The order of things: Inferring causal structure from temporal patterns

Neil R. Bramley[1] (neil.bramley.10@ucl.ac.uk), **Tobias Gerstenberg**[2] (tger@mit.edu),
David A. Lagnado[1] (d.lagnado@ucl.ac.uk)
[1]Department of Cognitive Perceptual & Brain Sciences, UCL, 26 Bedford Way, London, WC1H 0DS, UK
[2]Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA

Abstract

The timing and order in which a set of events occur strongly influences whether people judge them to be causally related. But what do people think particular temporal patterns of events tell them about causal structure? And how do they integrate multiple pieces of temporal evidence? We present a behavioral experiment that explores human causal structure induction from multiple temporal patterns of observations. We compare two simple Bayesian models that make no assumptions about delay lengths, assume that causes must precede their effects but differ in whether they assume simultaneous events can also be causally connected. We find that participants' judgments are in line with the model that rules out simultaneous causation. Variants of this model that assume people update their beliefs conservatively provide a close fit to participants' judgments. We discuss possible psychological bases for this conservative belief updating and how we plan to further explore how people learn about causal structure from time.

Keywords: causal learning; sequential learning; structure; Bayesian modeling; conservatism; time; memory; belief updating

Introduction

Hume's (1748/1975) claim that people infer causal connections when they find temporal precedence, contiguity and constant conjunction has largely been embraced by psychology. Associative learning theories predict that, ceteris paribus, the closer in time two events occurred, the more likely people are to believe that they are causally related (Shanks & Dickinson, 1987). However, much recent work has shown that for many real world scenarios, people's causal judgments are influenced by their expectations about delay length (Buehner & May, 2004; Schlottmann, 1999), and delay variability (Greville & Buehner, 2010) such that shorter-than-expected delays can also reduce causal judgments. On the other hand, a related line of work suggests that consistency of temporal order with a causal structure (over and above specific delay length), may be an even more important factor in how people induce causal structure (Lagnado & Sloman, 2002, 2004; Rottman & Keil, 2012). People appear to draw causal conclusions based on temporal order even when the mechanisms underlying the causal system are completely unknown, or when temporal order contradicts other sources of information such as covariation and the outcomes of interventions (Lagnado & Sloman, 2006). Several recent studies also suggest that people are reluctant to endorse causal connections between events which appear to occur at the same time (Burns & McCormack, 2009), even when the causal mechanism is plausibly instantaneous (Lagnado & Sloman, 2006; McCormack, Bramley, Frosch, Patrick, & Lagnado, under review; McCormack, Frosch, Patrick, & Lagnado, under review).

These findings suggest that causal inference and event timing are tightly coupled (Lagnado & Sloman, 2006; Rottman & Keil, 2012; Sloman, 2005), with causal inference from temporal information appearing to be more automatic (Michotte, 1946) and more developmentally basic (McCormack, Frosch, et al., under review) than other modes of causal learning. However, to date there has been little work on the role of temporal order[1].

The learning problem

Here we explore the general problem of how people induce causal structure from temporal patterns of activation. We investigate whether people make a default assumption that causes must precede their effects, or merely a weaker assumption that causes either precede or happen at the same time as their effects. We focus on identification of the causal structure of a simple system with two candidate cause components A and B, and a single effect component E (Figure 2). To keep the problem space manageable, we restrict the systems to binary (active/inactive) components with causal relationships that are generative and deterministic and where there are no spontaneous component activations. However, delays between causes and their effects are variable, such that the same causal structure can generate more than a single type of temporal activation pattern. We also restrict the evidence to temporal patterns in which all components activate. This means that people cannot rely on contingency information and have to base their causal judgments on temporal order information only (Figure 3 a).

Baseline models

In order to formalize the idea that people expect causes to precede their effects in a Bayesian framework, we created likelihood functions for the seven causal structures in the problem space. We assumed that the probability of seeing a particular temporal pattern of activations given a causal structure is $1/N$, where N is the number of distinct temporal orderings consistent with that structure (Figure 3 a). For example, pattern 4 in which component A activates before component E and B, is consistent with structure IV ($B \leftarrow A \rightarrow E$) but inconsistent with structure V ($A \leftarrow B \rightarrow E$) because it is impossible in this common cause structure for A to activate before B. This approach is simple because it makes no assumptions about the exact length of the time delays between causes and effects but only considers the qualitative ordering in which

[1]An exception is Pacer and Griffiths (2012), but their work focused on induction of connections between continuously varying variables, while we will focus here on sequences of point events.

the different events occur. This means that, for example, the common cause structure IV ($E \leftarrow A \rightarrow B$) is consistent with the time sequences where A activates first followed by either B then E (pattern 2), E then B (pattern 4) or E and B at the same time (pattern 3). Whether this structure is also consistent with A and B occurring at the same time followed by E (pattern 1), depends on whether or not one assumes that causes and effects can happen simultaneously. We formalize two baseline models which differ in terms of whether or not they rule out the simultaneous activation of causes and effects (cf. Figure 3b and c).

Sequential belief updating

Real world causal learning typically takes place incrementally, often over a whole lifetime, with causal beliefs evolving as new evidence is observed. We are therefore interested in exploring how people update their beliefs as they observe a causal system exhibiting different temporal patterns over multiple occasions.

Our baseline Bayesian learning models yield predictions about how people update their beliefs with increasing evidence, by maintaining a probability distribution over possible structures $\mathcal{S}$, and updating this distribution using Bayes theorem and the likelihood function $p(\mathcal{T}|\mathcal{S})$ for each temporal pattern $\mathcal{T}$ that is observed.

$$p(\mathcal{S}|\mathcal{T}) \propto p(\mathcal{T}|\mathcal{S}) \cdot p(\mathcal{S}) \quad (1)$$

Starting from a flat prior, representing complete ignorance about the structure of a system, we found that our two baseline models make rich and distinct predictions about the posterior distribution over causal structures given different sequences of temporal activation patterns (Figure 4).

Research has generally found that what people learn from sequentially presented information is dependent on the order with which information is presented (Hogarth & Einhorn, 1992). Existing models, such as Hogarth and Einhorn's (1992) belief adjustment model, mimic these phenomena and provide workable approximations to normative judgments by positing simple sequential updating strategies. For instance, by repeatedly updating a point estimate about a particular hypothesis to part way between a current estimate and each new data point, Hogarth and Einhorn's model explains primacy effects in point estimation. Bayesian models meanwhile, typically do not predict order effects due to the time-independent nature of Bayesian belief updating, although they can be captured by assuming some form of approximation or factorization (e.g. Fernbach & Sloman, 2009; Sanborn & Silva, 2013). Sticking with the purely Bayesian framework though, there are several ways of capturing biases and order effects in people's belief updates. Recency effects, in which people's beliefs are more strongly affected by later pieces of evidence, can be modeled via the addition of noise to the learner's posterior between learning instances (Bramley, Lagnado, & Speekenbrink, submitted; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). We can also model conservatism,

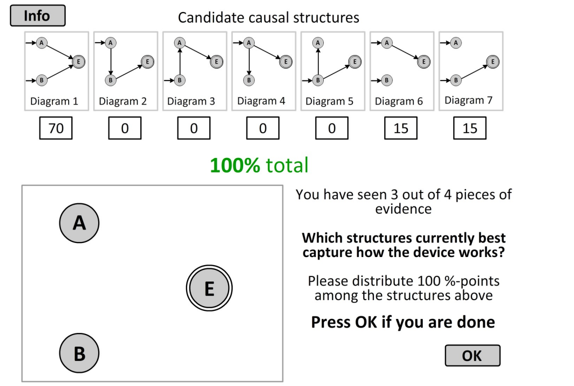


Figure 1: The experiment interface. Clips are shown in the bottom left panel.

e.g. sticky or slower than expected belief updating, by adding noise to the learner's likelihood function (Edwards, 1968).

In the current experiment, we will explore biases in people's sequential causal belief updates relative to our baseline models.

Overview of experiments

The task

To test whether one of our baseline models provides a good description of what people infer from temporal information, we designed an online task in which participants observe a causal device exhibiting several patterns of activation, and then make judgments about how they think the components of that device are causally connected. Evidence about each device was presented in the form of short movie clips. Each clip simply showed the three components, A, B, and E, which were represented by circles and arranged in a triangle (Figure 1). During each clip, all three components activated by turning from white to gray. The activated components remained gray until the end of the clip. We manipulated the order in which components activated during each clip. Each clip was instantiated one of the temporal patterns shown in Figure 3a. For a single device, participants would see a selection of activation patterns and be asked to make multiple judgments about which of the candidate causal structures captures how the device works.

We opted for a relatively abstract paradigm, in which participants are not told anything about the type of causal device or system they are identifying. This allows us to sidestep, as far as possible, complications due to the expectations people may have about the prior probability of different causal structures, delay lengths and delay variability. By keeping the task abstract we were able to focus on identifying default assumptions people make when they are facing a relatively new or unknown system.

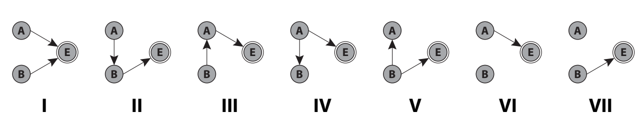


Figure 2: Possible causal structures.

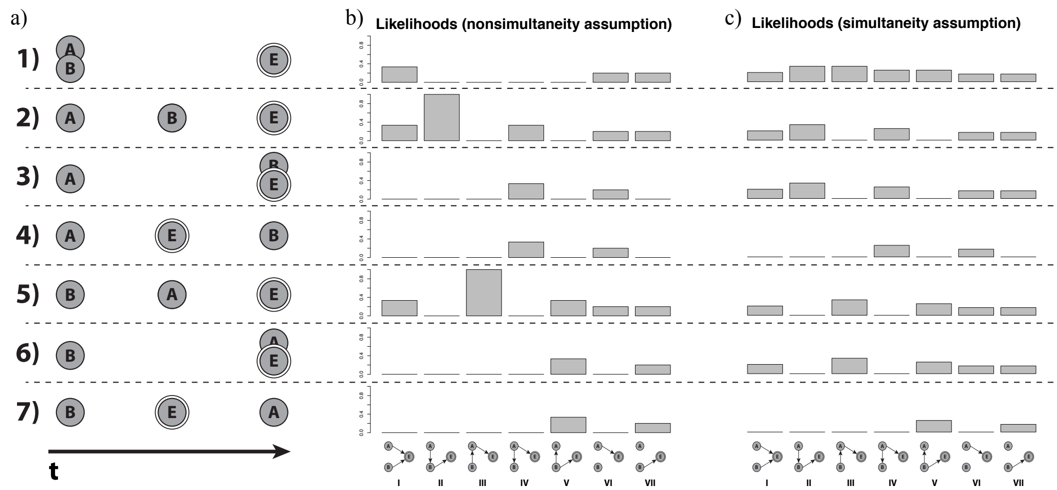


Figure 3: a) Seven possible temporal patterns of three events A, B, and E. Likelihood functions for the temporal patterns given the seven different causal structures with non-simultaneity assumption (b) or simultaneity assumption (c).

Possible causal structures

As discussed in the introduction, we restricted the space of possible causal structures to seven (Figure 2), each with two candidate causes A and B and an effect E. Participants were instructed that any parentless components in a causal structure diagram were caused by other unobserved components but that none of the other components would activate without being caused to by another component. This was done to minimize participants' expectations about when or why parent components of the device might activate. Participants were informed that structure I is conjunctive, meaning that both A and B must activate in order for E to occur.

Eliciting judgments

In order to have a fine-grained measure of people's beliefs, we asked participants to distribute 100 percentage points over the seven candidate causal structures, such that each value indicated their belief that the given structure captures how the device operates. They could not move on unless their indicated answers summed to 100%. Based on these percentage points, we get a subjective probability distribution over the seven structures from each participant, allowing us to directly compare their responses with the predictions of our models.

Variation in delays

We instructed participants that the delay between the activation of a cause and the activation of its effect was variable even though the causal links themselves were reliable. Thus, it was possible that the same device would exhibit different patterns of activation on different occasions. For example, structure IV ($B \leftarrow A \rightarrow E$) was compatible with both the $A-B-E$ temporal pattern and the $A-E-B$ temporal pattern, because the $A \rightarrow E$ link can sometimes occur more quickly than the $A \rightarrow B$ link.

Manipulating memory demands

In order to test whether participants' judgments are affected by memory effects such as forgetting of earlier pieces of evidence, we ran three different conditions between subjects. In condition i) participants were unable to remind themselves of what patterns they had previously seen. In condition ii) the sequence of activations seen on the previous clips of the current device was available throughout the experiment. In condition iii) the summary time lines showed, not just the order in which events had occurred, but also the exact delays separating events.

Methods

Participants 86 participants (39 female, $M_{age} = 35.3$, $SD_{age} = 12.3$) were recruited via Amazon Mechanical Turk. They were each paid \$1 and the task took 25.0 minutes on average ($SD = 9.8$). There were 25 participants in condition i), 32 in condition ii) and 29 in condition iii).

Stimuli and model predictions

Participants saw 16 causal devices in total (Table 1). For each device, participants were presented with four patterns of evidence. They were asked to provide a first judgment after they had seen the first three clips and were then given the chance to update their judgments after having seen a fourth piece of evidence. We selected clips such that, for half of the devices, our

Table 1: Patterns of observations (1st - 4th piece of evidence) for the 16 different devices. *Note*: Clip numbers refer to the temporal patterns in Figure 3a. The roles of components A and B were counterbalanced (e.g. pattern 2 A-B-E becomes pattern 5 B-A-E) and responses re-coded.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1st	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1	1
2nd	2	2	2	2	2	2	2	3	3	2	2	2	3	1	1	1
3rd	5	5	2	2	2	2	2	4	4	3	1	1	5	1	1	1
4th	4	2	1	2	4	5	6	3	5	4	2	6	4	1	2	3

models predicted a strong shift in belief between the first and the second judgment, while for the other half, little or no shift was predicted. For example, for device 1 (Figure 4a), participants first saw patterns 1, 2, and 5 ($AB-E$, $A-B-E$ and $B-A-E$) resulting in a strong prediction by both the simultaneous and the non-simultaneous models that participants will favor structure I: $A \rightarrow E \leftarrow B$.[2] Finally, participants saw pattern 4 ($A-E-B$) which is incompatible with the (conjunctive) common effect model, meaning that both models predict a dramatic shift to structure VI ($A \rightarrow E$) which both models consider to be the only structure consistent with all four clips (Figure 4c). For four of these eight devices the same shift was predicted by the simultaneous model as the non-simultaneous model, for three a different shift was predicted while for the other nine devices, the simultaneous model predicted no shift. In no case did we use a set of clips that resulted in all of the causal structures being ruled out.

In addition to whether each set of clips led to a large predicted shifts between participants' first and second judgments, we also selected sets of evidence for which the most likely structure differed depending on whether or not one made the simultaneity assumption. Thus the simultaneous and non-simultaneous baseline models disagreed about the most likely structure for one or both judgments on 8 of the 16 devices (see figure 4b and d for an example in which both judgments differ).

Procedure After reading the instructions, participants needed to successfully answer a comprehension check questions to proceed. The order in which the devices were presented was randomized between participants. However, the order of clips for each device was always as shown in Table 1. The delays between activations were generated randomly but the same values were used for all participants. We varied the delay between each activation, drawing each delay from a seeded uniform distribution between 200 and 1200 ms. The clips used in the experiment varied in total length between 566 and 2159 ms depending on these delays and whether there were three staggered component activation events (patterns 2, 4, 5 and 7) or only two (patterns 1, 3 and 6). We counterbalanced two reversed presentation orders of the seven structures shown at the top of the screen between participants (Figure 1a).

A demo of the task can be found at bit.ly/19WamZO.

Results and discussion

Condition differences Patterns of responses in the three memory conditions were very similar with pairwise correlations of .94 between conditions i) and ii) and .92 between conditions i) and iii). Multiple corrected one-way ANOVAs confirmed that there were no main effects of condition on participants' judgments. Therefore we conclude that memory load had little effect on peoples' judgments in this task and collapse the data over the three conditions for subsequent analyses.

[2]Both models predict people will prefer this model to the single link structures VI: $A \rightarrow E$ and VII: $B \rightarrow E$ because, by being consistent with fewer temporal patterns, structure I's likelihood is less widely dispersed.

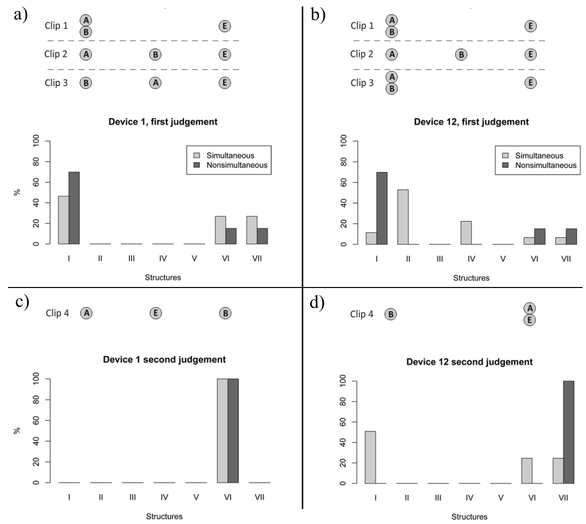
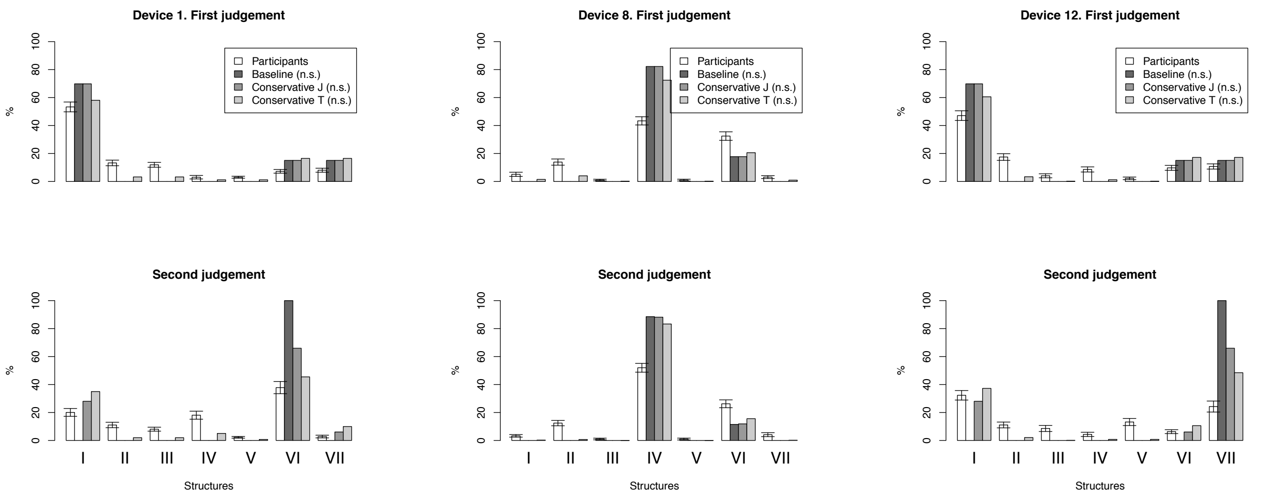


Figure 4: Baseline model predictions for devices 1 and 12 after the first three clips (a and b), and after the fourth clip (c and d).

Comparing the baseline models To assess qualitative correspondence with the models, we started by checking if judgments averaged over participants had the same modes as the predictions of one or other of the baseline models. For example, in Figure 5 we see that on average participants assign the most probability mass to the same structure as the non-simultaneous baseline model on both judgments for device 1 and 8, but diverge on the second judgment for device 12. Participants' modal response matched that of the non-simultaneous model 81% on the first judgment and 88% on the second, and the simultaneous baseline model 63% of the time on both the first and second judgments. For the trials where the two models made different predictions, participants' gave most points to the structure predicted by the non-simultaneous model 5/7 and 5/8 times first and second judgments respectively, but only 1/7 and 2/8 times for the si-

Table 2: Conservatism parameter c, fit of the different models based on RMSE, pearson's r correlation with averaged participant responses, and number of individuals with the highest correlation for each model. *Note*: Conservative J = conservative judgment model, Conservative T = conservative throughout model, s. = simultaneity assumption, n.s = nonsimultaneity assumption.

Model	c	RMSE	Correlation (r)	N highest r
Random		16.0	0	2
Baseline (s.)		18.3	0.56	5
Conservative J (s.)	.15	14.7	0.58	6
Conservative T (s.)	.10	14.7	0.58	21
Baseline (n.s.)		20.2	0.78	24
Conservative J (n.s.)	.12	16.3	0.84	4
Conservative T (n.s.)	.28	12.1	0.88	24



(a) 1. (AB)–E, 2. B–A–E, 3. A–E–B, 4. A–E–B (b) 8. A–B–E, 2. A–(BE), 3. A–E–B, 4. A–(BE) (c) 12. (AB)–E, 2. A–B–E, 3. (AB)–E, 4. B–(AE)

Figure 5: Comparison of the participants judgments with the non-simultaneous models. Devices (a) 1 and (b) 8. (c) 12. *Note*: Conservative J = conservative judgment model, Conservative T = conservative throughout model, n.s = nonsimultaneity assumption.

multaneous model (Figures 5).

For the 8 trials in which the non-simultaneous model predicted a large shift of probability between participants' first and second judgments, we found that in all cases, participants' judgments shifted in the expected direction. This was true both for the structure which became much less likely, and for the structure which became much more likely. For example with device 1, going from clip 3 to 4 both models predicted a shift from the common effect structure I to the singly connected structure VI and correspondingly, mean percentage attributions to structure I go down from 53% to 20% while attributions to structure VI rise from 7.3% to 37%. In two of the three cases where the simultaneous baseline model predicted a different shift than the non-simultaneous case, participants' judgments went against the simultaneous model's predictions. Therefore, based on modal choices and judgment shifts, participants appear to be acting more in line with the non-simultaneous baseline model (Figure 5).

Looking at the Pearson's r correlations between the average participant judgments and model predictions (Table 2), we see that the non-simultaneous baseline model correlates higher with participants' responses than the simultaneous model (.78 compared to .57), again suggesting that for the most part, participants made the non-simultaneity assumption in their inferences.

Conservatism While the non-simultaneous baseline model predicts participants' modal judgments for almost all trials and is well correlated with their overall judgments, quantitatively it is a poor fit. The RMSE between averaged participant judgments and the non-simultaneous baseline model (20.2) is actually slightly higher than for a random model that predicts participants will divide their percentage points evenly over all seven structures regardless of what they see (16.0, Table 2). Looking at the graphs, we see that this is due to participants typically being much less certain than our baseline model predicts. For example, the prediction for structure VI for device 1 was that it should rise to 100% on judgment 2, but on average participants redistribute barely half of the probability mass from structure I to VI (Figure 5a).

Having found no between-condition differences as evidence that forgetting could be driving participants' deviations from baseline predictions, another possible explanation for their divergence is that they might be *conservative* in their belief updating.

This might be due to participants' first judgments acting as an anchor leading them to shift less dramatically than a Bayesian model predicts when they see clip four. Alternatively, people might be conservative in their belief updating throughout the task due to more general skepticism or uncertainty about the evidence. For example, some participants may have suspected that components could sometimes activate spontaneously, or that a cause could generate its effect even when inactive. Alternatively, conservative updating could be a more general heuristic strategy for avoiding overly rapidly changes in belief as it is generally hard to be confident that evidence in the real world is perfectly reliable. Causal knowledge is widely thought to play a central role in guiding and constraining our everyday inferences (e.g. allowing predictions or diagnoses to follow the observation of the states of variables; Pearl, 2000). As such, it seems sensible to expect causal beliefs to be slower to change in the light of new evidence than the other less entrenched beliefs that they support.

In a Bayesian framework we can formalize conservatism by adding some unbiased noise into the likelihood functions.

If participants are generally conservative, we would expect a model that incorporates noise into the likelihoods for each observation to better explain participants judgments.

We created noisy likelihoods by mixing the likelihoods in Figure 3 with uniform likelihoods (with 1/7 for each pattern of data for each structure) to a degree controlled by a parameter c (i.e $\mathcal{L}_{conservative} = (1-c) \times \mathcal{L}_{baseline} + c \times \mathcal{L}_{uniform}$ where $\mathcal{L}$ stands for likelihood).

Fitting c to the averaged data by maximizing r, we augmented both baseline models with conservatism. For *conservatism after judgment (CJ)*, it was applied only once on final update, after the first judgment. For *conservatism throughout (CT)*, it was applied on all four belief updates. The results of these model fits are in Table 2. Note that *CJ* makes the same predictions for the first judgment as the baseline model but does not adjust this judgment as much after clip four while *CT* makes different predictions for both judgments (Figure 5). In particular, for the *CT* model with the non-simultaneity assumption there is a marked reduction in error, from 16.0 to 12.1, and increase in correlation from .78 to .88. The *CJ* model is also an improvement over the baseline, but not to the same extent. This suggests that participants were consistently conservative throughout the task, rather than only after having made their explicit judgment, although the higher parameter estimate for *CJ* (.28 compared to .12) suggests that conservatism may have been greater following the initial judgment.

Remaining variance While the Bayesian method of conservatism captures participants' judgments well, there is still additional variance. If participants updated their beliefs in some approximate, non-Bayesian, manner then their deviations may relate to the order in which they saw the clips. We experimented with non-Bayesian methods of modeling sticky belief updating while retaining the idea that people base their judgments on these simple likelihoods. As an example, we found that if we assumed that participants beliefs moved only halfway between their prior and the posterior implied by the non-simultaneity likelihood of each new clip ($posterior_{subjective} = mean(prior, posterior_{objective})$) we achieved an even better fit with the data, with an RMSE of just 7.2 percentage points and correlation of $r = .91$.

While this approach is not fully in the spirit of the pure Bayesian perspective, it does produce primacy effects, which we might expect given existing research, and provides a very tight fit with the data. This is suggestive of the idea that, while people are broadly consistent with Bayesian predictions, their causal beliefs might be updated in some non-Bayesian, sequentially biased, manner. We will explore what heuristic updating strategies people may use in a causal context in future work.

An additional potential source of divergence is that participants may have made some delay-expectation-related inferences like those found by Buehner et al (Buehner & May, 2004), despite having no task framing to work with. For example, participants may have been less inclined to endorse a $A \rightarrow E \leftarrow B$ or $B \leftarrow A \rightarrow E$ to the extent that A and B occurred at different times (despite the instructions that delay length was variable in this case). In ongoing work we are looking into more fine-grained likelihood functions that might capture assumptions about homogeneity of delays at the level of individual causal components and of whole systems.

Conclusions

In this paper we explored the dynamics of human causal learning from temporal order information. We found that rich predictions resulted from very minimal assumptions about temporal precedence of causes and that these predictions were borne out by participants' judgments in an online experiment. In particular, we found evidence that people typically assume that causes take time to cause their effects, and infer causal structure from order information in a broadly Bayesian but conservative way. Remaining deviations from this model suggest that people may be performing some form of approximate belief adjustment procedure or bringing in expectations about delay lengths.

Acknowledgments

We thank Ralf Mayrhofer for his involvement in this work, and Adam Sanborn for helpful comments. TG was supported by the Center for Minds, Brains and Machines (CBMM), funded by NSF STC award CCF-1231216. DL was supported by ESRC grant RES-062330004.

References

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (submitted). Conservative forgetful scholars: how people learn causal structure from sequences of interventions.

Buehner, M. J., & May, J. (2004). Abolishing the effect of reinforcement delay on human causal learning. *Quarterly Journal of Experimental Psychology Section B*, *57*(2), 179–191.

Burns, P., & McCormack, T. (2009). Temporal information and children's and adults' causal inferences. *Thinking & Reasoning*, *15*(2), 167–196.

Edwards, W. (1968). Conservatism in human information processing. *Formal representation of human judgment*, 17–52.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 678–693.

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

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1–55.

Hume, D. (1748/1975). *An enquiry concerning human understanding*. Oxford University Press.

Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. In W. Gray & C. D. Schunn (Eds.), *Proceedings of the twenty-fourth annual conference of the cognitive science society, mahwah,*. NJ: Erlbaum.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 856–876.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of experimental psychology. Learning, memory, and cognition*, *32*(3), 451–60.

McCormack, T., Bramley, N. R., Frosch, C., Patrick, F., & Lagnado, D. A. (under review). Children's causal structure learning.

McCormack, T., Frosch, C., Patrick, F., & Lagnado, D. A. (under review). The development of causal structure learning.

Michotte, A. (1946). *La perception de la causalité*. Inst. Sup. De Philosophie.

Pacer, M., & Griffiths, L. (2012). Elements of a rational framework for continuous-time causal induction. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 833–838).

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.

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

Sanborn, A. N., & Silva, R. (2013). Constraining bridges between levels of analysis: A computational justification for locally bayesian learning. *Journal of Mathematical Psychology*.

Schlottmann, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental psychology*, *35*, 303–317.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In *The psychology of learning and motivation* (Vol. 21, pp. 229–261). San Diego, CA: Academic Press.

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford University Press, USA.

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