Title
Causal Learning from Trending Time-Series

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

Authors
Soo, Kevin W
Rottman, Benjamin M

Publication Date
2018

Causal Learning from Trending Time-Series

Kevin W. Soo (kevin.soo@pitt.edu)
Benjamin M. Rottman (rottman@pitt.edu)
Department of Psychology, University of Pittsburgh
3939 O'Hara Street, Pittsburgh, PA 15260 USA

Abstract

Two studies investigated how people learn the strength of the relation between a cause and an effect in a time series setting in which both variables exhibit temporal trends. In prior research, we found that people control for temporal trends by focusing on *transitions*, how variables change from one observation to the next in a trial-by-trial presentation (Soo & Rottman, 2018). In Experiment 1, we replicated this effect, and found further evidence that people rely on transitions when there are extremely strong temporal trends. In Experiment 2, we investigated how people infer causal relations from time series data when presented as time series graphs. Though people were often able to control for the temporal trends, they had difficulty primarily when the cause and effect exhibited trends in opposite directions and there was a positive causal relationship. These findings shed light on when people can and can't accurately learn causal relations in time-series settings.

Keywords: causal learning, temporal trend, time-series

Introduction

Much real-world causal induction involves learning about relationships between causes and their effects as they unfold over time. This can be a complex task because variables may undergo temporal trends that obscure the underlying causal relationships (Yule, 1926). For example, a patient may experience increasing pain from a chronic disease over several months and take increasing amounts of pain medication to cope (a positive correlation), even though the medicine reduces pain on shorter timescales.

Recently, there have been a couple studies focusing on how people learn causal relations from time-series data that exhibit trends (Rottman, 2016; Soo & Rottman, 2018; White, 2015). In the present research, we investigated how people learn the strength of causal relationships when the cause and effect are continuous-valued, and exhibit strong trends over time. We evaluated if people are able to make correct causal inferences despite the trends, and what factors affect their ability to make accurate causal inferences.

Learning Causal Strength from Time-Series

The problem of causal learning from time-series data is that when a cause (X) and an effect (Y) exhibit trends over time, time is a confound. Due to the confound, the simple correlation of the absolute states of X and Y, cor(X, Y), often fails to capture the true causal strength. We refer to this model of causal strength induction from *states*, cor(X, Y), as r_{States}.

Soo and Rottman (2018) proposed another model of how people estimate causal strength from time series data, which is called $r_{\text{Transitions}}$. $r_{\text{Transitions}}$ estimates causal strength by taking the correlation of the changes in X and the changes in Y, cor(ΔX, ΔY). Δ refers to the first order difference score, the change in a variable from one observation to the next.[1]

Unlike r_{States}, we have argued that $r_{\text{Transitions}}$ uncovers the true causal influence of X on Y when a linear temporal confound is present. Soo and Rottman (2018; Appendix A) provide proofs and simulations demonstrating how $r_{\text{Transitions}}$ partials out linear temporal trends in the variables. This is the same reason that time series analysts use difference scores to control for non-stationarity (Shumway & Stoffer, 2011). For this reason, we say that using $r_{\text{Transitions}}$ helps people accurately estimate causal strengths.

Soo and Rottman (2018) demonstrated that when assessing the influence of a cause on an effect from time series data, peoples' judgments are sensitive to both r_{States} and $r_{\text{Transitions}}$, though they are more sensitive to $r_{\text{Transitions}}$, meaning that on the whole people tend to correctly infer causal strength despite temporal trends. (We regard these inferences as "correct" because we have previously argued that $r_{\text{Transitions}}$ actually uncovers the true causal strength when there are linear temporal trends.)

Rottman and Soo (2018) presented participants with observations of a cause and effect. All stimuli had r_{States} = .70 or -.70. The datasets were reordered to create versions with all positive transitions (from one observation to the next, X and Y changed in the same direction, producing a strongly positive $r_{\text{Transitions}}$), or all negative transitions (X and Y changed in opposite directions, creating a strongly negative $r_{\text{Transitions}}$), or the trials were randomized, producing a mix of transitions. (The first three columns of Figure 1 display the different orderings of the 20 data points.) The data were presented to participants in a trial-by-trial fashion with X and Y represented using vertical gauges (Figure 2). Although participants' estimates of causal strength were influenced by both transitions ($r_{\text{Transitions}}$) and states (r_{States}), the effect of transitions was considerably stronger.

Rottman and Soo (2018) also investigated two different versions of $r_{\text{Transitions.}}$ One captured the magnitude of the change, and another captured only the direction of the change (ΔX and ΔY are encoded as +1 for increases and -1 for decreases).

[1] This model assumes that the influence of X on Y occurs without a delay; the model associates the change in X from Time 1 to 2 with the change in Y from Time 1 to 2. These studies did not investigate situations in which the change in X produces a change in Y at a later time, which we leave for future research.

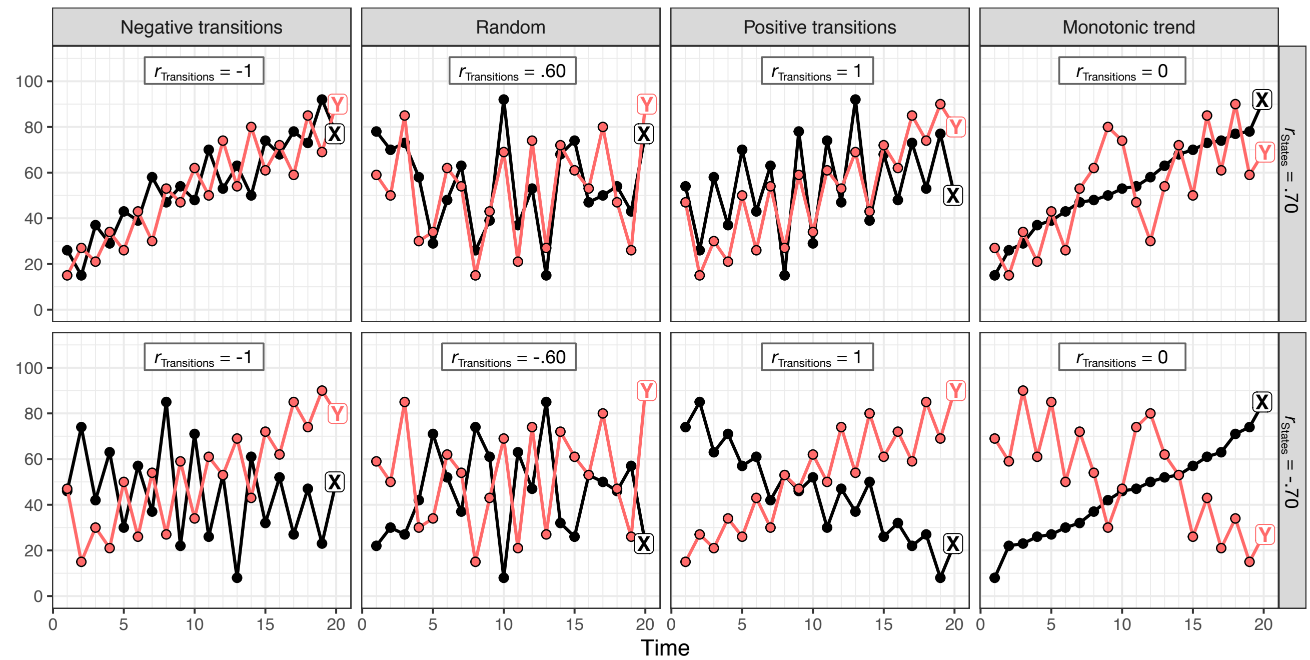

Figure 1: Time-series graphs of an example dataset used in Experiment 1, and Experiment 2 (without the monotonic trend condition). All datasets within the top row, and within the bottom row, have the exact same 20 states, but different orderings.

Though the two models are often highly correlated, we found that only the version that captured the direction of the changes explained unique variance in participants' judgments. Thus, in the current studies we only focus on this simpler version of $r_{Transitions}$.

What Happens with Extremely Strong Trends?

In Experiment 1, we sought to further test the theory that people focus on transitions for causal learning with time series data. In particular, we examined situations in which one variable exhibits an extremely strong monotonic trend such that from one trial to the next, it always increases (or always decreases) across time. We used the same datasets from prior studies, but sorted them by X or Y, such that one of those variables always increased or decreased.

One motivation for studying this case is theoretical. As explained below, the $r_{Transitions}$ model predicts a causal strength of zero despite there being very strong r_{States} predictions, which further helps to discriminate these two models. Furthermore, the case of monotonic trends can be investigated using the same datasets from prior studies by reordering the trials, allowing us to hold r_{States} constant.

Another motivation is revealed upon looking at the "monotonic trend" condition in Figure 1; it is a situation in which one variable exhibits a smooth monotonic trend and the other exhibits a noisy but roughly linear trend. Given that smooth processes exist in the real world, we thought this was an interesting time series case to study.

The rightmost column in Figure 1 shows the new monotonic trend condition with a monotonic trend in X. Consider the top-right panel in Figure 1. Both X and Y increase over time and the correlation between the two is quite strong (r_{States} = .70). However, just because two variables increase together does not mean that one causes the other; it could be that they are both just exhibiting trends. In contrast, consider the positive transitions condition in which r_{States} = .70. In that graph, both X and Y increase overall. Additionally, within a shorter timescale, increases in X are accompanied by increases in Y, and vice versa. This pattern provides strong evidence for a positive causal relation.

Based on the $r_{Transitions}$ model, the monotonic trend condition does not provide evidence for a causal relation. If X always increases from one observation to the next, there is no variance in ΔX (i.e. all the ΔX scores are +1), so cor(ΔX, ΔY) cannot be computed. For this reason, we treated the prediction of $r_{Transitions}$ as zero in the trend condition.[2] We predicted that their causal strength judgments would be close to zero because the variable exhibiting the monotonic trend does not exhibit much variance after accounting for the trend. Experiment 1 investigated whether participants would give causal strength judgments close to zero in the monotonic trend condition, or whether their judgments would be influenced by the strong correlation of the absolute states (r_{States} = .70 or -.70).

[2] For the version of $r_{Transitions}$ in which the magnitude of change scores are encoded, not just the direction, the value of cor(ΔX, ΔY) is close to zero for all the datasets in this study, so the predictions are essentially the same for both versions of $r_{Transitions}$.

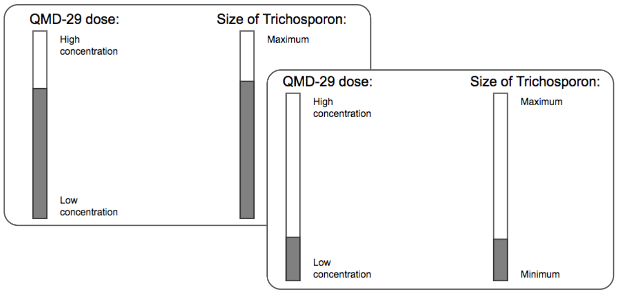

Figure 2: Visual presentation of stimuli in Experiment 1 and in the "trial-by-trial gauges" condition in Experiment 2.

Presentation Format Moderates Use of Transitions

How can time-series data be presented to people to optimize their causal inferences? In our previous research, we found that people were able to make fairly accurate judgments from a trial-by-trial presentation that involved two gauges representing the magnitude of the cause and the effect (Figure 2). Peoples' ability to infer causal relations fairly well in a trial-by-trial format is good news because it mimics how we often experience events in our daily lives.

We further demonstrated that people are much better in a naturalistic format (in which the gauge level represents magnitude) than when presented with trial-by-trial numbers representing magnitude, presumably because transitions are more salient when presented naturalistically.

However, lay people also often need to reason about data presented graphically, such as economic data in news reports (Fox & Hendler, 2011), and time-series graphs are the standard visualization for such data (Friendly, 2006; Javed, McDonnel, & Elmqvist, 2010). Experiment 2 investigated how well people infer causal relations controlling for temporal trends from time-series graphs.

We predicted that observing the data in a time-series graph would decrease the salience of transitions relative to the trial-by-trial presentation, making it harder for people to accurately infer causal relations. Instead, they might focus on the correlation between the absolute states of X and Y.

In Experiment 2, we compared causal judgments from participants observing stimuli presented in a trial-by-trial visual format (Figure 2) vs. a static time-series graph format (Figure 1). In addition, we created an intermediate format which involved a times-series graph, but instead of participants viewing the entire graph at once, the 20 observations were revealed sequentially. We hypothesized that revealing the data sequentially could make the transitions more salient, thereby leading to more accurate judgments than the static graph format.

Experiment 1

Experiment 1 tested whether learners used transitions in addition to states to estimate the causal strength between a cause (X) and an effect (Y) in a time-series setting. We used datasets in which the states were held constant, but the order of observations was manipulated to produce varied patterns of transitions (Figure 1). We predicted that participants' causal strength judgments would be strongest for datasets with all positive or negative transitions, followed by datasets with random orderings, and would be weakest for datasets with a monotonic trend in either the cause or effect because the $r_{\text{Transitions}}$ value was zero.

Method

Subjects 50 participants were recruited on MTurk and were paid $1.40. The experiment lasted between 7-10 minutes.

Design and stimuli Each learning dataset had 20 observations of X and Y, and each variable could take on values between 0 and 100. The design was a 2 (positive vs. negative r_{States}) × 5 (negative transitions, random order, positive transitions, monotonic trend in X, or a monotonic trend in Y) within-subjects design (see Figure 1).

The datasets were created in the following way. Using the *corgen* function from the R package *ecodist*, we generated 20 datasets with $r_{\text{States}} = .70$. Copies of each dataset with $r_{\text{States}} = -.70$ were made by flipping the values of X around the midpoint of the scale (X = 50). For the random order conditions, these datasets were presented with the trials in a random order. In the positive states random order condition most of the transitions were positive (Mean $r_{\text{Transitions}} = .59$, $SD = .12$), but in the negative states random order transition most of the transitions were negative (Mean $r_{\text{Transitions}} = -.59$, $SD = .12$). This is because with random trial orders, r_{States} and $r_{\text{Transitions}}$ are correlated.

The 20 trials of these datasets were reordered to produce the other four conditions**.** In the positive transitions conditions, the trials were reordered so that increases in X were always accompanied by increases in Y ($r_{\text{Transitions}} = 1$). In the negative transitions conditions, X and Y always changed in opposite directions ($r_{\text{Transitions}} = -1$). Finally, the conditions with monotonic trends in X or Y were ordered such that either X or Y always increased or decreased across the 20 trials; all these datasets had $r_{\text{Transitions}} = 0$. (Datasets that had repeated values of X or Y (e.g., X was exactly 56 on two trials) were slightly modified (e.g., one trial was changed to 58) so that the trials could be ordered to increase monotonically, and this change in the dataset was made to all the conditions).

Lastly, the observation order was counterbalanced to be presented forwards or in reverse (i.e. from 1-20 or 20-1 in Figure 1) randomly for each scenario a participant viewed.

Procedure Participants evaluated how the dosage of a drug (X) influenced the size of a microorganism (Y) over 20 observations ("days"). On each new day, participants clicked on a button to view a new drug dosage that was injected and the microorganism's resulting size. X and Y were displayed using gauges (Figure 2). After clicking the button on each day, the button was disabled for two seconds before the participant was allowed to advance.

After 20 days, the gauges disappeared and participants judged the causal strength of the drug on a scale from 8 ("high levels of the drug strongly cause the microorganism to increase in size") to -8 ("high levels of the drug strongly cause the microorganism to decrease in size"), with zero indicating there was no causal relationship.

Participants viewed all 10 conditions in randomized order. Each scenario involved a different drug-microorganism pair.

Results and Discussion

The means for the conditions are displayed in Figure 3. There was no difference in causal judgments for conditions with monotonic trends in X vs. Y (p = .32), so we collapsed those two conditions into a general "monotonic trend" condition. In the positive states conditions (gray triangles), the random order condition is displayed to the right of the monotonic trend condition, because it has positive $r_{\text{Transitions}}$. In the negative states conditions (black circles), the random condition is to the left of the monotonic trend condition because it has positive $r_{\text{Transitions}}$.

There is an effect of r_{States} such that the r_{States} = .70 conditions are judged more positively than the r_{States} = -.70 conditions. The negative transitions, random order, and positive transitions conditions exhibited the same pattern as past research; there is a strong effect of transitions such that the negative transitions conditions are judged much more negatively than the positive transitions conditions. In the random order conditions, the judgments are strongly positive when r_{States} = .70 but fairly negative when r_{States} = -.70. In conditions in which the states and transitions conflicted, the transitions "overrode" the states.

The new finding in this study is seen in the monotonic trend condition; participants' judgments were quite close to zero and considerably attenuated compared to the random transitions conditions. We compared judgments between these conditions using regressions with by-subject random intercepts and slopes for r_{States} and transitions. In the negative r_{States} condition, the judgments in the monotonic trend condition were significantly higher than the random order condition (B = 2.40, SE = 0.56, p < .001, partial-R^2 = .11). In the positive r_{States} condition, judgments in the monotonic trend condition were significantly lower than the random order condition (B = -3.15, SE = 0.43, p < .001, partial-R^2 = .23). That said, comparing the two monotonic trend conditions still revealed a difference between the positive vs. negative r_{States} conditions (B = 1.74, SE = 0.44, p < .001, partial-R^2 = .09). In sum, the effect of states was partially attenuated in the monotonic trend condition.

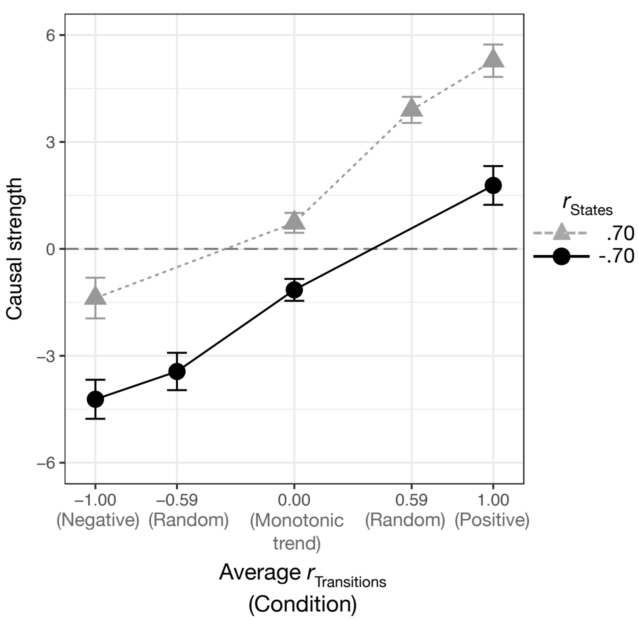

Figure 3: Experiment 1 condition means.
Error bars represent standard errors.

We also fitted an overall regression predicting all judgments with each dataset's r_{State} and $r_{\text{Transition}}$ value using a multivariate regression with a by-subject random intercept for repeated measures, as well as by-subject random slopes for r_{States} and $r_{\text{Transitions}}$. Controlling for transitions, there was a significant effect of states (B = 1.95, SE = 0.22, p < .001, partial-R^2 = .14). However, the effect of transitions after controlling for states was two times stronger than the effect of states in terms of variance explained (B = 3.27, SE = 0.31, p < .001, partial-R^2 = .30).

In summary, Experiment 1 provides additional evidence that people use transitions for inferring causal relations, and in particular, when one variable exhibits a very strong monotonic trend, there is only a small effect of states.

Experiment 2

In Experiment 2, we compared how well people learn causal relations from time-series data when presented in a trial-by-trial format like Experiment 1 vs. a static time-series graph vs. a hybrid in which the time-series graph was sequentially revealed. We predicted that in both the trial-by-trial and gradually-revealed graph conditions, participants' judgments would be more strongly influenced by transitions compared to in the static graph condition.

Method

Subjects 150 participants were recruited on MTurk and paid $0.90. The experiment lasted between 4-7 minutes.

Design and stimuli We used a 2 (positive vs. negative r_{States}) × 3 (positive transitions vs. random order vs. negative transitions) × 3 (trial-by-trial gauges vs. gradual graph vs. static graph) design. States and transitions were manipulated within-subjects, like in Experiment 1. We omitted the conditions with monotonic trends in X and Y for simplicity. Presentation format was manipulated between-subjects.

The datasets used in the present experiment were the same as those used in Experiment 1.

Procedure The procedure in the trial-by-trial gauges condition was identical to Experiment 1. Data in both graph conditions were shown to participants using line graphs programmed using d3.js. These graphs were similar in appearance to those in Figure 1.

In the gradual graph and the static graphs conditions, the graphs looked very similar to Figure 1 except without gridlines and labels. In the gradual graph condition, participants clicked on a button to reveal the observations for the next "day", gradually revealing the entire graph. After the final day's observations were revealed, the graph disappeared from the screen and participants made a causal judgment for that scenario.

In the static graph condition, the entire graph was revealed at the beginning of the scenario and kept on the screen for 40 seconds (this was the time it took to advance through 20 observations in the trial-by-trial gauges and gradual graph conditions if participants always advanced immediately to the next observation after viewing each transition). After 40 seconds, the graph disappeared from the screen and participants made a causal judgment for that scenario.

Attention check We included an attention check because the static graph condition did not require participants to actively watch the screen (in the other two conditions, participants had to wait for the button to become active and then click it to advance to the next trial, keeping their attention on the screen). Between 25-35 seconds after the start of the scenario, a common five-letter word was flashed on the screen for three seconds. Participants had to report the word prior to making their causal judgment at the end of the scenario. Participants knew that the word would appear sometime during the scenario, but not when, so they had to remain attentive. Different five-letter words were used on each scenario. For consistency, the attention check was also implemented in all three presentation format conditions. In the other two conditions, the word was flashed between observations 15-18. Almost all the participants noticed and could report the word correctly. Two correctly recalled 4 of the 6 words, 21 recalled 5 out of the 6, and the remaining 127 participants correctly recalled all 6 words. Because attention seemed to be close to ceiling, and there was not differential attention across conditions, we did not exclude any of the data.

Results and Discussion

The means for all conditions are displayed in Figure 4. We first analyzed data from each presentation format condition separately. We tested the relative fits of r_{States} and $r_{\text{Transitions}}$ to participants' judgments using multivariate regressions with by-subject random intercepts and by-subject random slopes for each model for repeated measures.

In the trial-by-trial gauges condition (the same format as Experiment 1), there was a significant effect of both states (B = 2.74, SE = 0.35, p < .001, partial-R^2 = .22) and transitions (B = 2.13, SE = 0.27, p < .001, partial-R^2 = .22), but no interaction between them. The effect of transitions was weaker than in Experiment 1, possibly because the attention-check distracted participants from the transitions.

In the gradual graph condition, the effect of transitions (B = 3.24, SE = 0.40, p < .001, partial-R^2 = .31) was much stronger than the effect of states (B = 1.76, SE = 0.36, p < .001, partial-R^2 = .07), and there was a small but significant interaction (B = 1.18, SE = 0.31, p < .001, partial-R^2 = .03). People are able to learn from transitions in time-series graphs if observations are revealed sequentially.

A regression including data from both the trial-by-trial gauges and gradual graph conditions (including by-subject random intercepts and slopes for r_{States} and $r_{\text{Transitions}}$) revealed a significant interaction between transitions and presentation format (p = .02); the effect of transitions was stronger in the gradual graph condition than the trial-by-trial gauges condition. This was due to the smaller effect of transitions in the gauges condition compared to Experiment 1 and past experiments (Soo & Rottman, 2018), in which the effect of transitions was stronger than that obtained in the gradual graph condition here.

In the static graph condition, the effect of states (B = 3.29, SE = 0.47, p < .001, partial-R^2 = .19) was much stronger than transitions (B = 1.06, SE = 0.34, p = .003, partial-R^2 = .04). There was a significant interaction between states and transitions (B = 1.68, SE = 0.38, p < .001, partial-R^2 = .05). In the static graph condition in Figure 4, the effect of transitions is visibly present in the positive states conditions, but not in the negative states conditions.

As predicted, the effect of transitions in the static graph condition was *weaker* than both the trial-by-trial gauges (p = .01) and the gradual graph conditions (p < .001). This

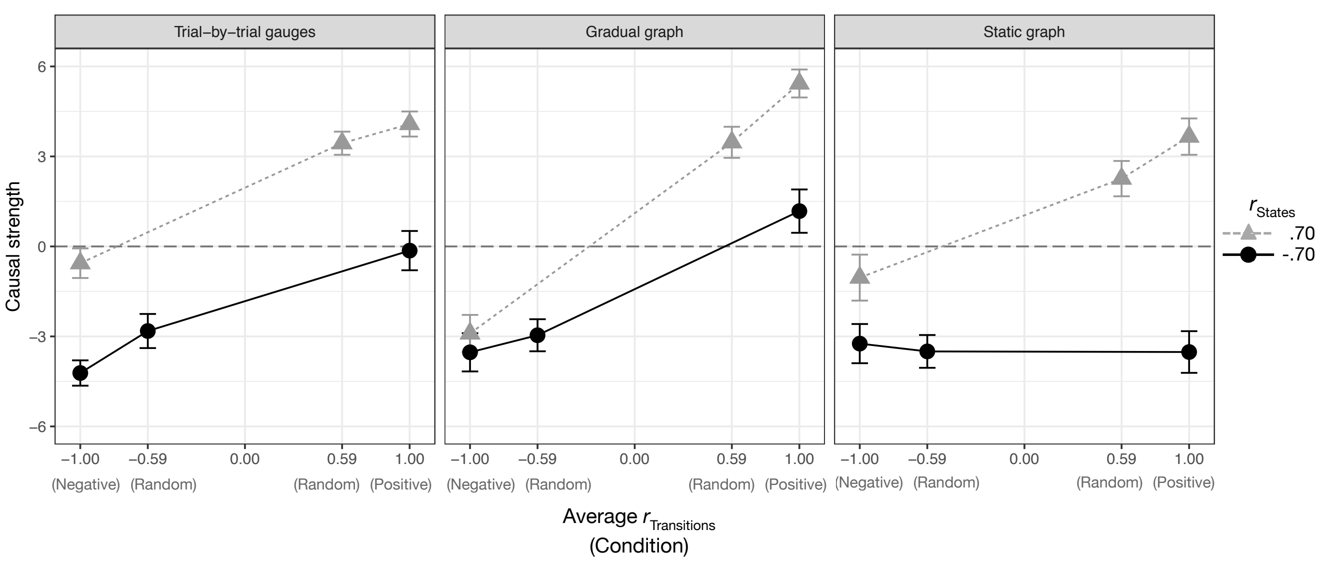

Figure 4: Experiment 2 condition means. Error bars represent standard errors.

finding is intuitive; the static graph presentation does not highlight transitions by sequentially revealing observations. However, it was surprising that the effect of transitions was attenuated exclusively in the negative states conditions. Note, it is not that participants were biased to see a positive relation (see function learning research by Kalish, Lewandowsky, & Kruschke, 2004). Instead, the issue is that they continued to infer a negative relation even when the causal relation, as revealed by the transitions, was positive.

In the condition with both positive transitions and states, the fact that X and Y both increase and decrease together is fairly salient in the graph since the X and Y lines overlap so strongly (Figure 1). However, in the condition with positive transitions and negative states, the lines only briefly overlap, making it harder to see if the transitions in X and Y are in the same or opposite directions. Participants may be focused on how X and Y are negatively correlated and infer a negative causal strength based on r_{States}.

The effect of presentation format found here suggests that different methods for communicating temporal data can greatly influence the inferences people draw from the data. In particular, the difference between dynamic and static visualizations can determine the relationships people perceive between variables.

General Discussion

Prior research on how people infer causal strength from time series data has found that people rely on both the absolute *states* of X and Y at individual points in time, and even more so the changes in X and Y from observation to the next. Furthermore, attending to changes or *transitions* allows people to control for temporal trends in X and Y, which can otherwise obscure the underlying causal relation.

In Experiment 1, we gathered further evidence showing that in time-series settings, people are able to control for temporal trends by attending to transitions. In particular, when one of the variables increases or decreases monotonically over time and the other variable increases or decreases but with more noise, people give fairly weak causal strength judgments. This makes sense in that the monotonic increase could be due simply to a trend over time rather than any relation with the other variable. Overall, this finding suggests that attending to the changes from one observation to the next is a simple heuristic that people often use and helps them make accurate causal inferences.

In Experiment 2, we tested if people control for trends when inferring causal strength from data presented in time-series graphs, which are ubiquitous in scientific and popular communications of temporal data. We found that people still attend to transitions when the graph was revealed gradually. However, presenting data in a static graph revealed an interesting boundary condition: people do not learn from transitions when the states of X and Y are negatively correlated. This suggests there are limits to peoples' ability to correctly infer causal relations from time-series graphs, but we also found that a simple animation dramatically improved performance.

Why might subjects have performed well in the static graph condition in Experiment 2 when there was a positive r_{States} correlation but not when there was a negative correlation? We suspect that learning was still good in the positive states condition because the two lines are largely overlapping, which could highlight the transitions. In contrast, in the negative states condition, the two lines cross, which makes it harder to see how the sequential changes in one variable correspond to changes in the other variable.

This explanation raises the possibility of another boundary condition. Even if there is a positive r_{States} correlation, it is possible that the two variables could be on very different scales (e.g., X in the range of 1-50, and Y in the range of 200-300). In this case, it might again become difficult to notice the correlations of the transitions.

In sum, subtle and non-obvious features of graphs may make it easier or harder to notice and compare the transitions among two variables, which would affect how well people control for trends and how accurate their causal inferences are. Future research is needed to determine guidelines for graphs that are easy to interpret.

Acknowledgments

This research was supported by NSF 1430439. We thank Daniel Oppenheimer for suggesting the manipulation of presentation format.

References

Fox, P., & Hendler, J. (2011). Changing the Equation on Scientific Data Visualization. *Science*, *331*(6018),705–708.

Friendly, M. (2006). A Brief History of Data Visualization. In *Handbook of Computational Statistics: Data Visualization* (pp. 15–56). Berlin: Springer.

Javed, W., McDonnel, B., & Elmqvist, N. (2010). Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 927–934.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review*, *111*(4), 1072–99.

Rottman, B. M. (2016). Searching for the Best Cause: Roles of Mechanism Beliefs, Autocorrelation, and Exploitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1233–1256.

Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. New York: Springer.

Soo, K. W., & Rottman, B. M. (2018). Causal Strength Induction From Time Series Data. *Journal of Experimental Psychology: General*, *147*(4), 485–513.

White, P. A. (2015). Causal judgements about temporal sequences of events in single individuals. *The Quarterly Journal of Experimental Psychology*, *68*(11), 2149–2174.

Yule, G. U. (1926). Why do we Sometimes get Nonsense-Correlations between Time-Series? A Study in Sampling and the Nature of Time-Series. *Journal of the Royal Statistical Society*, *89*(1), 1–63.