Upsetting the contingency table: Causal induction over sequences of point events

Michael D. Pacer (mpacer@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract

Data continuously stream into our minds, guiding our learning and inference with no trial delimiters to parse our experience. These data can take on a variety of forms, but research on causal learning has emphasized discrete contingency data over continuous sequences of events. We present a formal framework for modeling causal inferences about sequences of point events, based on Bayesian inference over nonhomogeneous Poisson processes (NHPPs). We show how to apply this framework to successfully model data from an experiment by Lagnado and Speekenbrink (2010) which examined human learning from sequences of point events.

Keywords: causal inference; continuous time; stochastic processes; Bayesian models

Introduction

Only a single bullet is needed to end a life. One momentary event quickly ruptures a slew of causal mechanisms, having effects that persist long after the trigger was pulled. We understand this causal inference easily enough, but how do we manage to arrive at that inference? More precisely, what formal tools can characterize events that last an instant but leave profound consequences in their wake?

Learning from instantaneous events is one of the most important forms of causal induction in the real world, but is not the case considered in most models of human causal reasoning. More commonly, these models rely on the existence of aggregated statistics often in the form of a contingency table describing the frequencies with which different combinations of events occur (Griffiths & Tenenbaum, 2009). The most traditional case is the 2×2 contingency table that describes two variables that are either present or not. It is out of this statistical paradigm that many of the models of human causal inference and learning have arisen (Gopnik et al., 2004).

The world offers more than contingency tables and frequency counts. Different kinds of data warrant different kinds of inferences. For example, when Greville and Buehner (2010) coerce events embedded in continuous-time into contingency tables, they are challenged with identifying an (assumed) underlying trial structure. To fill a 2×2 table of event pairs, one needs to slice the continuum into trials; a variable is counted if an event occurred in the slice and not if it did not. But different slicing regimens can warrant different inferences about the same data. This arises because of a fundamental asymmetry between the times at which events occur and when then do not occur. In a contingency table both X and $\neg X$ are treated as the same kind of entity, but instantaneous events can be counted while the the continuous expanse of nothingness in which events are embedded by definition cannot be counted, only measured.

But there is another way to model sequences of point events occurring in continuous-time. Here, we build a model of causal induction on the framework described in Pacer and Griffiths (2012), which uses a model of events embedded in continuous-time. It computes probabilities not in terms of the frequency with which events co-occur, but directly from the temporal distances between cause and effect events. Importantly, we accomplish this *without* needing to uniquely match individual cause events with individual effect events. We rely on probabilistic graphical models to describe the structure of causal relations. We use Poisson processes and operations over these processes to describe the functional relationships between variables, which provide the structure by which we identify and organize kinds of events.

Our focus in this paper is on causal induction from sequential point processes — sequences of events that are transient moments in continuous time. This is an important case for modeling causal induction, describing a wide range of real-world settings in which people perform causal induction such as trying to infer which particular interaction with which particular other person led to coming down with the flu. Bramley, Gerstenberg, and Lagnado (2014) raise a concern that these cases present a challenge for the continuous-time causal network approach presented by Pacer and Griffiths (2012).

The paper will proceed as follows. We review Poisson processes, which form the foundation for the rest of the paper. We introduce a physical analogue of these formal structures to provide intuition to the underlying mathematics of nonhomogeneous Poisson processes. These processes form the basis of our framework when combined with probabilistic graphical causal models. We then successfully apply the framework to Experiment 2 from Lagnado and Speekenbrink (2010) — a paradigm for human judgments of causal structure from sequences of point events. We then discuss the implications of this work for statistical and mechanistic theories of human causal reasoning.

Formal Foundations

In this section we describe formal foundations for our framework (cf., Pacer & Griffiths, 2012). We explain a class of point-process models (processes that describe the occurrence of point-events which have a location, but no measurable duration) called Poisson processes that are defined in terms of a space and a positive rate function that gives the expected number of events to be found in any subspace. This rate-function will depend upon the identity and time of events that occur during the course of the processes' activity (cf.,

Simma & Jordan, 2010; Blundell, Beck, & Heller, 2012). To accommodate this kind of conditional intensity function we use nonhomogeneous Poisson processes (NHPPs), whose rate functions are not constant.

We will first describe two ways of looking at *homogeneous* Poisson processes: arrivals and rates. After developing an intuition for Poisson processes with constant rate functions, we will then consider the general case of NHPPs and how to model sequences of point events in terms of a generative model that iteratively analyzes individual arrivals relative to the rates induced by previous arrivals.

Poisson Processes: Arrivals and Rates

Poisson processes can be interpreted in a number of ways that lend themselves more or less easily to different applications. We will describe Poisson processes in two senses: the arrival process sense and the rate-of-events sense.[1] Sequences of point events can be described sequences of successive arrivals, making the arrival perspective convenient for computing point event likelihoods. But it is easiest to conceive of the causal effects of point events in terms of their altering event rates. Both perspectives prove useful.

Arrivals The arrival sense can be understood by anyone who has ever waited in a queue. You will have to wait some amount of time before your turn, and we can assign a probability that you will be served by time t. If you were next in the queue and it were governed by a homogeneous Poisson Process with rate λ, the waiting time distribution of being served by time t would be an exponential distribution with mean $\frac{1}{\lambda}$ ($t \sim \text{Exp}(\frac{1}{\lambda}) : p(t) = \lambda e^{-\lambda t}$). Interestingly, this distribution is *memoryless*, such that, regardless of how long we have waited, we still expect to wait the exact same amount of time — it has no memory of how long it has been since the last event. This memorylessness property does not hold for the general class of NHPPs.

Rates Equivalently, we can count the number of events that occur in a measurable time-period, rather than looking at the delays between each event. Poisson processes define a "rate" of events, which describes the expected count of events to occur in any interval. A homogeneous Poisson process has a constant rate, λ, and for a time interval with length $|\tau|$ we can expect to see event-count distribution governed by a Poisson random variable, with mean $(\lambda|\tau|)$. In the case of nonhomogeneous Poisson processes, we will have a rate-*function* defined over time $\lambda(t)$. Integrating this function over some time-interval defines how many events are expected to occur (i.e., for $\tau_{[a,b)}$ the expected event count is $\int_{\tau_{[a,b)}} \lambda(s)ds$).

[1] Poisson processes can be defined over higher dimensional spaces (e.g., $\mathbb{R}^3$) than the real line. This complicates the arrival perspective, which implicitly relies on the order that events "arrive". The event-rate perspective is unchanged in higher dimensions; in that sense it could be said to be more "fundamental" than the arrival perspective. In this paper, for simplicity we focus on processes defined over time ($[0, \infty)$).

Combining Perspectives The arrival perspective gives us a probability distribution over intervals of time (i.e., intervals defined from now until the next event arrives), while the rate-of-events perspective gives us an instantaneous measure of event likelihood which is comprehensible only in terms its integration over intervals of time. The former is more useful in cases where events are analyzed one at a time. For example, when simulating dependent event sequences or calculating the probabilities of event sequences in terms of the likelihood of each event's occurrence given the previous relevant occurrences. In our model of Lagnado and Speekenbrink (2010), we will use this perspective to define the likelihood of each inter-arrival period conditional on the previous events.

The rate-of-events perspective is useful when simulating many events when the rate is independent of the particular occurrence of the events. The rate perspective is also useful for calculating event likelihoods, when the interval during when the events occurred is known, but the exact occurrence times are unknown. Pacer and Griffiths (2012) use this technique to analyze the data given to participants in Greville and Buehner (2007) in which data were presented in a tabular form that described the day during which bacteria died but not the exact timing of the events. This property allows us to recover a trial structure from continuous-time by integrating over intervals of time and treating occurrences within those intervals as events that occurred in those trials.

Most importantly for our uses, it is most straightforward to see causes as altering the rate-of-events and then computing an expected wait-time distribution based on those altered event rates. Describing effects in terms of rate changes will be the key to the *causal* aspect of our framework. Fortunately, Poisson processes have two closure properties, superposition and thinning, that allow us to create continuous time analogs of noisy-OR and noisy-ANDNOT (for more details, see Pacer & Griffiths, 2012).

Superposition and Thinning in Poisson Processes

This section will develop a rough physical model to aid in thinking about NHPPs as formed by functions on homogeneous Poisson processes. Namely, we aim to provide an intuition for the superposition and thinning closure-properties of homogeneous Poisson processes from the rate-of-events perspective. We do so by sketching a mechanistic picture of a particle emission system that exhibits these properties.

First, consider a decaying radioactive material which releases particles at a constant rate, λ. With a particle detector around the material you record the time-stamp at which particles hit the detector. A particle is expected to hit the detector, on average, every $\frac{1}{\lambda}$s. This detector will then be recording a homogeneous Poisson process with rate λ.

Suppose you were to place a barrier to block some of the paths leading from materials to the detector (call the proportion blocked $\pi : 0 \leq \pi \leq 1$, as in the parameter associated with the orange filter in Figure 1). From the detector's perspective, events associated with particles blocked by a filter are events that never occurred. This process is known as filtering

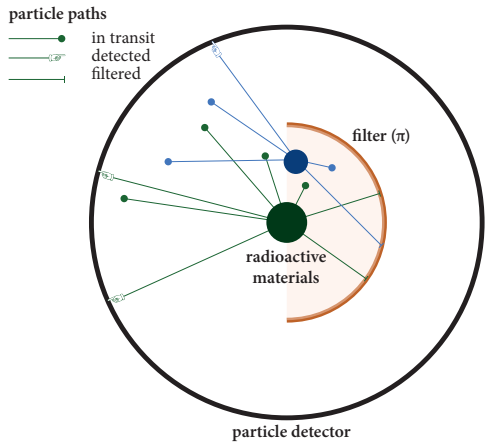

Figure 1: Particle emission detector model for visualizing superpositioning and filtering Poisson processes. Color distinguishes particle origins prior to detection, with color lost in detection. Filtered events are never detected.

the Poisson process, and if π is independent of the generating process, filtering gives a Poisson process with rate $(1-\pi)\lambda$.

Suppose you were to place another radioactive material in the detector, of a different kind than the original, but which did not interact with the original radioactive material (see the blue and green materials in Figure 1). From the perspective of the detector, there is no difference between the particles hitting it from different materials — it only knows *that* a particle hits it and *when* it hits it. If we suppose the rate of this new material's emitting particles is constant at λ_1, then the total set of events would be a Poisson process with rate $\lambda+\lambda_1$. This is the superposition property of Poisson processes: if jointly independent, the union of the events from two Poisson processes will be a Poisson process whose rate is their sum.

Superposition and thinning allow us to see how the rates of Poisson processes can change without altering their underlying structure. We can apply these transformations at particular times or intervals of time, thereby producing and increase (via superposition) or a decrease (via thinning) in the rate of events while at all times maintaining its identity as a Poisson process. Applying superposition or thinning as time-dependent functions thus allows one way to create NHPPs that nonetheless can be understood in terms of component processes and their transformations.

Causal Graphs and Sequential Point Processes

With superposition and thinning in our tool belts, we can discuss causal point processes feeding back into themselves. By applying thinning and superposition dependent on time we can build NHPPs from component processes. If we apply superposition and thinning functions relative to the occurrence of events at particular times (e.g., by scaling their effects relative to the distance from the events' occurrence time), then we can define NHPPs relative to the occurrence of particular events. If we then define the set of events capable of evoking changes in the Poisson process as themselves being the outcome of Poisson processes (see also Blundell et al., 2012), then we have successfully managed to create a system of causal relations in terms of Poisson processes.

It is worth noting that if we want to describe the causal effects of point events at all, we will need some form of influence function that lasts beyond the occurrence of the point event. Because point events are instantaneous, if a cause's effects last while cause persists the effects would have to occur at that same instant. But we are seeking to model causal relations that are not only embedded in a moment in time, but *across* time. We need to consider how events that occur at one point and time can influence events – or, more precisely, the underlying rate of events – at later time points and intervals.

We need a way to describe what kinds of things exist which things are related to one another and in what ways they are related — i.e., we need an ontology, plausible relations, and functional forms (Griffiths & Tenenbaum, 2009). Probabilistic graphical models are formal structures for expressing stochastic dependencies between variables. But are our events valid variables? To define a set of possible graphs over variables, we usually need to know what those variables are. Given that we do not know when events will occur events would make events a bad candidate for graphical nodes. Furthermore, there's no obvious way to say whether event at time t_0 counts as the "same" event (or even kind of event) as that which happens at t_1. Without some other information we would be left with a proliferation of variables equal to the number of events we observe. Instead, we will consider the case where events have signature identities that identify which sequence of events each event belongs to. These sequences will be our variables. Both events and variables feature in our ontology, but the graphs will be defined over potential relations between variables which will be associated with sequences of point events. Events will be the components through which variables actually affect one another.

As in Griffiths and Tenenbaum (2009), to compute posteriors over graphs, we will need a prior over the possible relations between variables. Often this will be made easier by knowing (from our ontology) which things are potential causes of which effects. When all possible binary relations could be graphs it is a challenge not because it is difficult to impose a prior *per se*, but because of how rapidly the number of graphs grows in terms of the number of variables. This is especially pernicious because by using continuous-time causal networks, we can represent directed cycles as easily as any other relationship, which allows even more graphs than in the traditional causal Bayes nets framework.

For each of these relationships, we will express a functional form defining how variables relate to one another. In particular, we need to relate the kind of relationship between variables (e.g., generative, preventative) to the how the influence of particular events associated with those variables changes over time. Our approach is to see an event(t') in the cause-variable's event set as inducing a NHPP on the variable's child nodes. For generative causes, the NHPP starts with a maximum rate(ψ) immediately when it is triggered

with the rate decaying exponentially according to a decay rate(ϕ) scaled distance from the cause event ($(t-t')$). This produces a NHPP with rate function $\lambda(t) = \psi e^{-\phi(t-t')}$.

We compute likelihoods for events sequences as follows. We compute a likelihood for an event using the rate function of its associated variable during its wait-time interval. We then update the rate functions for variables that are effects of the variable associated with the event, and iterate this process until we exhaust the event sequence. The waiting-time likelihood for each event can be further broken down into waiting-time likelihoods for the Poisson process composing its variable's rate function at the instant the event occurs. Note that from the superposition perspective an event's waiting-time implies that *none* of the variable's processes produced events until the point when the event in question occurs.

The general form of the cdf (cumulative density function) of the waiting time distribution until the first event of a NHPP with time-dependent rate function $\lambda(\cdot)$, is:

$$F(T \leq t) = F(t) = 1 - \exp\Big(-\int_0^t \lambda(s)ds\Big).$$

This can be converted to a pdf by taking the derivative according to t, which (assuming that the derivative exists) becomes,

$$p(t) = \lambda(t)\exp\Big(-\int_0^t \lambda(s)ds\Big).$$

Modeling Continuous Event Streams: Lagnado & Speekenbrink, 2010

In this section we will model Experiment 2 from Lagnado and Speekenbrink (2010) using the formal elements described in Pacer and Griffiths (2012) and above. As mentioned, Bramley et al. (2014) raised a concern that the framework described in Pacer and Griffiths (2012) does not address "sequences of point events". Here, we apply this framework to modeling data from real-time event streams, characterizing sequences of point events and using them for computing the same inferences Lagnado and Speekenbrink (2010) asked of their participants. Moreover, our models' judgments closely match average human responses, which suggests the framework succeeds both at characterizing the sequences of point events and at producing models capable of causal inference that comparable to that of human beings facing the same problems.

Experiment Description

Experiment 2 of Lagnado and Speekenbrink (2010) has the form of participants observing a continuous sequence of events (as a video) that represent the time-course of various kinds of seismic activity, specifically three kinds of seismic waves (which we shall refer to as A, B, C) and earthquakes (E). The goal of the participants was to infer which (if any) of the seismic waves were the cause of the earthquakes.

Earlier work suggests people will lessen their judgments of causal attribution between two variables if there is a longer(or more variable) delay between the occurrence of two events. However, this could either be because there is something specific about long delays between causes and effects, or that longer delays allow more opportunities during which other events could occur that are not causally related, thus weakening the connection between the original two variables of interest.[2] According to the design of the experiment – unknown to the participant – only one of the types of wave(A) was a cause of earthquakes, but sometimes non-causal waves would occur in the interval of time between the cause and its effects. This allows us to disentangle the two explanations for reduced causal strength due to longer delays. The length of time between the cause and its effect and the commonness of mid-interval events' sometimes were the primary differences between the experimental conditions.

The experiment had a 2×2 structure. DELAY-LENGTH could be LONG (mean delay between cause and effect = 6s) or SHORT (mean delay between cause and effect = 3s) — in both the standard-deviation is 0.1s. The probability a non-cause event occurred between the cause and the effect was LOW ($\approx 35\%$ of the time a non-cause event would occur between a cause and its effect) or HIGH ($\approx 65\%$). These probabilities are approximate because the event sequences were randomly sampled and so cannot be expected to exactly match expected percentages. The authors chose delay distributions between the occurrence of cause and lure events to produce these probabilities in aggregate across samples.

They generated sixty datasets per condition that represented the time-stamps and identities of events that occurred in the movie. Of these, the first twenty datasets were used, with each of twenty participants participating in all four conditions exactly once. They were told that each animation would last no more than 10 minutes.[3]

After each video participants were asked to provide judgments about the seismic waves that they had just observed. Participants were first asked to rate the extent to which each wave was a cause of earthquakes on a scale of "0 (does not cause the effect) to 10 (completely causes the effect)". This provides an "absolute" judgment of each wave's causal properties since the rating provided for one of the waves did not constrain the rating provided for the other waves. Participants were then asked for "comparative ratings, in which they divided 100 points amongst the three types of cause."

Building the Model

We treated the problem as one of structure induction. That is, given the knowledge that there are three possible cause variables ($\{A,B,C\}$) of the effect in question (E) and the data $\mathcal{D}$, we want to infer a posterior over the possible graphs linking

[2] We should note that "more opportunities" is actually somewhat misleading as opportunities in plural form suggests that there would be a countable number of opportunities during which these events could intervene. It is more accurate to say that long delays allow for a larger, continuous amount of "opportunity"(a mass noun).

[3] Though participants did see multiple conditions, we treat each trial independently rather than attempting to detect order effects. While we acknowledge its potential usefulness, addressing this is outside the scope of our analysis.

the causes to the effects. Then we will use this posterior to compute measures analogous to those given by participants.

Graphs and Parameterization We considered graphs with any subset of three potential independent causes $\{A,B,C\}$. All causal links were generative and non-interacting with the other causal links. Thus the total rate of effects under a graph would be the superposition of all Poisson processes induced by the activity of cause-events according to the graph. As described in Pacer and Griffiths (2012) this is the continuous-time analog to the Noisy-OR parameterization of a causal graph. Because causes were independent according to all graphs, the likelihood of their occurrences can be removed from our likelihood calculations.

In addition to a base-rate process $\text{PP}_{\lambda_\varnothing}$, which we assumed was a homogeneous Poisson process with rate $\lambda_\varnothing$, we allowed each cause-event ($t_d^{[X]}$) to initiate a NHPP with maximum rate (ψ_X) that decays exponentially(ϕ_X) relative to the the distance from the cause event($|t - t_d^{[X]}|$).

We sampled these parameters in a similar manner to Pacer and Griffiths (2012). We use uniform random variables ($u \sim U(-10^{\cdot 1}, 10^{\cdot 1})$) under a transformation ($\lambda_\varnothing = e^u$) to determine our initial timescale (in seconds), which acts as our base-rate PP. This creates a approximate scale-free baseline parameter ($\lambda_\varnothing \sim \frac{1}{\lambda_\varnothing}$) from which other parameters can be sampled. We sample $\psi_X \sim \Gamma(\lambda_\varnothing, 1)$ (the maximum rate induced by a single event of type X occurring) and $\phi_X \sim \Gamma(\lambda_\varnothing, 1)$ (the rate at which the intensity decays according to the distance in time from that instance) associated with each potential cause $X \in \{A,B,C\}$. Each cause instance($t_d^{[X]}$) produces a NHPP with rate function $\psi \exp(-\phi(t - t_d^{[X]}))$.

Because the baseline distribution is scale-free and defines other scales, these parameters are not "fit to the data". A "misfit" baseline scale produces overflow, underflow or other numerical and computational issues that result in model failure. But, any success can only stem from the model's structural commitments and the relation to the modeled data.

Data The data used to generate stimuli in Lagnado and Speekenbrink (2010) are organized by the time-step (in milliseconds) that an event occurred and the identity of the kind of event (i.e., A,B,C or E). Though there were sixty generated sequences consistent with the design principles of their experiment, we used only the first twenty which corresponded with the conditions that they ran in their study.

Structure Inference For each graph ($G_\alpha \in \mathcal{G}$) and dataset (D) we take our sampled parameters($\{\Theta\}_{m \in \{1,\ldots,M\}}$; for us, $M = 200000$) and compute:

$$\mathcal{L}(D|G_\alpha) \approx \frac{1}{M} \sum_{m=1}^{M} \exp(\ell(D|G_\alpha, \Theta_m))$$

We compute log-likelihoods ($\ell(\cdot)$) under G_α and Θ_m as follows. For computational efficiency, we eliminate events do not alter other events (e.g., under graph $B \to E$, we consider

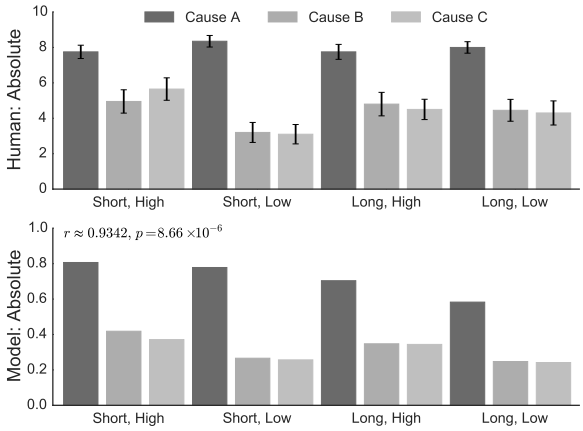

Figure 2: **Top**: Mean absolute judgments, Experiment 2 of Lagnado and Speekenbrink (2010). **Bottom**: m_{abs} model.

likelihoods of E- and B-events but eliminate A- and C-events). Using the reduced event set ($\{0, t_1, \ldots, t_i, \ldots, t_n\}$), we can partition the observation interval and considering each interval $\times$ event-identity pair $\tau_{t_j,t_{j+1}} \times (X_j, X_{j+1}) \equiv \tau_j$ conditioned on previous events associated with cause X ($t_d^{[X]} \forall d \le j$) we can calculate the log-likelihood. The total log-likelihood:

$$\ell(D|G_\alpha, \Theta) = -\Lambda_{(t_0,t_n)} + \lambda_{\{t\}_0^n},$$

where $\Lambda_{(0,t_n)}$ is the log-likelihood component of the intervals,

$$\Lambda_{(0,t_n)} = \lambda_\varnothing (t_n - t_0) + \sum_{\tau_j \in D} \Big[\sum_{\substack{X, \\ X \to E \in G_\alpha}} \Big[\frac{\psi_X}{\phi_X} \big(e^{-\phi_X t_j} - e^{-\phi_X t_{j+1}} \big) \sum_{t_d^{[X]} \le t_j} [e^{\phi_X t_d^{[X]}}] \Big] \Big],$$

and $\lambda_{\{t\}_0^n}$ is the log-likelihood component of the point events,

$$\lambda_{\{t\}_0^n} = \sum_{t_j \in t_j^{[E]}} \Big[\log(\lambda_\varnothing + \sum_{\substack{X, \\ X \to E \in G_\alpha}} \big[\psi_X (e^{-\phi_X t_j}) \sum_{t_d^{[X]} \le t_j} [e^{\phi_X t_d^{[X]}}] \big]) \Big].$$

Using the likelihood estimate($\mathcal{L}$) plus the prior for each graph (in our case, uniform over graphs $p(G_\alpha) \propto 1, \forall \alpha$), we can then compute the posteriors for all graphs.

$$p(G_\alpha|D) = \frac{\mathcal{L}(D|G_\alpha) \times p(G_\alpha)}{\sum_{G_\alpha \in \mathcal{G}} \big(\mathcal{L}(D|G_\alpha) \times p(G_\alpha) \big)}$$

Comparison to Human Responses People judged causes, not graphs; we need a way to map posterior probabilities $p(G_\alpha|D)$ to causal judgments. Lagnado and Speekenbrink (2010) asked participants for absolute measures (assign each potential cause a value on scale from 0 to 11) and comparative measures (assign a total of 100 points to the three causes).

We will model judgments as statements about structure inferences (not strength estimations). We interpret the absolute score in terms of a probability that a particular variable is thought to be present by marginalizing over the probabilities given to the graphs that include that variable is a cause. I.e.,

$$m_{\text{abs}}\big(X \in N; p(G|D)\big) = \sum_{G_\alpha : (X \to E) \in G_\alpha} p(G_\alpha|D)$$

Note if the complete graph were to receive all the probability measure, then $m_{\text{abs}}(A)$, $m_{\text{abs}}(B)$, and $m_{\text{abs}}(C)$ would each equal 1, and so their sum would equal 3. Thus, this is not a probability measure in the usual sense because we have not defined probabilities over causes, but over graphs.

However, by normalizing by the sum of these measures over all nodes, we can adapt the absolute measure to saying something about the comparative importance of the different nodes in producing the effect. This will sum to 1, but still should not be interpreted as anything like a direct probability of the cause being present.

$$m_{\text{comp}}\Big(X \in N; p(G|D)\Big) = \frac{\sum_{G_\alpha:(X\to E)\in G_\alpha} p(G_\alpha|D)}{\sum_{x\in N}\sum_{G_\alpha:(x\to E)\in G_\alpha} p(G_\alpha|D)}$$

Finally, we could consider the comparative prompt as implying that there is only one cause, and so we should only consider those graphs which attribute a single cause for producing the effect in question. In fact we can say that under the restriction that only one cause may exist the graph including X as its sole cause *is* the measure of the comparative importance of X (since it is the only graph with that cause).

$$m_{\text{unique}}\Big(X \in N; p(G|D)\Big) = \frac{p(X\to E|D)}{\sum_{X\in\{A,B,C\}} p(X\to E|D)}$$

Results

We find an excellent fit between our models' predicted values and average human judgments for both absolute ($r \approx 0.93,\ p < 10^{-5}$, see Figure 2) and comparative judgments ($m_{\text{comp}} : r \approx 0.96,\ p < 10^{-5}$; $m_{\text{unique}} : r \approx 0.98,\ p < 10^{-7}$, see Figure 3) of the different kinds of waves' causal importance.

Discussion

Learning from streams of events unfolding in continuous time is an important case to consider in studying human causal induction. We have shown how this case can be accommodated within the framework of Pacer and Griffiths (2012), successfully predicting mean human judgments of real-time event streams used in Lagnado and Speekenbrink (2010). We know of no other framework that has been brought to bear on real-time event streams.

Often, *statistical* frameworks of theories of human causal inference are contrasted with *mechanistic* theories. The kinds of arguments used against statistics are that they neglect to include information about temporal and spatial characteristics of the modeled processes. We have seen success in modeling causal induction that relies explicitly on the temporal aspects of events. We see this as a step towards reconciling mechanism and statistics. More sophisticated graphical models can be built that articulate complex causal mechanisms that are sensitive to particular temporal and spatial properties. This gives a hope for dissolving statistical-mechanistic divide and the birth of a new synthesis with statistical inferences computed over mechanistic theories.

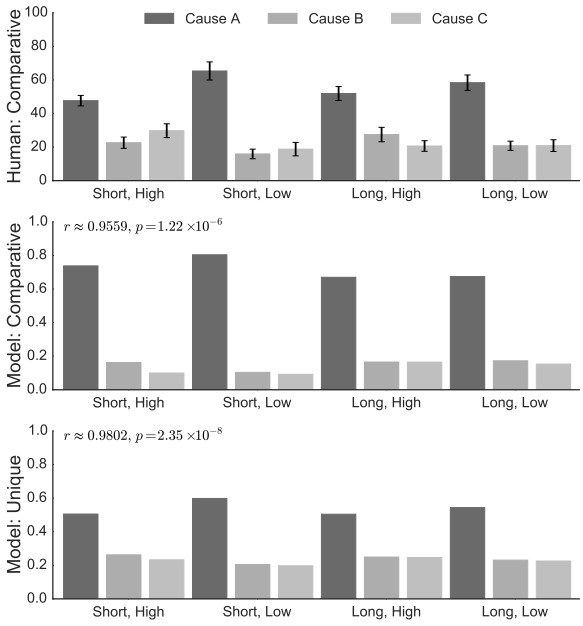

Figure 3: **Top**: Mean human comparative judgments from Lagnado and Speekenbrink (2010). **Middle**: m_{comp} model. **Bottom**: m_{unique} model.

Acknowledgments. We thank Maarten Speekenbrink and David Lagnado for providing their stimuli and results. Support for this work was provided by the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program awarded to MP and grant FA9550-3-1-0170 from the Air Force Office of Scientific Research awarded to TG.

References

Blundell, C., Beck, J., & Heller, K. A. (2012). Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems* (pp. 2600–2608).

Bramley, N. R., Gerstenberg, T., & Lagnado, D. A. (2014). The order of things: Inferring causal structure from temporal patterns. In *Proceedings of the 36th Annual Conf. of the Cognitive Science Society.*

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1-31.

Greville, W., & Buehner, M. (2007). The influence of temporal distributions on causal induction from tabular data. *Memory & Cognition*, *35*, 444453.

Greville, W., & Buehner, M. (2010). Temporal predictability facilitates causal learning. *Journal of Experimental Psychology: General*, *139*(4), 756.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological review*, *116*(4).

Lagnado, D. A., & Speekenbrink, M. (2010). The influence of delays in real-time causal learning. *The Open Psychology Journal*, *3*(2), 184–195.

Pacer, M., & Griffiths, T. (2012). Elements of a rational framework for continuous-time causal induction. In *Proc. of the 34th Conf. of the CogSci Society.*

Simma, A., & Jordan, M. (2010). Modeling events with cascades of poisson processes. In *International conference on uncertainty in artificial intelligence.*