Doozandeh, P. (2015). *Categorization of real-world causal events for causal modeling* [unpublished manuscript]. Accessible in p-doozandeh.com

Categorization of Real-World Causal Events for Causal Modeling

Pooyan Doozandeh

Author note:

This manuscript was completed in June 2015, and uploaded online in February 2021.

No funding agency or university supported the creation of this manuscript. Correspondence should be directed to Pooyan Doozandeh (pooyan.doozandeh@gmail.com).

Abstract

Designing models to infer causal rules from certain types of observations has been a topic of interest for cognitive scientists. So far, researchers first thought of a causal event in the world and then proposed a model for it. After some time, by considering the shortcomings of the previous model for other causal events, researchers proposed another model and this is repeated several times in the last few decades. The problem with this evolution of models is, for each model, there are always causal events that it is unable to address; since it is best suited for a certain instance of causality that designers had in mind. I think this problem stems from researchers' idea that their models are designed for the "causality". Causality, in the philosophical sense, does not exist, at least with respect to modeling real-world events. There are instead a number of entirely different *types of events* that are commonly known as "causal". Therefore, I propose that, before attempts at modeling, we should categorize different types of events that we know as causal. After categorization, every model can address one, or a limited number, of these types. The benefit of the categorization is, at each attempt of modeling, researchers limit their concerns to one type of events, instead of worrying about how to extend their model to other types. It also gives us a big picture of how it would be possible to design a unified system that distinguishes between different types of causal events and assign them to their own model.

Keywords: Causal events; Categorization; Modeling; Types of events

The Need for the Categorization of Real-World Causal Events

Attempts in Modeling

We have seen numerous efforts from different perspectives in the last few decades to formally model human's tendency to establish causal rules by observation. Researchers in different fields, from psychology and cognitive science to artificial intelligence, focused on how we humans learn causal rules and then tried to formally and computationally define this ability, with the view to make machines capable of learning by observing causal events. If we trace the history of activities in causal modeling, we face researchers in different camps, each working on modeling instances of causality that first came to their mind. After several years, another group came and by noticing the fact that some instances of the real-world causality cannot be included in the previous model, they tried to introduce another model. This pattern of criticism and introducing another model does not seemingly contain any problem per se, and it is the standard of evolution of philosophy and many fields of science.

To mention a few cases in modeling in such evolution, we have associative models as one of the early models of the field. Associative models focused on the co-occurrence between the cause and effect to assess whether there exists and how strong is the relationship between the cause and effect. This perspective has its root in Hume's (1739/2004) definition of causality and Rescorla-Wagner's (1972) research. Subsequently, other researchers, inside the associative camp, tried to present a more comprehensive model by trying to solve many problems facing the original simple associative model (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998).

Associative models had as their primary tool of evaluation of causal relationship, the rate of contingency between the cause or cue and effect or outcome. By realizing that there are many instances of causality that cannot be easily dealt with by such simple variable, researchers in a different camp introduced another model that does no longer depend on measuring the association with sample-by-sample measurement of the co-occurrence of cause and effect. ΔP and PowerPC (Shanks, 1995; Cheng, 1997, 2000) were models that aimed to measure the strength of the relationship between the cause and effect, by using a set of data that demonstrate

the binary values of the presence or absence of cause and effect in four possible states in a contingency table.

While these models, which are known as “rational models”, have extensive usages in many standard scientific experiments, they still fail to catch some aspects of human tendency in many instances of real-world causality, like the importance of prior knowledge in inferring causal relationships. That is when still another camp comes and introduces the Bayesian modeling (Griffiths et al., 2011). This approach, that can incorporate prior knowledge with the present data and is also capable of defining different causal structures in a given problem, is the leading model at the time of writing this article and many scientists are trying to make it capable of including other instances of the real-world causality (Vul et al., 2014). Some researchers, like Mozer et al. (2008), had already begun to criticize the Bayesian approach and argued that in many cases, Bayesian inference is inconsistent with human behavior in inferring causal relations. They are probably thinking of a new camp to introduce a model with the ability to deal with causal events that the Bayesian model cannot address!

I have so far presented a very general and abstract picture of some attempts in the history of activities in modeling causality. Some would argue that, like other fields, this evolution of models is natural within researches in causal modeling; but I claim that there is an obvious confusion among researchers and they tend to ask why the scope of their model, with respect to various causal events, is so limited. I think if researchers continue in this way, there will always be instances of causal relations that the newly presented model cannot explain and either previous or future models can outdo. As an example of such problem, Griffiths et al. (2011) ask “Why do we need only a handful of observations to learn a new causal relation in some situations, but dozens or even hundreds in other situations?” Instead of turning to prior knowledge to answer this question, as what the authors of that article did, I think the situation in which only a few observation sufficed was entirely a different event, with different mechanisms in the observer’s mind, than the other which required more observations.

Causality as a Vague Term in Modeling

The confusion, as I mentioned above, stems from the fact that researchers of the field have never given an abstract picture of what they were always trying to model. The ostensible aim is to model human's tendency of establishing causal relations from observations. First we have to define "causality" as the goal of all of these models.

Without going to philosophical discussions, the term "causality" has been used in philosophy for centuries. Most philosophers who used the term wanted to investigate a number of events in the world that shared one feature: having two distinct elements which are called cause and effect. The word "causality" has assisted philosophers in their argumentations concerning the nature and validity of inferring causal relations with empirical data. I think, however, that such general usage of the term "causality" is only suitable for those philosophical investigations; using the term in that manner is the source of confusions in our modeling of world-level causal events.

What is evident from the attempts of cognitive scientists in modeling causality is that they have used the term "causality" for a number of events in the real world, similar to how philosophers use it. I think those events in "causality", at least with respect to our discussions in modeling, are entirely different occurrences with distinct characteristics. Each group, or camp of researchers, thought of a number of similar instances of causality and tried to accommodate their models with those events; this is repeated several times in other camps with their own instantiation of causality. The word "causality" is unreal when we assign it to events in the world. In other words, researchers have presented a number of models, each for a different event in the world, not the "causality". Each model, which has been presented in the field, focused on a single event that does not share much with other events for other models. In this perspective, we now have a number of models, each suitable to deal with one kind of events in the world, but each claiming to model the "causality". This creates confusion for two reasons. First, none of them are actually modeling the "causality"; since in this sense, there is no external personification of the "causality". Second, researchers compare those models with each other, while in reality there is no shared ground or gauge to measure which model is better. One model is designed for one kind of event and another model for a different kind. From the perspective of

one event, one model looks better than others and if we change the perspective, another model would look more suitable. If we compare two of these models, it is as though comparing a sport car with a station wagon; each designed for a different purpose. That is why in every attempt in modeling, one model works best only with one or a few instances of causal events and there are always events that cannot be explained with the presented model.

The term “causality” is too abstract, vague, and unreal to be used in cognitive science. Instances of events in the world are real and every presented model so far tried to catch human’s performance when presented with one causal event. These events are, to a large extent, different occurrences that merely share what we know as two elements of cause and effect. Otherwise, they are distinct and therefore the nature of human’s mental mechanisms in inferring causal relation is entirely dissimilar among those events.

The models that researchers presented so far tried to address different causal events, not the “causality” in general sense. Up until now, the order of activities of researchers was first to think of one causal event in the world and then making a jump in trying to present a model for that event, claiming to model the “causality” or “causal induction”. I think to rid ourselves of confusions, instead of focusing on models, we must start from events that we know as instances of causality and, before any attempt at modeling, categorize those events under the rubric “causality”. The greatest advantage of such categorization is, when trying to build a model, instead of trying to deal with the general “causality”, the researcher knows what “kind” or “category” of causality he/she is trying to address and, in turn, limits his/her concerns only about that category, not worrying about the limitations of that model with respect to other categories of events. I will continue by giving an example of how such categorization would be.

An Example of the Categorization of Causal Events

My goal in this article is to emphasize the importance of categorization, not to categorize all instances of causality. I would only suggest more research to be done on this subject and here, I will present a very simple example of such categorization. We can think of different events and find similarities between them and see if we can put them in one category.

The first category belongs to the instances of causality in which causal events happen one-by-one and in every step, each single event changes human’s judgment of the existence and

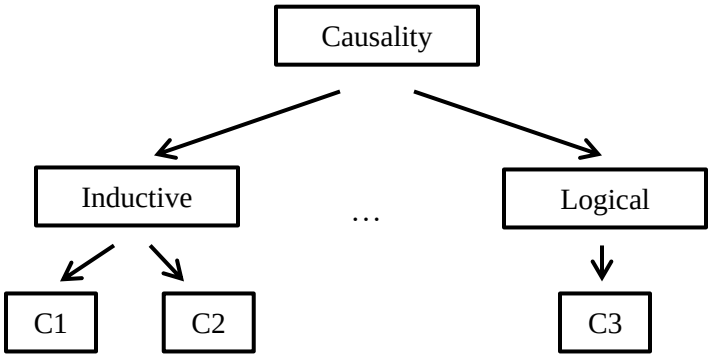
strength of relation between cause and effect. At first, the association of cause and effect would look novel, and in the first few conjunctions, change in our judgment is sharp, depending on whether the effect is followed by the cause. After some repetitions, these changes become more stable and we tend to depend on our judgment with more confidence. At present, we can call this category C1 and we can think of real-world instances that can be subsumed under C1. Imagine you are presented with a strange drink and you want to test what psychological effect the drink brings. You try it and after some minutes you tend to laugh more than usual. Tomorrow you drink the same mixture and again you begin to laugh. You repeat the same experiment and depending on the outcome, especially in the first few trials, you establish a new causal rule between the drink and laughing. As another example, imagine you are assigned as a teacher of a new class. The first time after you open the door of the classroom, it bleeps. You close and open the door again, and it bleeps again. Like previous example, depending on what happens in one-by-one trials, you tend to establish a causal relation between the act of opening the door and the sound of bleep. We can continue and think of many other examples that can be categorized as C1.

Next category, C2, represents instances that are not extended in time (like C1) and the data of the relationship between binary cause and effect is collected in a table and presented at once. The table is called “contingency table” and is widely used in many standard scientific experiments. Binary values for the presence and absence of cause and effect require four possible states and the numbers of those occurrences fill the table. Having formed the table, we can think of a simple formula in probability to calculate the strength of the relation between the cause and effect. This simple idea is originated from human’s tendency in forming causal rules by such observations and the table and formula tend to catch this natural tendency. Imagine the example of injecting a mouse with a certain chemical and testing whether a certain gene appears in it (Griffiths & Tenenbaum, 2005). We can consider two equal groups and we inject the first group with the chemical and leave the second group without injection. After several days, we check the proportion of mice in the first and the second group which expressed the gene. Based on the results we tend to establish a causal relation between the chemical and gene expression with

certain strength or not to establish any causal relation. The famous example of smoking and lung cancer also belongs to this category.

The last category of instances that we mention here, or C3, includes those cases that usually do not require much repetition and we soon tend to logically infer the causal relationship. To make it clear, imagine that your stomach hurts after you had a meal. You have had roasted chicken with peas and mushroom. You suspect which of these three you may be allergic to. After some days, your stomach hurts again and this time you had beef, carrot, and mushroom. You, like any rational person, would deduce that it is highly probable that the mushroom, the only food present in both meals, causes your stomach to hurt. Or think of the example of Blicket Detector (Gopnik et al., 2004). We have a strange gizmo, like a black box, and two objects, A and B. We put both objects on the black box and it starts making sound and lights. We remove them and it stops. In the next step, we put only object A and it does not start. If we stop the experiments in this step, we tend to think that object B was the cause of the black box to start. These and similar samples, which depend on human logical conclusions, belong to the category C3.

Having assigned different instances of causality to their categories, we can further group them in higher levels of abstractions. For example, we can notice that in categories C1 and C2, the repetition of instances plays a crucial role in the existence and strength of the relation between cause and effect, while in C3, repetition does not matter much, since our inference is logical. In fact in C3 and the Blicket Detector example, although the object B was never tested on the black box, we inferred that it is probable that B was the cause. Without going into much detail which is outside the scope of the present article, the third example does not seem to be an example of “causal induction”; for induction requires enumeration. So, a diagram of our example in categorization would look like the figure below.



Now, my simple suggestion is that researchers should design one model for one (or several) of these categories. I have to mention again that the categorization that is presented here is neither true nor complete; it is merely an example of how it would be possible to perform it and that, with respect to one or more variables, there can be many levels of groupings inside the categorization. I believe that, in this way, researchers in cognitive science who aim to model human's ability of establishing causal rules, can have a big picture of their activities and they can also investigate to what extent their proposed model can account for instances of causality. In addition, it would be possible that one model can address more than one category of causal events.

Discussion

The categories C1, C2, and C3 are distinct groups of events that are happen to be called instances of "causality". In other words, until now, the shared perspective of researchers was to have a top-down view that considers causality on top, which includes instances of the real-world events at the bottom. I propose the opposite view, which is bottom-up, and in this perspective, there are instances and categories of events which are, to a large extent, different from each other, but in common sense are called instances of causality. For example, if we compare the categories C2 and C3, it becomes evident that these are two entirely different groups of happenings and the mental mechanisms in response to these events are basically different.

Subsequently, they need different models with their own mechanisms for each of these groups of happenings.

The natural question that appears is: what about the models that have been presented so far? Does each of those models account for one category of causal events? We can think, for example, that if the PowerPC (Cheng, 1997) is modified in a way to consider the importance of sample size, it would probably be a model to address our category C2. Or Bayesian approach, as is presented in Griffiths et al. (2011), can suitably model human's judgments for our category C3. My point is that at first, researchers must make it clear what their target category is and then, they can more easily design their model with respect to that category. It must also be noted that it is possible that in one causal event, two categories overlap. For instance, in the example of the Blicket Detector, if we put objects A and B on the black box and then put object A on the black box for ten consecutive steps, regardless of whether the black box is activated or not, the category would shift from C3 to C1. This happens more often in the complex real-world instances of causality.

Another question that might come into mind is: what if researchers design a model that can address most or even all of the categories that we define? In that case, is there any justification for categorization? Most researchers in different camps, in particular some in Bayesian approach (Vul et al., 2014), are convinced that by future research, their proposed model can address all instances of causality. Even if we accept that lofty claim, we would still need categorization. It is evident that, if they are to simulate mind's performance, human mind works differently in response to different categories of causality. Such general model, which claims to address all categories, would be composed of different mechanisms to deal with different instances of causality, like two different mechanisms for categories C2 and C3. Upon arrival of the data, the system would need to distinguish between different instances, and model them with different mechanisms, and that is when they need categorization.

Regarding the previous paragraph, an important challenge appears that is not solely restricted to our present subject. Imagine we successfully created a full categorization of instances of causality and we also developed appropriate models to address those categories. We can then design a single unified system that receives the data from the world, categorize them,

and process them with their suitable models. The problem is, upon receiving the data, how to formally distinguish the incoming causal data and assign them to their suitable categories? In other words, how do we humans know that, for example for the Blicket Detector (C3), we must use logical inference or for the example of drinking (C1), we must reason by inductive repetition? It seems to be the problem of the future and it requires a more profound investigation which extends to other issues and problems in psychology and philosophy. For the moment, we would better focus on more down-to-earth challenges like how many categories of causality we can define, how to group those categories, and how to create satisfactory models or modify existing models for one or more of those categories.

References

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

Cheng, P. W. (2000). Causal reasoning. In R. Wilson & F. Keil (Eds.), *The MIT encyclopedia of cognitive sciences* (pp. 106–108). Cambridge, MA: Bradford, MIT Press.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology, 49B,* 60–80.

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,* 1–30.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and Blickets: Effects of Knowledge on Causal Induction in Children and Adults. *Cognitive Science, 35,* 1407-1455.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 285–386.

Hume, D. (1738/2004) *Treatise of Human Nature*, Oxford: Clarendon Press.

Mozer, M., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science, 32,* 1133–1147.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). New York: Appleton-Century-Crofts.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48,* 257–279.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25,* 127–151.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science, 38,* 599-637.

Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: The role of within-compound associations. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology, 51,* 121–138.