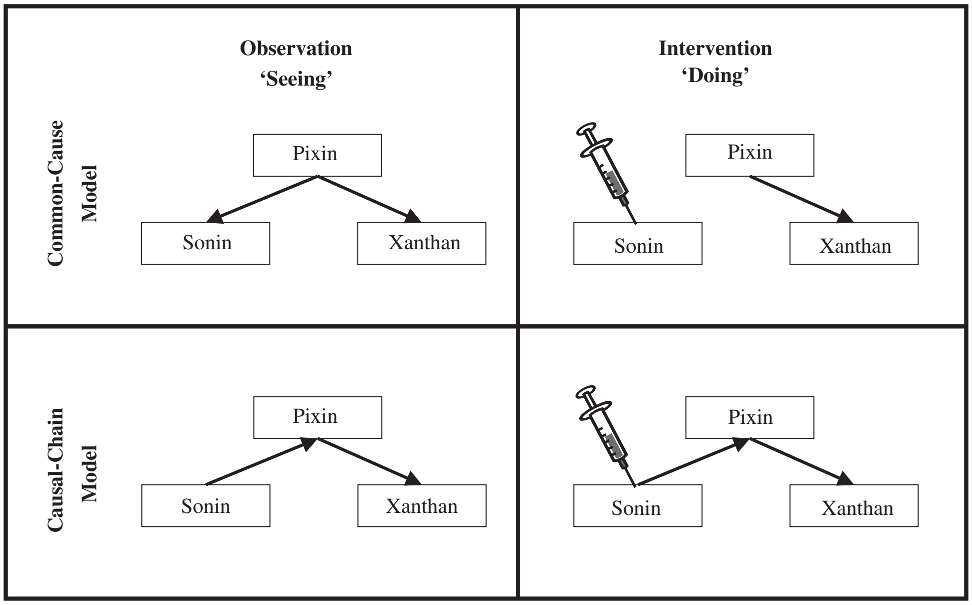

Fig. 2. Observational (‘‘seeing’’) and interventional (‘‘doing’’) predictions in a common-cause and causal-chain model, in which three fictional hormones are causally connected. The left side shows the models presented in the learning phase, which could be used for observational predictions. The right side depicts the models underlying the predictions of the outcomes of interventions. A hypothetical intervention in which humans or animals are inoculated with sonin requires the deletion of a causal arrow in the common-cause model but not in the causal-chain model.

sponded to these predictions remarkably well. They were capable of predicting patterns they had never observed, which indicates that they used causal-model representations to transform identical covariational information into different types of predictions. In several further experiments we manipulated the parameters of the model (base rates of the events, causal strength) and showed that participants' predictions were not only driven by the structure of the causal models but also by the learned parameters.

Causal Reasoning in Rats

We claim that humans have the natural capacity to form causal representations. How about nonhuman animals? A number of researchers have asserted that causal reasoning and learning are faculties that form a dividing line between humans and nonhuman animals. Recent research by us (Blaisdell, Sawa, Leising, & Waldmann, 2006) casts doubt on this conclusion.

In one experiment, rats went through a purely observational learning phase in which a light was sometimes followed by a tone and at other times followed by food. Importantly, no instrumental learning took place. When in the subsequent observational test phase the rats again heard the tone ("seeing"), they showed that they expected food in the niche in which it was typically delivered. Apparently they reasoned through the causal model link-by-link, from the tone through the light to the probable presence of food. By contrast, in a second test a lever that the rats had never seen before was introduced into the cage ("doing"). Whenever the rats curiously pressed the lever, the same tone was presented. Now, although tone and food had been associated by the rats in the learning phase as indicated in the observational test phase, they were less inclined to search for food after the lever presses. Apparently they reasoned that they—and not the light—were the cause of the tone, which led to their reluctance to expect food.

In a second study, a causal chain was presented in which the tone preceded the light, which in turn preceded food. Consistent with causal-model theory, the rats expected food regardless of whether they observed the tone or generated it with the lever. This shows that they were not generally reluctant to expect food after a novel intervention. The results revealed a deep understanding of causal relations and demonstrate that rats correctly differentiated between seeing and doing and among different causal models.

Limitations of Causal Reasoning

Although people exhibited a sophisticated ability to reason with causal models, there is also evidence for limitations. For example, Waldmann and Walker (2005) have shown that people have difficulties with transforming covariation information into causal-model representations when the task is complex or when the learner operates at her information-processing limit. Reips and Waldmann (in press) have similarly found that base rates may be neglected in learning tasks in which those rates are not crucial for error-free performance. Finally, Waldmann (in press) has discovered that domain assumptions affect how multiple causes are combined in the prediction of an effect. However, the combination rule is also influenced by the way the task is presented, which again shows that the capacity to form causal-model representations is also affected by task characteristics.

CONCLUSION

Hume has presented us with the puzzle: How do we acquire causal knowledge when we only observe covariation information? We have reported a number of studies showing that both human and nonhuman animals have a natural tendency to translate covariations into causal-model representations.

One important question for future research is to explore the generality and differences of causal-reasoning capacities across species. Another interesting question will be to analyze the relation between causal reasoning and rational models, such as causal Bayes nets (Gopnik et al., 2004). Our findings on limitations of causal reasoning suggest that such models, if interpreted as psychological theories, may often exaggerate what human and nonhuman animals can do. Answers to these questions hold profound implications concerning the structure, origin, and evolution of causal reasoning, an invaluable cognitive tool for exploiting one's world.

Recommended Reading

Buehner, M., & Cheng, P. (2005). (See References)

Lagnado, D.A., Waldmann, M.R., Hagmayer, Y., & Sloman, S.A. (in press). (See References)

Waldmann, M.R. (1996). (See References)

REFERENCES

Allan, L.G. (2005). Learning of contingent relationships [Special issue]. *Learning & Behavior*, *33*(2).

Blaisdell, A.P., Sawa, K., Leising, K.J., & Waldmann, M.R. (2006). Causal reasoning in rats. *Science*, *311*, 1020–1022.

Buehner, M., & Cheng, P. (2005). Causal learning. In K.J. Holyoak & B. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 143–168). Cambridge, England: Cambridge University Press.

Fenker, D.B., Waldmann, M.R., & Holyoak, K.J. (2005). Accessing causal relations in semantic memory. *Memory & Cognition*, *33*, 1036–1046.

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.

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

Reips, U.-D., & Waldmann, M.R. (in press). When learning order affects sensitivity to base rates: Challenges for theories of causal learning. *Experimental Psychology*.

Satpute, A.J., Fenker, D., Waldmann, M.R., Tabibnia, G., Holyoak, K.J., & Lieberman, M. (2005). An fMRI study of causal judgments. *European Journal of Neuroscience*, *22*, 1233–1238.

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

Waldmann, M.R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 53–76.

Waldmann, M.R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychological Bulletin & Review*, *8*, 600–608.

Waldmann, M.R. (in press). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science*.

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

Waldmann, M.R., & Hagmayer, Y. (2005). Seeing vs. doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 216–227.

Waldmann, M.R., & Holyoak, K.J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*, 222–236.

Waldmann, M.R., & Walker, J.M. (2005). Competence and performance in causal learning. *Learning & Behavior*, *33*, 211–229.