The Temporal Logic of Token Causes

Samantha Kleinberg and **Bud Mishra**
New York University
715 Broadway, 10th floor
New York, NY, 10003
samantha@cs.nyu.edu, mishra@nyu.edu

Abstract

While type causality helps understand general relationships such as the etiology of a disease (smoking causing lung cancer), token causality aims to explain causal connections in specific instantiated events, such as the diagnosis of a specific patient (Ravi's developing lung cancer after a 20-year smoking habit). Understanding why something happened, as in these examples, is central to reasoning in such diverse cases as the diagnosis of patients, understanding why the US financial market collapsed in 2007 and finding a causal explanation for Obama's victory over Clinton in the US primary. However, despite centuries of work in philosophy and decades of research in computer science, the problem of how to rigorously formalize token causality and how to automate such reasoning has remained unsolved. In this paper, we show how to use type-level causal relationships, represented as temporal logic formulas, together with philosophical principles, to reason about these token-level cases. Finally, we show how this method can correctly reason about examples that have traditionally proven difficult for both computational and philosophical theories to handle.

Introduction

When we want to determine what is responsible for a patient's symptoms, why a stock plummeted in value, or the reason a particular candidate won an election, what we want to know is what *caused* these particular events. But rather than finding a general relationship, such as "smoking causes lung cancer", we want to find whether a particular one, such as "Bob's smoking caused his lung cancer" is true. In order to do this in an automated way, we need an understanding of the general relationships (called type-level causality) and how these relate to the singular cases (called token-causality). We also need a system for combining this knowledge in a rigorous, automated, way.

While the problem of general causal inference has been studied in our prior work (Kleinberg and Mishra 2009) as well as that of other computer scientists (Pearl 2000), we cannot immediately use these inferences to explain token cases. A type-level relationship may indicate that a token case is likely to have a particular cause, but it does not necessitate this. Just as the relationship between smoking and lung cancer does not mean that all lung cancers are caused by smoking, we cannot immediately propose that a type-level cause is a token-cause. We must first establish whether the type-level relationship has been instantiated and take into account that we may wish to also assess the role of other hypotheses, including rare factors. In some cases we will not even know if the potential cause occurred, we may only have indirect information such as whether other causes and effects of the potential cause occurred. Finally, our intuition would tell us that a type-level relationship between smoking and lung cancer does not allow for smoking to cause lung cancer within a matter of hours. However, without such time constraints explicitly included in the inferred relationships – or such instances explicitly excluded in the token case (using this background knowledge and intuition), we will fall prey to potential misinterpretations.

The proposed methodology will incorporate solutions to all of these difficulties. Since the inferred logical relationships include time windows between cause and effect, we will infer relationships of the form "smoking causes lung cancer in 15-30 years", preventing cases such as a person beginning smoking and developing lung cancer within two days from being an instance of this type-level relationship. Methods that only test whether a general relationship is fulfilled, and do not have detailed time information as part of this general relationship, cannot avoid such scenarios without much manual filtering of the events and relationships tested. In general, our approach will be to infer the type-level causes as proposed in prior work (Kleinberg and Mishra 2009), and use the fact that these are represented by probabilistic temporal logic to our advantage – determining whether the facts of a token-case fit the known type-level relationship (computing the probability of this being the case when it is unknown whether causal relationships were instantiated) and using the known type-level strength of the relationships to rank the possible explanations for a token case.

We will first discuss related approaches in both philosophy and computer science then briefly review our type-level inference procedure before discussing our method for token-level reasoning and some examples to see how they are handled by the theory. Our approach will give results consistent with intuition in difficult cases where time is important, and will do this without full knowledge of the token events or a priori knowledge of a model.

Related work

Philosophy

There have been two facets to the problem: how can we (and should we) combine type- and token-level information; and once we make such a choice, how can we reason about token-causality? Among philosophers, there is no consensus on the solution to the first problem: we may learn type-level claims first and then use these to determine token level cases (Woodward 2005); the type-level relationships may follow as generalizations of token-level relationships (Hausman 2005); or they may be treated as entirely different sorts of causation (Eells 1991). For the second problem, there have been a number of approaches, each with its own advantages and drawbacks. Counterfactual methods, introduced by Lewis (1973) ask whether the effect would have occurred in the absence of some causal factor. If not, then that factor causes the effect. However, in cases where there are two events that both occurred, where each alone could have caused the effect, we then find that neither caused it. In later work, Lewis amended this to mean that dependencies are not based solely on *whether* events occur, but rather *how*, *when* and *whether* one event occurs depends on *how*, *when* and *whether* the other event occurs (Lewis 2000).

Another approach, due to Eells (1991), uses probability trajectories. Here we compare the probability of the effect before and after the cause occurs and up until the effect finally occurs in order to find a variety of relationships such as "because of", "despite", or "independently of". This approach is difficult to implement in practice, as it's rare to have enough information to be able to construct such a trajectory. In the philosophical approaches, a primary problem has been the practical implementation of these reasoning systems. Except in simple cases, being able to know the cause in a token case requires extensive background knowledge. Thus it has continued to be desirable to see what use type-level inferences can be for these token cases.

Logic and AI

Computational approaches have traditionally looked at the problem of beginning with a type-level model, and then using this to assess a particular case. These models may take the form of Bayesian networks or logical specifications of the system.

Approaches in logic have focused on the problem of reasoning about the results of actions on the system (Lin 1995; Thielscher 1997) or diagnosing the causes of system malfunctions based on symptoms (visible errors)(Poole 1994; Lunze and Schiller 1999). In particular, there has been a focus on reasoning about the indirect effects (ramifications) of actions. That is, how to take into account the effect of an action and propagate its changes on the world. Much work in this area stems from that of McCarthy and Hayes (1969), who introduced the situation calculus as a method of reasoning about causality, ability and knowledge, bringing together philosophical and logical representations of the world. Since then a number of modifications have been proposed (Lin 1995; Giordano, Martelli, and Schwind 2000), all of which aim to determine what could follow from an event or action. However, our problem is to look backward and find why what happened happened. Further, it is limiting to have to begin with a model, as we are rarely given any model in the cases of interest, and the problem of model inference is nontrivial. Work on fault diagnosis may seem closer to our approach, and generally allows for uncertainty about whether or not faults occurred and probabilistic relationships between faults and symptoms. These methods seek an explanation for something unusual and assume we begin with a set of causal knowledge or model specification. Here causality is usually interpreted in the sense of conditional dependence, and is most similar to the definitions employed in graphical models (Pearl 2000), where (absent) edges between nodes indicate conditional (in)dependence. The notion of when these events occur and how much time may be between them is not captured, though the output, like ours, is a ranking of possible causes for a fault.

Most recently, Hopkins and Pearl (2007) have proposed a framework drawing on earlier work on structural models (Halpern and Pearl 2001) as well as the work on situation calculus. Structural models had previously been used to link graphical models (Pearl 2000) to the counterfactuals introduced by Lewis. In this more recent adaptation, it is shown that counterfactuals may instead be modeled using the situation calculus, however one must still specify all dependencies - including those of counterfactuals. Here, a causal model is a situation calculus specification of the system (including preconditions of actions, etc.) and a potential situation and one may test whether a formula (here, it may be given a counterfactual interpretation) holds given the constraints on execution of the system (e.g. action preconditions).

The relationship between type and token

Previous approaches require that one must either begin with a model, know the truth values of all variables, or have a deterministic system. In contrast, we will infer relationships (temporal logic formulas with a causal interpretation) from time series data and then assess the support of each of these hypotheses for a token case.

Type-level inference

We will give a brief overview of our approach to type-level inference before discussing how to use these type-level causes for token-level cases. In prior work (Kleinberg and Mishra 2009) we created a new framework for causal inference, where cause and effect are described in terms of probabilistic computation tree logic (PCTL) formulas (Hansson and Jonsson 1994), and checked to see if they are satisfied in time series data (traces) using model checking. Then, to determine which of these possible inferred causal relations are significant, we compute the average difference a cause makes to its effect, using the concept of multiple hypothesis testing to determine at what level something is statistically significant (Efron 2004).

Temporal logic The logic used, PCTL, allows us to reason about formulas with probabilities as well as deadlines. With a set of atomic propositions A, formulas are defined

relative to a structure (called a discrete time Markov chain (DTMC)) $K = \langle S, s_i, \mathcal{T}, L \rangle$. This structure consists of a finite set of states, S; an initial start state, s_i; a total transition function, $\mathcal{T}$, giving the probability of transition between pairs of states; and a labeling function, L, giving the propositions true at each state. Note that in practice we will not usually have or infer these structures, but will perform the model checking procedure relative to a trace or set of traces.

Then the two types of formulas, state (those that hold within a state) and path (those that hold along a sequence of states) are defined as:

1. Each atomic proposition is a state formula.
2. If f and g are state formulas, so are $\neg f$, $f \wedge g$, $f \vee g$, and $f \rightarrow g$.
3. If f and g are state formulas, and t is a nonnegative integer or ∞, $f\mathcal{U}^{\leqslant t}g$ and $f\mathcal{U}^{\leqslant t}g$ are path formulas.
4. If f is a path formula and $0 \leqslant p \leqslant 1$, $[f]_{\geqslant p}$ and $[f]_{>p}$ are state formulas.

The "Until" ($\mathcal{U}$) formula in (3) means that the first subformula (f_1) must hold at every state along the path until a state where the second subformula (f_2) holds, which must happen in less than or equal to t time units. The modal operator "Unless" ($\mathcal{U}$) is defined the same way, but with no guarantee that f_2 will hold. In that case, f_1 must hold for a minimum of t time units. Path quantifiers analogous to those in CTL may also be defined, allowing use of the "always" (A), "exists" (E), "globally" (G) and "finally" (F) operators:

$$Af \equiv [f]_{\geqslant 1},$$
$$Ef \equiv [f]_{>0},$$
$$Gf \equiv f\mathcal{U}^{\leqslant\infty}false, \text{ and}$$
$$Ff \equiv true\ \mathcal{U}^{\leqslant\infty}f.$$

We will also make use of the "leads-to" operator, to which we have added a lower time bound:

$$f \rightsquigarrow_{\geqslant p}^{\geqslant t_1, \leqslant t_2} g \equiv AG[(f \rightarrow F_{\geqslant p}^{\geqslant t_1, \leqslant t_2} g)]. \quad (1)$$

This is interpreted to mean that for every path, from every state, if f holds, then g will hold in between t_1 and t_2 time units with probability p. If $t_1 = t_2$, this case simply says it takes exactly t_1 time units for g to hold after f holds. For a full discussion of PCTL as well as the problem of model checking PCTL formulas, see the original paper by Hansson and Jonsson (1994).

Conditions for causality We will infer causal relationships – leads-to formulas where c and e may be any state formulas – from time series observations (called traces). We assume there is some underlying true model for a system and the traces we observe are sequences of states the system has occupied. Then, the basic condition for causality is that c must be earlier than e and c raises the probability of e. Note that these are the minimum conditions, meaning that there could be a smaller window of time between c and e.

Definition 1. We say c is a *prima facie* cause of e if the following conditions all hold (relative to a trace, set of traces, or model):

1. $F_{>0}^{\leqslant\infty}c$,
2. $c \rightsquigarrow_{\geqslant p}^{\geqslant 1, \leqslant\infty} e$, and
3. $F_{<p}^{\leqslant\infty}e$.

This captures the primary feature of probabilistic theories of causality, but this simple definition erroneously admits many spurious causal relations. For instance, a barometer falls before it rains and seems to raise the probability of rain, but it does not cause the rain. Thus we need to further assess which of these potential causes are significant, comparing them with other possible explanations for the effect.

Significance of causes In order to determine whether a *prima facie* cause is significant, we will compare the average difference it makes to its effect given, pair-wise, each of the other *prima facie* causes of the same effect. That means that if there is only one other factor with respect to which the potential cause makes only a small difference, it may still have a high average value. With X being the set of *prima facie* causes of e, we compute

$$\epsilon_{avg}(c, e) = \frac{\sum_{x \in X \setminus c} \epsilon_x(c, e)}{|X \setminus c|}, \quad (2)$$

where

$$\epsilon_x(c, e) = P(e|c \wedge x) - P(e|\neg c \wedge x). \quad (3)$$

Finally, we use this ϵ_{avg} to determine c's significance.

Definition 2. A prima facie cause, c, of an effect, e, is an ϵ-*insignificant cause* of e if $\epsilon_{avg}(c, e) < \epsilon$.

Definition 3. A prima facie cause, c, of an effect, e, that is not an ϵ-*insignificant cause* of e is an ϵ-*significant*, or *just-so*, cause.

Then, the primary problem becomes one of determining an appropriate threshold for ϵ. In prior work (Kleinberg and Mishra 2009) we have shown how this problem may be treated as one of multiple hypothesis testing, where we aim to control the false discovery rate. Here, each prima facie cause is a hypothesis, and we want to find the level at which an ϵ_{avg} is statistically significant. Since we are testing a multitude of hypotheses (normally from hundreds to thousands), we will also infer the null hypothesis from the data using the empirical Bayesian formulation introduced by Efron (2004). We assume that the number of true positives is small relative to the total number of hypotheses tested, thus the ϵ_{avg} values will mostly fit a normal distribution, with deviations from this distribution indicating non-null tests. Thus once we determine an ϵ' such that for all $\epsilon_{avg} \geqslant \epsilon'$, the FDR is less than some small threshold (such as 0.01), ϵ' becomes our threshold.

The connecting principle

We will now use the strength associated with our type-level causes to assess the strength of the token-level claims. One way of relating these two levels of causality is by using the Connecting Principle, introduced by Sober (1986). The basic idea is that the support of a particular token hypothesis

(such as Bob's smoking caused his lung cancer) is proportional to the strength of the type level relation (such as smoking causes lung cancer).

Definition 4. *Connecting Principle*: if C is a causal factor for producing E in population P of magnitude m, then its support is given by:

$$S\{C(t_1) \texttt{ token caused } E(t_2) | C(t_1) \texttt{ and } E(t_2) \texttt{ token occurred in } P\} = m.$$

The value of the support, $S(H|E)$, which measures the support of H given E, can range from -1 to $+1$. Here C and E are types of causes and effects and the time-indices indicate the token events that occur in a particular place and time, represented by t_i. The measure of m used by Sober is[1]:

$$m = \sum_i [P(E|C \wedge K_i) - P(E|\neg C \wedge K_i)] \times P(K_i), \quad (4)$$

where the K_is are the background contexts and this measurement denotes the magnitude of causal factor C for effect E in population P. The background contexts are formed by holding fixed all factors in all possible ways. With factors x_1, x_2, x_3, one possible background context is $x_1 \wedge x_2 \wedge \neg x_3$.

For a particular token case, according to Sober, the relevant population may be defined using whatever is known about the case. So, if a person's age and weight are known, then the population is one comprised of individuals with those properties. If less is known, perhaps only that he is a U.S. citizen, then the relevant population is U.S. citizens. Note that in our method there could be separate structures for these populations (with this explicitly noted), or each bit of information defining the population could simply be a proposition, thus creating one structure that allows varying results based on additional properties that must hold. [2]

The main principle here is that a known type-level relationship between some c and e is good evidence for c causing e, if we see that both c and e have occurred. Clearly, the type-level relationship alone is not enough, the relata must actually be instantiated. In both Sober's method and ours, the type-level causes are precisely such because of their frequency of observation in some population. That is, if we find that 80% of people who develop disease X die shortly after, then this gives us reason to believe that if we observe a new patient who contracts disease X and dies, this is another instance of the disease being fatal.

Token-level reasoning

We assume that we begin with some set of inferred type-level causes and their significance scores, and some set of facts about the scenario such that these tell us which propositions are true and false at which times. Then, our goal is to assess the significance of the type-level causes for the token case. The result will be a weight for each of the potential causes, corresponding to the hypothesis that it was a token-level cause. In some cases a model might be known, but when it is not we assume we have the original time series traces used to infer the type-level causes. We will now reframe Sober's principle for our purposes, using our measure of significance and allowing incomplete information.

The set of possible token causes

We start with the question of selecting the hypotheses that should be examined in the token case. First, we note that an insignificant type-level cause can be a token-level cause. In fact, a token-level cause does not have to be even a *prima facie* type-level cause. We want to be able to consider such cases, and not immediately rule out factors that are not causes at the token level. Let us recall that we are calculating the support of token causal claims - with the presumption that we are interested in those with high levels of support. If two possible token causes took place on a particular occasion and one is a type-level genuine cause while the other is a type-level insignificant cause, the more likely explanation for the effect is that it was token caused by the type-level genuine cause. That is, if we have a number of token causal hypotheses, those with the highest support will be those with the highest value for ϵ_{avg} – our just-so or genuine causes. Thus, if we know that a just-so cause of the effect in question took place, we do not need to examine any insignificant or non-prima facie causes of the effect, as the only other causes that may have higher significance for the effect are other just-so or genuine ones. If none of the just-so or genuine causes occurred, then at that point we would have to examine alternative hypotheses.

Support of a causal hypothesis

Since we may not always know whether a cause occurred, we are thus interested in:

$$S(H) = S(H|E) \times P(E), \quad (5)$$

where $S(H|E)$ is a measure of support defined by Sober. That is, we want the support of a causal hypothesis, which is the support for it given the evidence times the probability of the evidence. However, since the evidence is that there is a type level relationship between c and e and that c and e occurred, then all of these are known to be true except for whether c occurred. Thus the probability of the evidence is really the probability of c. Then, using our measure of significance, ϵ_{avg}, we define support as follows. We will use the notation of $c \rightsquigarrow e$ to denote our hypothesis H that c "led-to" e in the token case. Note that this is not the leads-to operator introduced earlier, but simply a shorthand. Keeping in mind that there is a type-level relationship between c and e and that these are actual events occurring at particular times and places, we omit numerical subscripts for the moment.

Definition 5. Assume that e token-occurred in population P; that the probability that c token-occurred in P is $P(c)$; and that $\epsilon_{avg}(c, e)$ is the strength of the type-level relationship

[1] This is also called the average degree of causal significance (ADCS), and was introduced by Eells (1991).

[2] Note that in this case, saying that something is true for a population, where the population is defined by properties $p_1, p_2 ... p_n$ means testing whether, in addition to the formulas for the causal relationships, $p_1 \wedge p_2 \wedge \cdots p_n$ holds.

between c and e. Then, the support for the hypothesis that c token-caused e in P is:

$$S(c \rightsquigarrow e) = \epsilon_{avg}(c, e) \times P(c). \quad (6)$$

In cases where we know that c and e have token-occurred we are then computing the posterior support, where $P(E)$=1, and this reduces to the case outlined by Sober, where the probability of the evidence was always one. This means that if we have full knowledge of a scenario, the support for each possible explanation will be exactly equal to the strength of the corresponding type-level relationship. However, when we have missing data and are unsure as to whether or not a possible cause occurred, the support for the hypothesis will be weighted by the probability of the cause having occurred, given what we have observed.

Calculating the probability of c

To calculate the probability of a particular cause token-occurring, we can go back to our original data, using frequencies (calculating the frequency of sequences of length t where the evidence holds). However, if we have or have inferred the structure of the system, we may use that instead. In either case, first note that we are computing the posterior probability of c, where our evidence is one sequence of observations, comprised of a conjunction of the facts about the scenario. We will refer to this evidence as E. It will be easier to later represent the probability of $\neg c$ than c and thus we are now interested in $P(c|E)$, which is by an application of Bayes' rule:

$$P(c|E) = 1 - \frac{P(\neg c \wedge E)}{P(E)}. \quad (7)$$

Note that the facts we have about the current scenario will be time-indexed such that we have facts at times t_1, t_2 and so on, indexed relative to the beginning of the event or at times such that we know their order and can calculate the elapsed time between them. These facts constrain the set of states our system has occupied (assuming our model of the system is correct, or our data is representative of the system). If q is true at $t = 3$ then at t_3 the system must be in a state labeled with q. Let us now construct the set F where each $f_i \in F$ is the conjunction of facts that are known to be true at time i, for $i \in [0..t]$, where time 0 is the beginning of the token event and the effect e occurred at time t. When for a particular i there are no known facts of that time then $f_i = true$. Otherwise, a particular f_i might be something like $(asbestos \wedge smoking)$. We may also limit the evidence considered to known causes and effects of c.

Remember that there is a previously inferred relationship such as:

$$c \rightsquigarrow_p^{\geqslant x, \leqslant y} e, \quad (8)$$

between c and e (and c and e may themselves be logical formulas) where we assume $y \geqslant x$ and that we are computing $P(c)$. Then, when computing the numerator of the fraction in Equation (7) we add to the original f_i's, forming a new set F' such that $f'_i \in F' = f_i \wedge c$ if $t - y \leqslant i \leqslant t - x$. For both numerator and denominator, we proceed in the same manner, with the only difference being the addition of $\neg c$ to the f_is in F'. The negated c means that c did not occur in such a way as to satisfy the formula representing the relationship between c and e. Thus we are calculating the probability of c not having happened during that time window - given e's occurrence and all other known facts about the case.

Thus, when we do not have a model, the probability, $\frac{P(\neg c \wedge E)}{P(E)}$, will be the number of times the sequence of facts in F' is true along the trace, divided by the number of times the sequence F is true. For a set of traces, these would correspond to the number of traces in which each set holds. When we have a model, the probability is as follows. With $K = \langle S, s_i, \mathcal{T}, L \rangle$ being the structure representing the system, and where states satisfying each $f_j \in F \bigcup F'$ have been labeled as such and all states are labeled with $true$, then for $0 \leqslant t < \infty$, the probability of the set of paths beginning in s_0 (the start state of the system) where each $s_j \models_K f_j$, and the paths are of length t, is given by the following recurrence, where we begin with $j = t$ and $s = s_0$:

$$\begin{aligned} P(j, s) = &\text{ if } j = 0 \text{ and } f_{t-j} \in labels(s) \text{ then } 1 \\ &\text{else if } f_{t-j} \notin labels(s) \text{ then } 0 \\ &\text{else } \sum_{s' \in S} \mathcal{T}(s, s') \times P(j-1, s'). \end{aligned} \quad (9)$$

We will repeat this procedure twice, once for the numerator and once for the denominator, thus calculating the probability of each cause having occurred using Equation (7).

Procedure for assigning support to causes

Recall that we have sets of type-level genuine, just-so, and insignificant causes of the token-effect in question. Then to determine the support for each we must first test which of these are satisfied by the token-level observations. For the causes whose truth value we cannot determine, we use the above procedure to determine their probability given the observations. Recall that the support for each hypothesis is the previously computed ϵ_{avg} - weighted by the probability of the evidence. That is, the largest possible value of the support for a token hypothesis is its associated ϵ_{avg} (since the probability can be at most one). If any genuine or just-so type-level causes have occurred, this means that they will have the highest values of this support. As our goal is to find the likeliest causes - those with the most support - we can begin by taking these sets and testing whether any of their members are true on the particular occasion.

That is, with C being the set of just-so and genuine causes of the token-effect, e, and F being the set of known time-indexed facts, we test whether each $c \in C$ is true on this occasion given the facts. This means determining whether the components of the formulas occurred in such a way as to satisfy the causal relationship. Thus if the formula is $q \rightsquigarrow^{\geqslant 1, \leqslant 2} e$ and we know q at t_1 and e at t_2, the formula would be true in this token instance, while if the facts were instead that q at t_1 and e at t_4, it would be false. Now note that q could have been a formula itself, meaning we would have to initially determine the times at which it is true. Let us recall the types of formulas and discuss their truth values:

1. An atomic proposition, g, is true at time t if it actually occurred at t.

2. With g and h being state formulas, $\neg g$ is true at t if g is not true at t. Then, $g \wedge h$ is true at t if both g and h are true at t; $g \vee h$ is true at t if at least one of g or h is true at t and $g \rightarrow h$ is true at t if at least one of $\neg g$ or h is true at t.

3. Where f and g are state formulas, and s is a nonnegative integer or ∞, the path formula $f\mathcal{U}^{\leqslant s}g$ is true for a sequence of times, beginning at time t if there exists an $0 \leqslant i \leqslant s$ such that at time $t + i$ the state formula g is true and $\forall j : 0 \leqslant j < i$ the state formula f is true at $t + j$. The path formula $f\mathcal{U}^{\leqslant s}g$ is true for a sequence of times beginning at time t if either $f\mathcal{U}^{\leqslant s}g$ is true beginning at t or $\forall j : 0 \leqslant j \leqslant s$, f is true at $t + j$.

4. With f being a path formula and $0 \leqslant p \leqslant 1$, the state formulas $[f]_{\geqslant p}$ and $[f]_{>p}$ are true at time t if there is a sequence of times, beginning at t that satisfy the path formula f.

Following this formulation, we may identify if any $c \in C$ is true on the occasion in question, in which case their support is simply the associated ϵ_{avg} values. However, if this set is empty - either none occurred or we do not have enough information to determine whether any occurred, we must then calculate their probabilities, as described in the previous section. Note that we cannot assume that if the probability of a genuine or just-so cause is non-zero, then the support for the corresponding token hypothesis will be greater than for any insignificant causes. We did not test whether any insignificant causes actually occurred, so it is possible that for a genuine cause, c, $P(c)$ is low enough that despite its higher value for ϵ_{avg}, an actually occurring (probability = 1) insignificant cause has a larger value for the support ($\epsilon_{avg} \times P(c)$). In the case where there are many insignificant causes, testing whether each occurred may be computationally intensive. It is possible to define a threshold such that if the support for a cause is below it, insignificant and other causes are examined, and to constrain the set of insignificant causes to those which are true in the token case.

In any case, we begin with the probabilities, and thus support, for all genuine and just-so causes. When these values are very low or zero, we must examine the other potential explanations including our previously discarded type-level insignificant causes, and perhaps even those that are not prima facie causes. Further, it is possible that a negative cause - one that normally prevents the effect - actually was the token cause. After examining all of these, the final result is a set of possible explanations ranked by their support, with those having the highest values being the preferred explanations for the effect. We can also test any hypotheses of interest to see how they relate to the token effect.

Examples

We first discuss a simple case to illustrate the proposed approach, and then present a few examples of the types of reasoning that have traditionally posed problems for theories of causality.

Basic example

We begin with Alice and Bob, who each have a highly contagious case of chickenpox, which their friend Chris has now contracted.[3] Let us assume we have already found one significant type-level cause of contracting chickenpox (with all other causes being insignificant). This is represented by:

$$T \rightsquigarrow^{\geqslant 10, \leqslant 21}_{\geqslant p} P. \tag{10}$$

That is, touching or other close contact with a person who has chickenpox (T), causes the person who had this contact to contract chickenpox (P) in between 10 and 21 days, with probability p. Since we have found this to be a type-level cause, we also have the associated value of ϵ_{avg}.

We have the following facts about the token case:

1. Observation begins at day 0, when both Alice and Bob developed chickenpox;
2. Alice had lunch with Chris on day 1;
3. Bob went to Chris's party, and greeted him with a hug, on day 5;
4. Chris developed chickenpox on day 14;
5. The only significant cause of chickenpox is that in formula (10).

Our type level relationship says that if T is true at some time t then it can lead to P being true between time $t + 10$ and $t + 21$. The facts we begin with are that Alice's instance of T is true at $t = 1$ and Bob's at $t = 5$. To satisfy the causal formula of (10), P would need to be true in the intervals $[11, 22]$ or $[15, 26]$. P is true at 14 and thus Alice's contact with Chris can be considered as a possible token-cause of P. Now, for Bob's contact to be a token cause of P, P would need to be true at a time in $[15, 26]$. However, P is true at $t = 14$, which means this causal relationship did not occur, and it is not a possible token cause (since it could not lead to P at the time at which P actually occurred). Thus in this case our only potential token cause is Alice's contact with Chris, and the support for this token cause will be $\epsilon_{avg}(T, P)$.

Varying efficacy of causes

Another case that is difficult to reason about is when two causes of an effect both occur, but one is much stronger and "trumps" the other. Continuing with the case of Alice, Bob and Chris, let us say we now have two type-level relationships:

$$T \rightsquigarrow^{\geqslant 10, \leqslant 21}_{\geqslant p_1} P, \tag{11}$$

$$C \rightsquigarrow^{\geqslant 10, \leqslant 21}_{\geqslant p_2} P. \tag{12}$$

The first is identical to equation (10), but we have now added a second relationship, denoting the fact that chickenpox may also be spread through the air by coughing and sneezing. Further let us say for the moment that $p_2 \gg p_1$. In this case we have the following facts:

[3] Readers familiar with the philosophical literature on token causality will notice that this case directly parallels the common example of Billy and Suzy throwing rocks at a glass bottle.

1. Observation begins at day 0, when both Alice and Bob developed chickenpox;
2. Alice had lunch with Chris on day 1, but was careful to avoid directly touching him;
3. Bob saw Chris in an elevator, and Bob coughed and sneezed through their entire conversation on day 5;
4. Chris developed chickenpox on day 16;
5. The significant causes of chickenpox are those in formulas (11) and (12).

Now, we see that Alice's contact with Chris is an instance of the relationship in (11), while Bob's is an instance of (12), with both occurring in the time windows so as to satisfy these formulas. Thus both Alice and Bob could have infected Chris according to our type-level relationships. However, the ϵ_{avg} associated with the relationship in (12) will be much higher than that for the relationship in (11), owing to the large difference in associated probabilities. Thus, the support for Bob causing Chris's chickenpox by coughing on him, will be much larger than the support for Alice causing Chris's chickenpox by being in close contact with him. While this case may seem as simple as the first, it can be difficult to reason about using theories that do not take into account the strength associated with the type-level relationships. For instance, in counterfactual theories we would reason that had Bob not infected Chris, Alice would have, and vice versa – thus fallaciously concluding that neither in fact caused Chris's illness. More detailed versions of these theories would find that both contributed, but would not provide the ranking that we are able to provide.

Difficult case

Continuing with our running example, we will make the case somewhat tricky. In this scenario, Bob once again has chickenpox and seems to pass it on to Chris by coughing near him. However, this time Chris received the chickenpox vaccine shortly before he saw Bob. Surprisingly, he developed the milder, less contagious form, that in rare cases is in fact caused by the vaccine. Thus something that usually prevents the effect apparently caused it in this case. The related type-level relationships are:

$$C \leadsto^{\geqslant 10, \leqslant 21}_{\geqslant p_1} P, \tag{13}$$

$$V \leadsto^{\geqslant 5, \leqslant 26}_{\geqslant p_2} P. \tag{14}$$

That is, as before, coughing and sneezing transmits the virus with a high probability (p_1). In general V causes $\neg$P, but with probability $p_2 = 2\%$ (which is far less than p_1) it causes P. Note that since p_2 is so low, V is likely a type-level negative cause of P (lowering its probability). The facts are as follows:

1. Observation begins at day 0, when Bob developed chickenpox;
2. Chris received the chickenpox vaccine on day 2;
3. Bob saw Chris in an elevator, and Bob coughed and sneezed through their entire conversation on day 5;
4. Chris developed chickenpox on day 16;
5. The related significant cause of chickenpox is that in formula (13).

Recall that our procedure is to first identify the significant type-level causes that actually occurred. Thus, we see that Bob's contact with Chris is an instance of the relationship in (13), as the contact and Chris's chickenpox occurred at such times as to fulfill that relationship. We would not automatically test the contribution the vaccine made to Chris's illness. If we want to test the hypothesis that the vaccine caused the chickenpox, we would see that that type-level relationship was instantiated, with the support for the hypothesis then being exactly equal to the associated ϵ_{avg}, which will be less than that for the relationship associated with Bob. Thus what we know to be the actual cause of the illness has less support than something that did not cause this particular instance. It is possible that if we had more specific relationships related to the particular forms of chickenpox (and involving more details such as genetic factors and so on), we could find that the vaccine is likelier to cause this type. However, note that in most cases we will not be omniscient and thus we would not know that the disease was caused by the vaccine, thus it would be quite plausible that the chickenpox occurred due to Bob's coughing and despite the vaccine.

Conclusion

We have shown how inferred type-level causes, represented by logical formulas, may be used to reason about token-level cases. This method captures information about the timing of the general relationship and occurrence of actual events, allowing automated reasoning about cases that were previously only correctly handled with intuition. Through examples we have shown how we may arrive at results consistent with common sense and in what cases this type of reasoning is not adequate. In future work we will discuss how to include other knowledge as well as the possibility that some "facts" may be conflicting or incorrect.

References

Eells, E. 1991. *Probabilistic Causality*. Cambridge University Press.

Efron, B. 2004. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association* 99(465):96–105.

Giordano, L.; Martelli, A.; and Schwind, C. 2000. Ramification and causality in a modal action logic. *Journal of Logic and Computation* 10(5):625–662.

Halpern, J. Y., and Pearl, J. 2001. Causes and Explanations: A Structural-Model ApproachPart 1: Causes. 194–202.

Hansson, H., and Jonsson, B. 1994. A logic for reasoning about time and reliability. *Formal Aspects of Computing* 6(5):512–535.

Hausman, D. M. 2005. Causal Relata: Tokens, Types, or Variables? *Erkenntnis* 63(1):33–54.

Hopkins, M., and Pearl, J. 2007. Causality and Counterfactuals in the Situation Calculus. *Journal of Logic and Computation* 17(5):939.

Kleinberg, S., and Mishra, B. 2009. The Temporal Logic of Causal Structures. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI-09)*, 303–312. Corvallis, Oregon: AUAI Press.

Lewis, D. 1973. Causation. *The Journal of Philosophy* 70(17):556–567.

Lewis, D. 2000. Causation as influence. *The Journal of Philosophy* 97(4):182–197.

Lin, F. 1995. Embracing causality in specifying the indirect effects of actions. In Mellish, C., ed., *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1985–1991. San Francisco: Morgan Kaufmann.

Lunze, J., and Schiller, F. 1999. An example of fault diagnosis by means of probabilistic logic reasoning. *Control Engineering Practice* 7(2):271–278.

McCarthy, J., and Hayes, P. J. 1969. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence* 4(463-502):288.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Poole, D. 1994. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence* 11(1):33–50.

Sober, E., and Papineau, D. 1986. Causal factors, causal inference, causal explanation. *Proceedings of the Aristotelian Society, Supplementary Volumes* 60:97–136.

Thielscher, M. 1997. Ramification and causality. *Artificial Intelligence* 89(1-2):317–364.

Woodward, J. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, USA.