Title

Casual Structure Learning with Continuous Variables in Continuous Time

Permalink

https://escholarship.org/uc/item/3k68c185

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

Davis, Zachary J
Bramley, Neil R
Rehder, Bob

Publication Date

2018

Causal Structure Learning with Continuous Variables in Continuous Time

Zachary J. Davis, Neil R. Bramley, Bob Rehder
{**zach.davis, neil.bramley, bob.rehder**} **@nyu.edu**
Department of Psychology, New York University
6 Washington Place, New York, NY, USA 10003 USA

Abstract

Interventions, time, and continuous-valued variables are all potentially powerful cues to causation. Furthermore, when observed over time, causal processes can contain feedback and oscillatory dynamics that make inference hard. We present a generative model and framework for causal inference over continuous variables in continuous time based on Ornstein-Uhlenbeck processes. Our generative model produces a stochastic sequence of evolving variable values that manifest many dynamical properties depending on the nature of the causal relationships, and a learner's interventions (manual changes to the values of variables during a trial). Our model is also invertible, allowing us to benchmark participant judgments against an optimal model. We find that when interacting with systems acting according to this formalism people directly compare relationships between individual variable pairs rather than considering the full space of possible models, in accordance with a local computations model of causal learning (e.g., Fernbach & Sloman, 2009). The formalism presented here provides researchers in causal cognition with a powerful framework for studying dynamic systems and presents opportunities for other areas in cognitive psychology such as control problems.

Keywords: causal learning, continuous time, continuous variables, intervention

Introduction

As observed by Jordan and Rumelhart (1992), "in many environments the effects of actions are not punctate and instantaneous, but rather linger on and mix with the effects of other actions". Acting effectively in the real world requires, in real time, learning and manipulating variables that are related by rich functional relationships. These functional relationships may take the form of complex systems such as feedback loops, cycles, and long chains of variables that result in unpredictable dynamics (Gleick, 2011; Rehder, 2017). For example, most people are naturally able to adapt the pace, volume, and pitch of their voice based on subtle social cues in conversation.

This paper outlines a formalism for inference in this kind of complex problem space. It then presents an experiment where people are allowed to intervene on an idealized system with unknown causal structure. Unlike previous research, learners can observe the effects of their interventions in real time, as moment-by-moment changes in the environment. We find evidence that people compare direct relationships between variables, rather than attempting to navigate the full space of possible causal models.

Past research

Research in causal cognition has focused primarily on causal relationships between binary variables and contexts in which temporal information is either unavailable or abstracted away.

From this paradigm we have learned much about how people are able to use covariational information to infer causation (e.g., Cheng, 1997, Griffiths & Tenenbaum, 2005), yet there are other factors highly relevant to causation that have not been combined in a single experiment. The factors we consider in this project are interventions, time, and continuous variables.

First, manipulations of causal systems, or "interventions", can be highly informative about causal structure (Pearl, 2000). Psychological work has established that people are somewhat adaptive in their intervention behavior, in ways predicted by information optimal norms (Bramley, Dayan, Griffiths, & Lagnado, 2017a; Coenen, Rehder, & Gureckis, 2015; Steyvers et al., 2003) although they also exhibit some biases and are subject to cognitive constraints.

Second, time has long been seen as a powerful cue for causation (Hume, 2000), especially in being informative about causal direction (i.e. people rule out backwards causation, assuming that effects cannot precede causes, Bramley, Gerstenberg, & Lagnado, 2014; Greville & Buehner, 2010). Recent work has demonstrated that people are capable of doing inference over causal models that explicitly encode a generative model for temporal delays between punctate events (Pacer & Griffiths, 2015; Bramley et al., 2014).

Finally, given the ubiquity of continuous valued variables, they have received surprisingly little attention in the study of causal cognition. One exception–Pacer & Griffiths (2011)–has shown that people are capable of learning individual cause-effect relationships between continuous variables. Soo and Rottman (2016) extended this work to discrete time, finding that people's causal strength judgments are based on the correlations between the *changes* in variables of time rather than the variables themselves.

While most aforementioned studies investigated one of these aspects in isolation, Bramley, Mayrhofer, Gerstenberg, and Lagnado (2017b) combined interventions and time in a study on people's learning of causal structure between components that exhibited occasional events. They found that people are sensitive to expected delays, especially when they also expect the true delays to be reliable.

Our interest in studying the simultaneous effects of interventions, time, and continuous variables has several motives. One is that continuous variables that vary in real time enable the study of a broader range of realistic causal relationships, including those that result in more interesting dynamics such as oscillations and feedback loops (see Figure 1 for examples). Another motive is to investigate the sorts of interventions that people choose when the effects of past interven-

tions are still reverberating through the system. In contrast, past studies of interventions on binary variables have involved systems that reach steady state. Finally, we ask how resource-limited learners process the vastly larger amount of information that flow from such systems.

Ornstein-Uhlenbeck Process

An Ornstein-Uhlenbeck (OU) process is a stationary Gauss-Markov process in continuous time that reverts to a stable mean (Uhlenbeck & Ornstein, 1930). It can be conceptualized as Brownian motion, except rather than being a martingale an OU process implements a "force" that biases expected value towards the mean of the distribution. Beyond having convenient mathematical properties that we will outline later, OU processes have been used to model a diverse array of phenomena, from physical systems (Lacko, 2012) to financial systems (Barndorff-Nielsen & Shepard, 2001) to work in perception (Vul et al., 2009) showing that people optimally allocate attention to track particles moving according to an OU process.

While the formalism that follows will precisely define our generative model, it may be helpful to provide a high-level overview of the model. The generative model is novel, to our knowledge, in that it has OU processes continuously interacting in a causal Markov graph. The Hookean spring is a helpful visualization for this type of causation. As the cause and effect diverge, the spring is stretched and more force is exerted on the effect to pull back towards the cause. When the cause and effect are near each other, the spring is compact and not much force is exerted. Causal networks are built simply by chaining together springs, such that in the network $X \rightarrow Y \rightarrow Z$, dragging X will first pull Y, which will subsequently pull Z.

Formalization

In an OU process, Δx_t—the change in x from time t to $t+1$—is defined as follows:

$$\Delta x_t = \theta[\mu - x_t] + N(0, \sigma) \tag{1}$$

where x_t is the value of the process at time t, μ is the mean that the process will revert to in asymptote, σ is the variance, and θ is a parameter greater than 0 that determines how sharply the process reverts to the mean.

OU processes can be generalized to track a variable that changes over time. Rather than defining a static μ that the process reverts to, the effect variable X can be defined as trending towards a linear function of the value of some cause Y:

$$\Delta x_t = \theta[\beta_{YX} \cdot y_t - x_t] + N(0, \sigma) \tag{2}$$

where β_{YX} is a real number multiplied by the value of the cause Y (subscripts denote $\beta_{\text{cause effect}}$). While the current project uses a linear function of the value of the cause, Equation 2 could of course be generalized to nonlinear functions.

We assume that β values and σ remain constant over time and value of either cause or effect (although these assumptions can be loosened, e.g. Barndorff-Nielsen & Shephard (2001)). Equation 2 defines the generative model we use for variable Y causing X.

Inference

When evaluating the hypothesis that Y causes X, the relevant parameter to be inferred is β_{YX}. If $\beta_{YX}=0$, Y and X are causally unrelated, if $\beta_{YX} > 0$, X tracks on to some positive multiple of Y's value, and if $\beta_{YX} < 0$, X tracks on to some negative multiple of Y's value. Because we assume that β stays constant throughout the trial, we can use multiple observations of the process over time to infer the value of β between cause and effect.

We assume that σ and θ are known to participants and do not need to be inferred[1]. For interventions, we implement Pearl's (2000) notion of intervention as graph surgery. If the effect variable X is intervened on, the likelihood of observing its state is 1 regardless of the βs or values of its parents. The remainder of this section, then, only deals with observations where the effect variable is not intervened on.

In essence, inference involves comparing the observed change in X to the expected change in X, where the expected change in X is defined as the mean of Equation 2:

$$\mathbb{E}(\Delta x_t) = \theta[\beta_{YX} \cdot y_t - x_t] \tag{3}$$

The likelihood of β_{YX} given some change in X and previous observation of Y, then, is a comparison of observed Δx_t and $\mathbb{E}(\Delta x_t)$ for a given β_{YX}:

$$P(\beta_{YX}|\Delta x_t, y_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\Delta x - \mathbb{E}(\Delta x_t))^2}{2\sigma^2}} \tag{4}$$

Equation 4 implements the intuitive idea of comparing the observed change in X to the expected change in X if it was tracking Y.

Multiple variables

For a theory of causal learning to be successful, it must of course be able to account for graphs with multiple variables. This is especially important for tasks with a time component, as timing information can be crucial for distinguishing between, say, a common cause and chain graph structure. The generative model for a single variable with multiple potential causes is:

$$\Delta x_t = \theta\left[\sum_{i=1}^{n} \beta_{Y^iX} \cdot y_t^i - x_t\right] + N(0, \sigma) \tag{5}$$

Simply put, the mean that the process reverts to is a sum of the other variables Y^i, multiplied by their β weights. This is another stipulation for the current project that could easily be

---

[1]Experimentally, we teach these parameters with familiarization trials that expose them to the values of θ and σ that obtain throughout the experiment.

loosened (e.g. in a continuous version of noisy-or, Pacer & Griffiths (2011)).

As in the single variable case, inference involves estimating $\boldsymbol{\beta}$ weights ($\boldsymbol{\beta}$ bolded because we must jointly estimate multiple β weights). The likelihood is the same as in the single variable case, with a slight modification to $\mathbb{E}(\Delta x_t)$:

$$\mathbb{E}(\Delta x_t) = \theta\Big[\sum_{i=1}^{n}\beta_{Y^i X}y_t^i - x_t\Big] \tag{6}$$

$$P(\boldsymbol{\beta}|\Delta x_t, y_t) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\Delta x - \mathbb{E}(\Delta x_t))^2}{2\sigma^2}} \tag{7}$$

Thus far, we have a formalism for estimating causal strength (β weights) but not causal structure. To infer structure, one must have a model for how causal relationships generate β weights (e.g. a "regular" causal relationship can only have positive β weights). The full process involves (1) defining a distribution of β weights for different types of causal connections, (2) sampling from these distributions, and (3) multiplying the likelihoods by prior belief in each causal relationship to yield a posterior over causal structures. In this paper we define three types of causal relationships: "regular" (β=1), "none" (β=0), and "inverted" (β=-1).

For each observed Δx, we jointly estimate the full space of beta values for possible edges. For example, for three variables there are six possible edges, $\boldsymbol{\beta} = \{\beta_{XY}, \beta_{XZ}, \beta_{YX}, \beta_{YZ}, \beta_{ZX}, \beta_{ZY}\}$.

Properties

OU processes can implement many intuitively appealing features of continuously varying causal relationships. For example, a negative β weight would correspond to a decrease in one variable driving up the value of another (e.g. decreasing interest rates is generally thought to increase inflation). Both positive and negative feedback loops are also naturally implemented as β weights of equal sign with absolute value greater than 1 or absolute value less than 1, respectively. Similarly, oscillations can be implemented with β weights of mismatched signs (such as 2 and -2). These feedback loops can be implemented between only two variables, or as part of a cyclic causal structure with potentially many variables.

Figure 1 demonstrates some of the dynamics that can be implemented simply by varying β parameters. Cells (A) and (B) demonstrate regular and inverse connections for a single cause of an effect. To show the dynamics, for both plots the mean that the cause trends to is 0 for the first 30 observations, and 100 for the next 70. This shows that changes in the effect follow changes in the cause. For cell (C) we simply initiated both variables at 0 (note that the values get so large that X and Y are indistinguishable in this plot). To show that negative feedback trends both variables towards 0, both variables were initiated at 100 in cell (D). For cell (E) we initiated both variables at 0. In cell (F) for the chain network $X \rightarrow Y \rightarrow Z$, the cause variable X spends 10 time points centered around 0, and is then intervened on to set its new value to 100. This was done to show that changes in Y precede changes in Z.
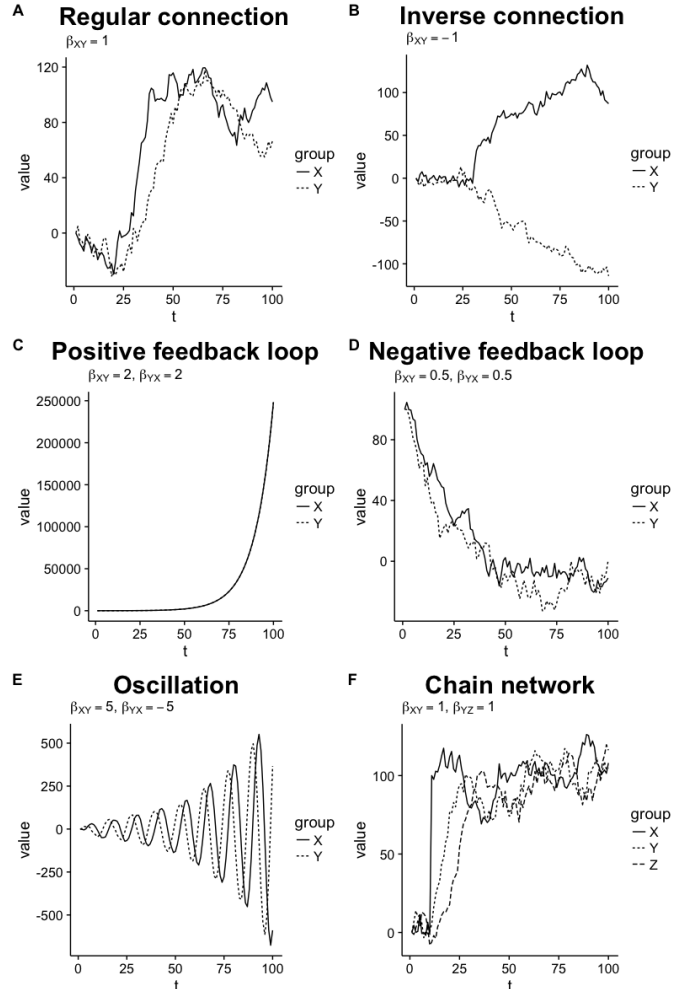

Figure 1: Examples of the dynamical phenomena resultant from varying β weights. **(A)** Regular connection, $\beta_{XY} = 1$. **(B)** Inverse connection, $\beta_{XY} = -1$. **(C)** Positive feedback loop, $\beta_{XY} = 2$, $\beta_{YX} = 2$. **(D)** Negative feedback loop, $\beta_{XY} = .5$, $\beta_{YX} = .5$. **(E)** Oscillation, $\beta_{XY} = 5$, $\beta_{YX} = -5$. **(F)** Chain network ($X \rightarrow Y \rightarrow Z$), $\beta_{XY} = 1$, $\beta_{YZ} = 1$.

Experiment: Causal Structure Learning

Method

Participants. 30 participants (13 female, mean age=37.5) were recruited from Amazon Mechanical Turk using the psiTurk framework (Gureckis et al., 2016), which has been shown to produce comparable results to lab experiments in cognitive science (Crump, McDonnell, & Gureckis, 2013). They were paid \$4 for approximately 30 minutes.

Materials and procedure. Each of the three variables was represented by a vertical slider that could be freely manipulated by clicking and dragging anywhere on the slider (see Figure 2)[2]. The top of the screen presented a timer counting

---

[2]For a demo, see https://zach-davis.github.io/publication/cvct/

down from 45 seconds, at which point the trial finished. Participants responded using six additional (3-state) sliders presented beneath the trial window, one for each potential causal relationship. Responses were constrained to be one of three options: 'Inverted', 'None', or 'Regular' (corresponding to βs = -1, 0, 1 respectively) and participants were pretrained on these terms in the instructions.
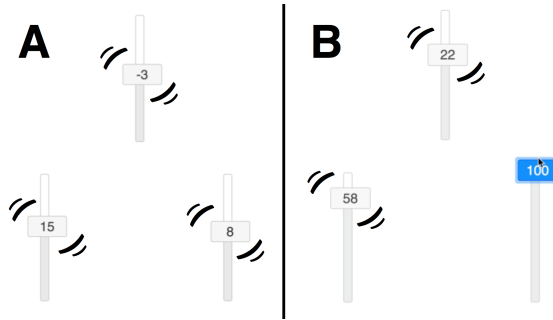

Figure 2: Sliders used by participants. (A) shows that the sliders all jitter if no interventions are made. (B) shows that the sliders do not jitter if intervened on.

Prior to the learning task, participants were shown four videos of an agent interacting with the structures to familiarize them with the interface. These videos informed participants of the underlying causal structure and demonstrated an agent interacting with the system. To prevent explicit instruction of any particular strategy, the videos displayed behaviors such as wobbling the intervened on variable, holding the variables at a constant level, and maxing out the intervened variable. They were shown examples of (1) a network with no causal connections, (2) a single "regular" causal relationship, (3) a single "inverse" causal relationship, and (4) a chain relationship with one "regular" and one "inverse" connection. Participants were not allowed to proceed to the test phase until they had watched all four videos. Participants were then presented with a five question comprehension check to ensure that they understood the task. Questions established that participants understood the duration of each trial, the difference between a regular and inverted connection, that there can be more than one connection per network, and that they must provide a response for all possible connections. Participants could not continue without answering all questions correctly. The parameters used during training and the learning task were $\theta = .1$, $\sigma = 5$, and βs were either -1, 0, or 1.

In the learning task, participants initiated the trial by pressing the "Start" button and the sliders started jittering according to an OU process, with unknown β weights driving the movement (there were no causes outside the network). The values of the sliders updated every 100ms. At any time, participants were free to click, hold, or drag anywhere on the slider. While the mouse was pressed down it fully determined the value of the slider, and once the mouse click was released the process would continue from that point according to the

OU process. The sliders were constrained to be between -100 and 100, and the buttons on the slider presented a rounded integer value in addition to moving up and down. Participants were instructed to make their judgments at the end of the trial, but were also free to make their judgments at any point after initiating the trial (see Figure 3). No feedback was provided at any point. After seeing a total of 25 causal structures, participants completed a brief post-test questionnaire.
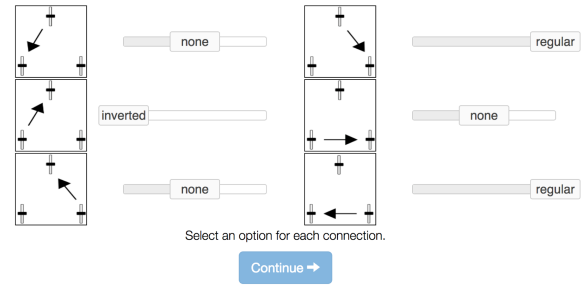

Figure 3: Judgment options for participants. Participants were presented with a trinary choice between "inverted", "none", and "regular".

Results

As a baseline, participants were significantly better than the chance probability of .33 for identifying causal relationships (M=.82, SD=.15); t(29)=17.36, p<.001. They were also above chance (.001) in selecting the correct causal model (M=.44, SD=.22); t(29)=10.81, p<.001.

Models. To analyze behavior in this task we will compare two competing accounts of causal structure learning. The first is that people are roughly normative in their structure learning behavior. Such an account would imply, for the current task, that people update their posteriors over all hypotheses at every time point. Importantly, learners acting in accordance with the normative model have to take into account dependencies between the connections, judging whether associations observed between variables can be explained by other graphs. For example, in the network $X \rightarrow Y \rightarrow Z$, the X–Z pair have many of the hallmarks of a direct causal relationship. They are correlated, changes in X precede changes in Z, and intervening on X later affects Z's value (but not vice versa). Only by understanding that Y mediates the relationship between X and Z (for example, through timing information) could a learner correctly uncover the underlying structure and avoid improperly concluding that X causes Z. There has been much work suggesting that adults and children are capable learners of causal structures and act roughly in accordance with the normative model, at least in sufficiently simple scenarios (Gopnik, Glymour, Schulz, Kushnir, & Danks, 2004; Griffiths & Tenenbaum, 2009).

We compare the normative model to a "local computations" (LC) model that has been advocated as a general-purpose account of causal learning behavior. The LC model

proposes that causal learning is structurally local, meaning that learners evaluate individual causal links without consideration of the entire graph. In the chain example, then, learners would not infer that the illusory causation between X and Z was mediated by Y, and draw a direct link between those two as well as the links $X \rightarrow Y$ and $Y \rightarrow Z$. A wide array of approaches to modeling causal structure learning have provided converging evidence for the LC model. Fernbach and Sloman (2009) introduced the local computations model, finding that it correctly predicted an overabundance of causal links and order effects in learning. Bramley et al. (2017a) extended this work to intervention decision-making, finding that people consider small changes to their currently held best hypothesis in both structure judgments and intervention decisions. Together, these studies suggest that people act in accordance with the LC model, testing and evaluating individual causal links rather than updating the full posterior space.

Model Formalism. Recall that calculating the likelihood of some β involves comparing observed and expected Δx:

$$P(\boldsymbol{\beta}|\Delta x_t, y_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\Delta x - \mathbb{E}(\Delta x_t))^2}{2\sigma^2}}$$

The normative and LC models both use this likelihood, but have different $\mathbb{E}(\Delta x)$. Because the LC model ignores the contribution of other variables, it estimates a single β at a time:

$$\mathbb{E}(\Delta x_t) = \theta[\beta_{YX} \cdot y_t - x_t]$$

In contrast, the normative model jointly estimates the contribution of multiple variables:

$$\mathbb{E}(\Delta x_t) = \theta\left[\sum_{i=1}^{n} \beta_{Y^iX} y_t^i - x_t\right]$$

As mentioned, participants were exposed to familiarization trials that showed that β values could be -1, 0, or 1. However, because this knowledge could only be approximate, we instead model them as believing that a "regular" or "inverse" connection is consistent with a range of βs. In particular, for "regular" connections we sample 1,000 β values from a distribution of positive values centered at 1 ($\Gamma(k = 5, \theta = 5)$), for "inverse" we do the same but negate the values, and for no connection we only sample β=0. Furthermore, we assume that participants have perceptual noise and therefore set the value of σ to 7, a higher value than the veridical 5. Note that neither of these decisions affect the forthcoming conclusions.

Model Results. We analyze results by comparing people's performance to the normative and local computations (LC) models. For both the normative and LC models, we calculate the posteriors for each of the 3^6 hypotheses that constitute every combination of the "inverted", "none", and "regular" judgments that participants made.

To evaluate the models, we performed a recovery experiment by feeding them the slider values for each trial for each participant, and evaluating the MAP estimate of each model

against the true generative network for that trial. As may be expected, the MAP estimate of the normative model recovered a higher proportion of the structures (.98) than the MAP estimate for the LC model (.58).

The first measure we use is to have the models perform the same task as participants (report the most likely structure, i.e. the MAP estimate), and compare whether participants and models draw the same links. In this coarse measure, the models were roughly equal (82% of links for the normative model, 80% for LC). The models pull apart, however, on the more sensitive quantitative measure of the log posteriors of participant judgments, given slider values, for each model. Of 30 participants, 10 were best fit by the normative model and 19 were best fit by the local computations model (1 participant could not be estimated due to errors in data collection).

The better quantitative fits of the LC model were strengthened by two qualitative predictions that distinguish it from the global model. The first qualitative prediction is an over-abundance of causal links (Fernbach & Sloman, 2009). Eighty-two percent (SD=.17) of the errors that participants made involved adding extraneous causal links, significantly greater than chance[3] (.59); t(29)=7.33, p<.001. The second qualitative prediction of the LC model as defined in this paper is an inability to distinguish between direct and indirect causes (e.g. in the network $X \rightarrow Y \rightarrow Z$, incorrectly also judging $X \rightarrow Z$). While in general participants were correct for eighty two percent of connections, they were dramatically worse for indirect effects (M=.16, SD=.21), performing below chance (.33); t(29)=-4.48, p<.001.

Discussion

This paper introduced a generative model of causal systems relating continuous variables in continuous time that exhibits familiar real-world dynamic phenomena such as feedback loops, and periodic oscillation. We explored active inference in these systems, allowing participants to freely manipulate the variables over a continuous trial window and comparing them to both a globally, and locally efficient inference model. We found that active learners who were free to interact with the system were better modeled as considering connections between each variable in isolation than updating beliefs over full graphs. Supporting the better quantitative fits of the local computations model, it also predicted two qualitative effects. Participants were much more likely to add extraneous causal links than remove existing ones, a surprising finding given the evidence for a bias for sparse causal models (Lu et al., 2008). The local computations model also predicted that participants would inappropriately draw connections between indirect effects. The evidence in this paper for the local computations model connects with a rich history in causal learning that suggests that people are limited in their ability to represent a full hypothesis space of possible models and

---

[3]For the structures used in this experiment, a hypothetical subject who responded "inverse", "none", and "positive" with equal probability would erroneously add a causal link 59% of the time.

instead consider narrow hypotheses and make alterations to a currently held model (Bramley et al., 2017a; Quine, Churchland, & Føllesdal, 2013). Further analysis and experimentation will be necessary to better understand the mechanisms and representations used by people in this new class of dynamical systems.

There are a number of limitations to the current project that could be addressed with further experiments. For one, we heuristically incorporated perceptual noise and assumed that people can track all three variables in real time. Extensions to the current project would be to fit perceptual noise or to account for the possibility that people cannot attend to all three variables simultaneously[4]. Both of these problems will likely become larger issues when more variables are added. Additionally, the presented analyses in this paper do not address intervention decision-making, a critical component of the active learning of causal structure. Future analyses would naturally involve, as a benchmark to compare against humans, models for selecting actions that maximize expected information gain.

Every day we must learn from and interact with systems that shift over time. Causal cycles, feedback loops, oscillations, inverse relationships, and all manner of dynamic properties are as common as they are difficult to learn. We hope that the formalism laid out in this paper will be helpful in studying the mechanisms for learning and action in this important class of problems.

References

Barndorff-Nielsen, O. E., & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 167-241.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36(2), 209-231.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017a). Formalizing Neuraths ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3), 301-338.

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 Meeting of the Cognitive Science Society* (pp. 236242).

Bramley, N. R., Mayrhofer, R., Gerstenberg, T., & Lagnado, D. A. (2017b). Causal learning from interventions and dynamics in continuous time. *In Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. London, UK: Cognitive Science Society.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367405

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive psychology*, 79, 102-133.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, 8(3), e57410.

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

Gleick, J. (2011). *Chaos: Making a new science (Enhanced edition)*. Open Road Media.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological review*, 111(1), 3.

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.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological review*, 116(4), 661.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior research methods*, 48(3), 829-842.

Hume, D. (2000). *An enquiry concerning human understanding: A critical edition* (Vol. 3). Oxford University Press.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive science, 16*(3), 307-354.

Lacko, V. (2012). Planning of experiments for a nonautonomous Ornstein-Uhlenbeck process. *Tatra Mountains Mathematical Publications*, 51(1), 101-113.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological review*, 115(4), 955.

Pacer, M. D., & Griffiths, T. L. (2011, December). A rational model of causal induction with continuous causes. In *Proceedings of the 24th International Conference on Neural Information Processing Systems* (pp. 2384-2392). Curran Associates Inc.

Pacer, M., & Griffiths, T. L. (2015). Upsetting the contingency table: Causal induction over sequences of point events. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press (2nd edition).

Rehder, B. (2017). Reasoning with causal cycles. *Cognitive science, 41*(S5), 944-1002.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive science, 27*(3), 453-489.

Quine, W. V. O., Churchland, P. S., & Føllesdal, D. (2013). *Word and object*. MIT press.

Soo, K. W., & Rottman, B. M. (2016). Causal Learning With Continuous Variables Over Time. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.*

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical review*, 36(5), 823.

Vul, E., Alvarez, G., Tenenbaum, J. B., & Black, M. J. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In *Advances in neural information processing systems* (pp. 1955-1963).

---

[4]Although Vul et al.'s (2009) finding that people optimally allocate attention to particles moving according to an OU process may ameliorate the latter concern