Solving the Flagpole Problem[*]

Abstract: In this paper I demonstrate that the causal structure of flagpole-like systems can be determined by application of causal graph theory. Additional information about the ordering of events in time or about how parameters of the systems of interest can be manipulated is not needed.

1. Introduction

The length of a flagpole's shadow can be causally explained by reference to the solar altitude and the flagpole's height, while neither the solar altitude nor the flagpole's height can be causally *explained* by the remaining two parameters. This explanatory asymmetry is due to the fact that the solar altitude and the height of the flagpole are causally relevant to the length of the flagpole's shadow and not vice versa. But how can we know this? Since it was always a central concern of philosophers to achieve the intended results without making more preconditions than absolutely necessary, this question can be rephrased as follows: Is there a way to grasp the correct causal structure on the basis of observation and without presupposing dispensable causal knowledge (this may be knowledge that the cause occurs always earlier than its effect or knowledge about how to intervene on specific parameters without directly manipulating some of the other parameters of the system under investigation)? This yet unanswered and frequently discussed question (cf. Schurz 2001, for instance) constitutes what I call the *flagpole problem*. In whichever way this problem might be solved, it is a matter of fact that the strict correlations between the solar altitude, the height of the flagpole, and the length of the flagpole's shadow alone are not sufficient for determining the causal structure of the system – if any two of these three parameters are known, then the third one can be computed, and thus there is no detectable asymmetry in terms of statistical regularities alone. In this paper I will provide a new approach of how the flagpole problem can be solved: The underlying causal structure (and hence, the causal asymmetry) of the flagpole scenario (and thus, the causal structure of all flagpole-like systems) can be uniquely determined by application of causal graph theory.

The paper is structured in the following way: In section 2 I will introduce the basic terms and principles of causal graph theory. In section 3 I will present a discovery algorithm for causal structures whose correctness (within a causal graph framework) can be proven. In section 4 I will demonstrate how the causal structure underlying the flagpole example can be uniquely determined by application of this algorithm.

2. Causal graph theory

Causal graph theory, as it was developed by researchers like Judea Pearl (Pearl 2009; Pearl 1997), Peter Spirtes, Clark Glymour, and Richard Scheines (Spirtes et al. 2000), is based on Hans Reichenbach's idea of connecting causality with probability (Reichenbach 1956; Reichenbach 1949). A causal graph G is an ordered pair $\langle V,E \rangle$, where V is a non-empty and finite set of so-called *vertices,* i.e., statistical variables representing events at a highly abstract level. Statistical variables shall be referred to via upper-case letters *'X', 'Y', 'Z'* etc., while lower-case letters *'x', 'y', 'z'* etc. shall stand for individual variables for specific values of *'X', 'Y', 'Z'* etc., respectively. Terms like *'X = x'* shall stand for 'variable X takes value x' and are interpreted as (still very abstract) type-level events*. E* is an asymmetric (and thus, irreflexive) relation among nodes in V ($E \subseteq V \times V$). *'E(X,Y)'* is a so-called *directed edge* that can (and, in the following, will) be graphically represented via a *causal arrow* (*'X→Y'*). In causal graphs*,* terms like *'X→Y'* are interpreted as 'X is a *direct cause* of Y'. A chain of causal unidirectional arrows from X to Y (e.g., $X \to Z_1 \to Z_2 \to Z_3 \to Y$) is called a *directed path* from X to Y ($X \to \to Y$). Given this characterization, $X \to Y$ turns out to be a special case of $X \to \to Y$. Further causal relations among variables in V can be captured via the following "family-terminology": Elements of the set of all variables X_i with $X_i \to Y$ ($Pa(Y)$) are called *parents* of Y. Elements of the set of all variables X_i with $X_i \to \to Y$ ($Anc(Y)$) are called *ancestors* of $Y.$ Elements of the set of all variables Y_i with $X \to Y_i$ ($Chi(X)$) are called *children* of $X.$ Elements of the set of all variables Y_i with $X \to \to Y_i$ ($Des(X)$) are called *descendants* of $X.$

When it comes to causal inference we have to distinguish between the "true" causal structure of the world and a huge amount of man-made causal graphs intended to capture some interesting parts of this highly complex causal structure. If there were a causal graph $G^t = \langle V^t, E^t \rangle$ that would capture the whole causal structure of the world, V^t would contain all in some way causally relevant variables and thus, indeed, all variables that are common causes of any pair of variables in V^t. If we are looking at any ordinary man-made causal graph $G = \langle V,E \rangle$, on the other hand, it is not so clear that all "true" common causes of pairs of variables in V are also included in V – a precondition needed for inferring

causal structures in a more accurate way (see also section 4). It is because of this that we have to distinguish between *causally sufficient* and *insufficient* variable sets *V*:

> *Causal Sufficiency:* A variable set *V* is *causally sufficient* if and only if all "true" common causes of all pairs of variables in *V* are also elements of *V*. (cf. Glymour et al. 1991)

Causal graphs can be connected to probability distributions via the following principle (cf. Spirtes et al. 2000) going back to Reichenbach 1956:

> *Causal Markov Condition:* A causal graph $G = \langle V,E \rangle$ and a probability distribution *P* over *V* satisfy the *Causal Markov Condition* if and only if for every $X \in V$: *X* is probabilistically independent of $V \setminus Des(X)$ conditional on $Pa(X)$.

Two variables (or sets of variables) *X* and *Y* are probabilistically independent conditional on a variable (or a set of variables) *M* ($X \perp_P Y | M$) if and only if $P(x|y,m) = P(x|m)$ for all values *x, y* and *m* of *X, Y,* and *M*, provided $P(y,m) > 0$. In other words: Additional information about the actual value of *Y* would not change the probability of any value *x* of *X,* given evidence *M = m*.

The *Causal Markov Condition* determines a set of conditional independencies which have to hold in every probability distribution compatible with a given causal graph. In most cases, though, there is a huge amount of causal graphs compatible with a given probability distribution. If, however, it can be guaranteed that a given probability distribution captures all and only the conditional independencies produced by the underlying causal structure, then the set of compatible causal graphs shrinks drastically. A probability distribution that fulfills this condition is called *faithful* (cf. Spirtes et al. 2000):

> *Faithfulness Condition:* A probability distribution *P* over a set of variables *V* is *faithful* if and only if it captures all and only the (conditional) independencies produced by its underlying causal structure.

3. The SGS algorithm

A multitude of more or less utile discovery algorithms for causal structures were developed over the last decades (Pearl 2009; Spirtes et al. 2000; Glymour et al. 1991). One of the more intuitive of these algorithms is the SGS algorithm, originally developed by Spirtes, Glymour, and Scheines (Spirtes et al. 2000, p. 82). Given a *faithful* probability distribution over a *causally sufficient* set of variables *V*, the SGS algorithm uniquely determines the skeleton of the causal graph $G = \langle V,E \rangle$ (which correctly

represents the "true" causal structure underlying the correlations among the variables in *V*), and, in addition, often uncovers a noteworthy amount of causal arrows between pairs of vertices of this causal graph:

SGS algorithm:

Presupposition: A *faithful* probability distribution over a *causally sufficient* variable set *V.*

Step 1*:* Form the complete undirected graph on the vertex set *V.*

Step 2*:* For all *X,Y∈V*: If *X* and *Y* are probabilistically independent conditional on any subset *M* of *V* \ {*X,Y*}*,* then remove the undirected edge between *X* and *Y* in the graph resulting from step 1.

Step 3*:* For all *X,Y,Z∈V*: If *X—Y—Z* (or *X→Y—Z*) and *X* and *Z* are non-adjacent (not *X—Z*), then orient *X—Y—Z* (and accordingly *X→Y—Z*) as *X→Y←Z* in the graph resulting from step 2 if and only if *X* and Y are probabilistically dependent conditional on every subset *M* of *V* \ {*X,Y*} with *Y∈M.*

Step 4*:* For all *X,Y,Z∈V*: If *X→Y—Z* and *X* and *Z* are non-adjacent, then orient *X→Y—Z* as *X→Y→Z* in the graph resulting from step 3*.* And: If *X—Y* and *X→→Y,* then orient *X—Y* as *X→Y* in the graph resulting from step 3*.*

The correctness of the SGS algorithm for *faithful* probability distributions over *causally sufficient* variable sets can be proven (Spirtes et al. 2000, p. 82). At this point we have all the formal tools to dare tackle the *flagpole problem*.

4. Solving the flagpole problem

Remember the *flagpole problem* as introduced in section 1: The solar altitude (*A = a*) and the height of the flagpole (*H = h*) are both causally relevant for the length of the flagpole's shadow (*L = l*). But how can we know this on the basis of observational evidence alone, i.e., without any additional causal information?

The *flagpole problem* can be solved in the following way: If we take many random observations (we look at many different flagpoles and not solely at a single one and take many different solar

altitudes) and write down the observed values of *A, H,* and *L* over a sufficiently long period (e.g., a year), we get a list of combinations of values of *A, H,* and *L.* By application of statistical methods, we can get a *faithful* probability distribution *P* over *V = {A,H,L}* containing the following and only the following conditional independence relation among variables in *V*: $A \perp_P H | \varnothing$ (to get the "right" dependence/independence relations in *P* it is important to measure *A, H,* and *L* at least approximately at the same time).

That *P* will contain this and only this conditional independence relation – i.e., that *P* is *faithful* – can be motivated via the following considerations: Specific solar altitudes are typically not accompanied by specific heights of flagpoles and hence, there are no *A*-values *a* which have a probabilistic influence on some *H*-values *h.* In addition, we can change *H*'s value however we wish to without changing the solar altitude. [1] Thus, the solar altitude *A* is without a doubt independent of the height of the flagpole *H* ($A \perp_P H | \varnothing$). But why are there no further independencies in the system under investigation? If the sun is very low (*A = low*), then the probability for a longer flagpole shadow (*L = high*) is increased – a certain range of *L*-values is probabilistically sensitive to a specific range of *A*-values. So the length of the flagpole's shadow *L* is probabilistically dependent on the solar altitude *A*. The same goes for the height of the flagpole *H* and the length of its shadow *L*: If the height of the flagpole is very low (*H = low*), then the probability for a very short flagpole shadow (*L = low*) is increased – a certain range of *L*-values is probabilistically sensitive to a specific range of *H*-values. So the length of the flagpole's shadow *L* is probabilistically dependent on the height of the flagpole *H*. The only remaining possible probabilistic independencies are those we could get among pairs of variables in *V* by conditioning on the third variable in *V*. Since the value of any variable in *V* can be computed given the actual values of the other two variables (see section 1), there are no such probabilistic independencies within the flagpole example, and hence *P* is *faithful*.

It is now time to apply the SGS algorithm to our example: Let us therefore assume that the system *V = {A,H,L}* is *causally sufficient.* Via applying step 1 *to V and P* we get the undirected graph consisting of *A—L—H* and *A—H;* so every pair of vertices in *V* is connected via an undirected edge after step 1. According to step 2, we have to remove the undirected edge between *A* and *H* due to the fact that conditioning on the empty set makes *A* and *H* probabilistically independent. Since *A* and *L* are probabilistically dependent conditional on *H, L* and *H* are probabilistically dependent conditional on *A*, and *A* and *L* as well as *L* and *H* are probabilistically dependent conditional on the empty set, the graph resulting from step 2 is *A—L—H.* With this intermediate result we have successfully revealed

---

[1] Note: In section 1 I promised to solve the flagpole problem (i.e., to infer the causal structure of the flagpole scenario on the basis of a given probability distribution) without reference to interventions. However, interventions are not used in *determining* the causal structure of the flagpole scenario and thus, the use of interventions in motivating the dependence/independence relations in *P* is harmless in regard to this promise.

the skeleton of the much sought-after causal graph $G = \langle V,E \rangle$ that correctly represents the underlying causal structure of the flagpole example. Since our graph $A—L—H$ fulfills the antecedent-condition of step 3 and since A and H become probabilistically dependent after conditioning on every subset M of $V \setminus \{A,H\}$ with $L \in M$ (actually, there is only one such subset M, namely $M = \{L\}$), step 3 instructs us to orient $A—L—H$ as $A{\rightarrow}L{\leftarrow}H$. Since all undirected edges have already been removed/replaced by causal arrows, step 4 can be skipped and the causal graph resulting from application of the SGS algorithm turns out to be $A{\rightarrow}L{\leftarrow}H$.

5. Conclusion

In this paper I demonstrated an alternative solution to the *flagpole problem*. I gave a sketch of causal graph theory, including its most important terms and principles. Afterwards I presented the SGS algorithm which is custom-built to reveal causal structures on the basis of probability distributions and whose correctness (within a causal graph framework) can be proven. I argued that there is one and only one probabilistic independence within the flagpole example: The solar altitude A and the height of the flagpole H are probabilistically independent ($A \perp_P H|\varnothing$). A probability distribution (over the system $V = \{A,H,L\}$), containing only the independence $A \perp_P H|\varnothing$, is, according to the SGS algorithm, compatible with only one causal graph: $A{\rightarrow}L{\leftarrow}H$. This causal graph mirrors our intuitions: The solar altitude A and the height of a flagpole H are causally relevant for the length of the flagpole's shadow L (and not vice versa). This result can be generalized for any flagpole-like system V (i.e., any *causally sufficient* system V containing exactly three variables X, Y, and Z with a *faithful* probability distribution P over V with exactly one (conditional) independence relation $X \perp_P Y|\varnothing$).

6. References

Glymour, Clark, Spirtes, Peter, and Scheines, Richard. 1991. Causal Inference. *Erkenntnis* 35: 151-189.

Pearl, Judea. 2009. *Causality.* Cambridge: Cambridge University Press.

Pearl, Judea. 1997. *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kaufmann.

Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley: University of California Press.

Reichenbach, Hans. 1949. *The Theory of Probability*. Berkeley: University of California Press.

Spirtes, Peter, Glymour, Clark, and Scheines, Richard. 2000. *Causation, Prediction, and Search*. Cambridge: The MIT Press.

Schurz, Gerhard. 2001. Causal Asymmetries, Independent versus Dependent Variables, and the Direction of Time. In *Current Issues in Causation*, ed. Wolfgang Spohn, Marion Ledwig, Michael Esfeld, 47-67. Paderborn: Mentis.