# Notes on the Book “Causality, Probability and Time Notes”, a Book by Samantha Kleinberg

D. Gueorguiev 9/9/2023

## Notes on Types of Causes and their Representation

### Notes on Prima Facie Causes

While most of the work on causal inference is in identifying spurious causes, the set of potential causes to be explored is still constrained by specifying minimal conditions for causality. For some and , to identify as a possible cause of , must be temporally prior to and must change the probability of . Prima facie causes are those that satisfy these basic requirements. Recall that when we describe a cause and effect , both may be arbitrarily complex logical formulas. In the following examples, we refer to and but note now that there are no conditions on them other than that they must be valid PCTL state formulas.

First, the temporal priority condition of the causal relationship is specified in terms of the time that elapses between cause and effect, rather than the occurrence times of the cause and effect. If occurs at some time and occurs at a later time , the relationship is characterized by the time that elapses between them, . To state that after becomes true, will be true with probability at least in or fewer time units – but with at least one time unit between and (so that occurs in a particular time window after ) – one may write:

Satisfying this formula requires at least one and potentially many transitions between and , as long as the sum of probabilities of the paths between and taking at least one time unit is at least .

The transitions are assumed to each take one time unit, but there is no restriction on the definition of a time unit. If we only want to say that c is earlier than e, the lower bound will be 1 and the upper bound . In most cases we aim to infer the timing of this relationship. Then, the bounds in the second condition that follows (1 and ) can be replaced with any arbitrary and where , and .

The probabilistic nature of the relationship between cause and effect can be described in terms of the probability of reaching and states, and of the paths between and states. We need to specify that must occur at some point and that the conditional probability of given is greater than the marginal probability of . Here is the formal definition:

**Definition**: *prima facie* cause

Where c and e are PCTL formulas, c s a prima facie cause of e if there is a p such that the following conditions hold:

1. (state where c is true will be reached with non-zero probability)
2. (the probability for reaching, within time bounds, a state where is true from a state where is true is greater than or equal to )
3. (state where is true will be reached with probability less than )

When making inferences from data, this means c must occur at least once, and the conditional probability of given is greater than the marginal probability of e (usually calculated from frequencies).

### Example of causal structure

*,*

Figure 1: Example structure containing states with their labels and transitions with their associated probabilities. is the initial state.

*The Example*: Find the cause and probability of Death () for a suicidal person () who jumps from a building () that may () or may not () have a safety net. The cause is probabilistic so the person may survive that jump attempt ().

Here we will only test minimal conditions for potential causality – that the causes occur at least one time unit before their effects (which may take infinite amount of time to occur) and change their probabilities.

With and we will test whether is a prima facie cause of .

We know that c meets the first condition 1. In the prima facie cause definition since satisfies , and the probability of reaching from is greater than 0.

To determine whether alters the probability of , we need the marginal probability of . We can deal with sets of states. These are those that have no path to a state where is true and where is not currently true (set ) and where the probability of reaching a state where holds is 1 (set ). A third set is needed when using a lower bound with the leads-to operator (set , states where is true or there is a path to that is shorter than the minimum time window). A state can be in both and in cases where every path from it leads to a state where holds, but where some of these paths are shorter than the minimum time on the leads-to formula.

The probability of is calculated as follows: the sets of interest are , , . Set is empty because there is no minimum time bound here. The only state from which it is not possible to reach is so . Clearly, the only state where success is guaranteed, and thus is the only member of , is . The probability of e occurring at any time, represented by the probability of is given by , since is the initial state of the system. The temporal operator means “finally”, so with an infinite upper bound it is satisfied from a state if there is some path that eventually leads to a state where is true. The probability of this satisfaction is the probability of the set of paths from that set to the state (or states) where holds. Thus, we can write:

(1)

In order to calculate and we write:

(2)

(3)

We recall that since and since . Thus, we can find , , and finally . Thus, we find that .

The calculation of is straightforward – there is only one path which leads to from so the probability of is exactly

### Insignificant Causes – Intuitions

The conditions for *prima facie* causality are insufficient to distinguish between causes and non-causes, and the primary difference between probabilistic theories of causality is in how exactly they make this distinction. The two main types of methods ar those based on information and those based on manipulation. The information-based theories use the idea that a cause provides some information about an effect that cannot be gained in other ways and set about finding evidence for that. Manipulation theories hold that a cause is a way of bringing about an effect and can be understood in terms of how the probability or value of the effect changes when manipulating the cause to be true. One approach is not inherently superior to the other – there are counterexamples to both manipulation and information-based methods and neither subsumes the other. Manipulation is not possible in many cases, and therefore it is undesirable to require it. On other side, methods that aim to infer causal relationships from observational data – because it is readily available or because manipulations are costly, unfeasible, or unethical – generally use some variant of the information-based approach. The distinguishing feature between all variants of the information-based methods is how to quantify the information provided about the effect. The basic idea is that of holding fixed some set of information and then see how likely the effect is when the cause is present and absent relative to that set of information.

Desiderata for quantifying causal influence

Consider the following set of relationships (an abstraction of DTMC):

This is a subset of a full system that we may have a model of or observe. Here can cause in 2 time units through two paths, directly or through d: of the time causes d at , of the time it causes directly at , and  of the time it does neither. Now assuming the marginal probability of is much lower than how can we determine how significant and are for ?

One approach is to find the earliest cause that best accounts for the effect. This is Suppes’ method which says that a cause is spurious if there is other information that predicts the effect at least as well and, once this information is known, the probability of the effect is unchanged by knowledge of the spurious cause. In this example this method would determine erroneously that is a spurious cause since occurs earlier than and .

Similarly, using Granger causality one would also fail to find as a cause of since the probability of is unchanged once the information about is included. This is incorrect since is in fact a cause of and accounts for exactly as well as does. Further, brings about more quickly than c does and with a higher probability ( versus when causes directly). In Suppe’s approach, though, as long as there is one such for which the relevant conditional probabilities are equal (or near equal, using his looser notion of -spuriousness), would be labeled as spurious. In this case, is not only useful for predicting but is known to actually cause it. If we included other variables in the example there would be many other cases where makes a significant difference to ’s probability. It is not similarly possible for to be spurious in Suppe’s approach since the only other factor that could make it seem so is and occurs later than . This is contrary to the approach taken here, which aims to infer the most direct causes of an effect, relative to the set of factors measured.

Since attempting to find a single factor that renders a possible cause irrelevant can lead to anomalous results, another approach is assessing the average significance of a cause for its effect. The idea is to measure overall how well a possible cause predicts its effect.

//TODO: Finish this

### Measuring Causal Significance

The general approach is to calculate the average difference in the probability of an effect for each of its prima facie causes, in relation to all other prima facie causes of the same effect. If the probability of the effect does not differ substantially when a potential cause is present or absent, then it may be insignificant.

//TODO: Finish the section on measuring Causal Significance

## Notes on Testing Prima Facie Causality

We want to adapt standard PCTL model checking procedures to validate formulas directly in a set of time series data without first inferring a model (as this can be computationally complex or infeasible in many cases)

### The set of hypotheses

The initial hypotheses are first tested to determine which meet the conditions for prima facie causality, being earlier than and raising the probability of their effects. These are a set of formulas of the form:

(1)

where and are PCTL state formulas, , and . To form this set, the simplest case is when we have some knowledge of the system and either explicitly state the formulas that may be interesting or use background information to generate the set. Data on risk factors for disease may include gender, race, and age group, but we can avoid generating formulas for scenarios that are mutually exclusive or which cannot change, so that while a person may be of multiple races, these cannot change over time and there are also constraints such that a person cannot be simultaneously elderly and a child. Similarly, we may not know the exact connection between neurons in a particular scenario but may have background knowledge on the timing between one firing and triggering another to fire. Here we could choose to generate increasingly large formulas and stop at some predefined size. Another

## Appendix

### Appendix A: Hume’s Regularity Theory

Hume’s interpretation of Causality is that the latter is routine occurrence of an effect following a cause. This is known as the Regularity Theory of Causality. There are many examples of *noncausal regularities* (*Example*: the presence of umbrella vendors and rain) and *causation without regularities* (*Example*: a drug causing death to a single patient). One of the major omissions in Hume’s Regularity Theory is that there is no method for separating which parts of a regular sequence of events are essential from those that are correlated with others.

*Example:* One may always open the door to my office, turn on the light and then begin writing papers or code on his computer. While the light is being useful for writing, it is not essential.

Many effects have multiple causes that are comprised of interacting components, such as the impact of environmental factors and genetic mutations on health. Hume’s approach does allow for reasoning with these types of causal complexes. Hume’s theory was updated by John Leslie Mackie formalizing the ideas of necessity and sufficiency of the causes considering multiple components of a cause and multiple causes of an effect.

**Definition A.1**: An event is a *necessary condition* of an event if whenever an event of type occurs, an event of type occurs. is a *sufficient condition* of if whenever an event of type occurs an event of type also occurs.

**Definition A.2** (*Mackie*): *INUS* condition: an insufficient but non-redundant part of an unnecessary but sufficient condition.

**Definition A.3** (*Mackie*): is an *INUS* condition of *iff*, for some and some is necessary and sufficient condition of , but is not sufficient condition of and is not sufficient condition of .

**Corollaries**:

1. is sufficient for
2. is not necessary since could also cause
3. alone maybe insufficient for
4. is a non-redundant part of

*Example*: A lit match ( ) may be a cause of house fires but there are many other situations when a match is lit and does not cause a fire ( ), and in which a fire occurs without a lit match ( ). For a match to cause a fire, some set of circumstances ( ) are required, where each of its components is necessary *and* they together with are sufficient for a fire to occur. This gives the minimum conditions for something to be a cause, so a cause may be more than an *INUS* condition as it may be necessary or sufficient.

This definition addresses types of events, but Mackie also gives a method for using INUS conditions to find causes of individual, actually occurring, events. To do this, we need two more definitions.

**Definition A.4**: is a minimal sufficient condition for if no conjunct is redundant (i.e., no part, such as is itself sufficient for ), and is sufficient for .

**Definition A.5**: is at least an *INUS* condition of *iff* either is an *INUS* condition for , or is a minimum sufficient condition for , or is a necessary and sufficient condition for , or is part of some necessary and sufficient condition for .

Then, for C to be a cause of E on a particular occasion (what is referred to as token, or singular, causality) according to Mackie, the following must be true:

1. is at least an *INUS* condition of
2. was present
3. The components of , if there are any, were present
4. Every disjunct in not containing as a conjunct was absent

Using the house fire example, a lit match was the cause of a specific fire if it was present; oxygen, flammable material, and the other conditions needed for a lit match to cause a fire were also present; and there was no unattended cooking, faulty electrical wiring, or other factors that cause fires in the absence of lit matches. The third and fourth conditions in the previous list ensure that the other factors needed for to cause are present, while avoiding the problem of overdetermination. If instead there was a lit match and the house was simultaneously struck by lightning, the fourth condition would be violated, and neither would be deemed the cause of the fire. Mackie’s rationale for this is that if two separate INUS conditions were present, either could have caused the effect and there is no way to assess their individual contributions to it using only regularities. This is one of the biggest limitations of the theory, but there are other impediments to practical usage. This approach does not easily allow for anomalous events, such as a factor that normally prevents an effect bringing it about in some unusual scenarios – such as a drug that normally cures a disease but causes death in small fraction of the people who take it.

### Appendix B: Counterfactuals

One of the key problems faced by regularity theories is distinguishing between a relationship that regularly occurs and a factor that made a difference to the occurrence of an effect on a particular occasion. Counterfactual approaches to causality usually aim to address these types of questions, assessing whether a particular instance of an effect would have occurred in the absence of the cause on that specific occasion. This is known as *token (or singular) causality*. Although counterfactuals can be applied to type-level cases, they are primarily used to analyze situations where the cause did actually happen and one wants to determine what difference it made to the effect.

When Hume defined causality, he wrote that *a cause is an object followed by another, and where all the objects similar to the first are followed by objects similar to the second*. Or in other words, *where if the first object had not been, the second never had existed*. The first part (known as the regularity definition) is quite different from the second part (the counter-factual definition). Using the first half, the presence of an umbrella vendor may still seem to cause rain due to regular occurrence. However, if the umbrella vendors went on strike, rain would still occur so we can see that there is no counterfactual dependence.

David Lewis developed the primary ***counterfactual theory of causality***, discussing how we can use these conditional statements to analyze token causality between events. Central to Lewis’s work is being able to reason about what would have happened using possible worlds, and comparative similarity between possible worlds.

*A possible world* may be thought of as a maximally consistent set of propositions true in that world, and a world is closer to actuality than another world is if it resembles the actual world more than any other world does. While there is no standard method for comparing possible worlds, we can still reason about their similarity. A possible world which takes less departure from reality is closer to the real world than another possible world which seems less like to materialize. Lewis introduced two constraints on the similarity relation:

1. Any two worlds may be compared, but it is weak ordering so they may be equally similar to the actual world
2. The actual world is closest to actuality since it is more similar to itself than any other world is.

One can use these types of comparisons to define counterfactual dependence, and then causal dependence. The *counterfactual* of two propositions and (represented by ) means that if were true, would be true. Where is true in the actual world if and only if (1) there are no possible worlds where A is true (vacuous ase), or (2) a world where both and are true is closer to the actual world than any -world where does not hold. As Lewis says “it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent”. One can then express statements such as “had someone not thrown a rock at the window it would not have broken”.

Causal dependence is represented using two counterfactuals, stating that had not occurred would not have either (), and that had occurred, would have occurred too (). While this causal dependence also implies causation, according to Lewis, causality is transitive, so one may have cases of causation without causal dependence. If there is a chain of causal dependence, such that depends causally on , and depends causally on , is a cause of even though may not depend counterfactually on . In Lewis’s theory is a cause of if is causally dependent on or if there is a chain of causal dependence between and .

The main problems with this approach: 1) *transitivity* and 2) *overdetermination*.

With respect to *transitivity*, we can find situations such that some event a would generally prevent some event c from occurring but in the actual events, a causes another event b, which in turn causes c to occur. Thus, the counterfactual account can lead to events counterintuitively being labeled causal. Thus, the counterfactual account can lead to events counterintuitively being labeled causal. (McDermot, 1995) gives one such counterexample. Suppose I give Jones a chest massage (), without which he would have died. Then, he recovers and flies to New York (), where he eventually has a violent death (). Here, was a cause of , as without massage he would not have been well enough to travel, and is a cause of , but did not cause . Whether or not occurred, Jones still would have died, but there is a causal chain between and . It is counterintuitive to say that preventing death in one manner caused it in another simply because it allowed the person to live long enough to die differently. However, transitivity is needed in this method for reasoning about cases where there is causation without causal dependence.

The second problem, *overdetermination*, implies redundant causation. Say there are two potential causes for an effect (both present) and the effect would have been the result of either, so that the effect depends causally on neither and the system is overdetermined. One common example is that of a firing squad: if one shot had not killed the prisoner, another would have. This redundant causation may be either symmetrical (each potential cause could equally well be called the cause of the effect, there is nothing to distinguish which was the actual cause) or asymmetrical (there was one cause that *preempted* the other). In the asymmetrical case, if we say was the preempting cause, the preempted and the effect, then had not occurred, would still have caused , and thus is not the cause of despite its causing . Imagine an alternate ending to Thelma and Louise, where the police have deemed them a danger. In this scenario, if the women had not driven off the cliff to their deaths, the police would have shot them instead. There is no causal dependence between their deaths and driving off the cliff yet it is clearly the cause of their deaths as the police did not actually shoot (but would have).

Many events may occur even in the absence of a particular cause but the manner in which they occur may differ, whether this means for example their location or time or intensity. There may be cases, particularly in biology, where systems have backup mechanisms that ensure the result is produced and the manner of its occurrence may not differ substantially between the various possible causes. In the example of Jones’s death, we took for granted that we were attempting to assess his death at that particular time, since had he not died violently, he would have died of old age or disease or in some other manner. The inconsistencies in this theory led to a later revision, where causes affect how, when, and whether their effects occur (Lewis, Causation as Influence, 2000).

### Appendix C: Granger causality

#### Notation:

denotes the value of the variable at time

denotes the set of measurements of up to time i.e.,

denotes the set of all possible knowledge up to time (including both and ). One can think of

**Definition A.1** Granger-causes if

(A.1)

As this is not exactly causality, it has come to be called Granger causality. Granger’s definition takes temporal priority as a given and does not make claims about how much of a difference makes to the probability of (or whether this difference is positive or negative). may not be the best or only predictor of , rather it is simply found to be informative after accounting for other information.

Granger’s test is understood usually as a *bivariate test*. In the bivariate test only two time series are included: that of the effect, and that of the cause, . One bivariate method is to use autoregressive model with the two variables where if the coefficients of the lagged values of are non-zero, then is said to Granger-cause

An -lag autoregressive model takes lagged values of time series up to when calculating the value of a variable at time . Each lagged value is weighted by a coefficient, so that a variable may depend more strongly on recent events than those that are more temporarily distant. More formally, can be represented as the following -lag linear autoregressive model:

(A.2)

The lags mean that the values of and at times influence that of at time . The coefficient indicates how much the value at depends on . Here means the influence of on , so that models the dependence of on itself and models that of on . Here can be finite or infinite – the latter was done in the original Granger work. The error term is assumed to be random variable with mean zero. Using this, Granger causality can be tested by whether non-zero values for lead to a smaller variance in the error term than when these are zero (and whether this reduction is statistically significant).

Eichler (2009) points out that this bivariate approach does not capture Granger’s original definition. Further, it cannot distinguish between causal relationships and correlations between effects of common cause. This can be seen in eq. (A.2) If and have common cause, , and these effects do not always occur simultaneously, then

will provide information about when has occurred and will thus significantly improve the prediction of . A more accurate approach is the multi-variate one, which includes other variables in the model of each time series. Using a vector autoregressive model with variables , now instead of a single variable is a vector representing the measurement of all variables in at time . The system is represented as:

(A.3)

Here is a matrix of coefficients and is a vector of error terms. Using this representation, Granger-causes if at least one of is non-zero.

While this model comes closer to causal inference than the bivariate test does, it has practical problems. Such a model quickly becomes computationally infeasible with even a moderate number of lags and variables. Using a model of order with variables leads to matrices

//TODO: Finish this

### Appendix D: Notes on Probabilistic Measures of Causal Strength, article by Brandon Fitelson and Christopher Hitchcock

#### Causation as Probability-raising

*Discussion on previous work in the article*

The idea that causes raise the probabilities of their effects is found in many different approaches to causation. In probabilistic theories of causation, of the sort developed by (Reichenbach, 1956), (Suppes, 1970), (Cartwright, 1979), (Skyrms, 1980), and (Eells, 1991), is a cause of if raises the probability of in fixed background contexts. We form a partition , where each is a background context. Then is a cause of in context just in case or equivalently, just in case . Note that both inequalities fail, albeit for different reasons, if . The idea is that each background context controls for confounding causes of , so that any correlation that remains between and is not spurious. According to (Cartwright, 1979) each background context should hold fixed (either as being present or as being absent), every cause of E that is not itself caused by C. (Eells, 1991) has similar proposal. If we construct the background contexts in this way, we would expect the conditional probabilities of the form and to take values of 0 and 1 if is caused deterministically. However as (Dupre, 1984) points out, this carves up the background conditions more finely than is needed if the goal is simply to avoid confounding. For this purpose, it suffices to hold fixed the common causes of and . If we construct the more coarse-grained partition in this way, the conditional probabilities and might take intermediate values even if determinism is true. An issue remains about what it means to say that causes *simpliciter*: whether it requires that raise the probability of in all background contexts (the proposal of (Cartwright, 1979) and (Eells, 1991)), whether it must raise the probability of in some contexts and lower it in none (in analogy to Pareto-dominance, the proposal of (Skyrms, 1980), or whether should raise the probability of in a weighted average of background contexts (this is essentially the proposal of (Dupre, 1984)). We will avoid this issue by confining our discussion to the case of a single background context.

In his paper (Lewis, Causal Explanation, 1986), Lewis offers a probabilistic version of his counterfactual theory of causation. Lewis says that *causally depends upon* just in case (i) and both occur, (ii) they are suitably distinct from one another, (iii) the probability that would occur at the time occurred was , and (iv) the following counterfactual is true: if C had not occurred, the probability that E would occur would have been substantially less than . Lewis takes causal dependence to be sufficient, but not necessary, for causation proper.

//TODO: Finish this

# Bibliography

Brian Fitelson, C. H. (2011). Probabilistic Measures of Causal Strength. In P. M. (ed.), *Causality in the Sciences* (pp. 600-627). Oxford, UK: Oxford University Press.

Cartwright, N. (1979). Causal Laws and Effective Strategies. *Noûs 13*, 419-437.

Dupre, J. (1984). Probabilistic Causality Emancipated. *Midwest Studies in Philosophy, IX*, 169-175.

Eells, E. (1991). *Probabilistic Causality.* Cambirdge CB2 1RP: Cambridge University Press.

Humphreys, P. (1980). Probabilistic Causality and Multiple Causation. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,Vol. 1980, Volume Two: Symposia and Invited Papers (1980), pp. 25-37* (pp. 25-37). Chicago, IL: The University of Chicago Press on behalf of the Philosophy of ScienceAssociation.

Lewis, D. (1973). Causation. *Journal of Philiosophy, Vol. 70, No. 17*, 556-567.

Lewis, D. (1973). *Counterfactuals.* Malden, Massachusetts: Blackwell Publishers.

Lewis, D. (1986). Causal Explanation. *Philosophical Papers Vol. Ii*, 214-240.

Lewis, D. (2000). Causation as Influence. *Journal of Philosophy*, 182-197.

Mariarz, M. (2015). A review of the Granger-causality fallacy. *The Journal of Philosophical Economics: Reflections on Economic and Social Issues, VIII: 2*, 86-105.

McDermot, M. (1995). Redundant Causation. *Brit. J. Phil. Sci. 46*, 523-544.

Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference, and Learning.* Berkeley, CA: UC Berkeley.

Otte, R. (1982). *Probability and Causality.* Ann Arbor MI 48106: University Microfilms International.

Peter Sprites, C. G. (1993). *Causation, Prediction, and Search.* Pittsburgh , Pennsylvania: Carnegy Mellon University.

Reichenbach, H. (1956). *The Direction of Time.* Mineola, NY: Dover Publications, Edited by Maria Rechenbach.

Skyrms, B. (1980). *Causal Necessity: A Prgramtic Invenstigation.* New Haven, CT: Yale University Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality.* Amsterdam: North-Holland Pub. Co.

Thomas Richardson, P. S. (1999). *Scoring Ancestral Graph Models, Technical Report No. CMU-PHIL-98.* Pittsburgh, Pensylvania: Carnegy Mellon University Press.

# Downloadable Links for the Bibliography

(Brian Fitelson, 2011): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/ProbablisticMeasuresOfCausalStrengthFitelson2011.pdf)

(Lewis, Causal Explanation, 1986): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/CausalExplanation_DavidLewis_1986.pdf)

(Lewis, Causation as Influence, 2000): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/Causation_as_Influence_lewis_2000.pdf)

(Lewis, Counterfactuals, 1973): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/Counterfactuals-lewis-1973.pdf)