What you didn't see: Prevention and generation in continuous time causal induction

Tianwei Gong (tia.gong@ed.ac.uk)
Neil R. Bramley (neil.bramley@ed.ac.uk)

Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK

Abstract

How do people use temporal information to make causal judgments? A number of studies have investigated the role of time in inferring generative causal structure, while few have examined prevention. Here, we focus on a challenging task in which participants learn the structure of several causal "devices" by watching the devices' patterns of activation over time. Each device potentially includes both generative (producing an activation of its effect) and preventative (blocking any effect activations within a short time window) causal relationships. We examine judgment patterns through the lens of a normative model which incorporates actual causation with considerations of prevention. We contrast this with a more computationally tractable feature-based approximation. Participants' performance was substantially above chance in all conditions. The majority of participants' causal judgments were best fit by the feature-based approximation based on delay and count heuristic cues.

Keywords: causal learning; time; prevention; structure induction; Bayesian modelling

We naturally think about the world in terms of the progression of events governed by the law of cause and effect. This means that we update our beliefs not just on the basis of unexpected events, but also based on the absence of expected events. If you see a FedEx employee at your door but you have not ordered anything recently, you might be surprised and seek an explanation such as an early birthday present. Conversely, if you order something on Black Friday but have not received it by Christmas, you may start to suspect that something has gone wrong with your order. We are good at using temporal cues to infer causal relations in these situations. The goal of this paper is to build a better understanding of how people make both generative and preventative causal inferences on the basis of observations of events occurring over time.

Humans make systematic use of temporal information to infer causal connections. Generally, people make stronger causal attributions for short temporal delays than long temporal delays (Buehner & McGregor, 2006; Shanks et al., 1989), but this is moderated by expectation, with shorter-than-expected delays also reducing causal judgment strength (Buehner & May, 2002). People are also sensitive to delay reliability with causal judgments decreasing as increasing interval variability increases between putative causes and effects (Bramley et al., 2018; Greville & Buehner, 2010). Recently, researchers have used gamma distributions (Bramley et al., 2018; Bramley et al., 2017; Lucas et al., 2015; Stephan et al., 2018) and Poisson-processes (Pacer & Griffith, 2012, 2015) to model human representation of temporal information and their sensitivity to delays between events in causal reasoning. This work shows that when observing a causal system over time, people can identify the causal structure linking multiple components on the basis of patterns of discrete events in time (Bramley et al., 2018).

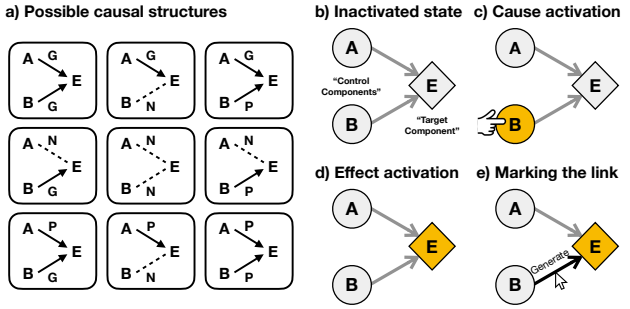

Figure 1: Example causal device from our task in different states: a) Inactivated; b-c) Activated; d) Participant response interface. e) The response hypothesis space (all possible causal structures where G = generative; N = non-causal; P = preventative).

All of the work above, exploring structure judgments based on temporal information, focuses on cases of *generative* causation, in which events such as an activation or change in a cause variable produce other events or changes in its effects. It is less clear how and when people will infer *preventative* causation, in which the occurrence of one event stops another event or change from occurring. To our knowledge, the only few studies related to preventative causal judgment given temporal information adopted a "rate-based" representation (Anderson & Sheu, 1995; Griffiths & Tenenbaum, 2005; Pacer & Griffith, 2012). In these tasks, participants see evidence about how frequently an an event occurs per unit time. Preventative causes are inferred when the rate of an effect temporarily decreases, while generative causes are inferred when its rate temporarily increases. A key property of this setting is that the lower the effect's base rate, the harder it is to reliably identify preventative influences while generative influences become more conspicuous when the effect's base rate is low (Pacer & Griffith, 2012). That is, there needs to be something there to prevent for a preventative cause to reveal itself.

While frequently recurring events may naturally be thought in terms of rates, daily life often involves sparse causal evidence with little in the way or regularity or re-occurrence. Hence, we must frequently engage in reasoning at the level of individual events, as in the FedEx example above. The distinction between generation and prevention becomes more profound in these sparse settings. Generation is about explaining an (otherwise unexpected) presence, but prevention is about explaining an unexpected *absence*. Prima facie, this makes preventative inference subordinate to the inference of a generative processes or base rate because people have to expect something to happen (i.e. to be generated) before they can be surprised that it did not. This kind of forward thinking and expectation dynamics has not been captured by previous models aiming to capture reasoning about generative causal structure (Bramley et al., 2018; Bramley et al., 2017). In this paper, we propose both computational- and algorithmic-level solutions to this problem. How do people use temporal cues to infer both generation and prevention in causal systems given sparse event data.

In the rest of this paper, we first formalise our chosen learning problem and its normative solution. Then, we collect human data and analyse performance. Finally, we contrast normative inference with a more computationally tractable heuristic proposal finding evidence suggesting people abstract simple count and inter-event delay features that are noisily diagnostic of causal structure.

Learning problem and normative inference

As a first foray into preventative causation in real-time causal structure induction, we constrain our learning problem to a space of 9 unconnected, singly connected and collider type (i.e., common effect) causal devices (Figure 1a). However, the experimental paradigm and computational models we introduce here can be directly generalised for learning in other (i.e., broader) causal hypothesis spaces. The devices investigated in this paper were made up of two *control* components (A, B) and one *target* component (E) (See Figure 1). The connection between each control component and the target component could be generative, preventative, or they could be unrelated.

We focus on causal relationships between *point events* occurring at a device's components in time. Preventive inference is only possible when there is an something to prevent (Griffiths & Tenenbaum, 2005), for example when a putative effect component has an above zero base rate. In our task, the target component is set to spontaneously activate with moderate regularity. Meanwhile, an activation of a generative component will produce an "extra" activation of the target component. An activation of a preventative component will block *all* activations of the target component within a short time window, no matter whether those activations are caused by a generative connection or by self-activation. Preventative connections are thus conceived as having a broad preventative scope (i.e., preventative connections are conceived as having a broad preventative scope; Carroll & Cheng, 2009; see Figure 2c for details). Activations of non-causal components have no impact on the target component.

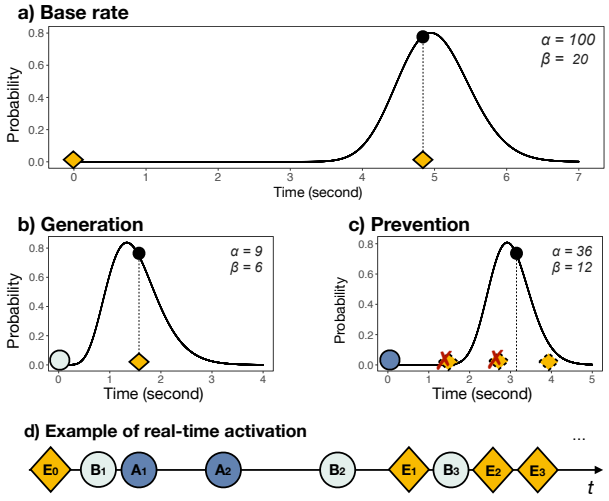

Figure 2: Gamma distributions: a) Base rate; b) Generation; c) Prevention (i.e., blocking expected events); d) Example events in continuous time (A: preventative; B: generative).

Participants' task is to observe short clips showing devices' patterns of activation over time and then judge which one of the nine potential structure hypotheses (Figure 1a) is correct for that device. The timing of self-activations, generative delays, and preventative blocking windows in the clips are sampled from gamma distributions according to the true underlying structure (Figure 2). The actual sampled values are unknown to the learner (human or model), but the generative gamma parameters are assumed to be known, since participants are trained on these in the instructions.

Normative inference

As a benchmark for performance in this task, we develop an ideal Bayesian model. The ideal reasoner is presumed to take all activation events within the observation interval as the basis of their inference. We write this as $\mathbf{d}_\tau\{d_X^{(1)},\ldots,d_X^{(n)}\}$, with events indexed in chronological order and X indicating the activated component. The learner then updates their prior, here over the nine structures $P(S)$, with a likelihood function $p(\mathbf{d}_\tau|S,\mathbf{w})$ to get the posterior distribution $P(S|\mathbf{d}_\tau,\mathbf{w})$, given the set of parameters $\mathbf{w}$:

$$P(S|\mathbf{d}_\tau,\mathbf{w}) \propto p(\mathbf{d}_\tau|S,\mathbf{w})\cdot P(S) \quad (1)$$

We assumed for simplicity that prior $P(S)$ is uniform over the potential structures (Figure 1a), and that the parameters $\mathbf{w}$ (i.e., gamma shapes α and rates β) are known.

Likelihood calculation (cf. Bramley et al., 2017) then depends on an enumerative actual causal attribution step (Halpern, 2016). The basic idea is that accurate judgments about *type-level* causal relationships (i.e., about the underlying causal structure) depend on detailed considerations about

the *token-level* causation giving rise to the observable evidence (which particular event actually caused which particular effect). There is often a very large number of possible ways that even a single causal hypothesis could have produced a particular pattern of observations. Therefore, in order to maintain rational beliefs about causal structure, the ideal reasoner considers all possible causal paths that could describe what actually happened given each possible structural hypothesis, summing up the individual likelihood of these possibilities to assess the overall likelihood of each "type level" structure hypothesis having produced the observations.

Following previous research using this model (Bramley et al., 2017), all possible actual paths of effect activations are enumerated consistent with a candidate structure $s \in \mathbb{S}$ producing the observed data[1]. Each generative path $\mathbf{z_g} \in \mathbf{z}$ should satisfy that each effect e' is not over-determined (i.e. has a single true cause), and each single cause g' does not produce its effect twice. Then, the probability of each path is calculated based on the delay between the effect and its supposed actual cause:

$$p(\mathbf{d}_\tau|\mathbf{z_g},\mathbf{w}) = \prod_{g',e' \in \mathbf{z_g}} p(t_{g'\to e'} = t_{g'e'}|\mathbf{z_g},\mathbf{w}) \tag{2}$$

Incorporating prevention Although the actual causal attribution enforces an exact one-to-one mapping from effect events to cause events and therefore guaranteed rational causal inference in the generative setting, it is cast as a diagnostic inference process (i.e. explaining the observed events under each candidate causal structure, Halpern, 2016), so does not directly capture potential surprising non-occurrences, here the consideration of the frequency of the target component's self-activation in the light of potentially *preventative* influences. Here, we expand the approach by adding two forward-thinking rules to the likelihood calculation. The probability of *preventative* causation is calculated based on the delay between the putative preventative control component's activation and its following subsequent target component activation. This follows the logic that the shorter the delay, the less likely it is that prevention assumption occurred:

$$p(\mathbf{d}_\tau|\mathbf{z_p},\mathbf{w}) = \prod_{p',e' \in \mathbf{z_p}} p(t_{p'\to e'} < t_{p'e'}|\mathbf{z_p},\mathbf{w}) \tag{3}$$

Another rational rule added here aims to represent the *reflective* thinking in actual causal attribution: for supposed causes that did not have their corresponding effect e in the path, we must attribute them as either occurring after the end of the clip *or* as prevented:

$$p(\mathbf{d}_\tau|\mathbf{z_r},\mathbf{w}) = \prod_{g',p' \in \mathbf{z_r}} [p(t_{g'\to e} > t_{end}|\mathbf{z_r},\mathbf{w}) + p(t_{g'\to e} < t_{g'p'} + t_{p'\to e}|\mathbf{z_r},\mathbf{w}) \cdot p(t_{g'\to e} \leqslant t_{end}|\mathbf{z_r},\mathbf{w})] \tag{4}$$

[1]Exhausively wherever feasible and approximated otherwise

Since there can only be one true generative actual path in the set of possible paths $\mathbb{Z}_s$, we then sum over all paths to get the likelihood of the data given a candidate model $s \in \mathbb{S}$:

$$p(\mathbf{d}_\tau|s,\mathbf{w}) = \sum_{\mathbf{z} \in \mathbb{Z}_s} \prod_{\mathbf{z_g},\mathbf{z_p},\mathbf{z_r} \in \mathbf{z}} p(\mathbf{d}_\tau|\mathbf{z_g},\mathbf{w}) \cdot p(\mathbf{d}_\tau|\mathbf{z_p},\mathbf{w}) \cdot p(\mathbf{d}_\tau|\mathbf{z_r},\mathbf{w}) \tag{5}$$

Experiment

Methods

Participants Sixty participants (26 female, aged 40 $\pm$ 13) were recruited via Amazon Mechanical Turk and were paid between \$1.00 and \$2.08 (\$1.78 $\pm$ 0.19) depending on their performance. The task took around 15 minutes[2][3].

Design We used the generative process described above to determine the timing of component activations in each clip, and the activations were displayed as the component lighting up yellow for 350 ms (see Figure 1c and 1d). The clips were created by sampling from the causal devices as follows. The target component activated every 5 $\pm$ 0.5 seconds by itself (gamma distributed with $\alpha = 100$, $\beta = 20$, Figure 2a). The control components activate every 5 $\pm$ 5 seconds ($\alpha = 1$, $\beta = 0.2$, which also belongs to memoryless exponential distribution, making them essentially randomly distributed in each clip). Each activation of a generative component then produced an extra activation of the target component after a delay of 1.5 $\pm$ 0.5 seconds ($\alpha = 9$, $\beta = 6$, Figure 2b). Each activation of the preventative component blocked all activations of the target component for 3 $\pm$ 0.5 seconds ($\alpha = 36$, $\beta = 12$, Figure 2c). The activation of control component was accompanied with a hand sign (Figure 1c) and participants were told that this implied control components were intervened on or caused by someone else, at random moments rather than following any informative pattern.

Eighteen 20-second clips were created (2 clips for each of the nine structures). Each clip began with one self-activation of the target component in order to help participants establish the base rate. We selected the clips to make sure that each control component was activated three times within 20-seconds and that no activation was masked by another in the clips.

Procedure Participants clicked a "Start" button to watch the clip in each trial, and then marked their answers for two connections during or after the clip by clicking the connection (Figure 1e). Each clip could only be played once. The order of 18 trials, as well as the click pattern (whether you would click once, twice or thrice to select generative, preventative or non-causal), and the vertical position of A and B components (above or below) were randomized independently between participants.

Prior to the inference task, participants were told about and trained on the timing of three types of connections as well

[2]Pre-registration available at https://osf.io/nrzpu/

[3]Material, data, analysis code available at https://osf.io/q8n72/

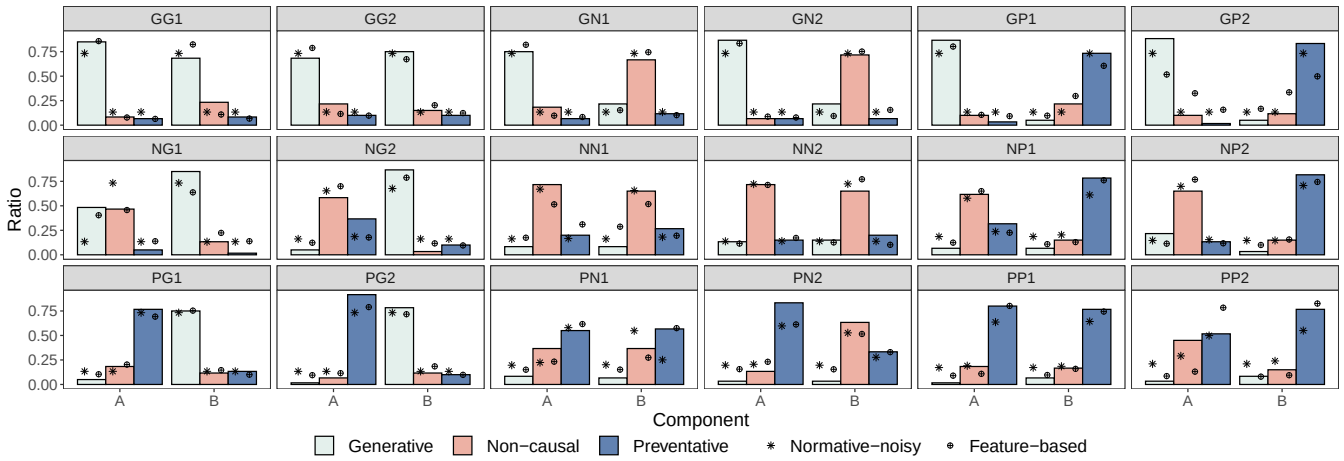

Figure 3: Participants' responses (bars) and results of model fitting (points). Titles indicate true structures (e.g., GN = A is generative; B is non-causal). Data were modelled at the device level and collapsed into connection level to simplify visualizations.

as the target component's self-activation. They also practiced how to respond, and completed one practice trial with a causal device that included one generative connection and one non-causal connection. Feedback was provided in the practice trials but not the test trials. Participants had to pass introduction check questions before starting the experiment. To properly incentivize judgments, a 3-cent bonus would be paid for each correctly identified connection during the main experiment in addition to the basic $1 payment.

Results

We did not find an effect of position or click pattern counterbalances, so we combined all participants' data in the following analyses.

Accuracy The accuracy for each participant per connection was 73 $\pm$ 17%, which is significantly above chance (33%), $t(59) = 17.58, p < .001$. Most participants (92%) performed above chance ($\geqslant$ 17 connections correct, binomial probability $< .05$). Accuracy was above chance levels for all three connection types taken separately (generative: 80 $\pm$ 20%, $t(59) = 17.77, p < .001$; non-causal: 62 $\pm$ 24%, $t(59) = 9.22, p < .001$; preventative: 76 $\pm$ 22%, $t(59) = 15.35, p < .001$, Figure 3). Most connections (97%) received above-chance correct answers ($\geqslant 26/60$ participants correct, binomial probability $< .05$).

The accuracy for each participant at the device level (1 = correct in both connections; 0 = otherwise) was 56 $\pm$ 22%, again substantially higher than chance (11%), $t(59) = 15.70, p < .001$. All but one participant showed an above chance device-level accuracy ($\geqslant$ 4 devices correct, binomial probability $< .05$). All device received above-chance correct answers ($\geqslant 11/60$ participants correct, binomial probability $< .05$).

The ideal Bayesian model's accuracy was significantly higher than participants' at the connection level (total: 97%, $t(59) = 10.90, p < .001$; generative: 100%, $t(59) = 7.64, p < .001$; non-causal: 94%, $t(59) = 10.21, p < .001$; preventative: 98%, $t(59) = 8.02, p < .001$) as well as the device level (95%, $t(59) = 10.53, p < .001$).

Choice Participants gave responses 33% times as generative, 31% as non-causal, 35% as preventative, so showed no overall bias towards particular edge types ($\chi^2(2) = 5.14, p > .05$). As shown in Figure 4, generative and preventative connections were mistaken for non-causal connections, and non-causal connections were mistaken for preventative or generative connections, while generative and preventative connections were less likely to be mistaken for one other, which means that participants were more certain in distinguishing between generation and prevention but sometimes confounded by non-causal connections.

In contrast, the normative ideal Bayesian model shows only a minor deficit in distinguishing between preventative and non-causal connections, but not between other connections. This deviation suggests that participants may rely on other heuristic strategies to solve the problem. In the following section, we describe one computationally tractable heuristic model and compare this with the normative model as well as human performance.

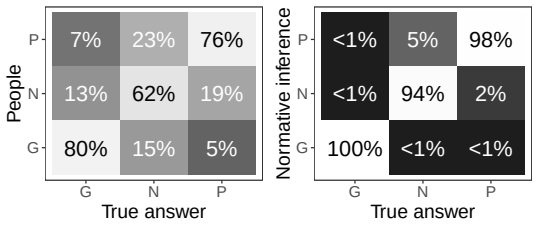

Figure 4: Overall choice patterns between different kinds of causal connections (participants vs. normative model).

Modelling heuristic inferences

Feature-based inference

One possible way learners might approximate Bayesian structure inference for the present learning problem is inspired by Ullman et al.'s (2018) Simulation-based approximations and Summary Statistics approach. This model assumes that people leverage cues based on rough and summarised statistics of different structure types to make inferences. The proposed cognitive process draws on (imagined) evidence under different causal structures (and/or historical data from known models) to learn statistical cues can be used to distinguish structure approximately but efficiently. Under this proposal, people select whatever hypothesis has the closest match in terms of these statistical cues to the observed data. Here we investigate two cues that people may apply to the current task:

1. Delay: the delay between control component's activation and its nearest target component's activation.

2. Count: the number of following target component's activations after one control component's activation and before any other activation of any control component.

These cues are certainly not exhaustive but they are relatively simple to track and reasonably discriminatory between the causal structures. For the delay cue, it is expected to see a short interval between its activation and the target component's next activation if the component is a generative component but a medium interval if there is no connection or a long interval if it is preventative (see Figure 5 for the probability density distribution). For the count cue, more than one effect activation is likely to follow the activation of generative component before the next control component event, which results from the existence of base rate activations, while most frequently, zero effect activation's will likely to follow the activation of preventative components before the next control component event (Figure 5). The former cue considers concrete delay information but ignores the possibility of different causal pathways, while the latter cue also ignores the exact temporal interval between events.

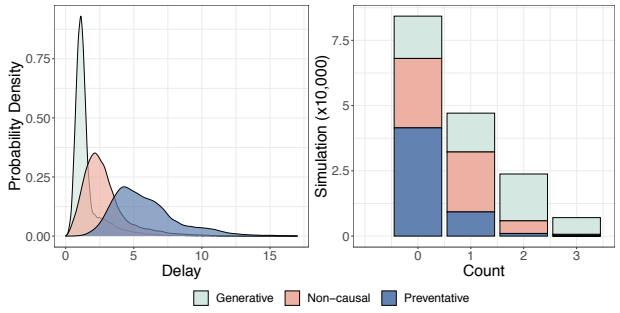

Figure 5: Empirical distributions of "delay" (left) and "count" (right) cues in the feature-based model.

Characteristic summary statistics for each structure hypothesis were constructed by simulating 20,000 twenty second sequences of point events from each structure type, and then calculating the empirical likelihood distribution for each intervention (A1, A2, A3, B1, B2, B3) in each structure. When observing a new control component's activation, the probability of each causal structure was estimated by the normalised likelihood of the summary statistic calculated on the observed data. For example, an initial intervention on A (i.e., A1) with an effect occurring 2 seconds later has a likelihood of e.g. $[.58,.40,.02]$ of having being produced by a generative, non-causal or preventative $A \rightarrow E$ connection respectively. Inherent to this heuristic approach is the radical simplifying assumption that the features of the evidence subsequent to each control component event are modular and independent, that is, ignoring that they also depend on the behaviour of the other causal components (Fernbach & Sloman, 2009). Thus, each connection was estimated independently and then combined to yield a probability for each causal structure. The two cues were .68 correlated in their structure predictions under the current trial set.

Model fitting with human data

We compared participants' choice distribution for each device to the behaviour of our normative and heuristic models as well as a random baseline. Excepting the random baseline, we assumed that participants chose their answer according to a softmax decision rule (Luce, 1959):

$$p(n) = \frac{\exp(\lambda \cdot v_n)}{\sum_{n' \in N} \exp(\lambda \cdot v_{n'})} \tag{6}$$

The "inverse temperature" parameter $\lambda \in [0, \infty]$ controls how consistent the participant is in selecting the answer with the largest v_n in choice n. Larger λ means that the participant's answer is better aligned with the model's answer with $\lambda = 0$ modeling random selection. For the combination of two cues in the feature-based model, we use two λ parameters to give weights to different cues.

Besides, we also included a noise parameter θ in our normative model to investigate whether people can follow the mechanism of normative inference but just have less precise temporal perception. It works by amplifying the original variance of 0.25 in gamma distributions for generative, preventative connections as well as self-activations into $0.25 \cdot \theta$, where $\theta \in \{0, 1, 2, \ldots, 8\}$. We called them normative-noisy models when $\theta > 1$. The accuracy of these models at the device level is summarised in the Table 1.

We fit each model to predict aggregate participant choices as well as to individual participants using BIC as our measure of fit. The results are shown in Table 1. The feature-based model that combines cues of "delay" and "count" best fit human aggregate results. 75% of individual participants were best fit by feature-based models included single or two features, 17% were best fit by normative-noisy models, and 8% were detected as responding randomly.

Table 1: Model accuracy and fitting results at device level.

Model	Accuracy	Parameters	BIC	N Best
NN:	83-95%	λ:2.67; θ:3	3378	10/60
FB:				**45/60**
delay	60%	λ:3.62	3431	(23)
count	43%	λ:5.49	3548	(12)
combine	43-60%	λ_d:2.36; λ_c:3.03	**3239**	(10)
RD:	11%		4768	5/60

Note: NN:Normative & Normative-noisy; FB:Feature-based; RD:Random. Model accuracy was calculated prior to the fitting of human data and under consideration of the noise (θ) parameter.

General Discussion

In this study, we investigated how people infer causal structures that contain generative and preventative connections on the basis of observing a handful of events occurring in continuous time. The fact that, in this setting, the effect has a base rate of spontaneous activation makes the likelihoods of observations non-deterministic (unexpected events could always be caused by the base rate) while the potential precense of additional generative influences complicates the evidence. Although our task is computationally challenging from the normative model perspective, performance was generally good, with judgments well above chance for preventative, as well as generative and non-causal connections. We thus provide empirical evidence that people can use real-time temporal information to learn causal structure that involves preventative as well as generative causation.

In terms of modelling, we for the first time introduced considerations of prevention into an "actual causal attribution" process model (Halpern, 2016). By exhaustively constructing possible actual causal paths given observed data, our normative model demonstrates that near-perfect performance is possible in this setting, at least given the correct delay assumptions and unlimited processing power. The normative and normative-noisy models had higher accuracy than the feature-based model, indicating that actual attribution, the top rung of Pearl's so-called "ladder of causation" (2018), is key for achieving benchmark levels of accuracy in this task. The inference and approximation we present in this paper are not restricted to the current paradigm but can be modified to handle a wide range of causal systems in future studies. Essentially, any system can be represented with a causal mechanism that produces point events over time can be inferred and reasoned about in this way.

Despite the accuracy win for the normative approach, human responses were better captured by our feature-based model that established and exploited two heuristic cues: delay and count. This model assumed that people tracked the delays between putative cause-effect activations and counted the effect events between putative cause activations, and made causal inference by comparing these observed patterns to the patterns characteristic to each edge hypothesis. This strategy used simulation to sidestep the complexity of exact cause-effect mapping and summarises the observations as cues to reduce the memory load during watching the evidence. The approach sacrifices precision in terms of actual causal attribution but may capture how people manage the information stream in real time caual induction settings, given the real time nature of the evidence presentation and their limited cognitive resources. The model also demonstrates one possible way that mental simulation could contribute to temporal causal reasoning, as an extension to current perspectives on mental simulation in physical reasoning (Gerstenberg et al., 2017; Ullman et al., 2018).

In sum, this paper investigated human causal structure induction from observation of real-time event patterns involving prevention as well as generation. People were capable of real-time causal structure induction in this setting and our modelling suggested they may achieve this via statistical cues such as average delays and counts that are much easier to establish than the exact generative model likelihoods. Every day we not only face surprising occurrences, but also wonder about things that would, should or could have happened. Preventative causation may be as pervasive as generative causation but better hidden, revealing itself only through violation of model based expectations. As such, an formal account of preventative inference is ripe for inclusion in an account of human causal reasoning.

References

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition*, *23*(4), 510–524.

Bramley, N. R., Gerstenberg, T., Mayrhofer, R., & Lagnado, D. A. (2018). Time in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1880.

Bramley, N. R., Mayrhofer, R., Gerstenberg, T., & Lagnado, D. A. (2017). Causal learning from interventions and dynamics in continuous time. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 150–155).

Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking & Reasoning*, *8*(4), 269–295.

Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking & Reasoning*, *12*(4), 353–378.

Carroll, C., & Cheng, P. (2009). Preventative scope in causation. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the cognitive science society* (pp. 833–838).

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of experimental psychology: Learning, memory, and cognition*, *35*(3), 678.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological science*, *28*(12), 1731–1744.

Greville, W. J., & Buehner, M. J. (2010). Temporal predictability facilitates causal learning. *Journal of Experimental Psychology: General*, *139*(4), 756–771.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334–384.

Halpern, J. Y. (2016). *Actual causation*. MIT Press.

Lucas, C. G., Holstein, K., & Pacer, M. (2015). Inferring causal structure and hidden causes from event sequences. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (p. 2937).

Luce, R. D. (1959). *Individual choice behavior*. Wiley.

Pacer, M., & Griffiths, T. L. (2012). Elements of a rational framework for continuous-time causal induction. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 833–838).

Pacer, M., & Griffiths, T. L. (2015). Upsetting the contingency table: Causal induction over sequences of point events. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1805–1810).

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic Books.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology*, *41*(2), 139–159.

Stephan, S., Mayrhofer, R., & Waldmann, M. R. (2018). Assessing singular causation: The role of causal latencies. In T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 1080–1085).

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive psychology*, *104*, 57–82.