Learning Hidden Causal Structure from Temporal Data

Simon Valentin
s.valentin@sms.ed.ac.uk
School of Informatics
University of Edinburgh

Neil R. Bramley
neil.bramley@ed.ac.uk
Psychology Department
University of Edinburgh

Christopher G. Lucas
clucas2@inf.ed.ac.uk
School of Informatics
University of Edinburgh

Abstract

Past research indicates that humans can infer hidden causes from covariational evidence, and readily use temporal information to infer relationships among events. Here we explore a setting in which people can attribute events to a common hidden cause or causal relationships among observed events, including causal cycles, purely on the basis of timing information. We present data from three behavioral experiments and extend previously proposed Bayesian models that makes use of order and delay information for causal structure learning. Our findings support the idea that people rely on the delays between events rather than order information alone. Meanwhile, deviations from our model predictions suggest that people have an inductive bias against common hidden causes and rely on heuristics to distinguish between causal structures, such as event overlaps, at least with the cover story considered in these experiments. Further, our data suggest that people have particularly flexible representations of cyclic relationships.

Keywords: causal; learning; temporal information; event cognition; Bayesian models; latent variables; particle filtering

Inferring causal relationships from observational data is a notoriously hard problem in machine learning but human learners have an impressive ability to do so, often making systematic judgments from sparse, noisy data despite their limited computational resources (Waldmann, Hagmayer, & Blaisdell, 2006). However, a ready ability to infer causal links between observations is only part of the picture. Frequently causes of observed phenomena cannot be observed directly, but rather must be induced to explain observed patterns.

Previous studies have shown that adults and even children as young as 10 months can use intervention and covariation information to learn about hidden causes (Kushnir, Gopnik, Lucas, & Schulz, 2010; Lucas, Holstein, & Kemp, 2014; Saxe, Tenenbaum, & Carey, 2005). In particular, Lucas et al. (2014) showed that adults can infer the presence of one or several hidden causes, as well as their functional forms on the basis of observed statistical contingencies. These studies focused on interventions and statistical contingencies, but other information, notably temporal order and delay, also informs people's causal inferences. The role of time in causal cognition has been studied extensively, with perceptual research going at least as far back as Michotte (1946). Indeed, findings indicate that human causal learning, is strongly driven by temporal considerations. For instance, people have been shown to make inferences that align with the temporal order of events, even if temporal information is at odds with covariation cues (Lagnado & Sloman, 2006; Rottman & Keil, 2012). Regarding more nuanced temporal information, it has been shown that longer delays between two events lead to weaker judgments of causality (Shanks, Pearson, & Dickinson, 1989), potentially because more events may have occurred in the meantime that could explain the effect (Lagnado & Speekenbrink, 2014). Meanwhile, people are also able to adapt their expectations to specific domains. For instance, we expect the delay between pressing the power button on a computer and seeing the device turn on to be short but between eating spoiled food and getting sick to be much longer.

Recent research has shown that people use event order to rule out incompatible causal structures (as effects cannot precede their causes), and delays as well as variability between events to shape more fine-grained judgments (Bramley, Gerstenberg, Mayrhofer, & Lagnado, 2018). However, little is known about when and whether people infer hidden causes on the basis of temporal data. In the present work, we approach this problem in the tradition of rational analysis (Anderson, 1991), comparing human judgments to the predictions of rational models with specific assumptions and inductive biases.

Our contributions are as follows. We reframe a previously proposed order model for causal structure learning to model common dynamic causal systems as finite state machines that may contain causal cycles and hidden causes. We further extend a previous model that incorporates delay information by using a variant of dynamic Bayesian networks (DBNs) as our state space representation, using particle filters for inference. This allows us to consider (hidden) structured stochastic point-processes as causal generative models, using a probabilistic state-space representation that permits online inference. We compare model predictions to human judgments in three experiments. Our experiments contain stimuli sampled from the generative models as well as matched conditions where we manipulate heuristic features – like event alternation or simultaneous event onsets – to provide strong signals about the underlying causal structure.

Approaches to modeling causation over time

Causal graphical models (CGMs, also known as causal Bayesian networks; Pearl, 1995) have become a dominant tool for causal inference both in data science and as a framework for modeling causal cognition. However, one of their major limitations is a lack of built-in semantics to represent temporal dynamics explicitly.

Bramley et al. (2018) proposed a simple Bayesian model that uses order information to distinguish between possible causal structures, by assigning equal probabilities to all event sequences that are consistent with a given causal structure (e.g., such that effects never precede their actual causes). This simple likelihood function naturally penalizes more flexible structures, since they are compatible with a wider range of data patterns. Bramley et al. (2018) also proposed a delay-based model, using exponential distributions to model base rates for (randomly occurring) independent variables and gamma distributions to model (more reliable) cause-effect delays. Three experiments and modeling suggested people use order information to rule out incompatible causal structure and among the compatible ones, prefer those that have more similar and reliable causal delays.

A Dynamic Bayesian Network (DBN; Dean & Kanazawa, 1989) provides another approach to encoding temporally dynamic causal relationships. A DBN is a CGM that can be "unrolled" over time to model the influence of variables going from one time point t to the next $t+1$. An important characteristic of DBNs is that they can be used to express and reason about systems containing cycles (since edges always travel forward in time, the unrolled graph remains acyclic which is a necessary property of a CGM). However, standard DBNs represent time using discrete steps, rather than as a continuous quantity, which is associated with conceptual and computational issues (Nodelman, Shelton, & Koller, 2002).

Continuous Time Bayesian Networks (CTBNs; Nodelman et al., 2002) extend DBNs to represent structured stochastic processes in continuous time, solving the problem of choosing a time scale. However, both standard DBNs and CTBNs implicitly assume that delays between events are memoryless, following exponential (or geometric) distributions (Murphy, 2012). This assumption means that the probability of an event does not depend on how much time has already elapsed. However, this does not hold with many real-world phenomena, such as the delay between pressing the power button on a computer and the device turning on, or on a longer time scale, incubation periods expressing the delay between exposure to a harmful agent and the first sign of symptoms. This assumption can be relaxed by positing additional states (Murphy, 2012), but the family of expressible distributions is still limited and the procedure is computationally costly. Another option is to introduce a "dwell-time" counter into each state (Murphy, 2012), but this approach does not readily extend to multivariate state spaces, which are considered here.

Normative framework for structure inference We now outline the rational analysis framework we assume. Prior beliefs about a set of causal structures S and their parameters θ are represented in the prior distributions $p(s)$ and $p(\theta \mid s)$, respectively. Given data $\mathcal{D}$, we update our belief by applying Bayes' theorem. That is, we obtain

$$p(s \mid \mathcal{D}) \propto p(s) \int p(\mathcal{D} \mid \theta, s) p(\theta \mid s) \, \mathrm{d}\theta. \qquad (1)$$

The integral in this expression, as well as the normalizing constant, are often intractable, necessitating the use of approximations. For all modeling considerations that follow, we distinguish X and Y from their (noisy) observations o_X and o_Y, which may occur after a delay. Common to all representations is the assumption that causes always produce their effects, though the delays can vary.

Order model

Our order model extends the order model proposed in Bramley et al. (2018). We formalize our observed events at time t as a tuple $(o_X^{(t)}, o_Y^{(t)})$, where $o_X^{(t)}$ and $o_Y^{(t)}$ are binary observed variables with 1 denoting an active state (light on) and 0 and *inactive* state (light off).

As videos require discretization into time frames, simultaneous event-onsets may occur. For modeling, we treat these events as point events, i.e. only the first frame is seen as active and the duration over multiple frames is not taken into account. Hence, we can have three legal states: $(o_X^{(t)} = 1, o_Y^{(t)} = 0)$, $(o_X^{(t)} = 0, o_Y^{(t)} = 1)$ and due to discretization $(o_X^{(t)} = 1, o_Y^{(t)} = 1)$[1]. We can think of the order model as looking at the onset-skeleton of a given event sequence, where the observed variables correspond to direct observations of the underlying system.

We construct generative models for each causal structure, compactly represented as probabilistic finite state machines (PFSMs; Vidal, Thollard, de la Higuera, Casacuberta, & Carrasco, 2005) (Figure 1). Depending on the causal structure, different state transitions are legal.

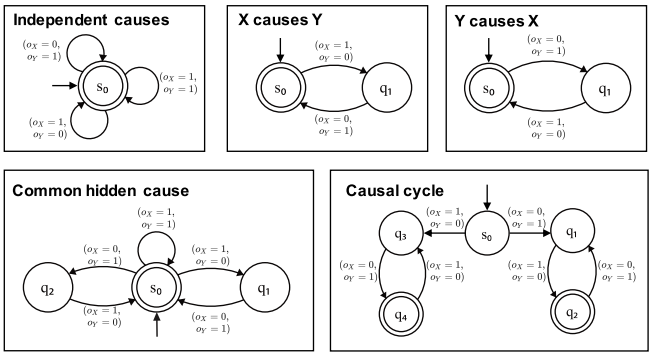

Figure 1: Order model formulated as finite state machines. Transition probabilities omitted for the PFSM representation for readability.

For independent causes, no restrictions on transitions between legal events apply. For X causes Y or Y causes X there is only one legal transition for each state, such that activations of X and Y always alternate. Whether X causes Y or Y causes

[1]Note that the state $(o_X^{(t)} = 0, o_Y^{(t)} = 0)$ is ignored under an order-only representation. A delay between two events can be thought of as observing "null-events" for that period of time. Taking this information into account would go beyond an order-only approach, since one could observe a varying number of such null-events, thereby introducing a notion of delay.

X is thus only determined by whichever variable activates first in an order-only scenario. A common hidden cause can produce a simultaneous onset of both observed variables or they activate in succession. As opposed to independent causes, however, we cannot have the same variable activate twice in succession. For causal cycles, either variable can start the observation sequence, after which the variables activate in turn.

In line with the order model in Bramley et al. (2018), we set the transition probability for a state k to $\frac{1}{\text{outdegree}(k)}$, where the outdegree is defined as the number of outgoing edges from a state. This has the effect that when two or more structures are compatible with a sequence of observations, the simplest one receives the highest posterior probability (given a uniform prior). The likelihood is here simply computed as the probability of a particular trace under the model. Identifiability experiments (see online supplement at `http://causalityandtime.com`) reveal that, X causes Y (and Y causes X) could always be identified, while independent causes and common hidden causes were recovered less often. Causal cycles cannot be recovered based on order information alone, with X causes Y (or Y causes X) being inferred instead, as they provide more parsimonious explanations of the data.

Delay model

We next describe our delay-sensitive model. Following Bramley et al. (2018), we assume gamma distributions represent beliefs about expected delays and their variability, while preventing backward-causation, as only forward-causation has support.

We represent causal structures as DBNs, where nodes denote the absolute time at which a particular event occurred. We look at the transition models for each structure from $t-1$ to t, with each step t tracking the occurrence of both variables, as presented in Figure 2. Edges represent parameterized gamma delays between occurrences of events, such that $\theta = [\mu, \sigma]^\top$ specifies a gamma distribution Gamma(μ, σ), reparameterized in terms of mean and standard deviation and expressed in seconds. For instance, if $X^{(t-1)}$ is connected to X_t by an edge, $X^{(t-1)}$ last occurred at some time $X^{(t-1)} = 11s$ and we sample a delay of $1.4s$, then the value of $X^{(t)}$ is $X^{(t)} = X^{(t-1)} + 1.4s = 12.4s$.

For the independent structure, there are no assumptions on dependencies between occurrences of X and Y. For X causes Y the occurrence of $X^{(t)}$ depends only on the previous occurrence of X (i.e. $X^{(t-1)}$), but is (conditionally) independent of the previous occurrence of Y. For the common hidden cause structure, the occurrences of X and Y depend on the last occurrence of H. For simplicity and comparability with other structures, we tie the parameters that describe the distribution of when X and Y occur after an occurrence of H. Lastly, the causal cycle resembles X causes Y (or Y causes X) but distinguishes itself by two characteristics: Expected delays between X and Y are symmetric, hence the parameters are tied, and both X or Y can start a sequence of observations.

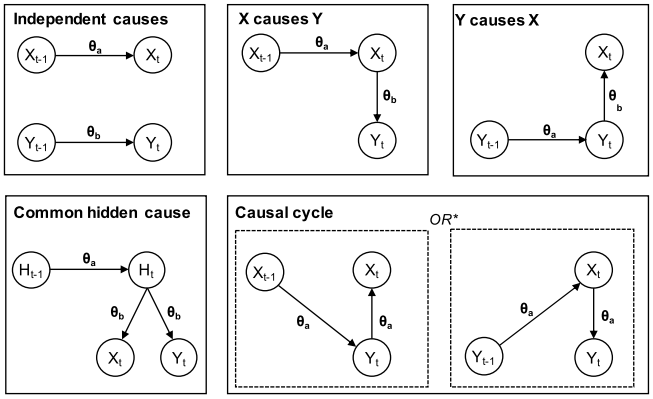

Figure 2: Delay model states as DBN models. θ denotes parameters $[\mu, \sigma^2]^\top$; OR* indicates that the cycle can either start with X or Y, depending on how the initial state was sampled.

We approach inference for this model as a tracking problem of the state of our system over time using a particle filter (also referred to as sequential Monte Carlo; Doucet & Johansen, 2011). Particle filters provide an approximate inference method that is well matched to this setting. Here, the posterior distribution is approximated via a set of *particles* (corresponding to a hidden state for a particular structure with a set parameterization) which are weighted by the probability of the observed data and occasionally resampled.

Our quantity of interest is $p(s \mid \mathbf{o}^{(1:t)})$, i.e. the marginal over parameters and trajectories of hidden states. Particles are initialized by sampling a structure $s \in S$ from a uniform multinomial distribution. The gamma distribution's mean has an Exp(0.5) maximum-entropy prior, and the variance has a Half-Cauchy(0.5) prior, following recommendations from Gelman (2006). Initial states are sampled from independent zero-mean Gaussians with Half-Cauchy(0.5) variances.

We assume that observations $(o_X^{(t)}, o_Y^{(t)})$, with $o_X^{(t)}, o_Y^{(t)} \in \mathbb{R}_{>0}$, are stochastic (noisy and/or delayed) reflections of underlying events $X^{(t)}$ and $Y^{(t)}$, respectively. Reflecting that stochastic relationship, $p(o_X^{(t)} \mid X^{(t)})$ is distributed according Gamma$(0.2, 0.01) + X^{(t)}$, likewise for $p(o_Y^{(t)} \mid Y^{(t)})$, loosely based on results from (Amano et al., 2006). This serves to provide a fairer comparison to how people perform the task and prevents single particles with large weights to dominate the particle filtering estimate. Particles were resampled using systematic resampling. In order to ensure the robustness of the approximations to the posterior, we assessed consistency across multiple runs with 10^5 particles in all experiments. Identifiability experiments (see online supplement at `http://causalityandtime.com`) showed that the generative model can most often be recovered successfully in all structures, while the independent structure appears most difficult to identify, due to its flexibility and consequent complexity penalization in the model comparison.

Model evaluation and comparison

We assess the extent to which people's individual judgments can be predicted using our order and delay models based on stimuli presented to participants. To this end, we compare the posterior distribution given by a participant on a particular video to a mixture of our order or delay model prediction. We use the root mean squared error (RMSE) between people's and model's judgements, following Bramley et al. (2018).

As our baseline, we use people's mean empirical judgments per structure, as a loose approximation to their priors and reflecting the performance of the best possible model (in terms of RMSE under a Gaussian error model) that can be achieved without predicting differences across conditions or participants. That is, our predictions are $\hat{y}_{bl}(s,j) = \mu_s$, where μ_s, $s \in S$ is the mean judgment per structure s computed over all $j \in J$ judgments contained in the training fold. Our alternative models are given by mixtures of the baseline and the order or delay model predictions. The prediction for a particular rating is $\hat{y}_\varepsilon(s,j) = \varepsilon\mu_s + (1-\varepsilon)\mathcal{M}(s,j)$, where $\mathcal{M}$ refers to either the order model $\mathcal{M}_o$ or the delay model $\mathcal{M}_d$. The free parameter $\varepsilon \in [0,1]$ determines the weight given to stimulus-specific order- or delay-model predictions. For fitting and assessing model generalizability to new conditions, we optimize the parameters on the judgments and predictions from $K-1$ of $K=12$ conditions and evaluate the performance on on the left-out condition. This provides a rigorous test of generalizability, as predictions are made on an unseen condition.

Experiments

We study the problem of inferring which causal structure generated a series of observations using a task where participants watched 35 second videos of two lights representing bacteria. The cover story involves that participants observe bioluminescent bacteria from different environments. All the bacteria glow dimly most of the time, but sometimes they briefly illuminate. The task was to learn whether there are relationships between the illumination of different bacteria, and what forms those relationships take.

Participants were first trained on how to interpret causal graphical models and their understanding checked with a pretest. It was instructed that additional hidden causes may also be present, and these might influence the behavior of the bacteria. After each video, participants gave posterior probability judgments for the causal structures (presented in Figure 3), ranging from 0 "impossible" to 100 "certain", which were normalized during pre-processing.

All data were collected from online experiments with Amazon's MTurk service for Experiments 1 and 2 and volunteer participants from Reddit for Experiment 3. Each participant was shown one video per condition in random order and conditions were counterbalanced for color and left-right differences.

Experiment 1

The data comprise judgments from $N = 38$ participants. It was hypothesized that forcing events to alternate would lead people to infer a causal cycle instead of independent events whenever delays between events were symmetric and a causal chain X–Y when X–Y delays were shorter than Y–X delays instead of a common hidden cause.

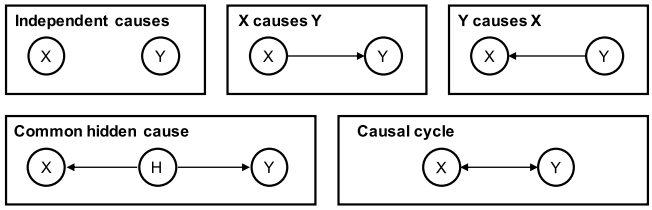

Figure 3: The causal structure hypotheses presented to participants.

Methods and materials The conditions were as follows:

[IO] Independent, original order: Bacteria illuminated independently and were presented in original order. Inter-illumination intervals were sampled from Gamma($\mu = 2s, \sigma = 0.63s$).

[IA] Independent, alternating: Events were generated as in the [IO] condition, but if the original sampled sequence contained two successive illuminations of the same bacterium, the identity of the second occurrence was switched, e.g., "*X X Y*" was changed to "*X Y X*".

[CO] Clustered, original order: A hidden cause was sampled as in the independent condition. X and Y illuminated after the onset of the hidden cause with delays sampled from Gamma$(0.5, 0.32)$.

[CA] Clustered, alternating: Events were generated as in the [CO] condition, but forced to alternate.

Results Figure 4 displays aggregate human, order and delay model judgments. Taken together, the [CO] condition was most surprising, as people gave similar probability judgments for multiple structures, descriptively slightly preferring causal cycles over a common hidden cause. We further investigate this in Experiment 3.

Experiment 2

The second experiment comprises a modified independent condition and three new generative processes. The data consist of judgments from $N = 21$ participants.

Methods and materials The conditions were as follows:

[I] Independent: Events for both bacteria were sampled independently. Delays between events had a larger expected duration and higher variance than the [IO] condition from Experiment 1 (Gamma$(2.67, 1.5)$).

[DL] Dependent, long CE delay: A cause event was sampled in as in the [I] condition. The delay between the end of the illumination of the cause bacterium and the onset of the effect bacterium was then sampled from Gamma$(1.1, 0.6)$.

[DS] Dependent, short CE delay: Same as the [DL] condition, but with a shorter cause-effect delay, distributed according to Gamma$(0.55, 0.3)$.

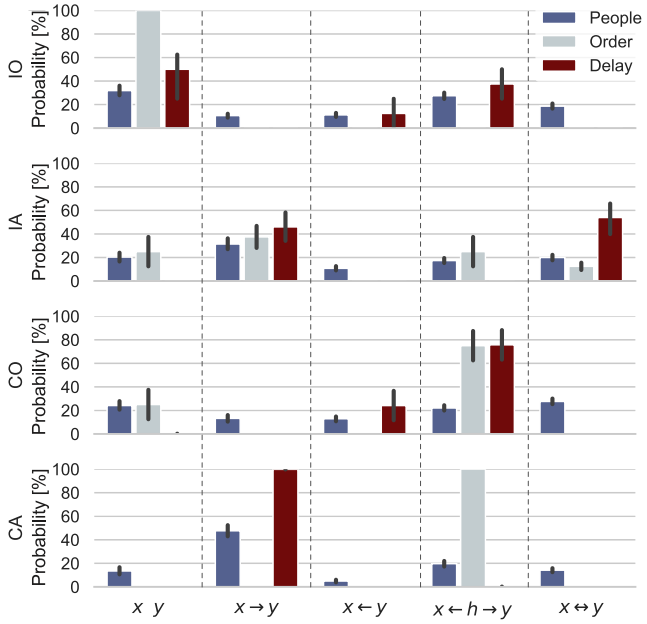

Figure 4: Experiment 1: Average people's, order model's and delay model's posterior probabilities for each structure. Error bars denote standard errors.

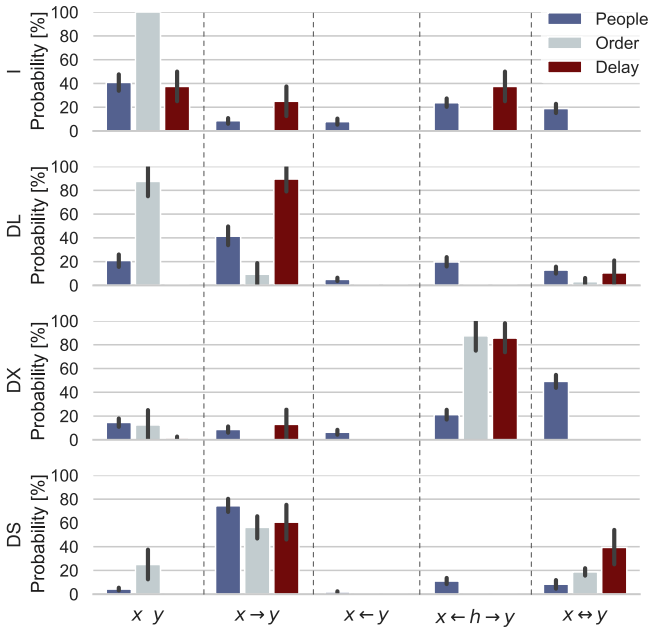

Figure 5: Experiment 2: Average people's, order model's and delay model's posterior probabilities for each structure. Error bars denote standard errors.

[DX] Dependent, short-delay, switching: Events were generated as in the [DS] condition, but the identity of the cause was swapped for each pair of observations with probability .50. This provides instances of [DS] event pairs for both X causes Y and Y causes X.

Results As presented in Figure 5, this experiment, with the [DX] condition, indicates that people may also infer cycles when delays are symmetric as well as when there are instances of causal chains in either direction, contrary to the assumptions included in our models. As in Experiment 1, we find that, descriptively, the order and delay models are more confident in their judgments than our participants.

Experiment 3

Building on Experiment 1, we ran a further experiment to assess the conditions under which people infer a common hidden cause. Experiment 3 includes additional manipulations of the stimuli that are ignored by the normative models, but provide information that people may use in a heuristic fashion to distinguish between structures. Here, we test for the effects of variance, event duration and onset-overlap. Total overlap (in terms of video frames) and variance are associated, as lowering the variance also means having more overlapping frames. In order to tease these factors apart, we manipulate event length between 4 and 8 frames, which changes the overlap but not the variance. For our normative models, this manipulation of frame length does not matter, since events are modeled as point events. An additional possible influence is given by whether events overlap at onset, which provides a different signal from overall frame-wise overlap. To assess the effect of onset-overlap on causal judgments, we deterministically set the number of onset overlaps to either zero or one. The data consist of judgments from $N = 21$ participants.

Methods and materials All conditions use common hidden cause structures, replicating and extending the [CO] condition from Experiment 1. Condition [CO] corresponds to the high-variance, short duration [HS] condition, with a standard deviation $\sigma = 0.32s$ for the delay from H to X and Y. For the low-variance short duration condition [LS], we halved the standard deviation to $\sigma = 0.16s$. The [HL] and [LL] conditions are given the same delay parameterizations as the short duration conditions, but events last for 8 instead of 4 frames. We present each participant with one video from each of the four conditions related to variance and event length in random order. Counterbalanced across participants, the first two videos then have no onset overlap, while video three and four have exactly one (or vice versa).

Results As shown in Figure 6, we replicate the findings from the [CO] condition in Experiment 1 and find that people still assign similar probabilities to multiple structures. In particular, people frequently report a causal cycle or independent cause rather than the generative common hidden cause structure. To tease apart the effects of variance, event duration and onset-overlap, we ran linear mixed models on people's judgments on the common hidden cause and the cyclic structure (for detailed results, see online supplement at `http://causalityandtime.com`). As the observations are nested within individuals, we included a random intercept for each participant.

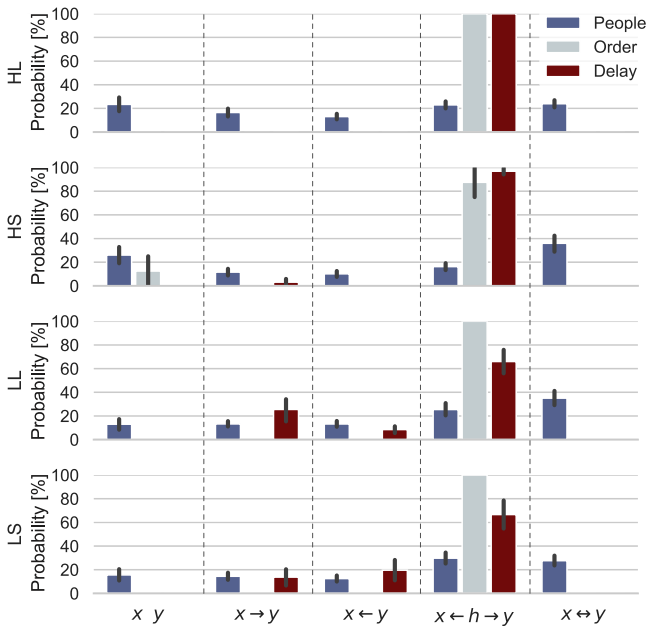

Figure 6: Experiment 3: Average people's, order model's and delay model's posterior probabilities for each structure. Error bars denote standard errors.

Regarding judgments for the common hidden cause structure, an effect of variance was found to be statistically significant. On average, participants gave $\Delta M = 7.15\%$ ($p = .031$, controlling for the effects of variance and duration) higher probability judgments for a common hidden cause when variance was low as compared to high. This indicates that people are sensitive to variance as an indicator for a common hidden cause. No statistically significant effects (at a level of $\alpha = .05$) were found for onset-overlaps or event duration.

For causal cycles, neither variance nor duration received weights that were significantly different from zero. However, there was a significant effect of onset overlaps, as people gave lower probability judgments for causal cycles when there was one onset-overlap versus zero, $\Delta M = -11.44\%$ ($p = .012$, controlling for the effects of variance and duration). This indicates that people are sensitive to onset-overlaps, and use this information to partially rule out cyclic relationships.

Global evaluation

As a more global assessment of our models on all three experiments, we evaluate the usefulness of the order and delay model inferences in predicting participants judgments on videos from an unseen condition. Table 1 presents aggregate RMSE values for the different models.

Fitted ε values indicate that only small weights are given to order or delay model predictions, though both outperform the baseline, with the delay mixture performing best. However, it should be kept in mind that the differences in RMSEs are comparably small.

	bl	bl + order	bl + delay
RMSE	22.07	21.95	**21.62**
SE	0.98	1.02	0.90
ε		0.91	0.90

Table 1: Aggregated results per model over all cross validation folds. bl denotes the baseline model; SE describes the standard error of the mean computed over conditions; Lowest RMSE in bold face; ε describes mixture weights.

Discussion

Our findings indicate that, overall, an order-only approach is useful but insufficient to account for how people infer hidden causal structure from temporal data. The delay model showed good identifiability results and helped most in predicting people's judgments. Overall, we find that people spread their probability judgments more broadly across structures and are more variable than our model predictions would suggest. One explanation for this finding is that people are conservative in their probability judgments (e.g. Bramley, Lagnado, & Speekenbrink, 2015; Edwards, 1968), thereby effectively "hedging their bets". That is, people may retain uncertainty in their judgments by assigning some probability even to structures that seem implausible based on the data.

What explains the differences between model predictions and people's judgments in particular conditions? We have seen in the [DX] condition from Experiment 2 that people may infer a causal cycle when seeing both instances of X causes Y and Y causes X (with short delays between the two variables). This is different from our assumption of symmetric delays, which was predicted in condition [CA] in Experiment 1. Hence, judgments for the [DX] condition point to considering more flexible state spaces than those provided by the DBNs considered in the present study. This could be achieved by introducing a probability into our model-dynamics that describes whether X or Y occurs first in the next time step, by sampling the connections of the corresponding edges. Sampling from this process would yield instances of X causes Y as well as Y causes X, as required.

The effects in Experiment 3 were surprising in that in in all conditions participants gave, on average, similar probability judgments to causal cycles and to common hidden causes, with a trend toward favoring the former. However, people appear to be sensitive to variance and do infer a common hidden cause in some cases. Further, our data suggest that people use the presence of onset-overlaps to distinguish between causal structures in a heuristic fashion.

In the present study, we only considered the cover story of the interaction between two bioluminescent bacteria. As people may have different a priori beliefs about hidden causes in other scenarios, we are currently running experiments where we investigate the influence of different cover stories on people's inductive biases towards hidden causal structure.

References

Amano, K., Goda, N., Nishida, S., Ejima, Y., Takeda, T., & Ohtani, Y. (2006, 4). Estimation of the Timing of Human Visual Perception from Magnetoencephalography. *Journal of Neuroscience*, *26*(15), 3981–3991.

Anderson, J. R. (1991, 9). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485.

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), 1880–1910.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 708.

Dean, T., & Kanazawa, K. (1989). A model for reasoning about real-time processes. *Computational Intelligence*.

Doucet, A., & Johansen, A. M. (2011). A Tutorial on Particle Filtering and Smoothing: Fifteen years later. In *The oxford handbook of nonlinear filtering* (pp. 656–704). Oxford University Press.

Edwards, W. (1968). Conservatism in human information processing. *Formal representation of human judgment*.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Kushnir, T., Gopnik, A., Lucas, C., & Schulz, L. (2010). Inferring hidden causal structure. *Cognitive Science*, *34*(1), 148–160.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(3), 451–460.

Lagnado, D. A., & Speekenbrink, M. (2014). The Influence of Delays in Real-Time Causal Learning. *The Open Psychology Journal*, *3*(1), 184–195.

Lucas, C. G., Holstein, K., & Kemp, C. (2014). Discovering hidden causes using statistical evidence. In *Proceedings of the annual meeting of the cognitive science society.*

Michotte, A. (1946). *La perception de la causalité*. Louvain: Publications Universitaire.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective (adaptive computation and machine learning series)*. MIT press.

Nodelman, U., Shelton, C. R., & Koller, D. (2002). Continuous Time Bayesian Networks. In *Proceedings of the eighteenth conference on uncertainty in artificial intelligence.* Morgan Kaufmann Publishers Inc.

Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika*, *82*(4), 669.

Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: Observations and interventions. *Cognitive Psychology*, *64*(1-2), 93–125.

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret Agents: Inferences About Hidden Causes by 10- and 12-Month-Old Infants. *Psychological Science*, *16*(12), 995–1001.

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 Section B*(2), 139–159.

Vidal, E., Thollard, F., de la Higuera, C., Casacuberta, F., & Carrasco, R. C. (2005). Probabilistic Finite-State Machines–Part I. *IEEE transactions on pattern analysis and machine intelligence*, *27*(7), 1013–1025.

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the Information Given: Causal Models in Learning and Reasoning. *Current Directions in Psychological Science*, *15*(6), 307.