

---

# Bayesian Causal Induction

---

**Pedro A. Ortega**

Max-Planck Institute for Biological Cybernetics  
Spemannstraße 38, 72070 Tübingen  
peortega@dcc.uchile.cl

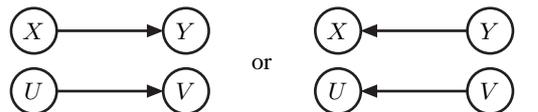
## Abstract

Discovering causal relationships is a hard task, often hindered by the need for intervention, and often requiring large amounts of data to resolve statistical uncertainty. However, humans quickly arrive at useful causal relationships. One possible reason is that humans extrapolate from past experience to new, unseen situations: that is, they encode beliefs over causal invariances, allowing for sound generalization from the observations they obtain from directly acting in the world. Here we outline a Bayesian model of causal induction where beliefs over competing causal hypotheses are modeled using probability trees. Based on this model, we illustrate why, in the general case, we need interventions plus constraints on our causal hypotheses in order to extract causal information from our experience.

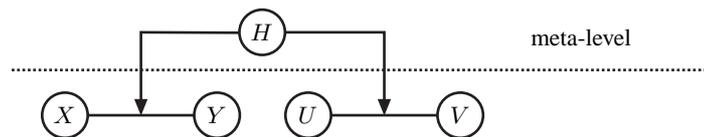
## 1 Introduction

A fundamental problem of statistical causality is the problem of *causal induction*<sup>1</sup>; namely, the generalization from particular instances to abstract causal laws [5, 4]. For instance, how can you conclude that it is dangerous to ride a bike on ice from a bad slip fall on wet floor?

In this work, we are concerned with the following problem: how do we determine from experience whether “ $X \rightarrow Y$  and  $U \rightarrow V$ ” or “ $Y \rightarrow X$  and  $V \rightarrow U$ ”? That is, which of the two causal hypotheses over  $X, Y, U$  and  $V$  is correct,



even in the case when both models represent identical joint distributions? Furthermore, if we collect evidence supporting the claim “ $X \rightarrow Y$ ”, how do we extrapolate this to the (yet unseen) situation “ $U \rightarrow V$ ”? The main challenge in this problem is that the hypothesis, say  $H$ , is a random variable that controls the very causal structure. That is, a more accurate graphical representation would be the model:



which cannot be analyzed using the framework of graphical models alone because the random variable  $H$  operates on a meta-level of the graphical model over  $X, Y, U$  and  $V$ .

---

<sup>1</sup>For a thorough treatment of non-causal induction, refer to [8].

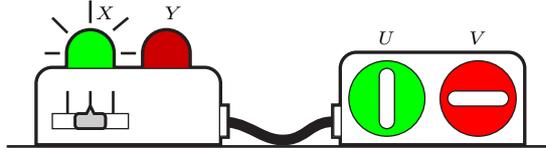


Figure 1: (Left) A device with a green and a red light bulb. A switch allows controlling the state of the green light: either “on”, “off” or “undisturbed”. (Right) A second device having a green spinner and a red spinner, both of which can either lock into a horizontal or vertical position. The two devices are connected through a cable, establishing thus a relation among their randomizing mechanisms.

In this work these difficulties are overcome by using a probability tree to model the causal structure over the random events [9]. Probability trees can encode<sup>2</sup> alternative causal realizations, and in particular alternative causal hypotheses. All random variables are of the same type—no distinctions between meta-levels are needed. Furthermore, we define interventions [7] on probability trees so as to predict the statistical behavior after manipulation. We then show that such a formalization leads to a probabilistic method for causal induction that is intuitively appealing.

## 2 Causal Induction in Probability Trees

Imagine we are given a device with two light bulbs, one green ( $X$ ) and one red ( $Y$ ), whose states obey a hidden mechanism that correlates them positively. Moreover, the box has a switch that allows us controlling the state of the green bulb: we can either leave it undisturbed, or we can intercept the mechanism by turning the light on or off as we please (Figure 1, left device). We encode the “on” and “off” states of the green light as  $X = x$  and  $X = \neg x$  respectively. Analogously,  $Y = y$  and  $Y = \neg y$  denote the “on” and “off” states of the red light. We ponder the explanatory power of two competing hypotheses: either “green causes red” ( $H = h$ ) or “red causes green” ( $H = \neg h$ ).

### 2.1 Representation

Assume that the probabilities governing the realization of  $H$ ,  $X$  and  $Y$  are as detailed in Figure 2a. In this tree, each (internal) node is interpreted as a causal mechanism; hence a path from the root node to one of the leaves corresponds to a particular sequential realization of causal mechanisms<sup>3</sup>. The logic underlying the structure of this tree is self-explanatory:

1. *Causal precedence*: A node causally precedes its descendants. For instance, the root node corresponding to the sure event  $\Omega$  causally precedes all other nodes.
2. *Resolution of variables*: Each node resolves the value of a random variable. For instance, given the node corresponding to  $H = h$  and  $X = \neg x$ , either  $Y = y$  will happen with probability  $P(y|h, \neg x) = \frac{1}{4}$  or  $Y = \neg y$  with probability  $P(\neg y|h, \neg x) = \frac{3}{4}$ .
3. *Heterogeneous order*: The resolution order of random variables can vary across different branches. For instance,  $X$  precedes  $Y$  under  $H = h$ , but  $Y$  precedes  $X$  under  $H = \neg h$ . This allows modeling different causal hypotheses.

While the probability tree represents our subjective model explaining the order in which the random values are resolved, it does not necessarily correspond to the temporal order in which the events are revealed to us. So for instance, under hypothesis  $H = h$ , the value of the variable  $Y$  might be revealed before  $X$ , even though  $X$  causally precedes  $Y$ ; and the hypothesis  $H$ , which precedes both  $X$  and  $Y$ , is never observed.

<sup>2</sup>Conditional independencies are also captured within a probability tree [9, Chapter 8.2].

<sup>3</sup>Note that the set of paths is the sample space of the experiment’s probability space.

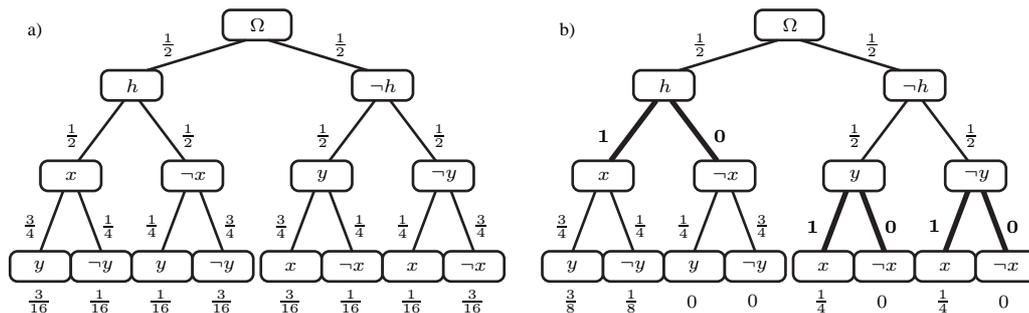


Figure 2: a) The probability tree representing the statistics of the device with two lights. The probability of a realization (written under the leaves) is calculated by multiplying the probabilities starting from the root until a leaf is reached. Note that the two hypotheses are statistically indistinguishable. b) The probability tree resulting from (a) after setting  $X = x$ .

## 2.2 Interventions

Suppose we observe that both lights are on. Have we learned anything about their causal dependency? A brief calculation shows that this is not the case because the posterior probabilities are equal to the prior probabilities:

$$P(h|x, y) = \frac{P(y|h, x)P(x|h)P(h)}{P(y|h, x)P(x|h)P(h) + P(x|\neg h, y)P(y|\neg h)P(\neg h)} = \frac{\frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2}}{\frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{2} = P(h).$$

This makes sense intuitively, because by just observing that the two lights are on, it is statistically impossible to tell which one caused the other. Note how the factorization of the likelihood  $P(x, y|H)$  depends on whether  $H = h$  or  $H = \neg h$ . How do we extract causal information then? To answer this question, we make use of a crucial insight of statistical causality:

To obtain new causal information from statistical data, old causal information needs to be supplied (paraphrased as “no causes in, no causes out” [2] or “to find out what happens when you kick the system, you have to kick the system” [1]).

Thus, we now repeat our experiment, but this time *we* turn on the green light ( $X = x$ ). We reflect this choice by changing all the mechanisms that resolve the random variable  $X$ , placing all the probability mass on the outcome  $X = x$  (see Figure 2b). Assume that we subsequently observe that the second light is on. Then, the posterior probabilities are

$$P(h|\hat{x}, y) = \frac{P(y|h, \hat{x})P(\hat{x}|h)P(h)}{P(y|h, \hat{x})P(\hat{x}|h)P(h) + P(\hat{x}|\neg h, y)P(y|\neg h)P(\neg h)} = \frac{\frac{3}{4} \cdot 1 \cdot \frac{1}{2}}{\frac{3}{4} \cdot 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \frac{3}{5},$$

where  $\hat{x}$  is Pearl’s notation to indicate a causal intervention of  $X$ . Since  $P(h) > P(h|\hat{x}, y)$ , we have gathered evidence favoring the hypothesis “green causes red”. This was only possible because our intervention introduced a statistical asymmetry among the two hypotheses that did not exist before.

## 2.3 Extrapolation

Let us now connect a second device to the first one (Figure 1, right device). This device carries two spinners, a green ( $U$ ) and a red one ( $V$ ). A hidden randomizing mechanism chooses their orientations (either horizontal or vertical) independently from the state of the colored lights. However, the connection and the mysterious color coding suggest that there must be a relation between the two randomizing mechanisms. Hence, we impose that the combined system either follows the law “green causes red” or “red causes green”—intentionally excluding the cases “ $X \rightarrow Y$  and  $U \leftarrow V$ ” and “ $X \leftarrow Y$  and  $U \rightarrow V$ ”.

The probability tree over the random variables  $X, Y, U$  and  $V$  extends the probability tree from Figure 2a by appending sub-trees over  $U$  and  $V$  having the restriction that the nodes resolving  $U$  precede the nodes resolving  $V$  under hypothesis  $H = h$ , and that the nodes resolving  $V$  precede the nodes resolving  $U$  in the case  $H = \neg h$ .

Note however, that for this new tree, the posterior probability over the hypothesis “green causes red” given that we turned on the green light and saw the red bulb lighting up is identical to the previous tree, namely  $P(h|\hat{x}, y) = \frac{3}{5}$ . The restriction we have imposed over the possible causal hypotheses has enabled us extrapolating causal information from our experience with  $X$  and  $Y$  to the yet unobserved variables  $U$  and  $V$ . This extrapolation would not have been possible if we had kept all four causal hypotheses. Hence, in the general case, causal extrapolation rests on constraints on our causal hypotheses.

### 3 Concluding Remarks

The problem of causal induction has been addressed relatively recently by the statistics and machine learning community, mainly under the context of graphical models [7, 11, 10, 3, 4, 6]. This has led to the development of many algorithms that propose a suitable causal graphical model explaining the data. Many of these algorithms rely on independence assumptions, and hence naturally they proceed by exploiting the independence relations found in the data to construct a causal model.

This work outlines a general method for causal induction that is Bayesian in nature and does not rely on independence assumptions. It is based on the idea of combining probability trees [9] with interventions [7] for predicting the behavior of a manipulated system with multiple causal hypotheses. We have seen that both the interventions and the (constraints on the) causal hypotheses introduce statistical asymmetries that permit the extraction and extrapolation of causal information. Of course, this means that the amount and the forms of causal relations that we can discover are determined (a) by our constraints on the set of causal hypotheses and (b) by the interventions that we are allowed to apply to the system (and essentially, to our hypotheses). In a sense, one could say that we are “imprinting our own causal laws onto our experience”. However, this raises more fundamental questions that we have not addressed here: where do these constraints on our causal hypotheses come from and what logic do they obey?

#### Acknowledgments

The author would like to thank Daniel A. Braun, Philip Dawid, David Balduzzi, Samory Kpotufe, Theofanis Karaletsos, Eleni Sgouritsa and Marcus Hutter for discussions of previous versions of this manuscript. This study was supported by the Emmy Noether Grant BR 4164/1-1, “Computational and Biological Principles of Sensorimotor Learning”.

#### References

- [1] G.E.P. Box. Use and abuse of regression. *Technometrics*, 8(4):625–629, 1966.
- [2] N. Cartwright. *Nature’s Capacities and Their Measurements*. Oxford University Press, 1994.
- [3] A. P. Dawid. Fundamentals of statistical causality. Research Report 279, Department of Statistical Science, University College London, 2007.
- [4] T.L. Griffiths and J.B. Tenenbaum. From mere coincidences to meaningful discoveries. *Cognition*, 103(2):180–226, 2007.
- [5] D. Hume. *A Treatise on Human Nature*. 1739-1740.
- [6] J.M. Mooij, O. Stegle, D. Janzing, K. Zhang, and B. Schölkopf. Probabilistic latent variable models for distinguishing between cause and effect. In *Advances in Neural Information Processing Systems*, 2010.
- [7] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2009.
- [8] S. Rathmanner and M. Hutter. A philosophical treatise of universal induction. *Entropy*, 13(6):1076–1136, 2011.
- [9] G. Shafer. *The Art of Causal Conjecture*. MIT Press, 1996.
- [10] R. Silva. *Automatic Discovery of Latent Variable Models*. PhD thesis, Machine Learning Department, Carnegie Mellon University, 2005.
- [11] P. Spirtes and R. Scheines. *Causation, Prediction, and Search, Second Edition*. MIT Press, 2001.