Cognitive Psychology 64 (2012) 93–125

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Causal structure learning over time: Observations and interventions ☆

Benjamin M. Rottman [a,*], Frank C. Keil [b]

[a] Department of Hospital Medicine, University of Chicago, 5841 S. Maryland Ave., MC5000, Chicago, IL 60637, United States
[b] Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520, United States

ARTICLE INFO

Article history:
Accepted 18 October 2011
Available online 7 December 2011

Keywords:
Causal structure
Causal learning
Time
Observation
Intervention

ABSTRACT

Seven studies examined how people learn causal relationships in scenarios when the variables are temporally dependent – the states of variables are stable over time. When people intervene on *X*, and *Y* subsequently changes state compared to before the intervention, people infer that *X* influences *Y*. This strategy allows people to learn causal structures quickly and reliably when variables are temporally stable (Experiments 1 and 2). People use this strategy even when the cover story suggests that the trials are independent (Experiment 3). When observing variables over time, people believe that when a cause changes state, its effects likely change state, but an effect may change state due to an exogenous influence in which case its observed cause may not change state at the same time. People used this strategy to learn the direction of causal relations and a wide variety of causal structures (Experiments 4–6). Finally, considering exogenous influences responsible for the observed changes facilitates learning causal directionality (Experiment 7). Temporal reasoning may be the norm rather than the exception for causal learning and may reflect the way most events are experienced naturalistically.

1. Introduction

In our attempts to explain and act on the world around us we almost invariably need to reason about causal structures. Scientists study the causal relationships between the variables involved in

☆ Experiments 1 and 3 were presented at the 32nd Annual Conference of the Cognitive Science Society (Rottman & Keil, 2010). Experiments 4 and 5 were presented at the 33rd Annual Conference of the Cognitive Science Society (Rottman & Keil, 2011).

* Corresponding author. Present address: Department of Hospital Medicine, University of Chicago, 5841 S. Maryland Ave., MC5000, Chicago, IL 60637, United States.
E-mail address: benjaminrottman@uchicago.edu (B.M. Rottman).

0010-0285/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2011.10.003

phenomena ranging from global warming to cognitive processes. Lay people understand phenomena like academic failure, unemployment, and loneliness in terms of complex webs of causes and effects (Lunt, 1988, 1989, 1991). Recently, cognitive scientists have investigated how people build their notions of causal structures. For example, how do people form a belief in the causal relationships between depression, anxiety, and insomnia; what causes what? Most prior work on causal structure learning has used events that are temporally independent, analogous to a between-subjects experimental design. For example, a clinical psychologist may try to learn the causal relationships between depression, anxiety, and insomnia, by observing 100 people who have different combinations of these disorders.

However, many if not most of our learning experiences involve repeatedly learning about one entity over time. For example, one might develop beliefs about the causal relationships between depression, anxiety, and insomnia, by observing a friend who experiences these disorders wax and wane over time, analogous to a within-subjects design. Furthermore, much of our causal learning involves learning about ourselves over time (e.g., do I really have an allergic reaction every time I eat wheat?). Given that we experience most events sequentially over time, this latter type of temporal or "within-subjects" situation may be the norm for causal learning.

To explore these different forms of causal learning, we will first introduce normative causal models for independent vs. temporally dependent scenarios; static and dynamic graphical causal models. Then, we describe different strategies people may use to learn causal structures for these two scenarios, and when manipulating vs. observing variables. Finally, we describe an approach to test whether people use a temporal strategy that is appropriate for learning causal structures in scenarios with temporally dependent events.

1.1. Static graphical causal models

1.1.1. Simulating causal models

Graphical causal models have mainly been developed to represent the causal relationships within a set of independent observations. We call this class of models "static." This section first describes how static causal models can be used to simulate a set of observations (e.g., 100 people who have different combinations of depression, anxiety, or insomnia). The next section describes how one can learn the causal structures that are most likely to have generated a set of observations.

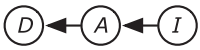
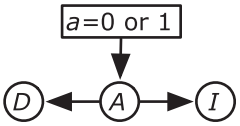
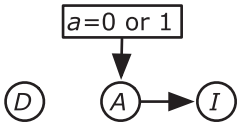
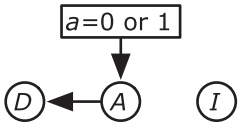
A graphical causal model consists of a set of nodes, which represent events, and arrows between the nodes, which represent causal relationships. In order to simulate how the model functions, one must also know the "parameters" of the model. For every node that does not have any known causes, one needs to know its base rate. Additionally, for every node that does have direct causes, one must know its conditional probability given its direct causes.

Table 1 presents three causal structures representing possible causal relationships between depression, anxiety, and insomnia. Consider the common cause structure, $D \leftarrow A \rightarrow I$, which asserts that anxiety influences depression and insomnia. The parameters for this structure assert that an individual person has a 10% chance of having an anxiety disorder. Out of people who are anxious, 75% are also depressed, and out of people who are not anxious 25% are depressed. Additionally, out of people who are anxious 75% are insomniacs, and out of people who are not anxious 25% are insomniacs.

Knowing the parameters and the causal structure, one can determine the percent of the population that has each of the eight possible combinations of depression, anxiety, and insomnia – the joint probability distribution. For example, the percent of the population who has depression and anxiety but not insomnia, can be determined by taking the product of the base rate of people who have anxiety, 10%, the conditional probability of having depression given that one has anxiety, 75%, and the conditional probability of not having insomnia given that one has depression, 25%. The product, shown on the second line of the joint probability table for observations, is 1.9%.

One can also simulate the causal structure under various interventions. For example, suppose that one was interested in the prevalence of depression and insomnia within an otherwise typical population that was taking anti-anxiety medications so that no one was anxious (A is set to 0 in the bottom section in Table 1). Under the graphical causal model framework, a perfect intervention severs all the relationships from other causes of the manipulated variable. For the common cause structure, after A

Table 1
Simulating three causal models for observations and interventions.

			Common cause	Chain 1	Chain 2
					
			Parameters for the three Markov equivalent structures		
			$P(a = 1) = .1$ $P(d = 1\|a = 1) = .75$ $P(d = 1\|a = 0) = .25$ $P(i = 1\|a = 1) = .75$ $P(i = 1\|a = 0) = .25$	$P(d = 1) = .3$ $P(a = 1\|d = 1) = .25$ $P(a = 1\|d = 0) = .036$ $P(i = 1\|a = 1) = .75$ $P(i = 1\|a = 0) = .25$	$P(i = 1) = .3$ $P(a = 1\|i = 1) = .25$ $P(a = 1\|i = 0) = .036$ $P(d = 1\|a = 1) = .75$ $P(d = 1\|a = 0) = .25$
			Joint probability tables		
D	*A*	*I*	$P(A)P(D\|A)P(I\|A)$	$P(D)P(A\|D)P(I\|A)$	$P(I)P(A\|I)P(D\|A)$	*Pr.*
1	1	1	$.1 \times .75 \times .75$	$.3 \times .25 \times .75$	$.3 \times .25 \times .75$	.056
1	1	0	$.1 \times .75 \times .25$	$.3 \times .25 \times .25$	$.7 \times .036 \times .75$	.019
1	0	1	$.9 \times .25 \times .25$	$.3 \times .75 \times .25$	$.3 \times .75 \times .25$	.056
1	0	0	$.9 \times .25 \times .75$	$.3 \times .75 \times .75$	$.7 \times .964 \times .25$	.169
0	1	1	$.1 \times .25 \times .75$	$.7 \times .036 \times .75$	$.3 \times .25 \times .25$	.019
0	1	0	$.1 \times .25 \times .25$	$.7 \times .036 \times .25$	$.7 \times .036 \times .25$	.006
0	0	1	$.9 \times .75 \times .25$	$.7 \times .964 \times .25$	$.3 \times .75 \times .75$	.169
0	0	0	$.9 \times .75 \times .75$	$.7 \times .964 \times .75$	$.7 \times .964 \times .75$	.506
			Graphical models and outcomes after interventions on anxiety		
Intervention					
Set $a = 0$			$P(d = 1\|a = 0) = .25$	$P(d = 1) = .3$	$P(d = 1\|a = 0) = .25$
Set $a = 1$			$P(d = 1\|a = 1) = .75$	$P(d = 1) = .3$	$P(d = 1\|a = 1) = .75$
Set $a = 0$			$P(i = 1\|a = 0) = .25$	$P(i = 1\|a = 0) = .25$	$P(i = 1) = .3$
Set $a = 1$			$P(i = 1\|a = 1) = .75$	$P(i = 1\|a = 1) = .75$	$P(i = 1) = .3$

is set to 0, the probability of D and I are easily determined from the parameters of the model; $P(d = 1|a = 0)$ and $P(i = 1|a = 0)$. For the chain structure $D \rightarrow A \rightarrow I$, the probability of D is simply the base rate, and D no longer has any influence on A. The probability of I is simply $P(i = 1|a = 0)$.

1.1.2. Inferring structures from joint probabilities

The fundamental idea behind learning causal structures is that some structures produce characteristically different sets of observations. For example, the common effect causal structure $D \rightarrow A \leftarrow I$ would likely produce sets of data in which D and I are independent (they are not correlated). In contrast, a common cause structure $D \leftarrow A \rightarrow I$ would likely produce sets of data in which D and I are correlated. If one comes across a dataset in which D and I are correlated, one can infer that $D \leftarrow A \rightarrow I$ is more likely the true causal structure (see Glymour & Cooper, 1999; Neapolitan, 2004; Pearl, 2000; Spirtes, Glymour, & Scheines, 2000).

However, it is not always possible to infer causal relationships from observations. Correlation really does not entail causation. It is impossible to distinguish $X \rightarrow Y$ from $X \leftarrow Y$ because both of these structures produce sets of data in which X and Y are correlated. More generally, causal structures cannot be distinguished when they have the same sets of dependencies and conditional dependencies between the variables, in which case they are said to be "Markov equivalent." The Markov condition is an assumption that any node, conditional on its direct causes, is independent of all other nodes that are not direct or indirect effects.

The three structures in Table 1 are Markov equivalent; all three variables are unconditionally dependent, but if A is held constant, D and I are independent. For Markov equivalent causal structures it is possible to choose parameters for the two structures that would produce the same joint probability distribution. In Table 1, we first chose parameters for the common cause model and then derived parameters for the other two models that would produce the same joint probabilities. The

joint probability of $D \leftarrow A \rightarrow I$, $P(A)P(D|A)P(I|A)$, can be proven to be equivalent to that of $D \rightarrow A \rightarrow I$, $P(D)P(A|D)P(I|A)$, by using Bayes' Rule; $P(A|D) = P(D|A)P(A)/P(D)$. Since the parameters are not known when inferring the causal structure, it is impossible to distinguish Markov equivalent structures.

Though it is not possible to distinguish Markov equivalent structures from observations, intervening on a causal structure can allow one to identify the causal structure. For example, consider hypothetical interventions in which people are either given a drug that makes everyone anxious or makes everyone calm (bottom of Table 1). When people are given the anti-anxiety drug (compared to a hypothetical anxiety-causing drug), they would have lower D and I for $D \leftarrow A \rightarrow I$, lower I but not D for $D \rightarrow A \rightarrow I$, and lower D but not I for $D \leftarrow A \leftarrow I$. Thus, by comparing samples with opposite interventions, one can infer the causal structure.

1.1.3. Psychology of learning static causal structures

Gopnik et al. (2004) and Steyvers, Tenenbaum, Wagenmakers, and Blum (2003) proposed that people use strategies akin to the statistical theories explained above for learning causal structures. For observations, only one study has demonstrated that people are able to learn causal structures at above chance levels (Steyvers et al., 2003), and the fairly weak performance has led others to conclude that there is little evidence that people learn causal structures from observations alone (Lagnado, Waldmann, Hagmayer, & Sloman, 2007). Mayrhofer and Waldmann (2011) demonstrated that people often use a heuristic that causes are often sufficient to produce effects, which can produce inferences that resemble the statistical strategies.

People can learn causal structures from interventions (Sobel & Kushnir, 2006; Steyvers et al., 2003; see also Hagmayer, Sloman, Lagnado, & Waldmann, 2007). However, Lagnado and Sloman (2004, 2006; see also Fernbach & Sloman, 2009) suggested that much of causal structure learning, both from interventions and observations, appears to be due merely to using a delay cue. For example, if one intervenes upon X, and then Y occurs, and then Z occurs, people tend to infer the chain $X \rightarrow Y \rightarrow Z$ even if they are told that the temporal delay information is not a valid cue for the causal structure.

1.1.4. Assumptions about prior states

In scenarios with independent events, as we have discussed so far, the states of variables are unknown before interventions. For example, the bottom section of Table 1 shows how to model a set of otherwise normal people who are given an anti-anxiety medication. However, why would a psychiatrist prescribe an anti-anxiety medication without knowing if a patient has anxiety?

Meder, Hagmayer, and Waldmann (2009) called interventions when the state is not known before the intervention "hypothetical interventions." From our perspective, hypothetical interventions are psychologically strange because interventions are typically performed to produce a desired change. Yet, much of the research on causal structure learning from interventions has investigated such hypothetical interventions (e.g., Lagnado & Sloman, 2004; Steyvers et al., 2003). In studies in which the states of variables were known before an intervention, the states of all the variables tended to be reset to the default "inactive" state before each intervention (e.g., Fernbach & Sloman, 2009; Gopnik et al., 2004; Sobel & Kushnir, 2006).

However, these kinds of event sets with independent trials, hypothetical interventions, or default states at the beginning of each trial are not broadly representative of how we normally experience strings of events. In daily life, interventions typically modify states that are temporally extended. When one crashes a car, the car *remains* broken until it is repaired. More in line with a causal structure learning task, consider a clinical psychologist learning the causal relationships between depression, anxiety, and insomnia by observing one patient over a course of 100 sessions. In this "within-subjects" scenario, it is natural to believe that each variable remains fairly stable over sessions, and when an intervention is enacted, it modifies a variable from its previous state.

1.2. Dynamic causal structures

1.2.1. Modeling causal structures with stable states

Dynamic Bayesian networks model scenarios in which the states of variables are temporally dependent (e.g., Fig. 1; see Neapolitan, 2004). At each time period, the state of a variable is determined both

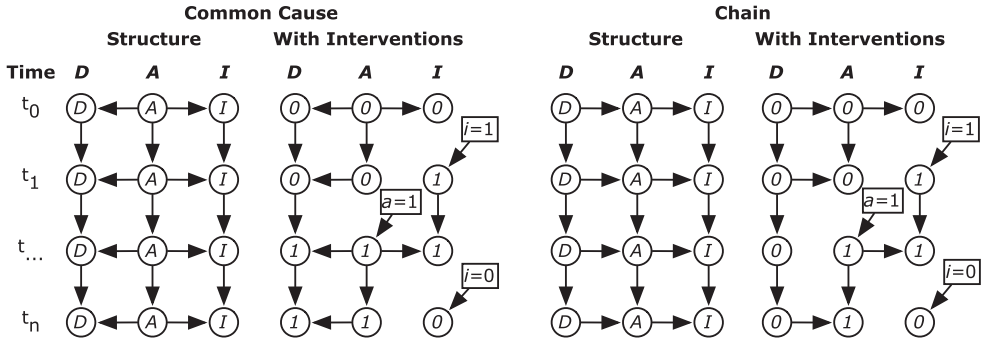


Fig. 1. Example dynamic causal models.

by the causal relationships within that time period and by prior states. Fig. 1 shows dynamic versions of the common cause and chain structures extended over time. Fig. 1 also shows how interventions can be modeled; when a variable is intervened upon at time *t*, all of that variable's other influences at time *t* and prior times are severed. The intervention propagates to effects at the same time period, and to the same variable at subsequent time periods. For example, in the common cause structure in Fig. 1, when *A* is manipulated and set to 1, *A* also remains at 1 in the next time period. Alternatively, when *I* is set to 1, *D* and *A* remain stable at 0.

Some computational algorithms can learn the structure of dynamic Bayesian networks (e.g., Friedman, Murphy, & Russell, 1998). However, as a first attempt at investigating how people learn causal structures with stable states over time, we have chosen not to compare human performance to a normative model in favor of exploring the breadth of the psychological phenomenon. We also believe that learning dynamic causal structures with stable states is intuitive and can be performed with simple strategies, as explained in later sections.

1.2.2. Learning dynamic causal structures from repeated interventions

We propose that when one intervenes, one usually *compares the outcome after the intervention against the prior state*, and assumes that any change in the outcome must be due to the intervention. This strategy consists of three simple heuristics, which we call *stability heuristics*.

A. A variable only changes state if it is directly intervened upon, or manipulated in some way, or if a variable that causes it (directly or indirectly) is manipulated. If it is not intervened upon, its state should remain stable.
B. If a variable *X* is intervened upon and *Y* changes state at the same time, then *X* causes *Y*. It is unlikely that *Y* happened to change due to some second unknown factor at the same time as *X*.
C. If a variable *X* is intervened upon and *Y* does not change state, then *X probably* does not cause *Y*. However, *X* may cause *Y*, but *Y* may not change after an intervention on *X* if *Y* is already at the state produced by an intervention on *X*. These are cases of *Y* already being at "ceiling" or at "floor."

To understand this strategy, consider an extended example in which a psychologist is learning the causal relationships between depression, anxiety, and insomnia by tracking a single patient, Jack, over time. Table 2 shows a summary of six events experienced by Jack that selectively manipulate either his anxiety, depression, or insomnia. The manipulated variable is in bold.

According to the temporal strategy, the six events that Jack experienced suggest that anxiety influenced depression and insomnia; $D \leftarrow A \rightarrow I$. Initially Jack started to take sleeping pills and his insomnia, depression and anxiety were low. The state of Jack's insomnia, depression, and anxiety are unknown before Time 1, so it is impossible to make any inferences about the effect of sleeping

Table 2
Example of interventions within one entity or many entities.

Time	Within-subjects conditions								Between-subjects conditions				
	Order 1: Jack				Order 2: Jill								
	Intervention	*D*	*A*	*I*	Intervention	*D*	*A*	*I*	Person	Intervention	*D*	*A*	*I*
1	Sleeping pills	0	0	**0**	Vacation	0	**0**	0	Bill	Dog dies	**1**	0	0
2	Paper rejected	1	**1**	1	Coffee each day	1	0	**1**	Nick	New romance	**0**	0	1
3	Vacation	0	**0**	0	Paper rejected	1	**1**	1	Sarah	Paper rejected	1	**1**	1
4	Dog dies	**1**	0	0	Sleeping pills	0	0	**0**	Anne	Vacation	0	**0**	0
5	Coffee each day	1	0	**1**	Dog dies	**1**	0	0	Jim	Coffee each day	1	0	**1**
6	New romance	**0**	0	1	New romance	**0**	0	1	Laura	Sleeping pills	0	0	**0**

Note: Boldface represents an intervention. 1 represents depression, anxiety, or insomnia and 0 represents the absence of depression, anxiety, or insomnia.

pills. At Time 2, Jack found out that his manuscript was rejected, which made him anxious. At the same time, he became depressed and started to have trouble sleeping. At Time 3, Jack went on vacation, a manipulation to decrease his anxiety, and his depression and insomnia decreased. Both Times 2 and 3 suggest that anxiety influenced depression and insomnia. During Times 4–6, Jack experienced events that manipulated his depression and insomnia. However, during all these manipulations, the other two variables remained stable, suggesting that neither depression nor insomnia influence other variables.

This strategy involves comparing the states of variables before and after an intervention. Thus, the order of events is critical. To show the impact of the order of the trials, consider Order 2 for Jill, who experienced the same six events (rows) as Jack, but in a different order. Here we just highlight two key trials to show the different inferences compared to Jack. At Time 2, Jill started to drink caffeinated beverages every day, which gave her insomnia, and at the same time she also became depressed. At Time 4, Jill started to take sleeping pills, which cured her insomnia, and at the same time her depression and anxiety receded. Times 2 and 4 suggest that insomnia influenced depression and anxiety. In sum, the temporal strategy can produce very different inferences depending on the order of the events.

The temporal strategy comparing variables before and after an intervention is very different than the non-temporal intervention strategy used for independent events and static causal structures. For the non-temporal strategy consider the same six events experienced by Jack and Jill, but instead of occurring to one person, they occurred to six separate people (Table 2, between-subjects condition). For each person, the states of the variables are only known after the intervention, not before (“hypothetical interventions”). In order to learn the causal structure, the appropriate strategy is to compare people who had opposite interventions. For example, Sarah and Anne had opposite interventions on their anxiety, and their depression and insomnia were also different, suggesting that anxiety influences depression and insomnia.

1.2.3. Learning dynamic causal structures from repeated observations

As explained earlier, the direction of causal relationships (e.g., $X \rightarrow Y$ vs. $X \leftarrow Y$) cannot be determined from a set of independent observations. However, when observing a single entity over time with stable states, the direction of causal relationships can be learned. Consider Table 3 describing eight events experienced by Jack. Here one merely observes the states of Jack’s anxiety and depression and may posit events that Jack may have experienced that manipulated his anxiety or depression. The bolded variables are posited to have been influenced by exogenous factors.

According to the temporal strategy for Jack, depression influences anxiety. There are a number of transitions (Times 4–5, 7–8) in which Jack’s anxiety changed but his depression remained stable. If anxiety influences depression, then presumably his depression would have changed as well, suggesting that anxiety does not influence depression. Jack also experienced a number of transitions in which both his depression and anxiety changed at the same time, suggesting some causal relationship between the two (Times 1–3, 5–6). A plausible interpretation of these transitions is that Jack experienced an exogenous event that influenced his depression, which in turn influenced his anxiety. It is also

Table 3
Example of observations within one entity or many entities.

Time	Within-subjects conditions				Between-subjects conditions		
	Order 1: Jack		Order 2: Jill				
	D	*A*	*D*	*A*	Person	*D*	*A*
1	1	1	1	1	Bill	1	1
2	**0**	0	0	**0**	Nick	1	1
3	**1**	1	1	**1**	Sarah	1	1
4	1	**0**	**0**	1	Anne	1	0
5	1	**1**	**1**	1	Jim	0	1
6	**0**	0	0	**0**	Laura	0	0
7	0	**1**	**1**	0	Matt	0	0
8	0	**0**	**0**	0	Lisa	0	0

Note: Boldface represents *inferred* exogenous influences on the bolded variables. 1 represents depression or anxiety, and 0 represents the absence of depression or anxiety. Across Orders 1 and 2, different interventions are inferred on *D* or *A* to explain the events.

possible that Jack experienced an exogenous event that influenced his anxiety, which succeeded in producing a change in his depression; however, this inference would be less likely because it runs contrary to Transitions 4–5 and 7–8. In sum, effects may frequently change without changes in other observed variables, but when causes change, they usually produce a change in their effects.

The order of these observed events is critical for inferring the causal structure. Consider the eight events experienced by Jill in Table 3, which are the same eight events experienced by Jack except that Times 4 and 7 were switched. For Jill, there were transitions in which her depression changed while her anxiety remained stable (Times 4–5, 7–8). By the same logic explained for Jack, one might infer that anxiety causes depression for Jill.

Finally, consider the same eight events experienced by Jack and Jill but now occurring to eight separate people (between-subjects condition Table 3). There is clearly a correlation between anxiety and depression, but because there is no temporal information, it is impossible to make any inferences about the direction of the relationship.

1.2.4. Psychology of learning causal models over time

The literature on controlling complex systems has investigated situations with repeated interventions in which the variables are dependent over time (e.g., Brehmer, 1992; Gonzalez, Vanyukov, & Martin, 2005; see Osman, 2010, chap. 7, for a review). Much of that literature concluded that people understand such scenarios merely in terms of which action to perform to produce a desired effect given the current states of the variables rather than understanding the structural relationships between the variables (e.g., Dienes & Fahey, 1995, 1998). However, those experiments often used large numbers of variables, complex dynamics, feedback loops, and required multiple simultaneous interventions. Additionally, some of these tasks required weeks of training, so it is perhaps not surprising that people did not learn the structure. In contrast, Hagmayer, Meder, Osman, Mangold, and Lagnado (2010) demonstrated that people sometimes spontaneously learn about the causal structure underlying the scenario in dynamic control tasks. However, in their study, only one variable was temporally dependent, and participants could only learn one of two possible causal structures. The current studies build on Hagmayer et al.'s (2010) study by examining repeated interventions on temporally dependent variables.

We know of only one study that has investigated how people learn causal structures from observing variables over time. White (2006) investigated how people learn causal relations in food webs; if species *X* eats species *Y*, then whenever *X*'s population changes in size, *Y*'s population will change in size. Though White's study was framed as testing static causal structures (i.e. Gopnik et al., 2004; Steyvers et al., 2003), the cover story was actually about how population sizes *change* over time. However, White found that people primarily used delay cues (if *X* changed one season, and then *Y* changed the following season, *X* eats *Y*; see Lagnado & Sloman, 2004, 2006), not the strategy suggested here.

White may have failed to find an effect for a number of reasons. First, people might have a strong belief that food webs have delays built in. Additionally, food webs typically involve bidirectional effects between predator and prey, making the stimuli difficult to comprehend. Finally, the abstract nature of the task and use of verbal summaries rather than trial-by-trial observations may have hindered reasoning about exogenous factors responsible for the observed changes (see Section 9).

1.3. Overview of experiments

In seven experiments, we investigated how people learn causal structures over time. Our primary experimental manipulation involved changing the order of trials to test whether people use the temporal strategies proposed here. Within alternate orders we kept the same set of events to ensure that any differences between participants' responses cannot be explained by the atemporal strategies designed for static causal structures. Our goal here was not to compare atemporal and temporal strategies to see which one better predicts participants' learning in general, because normatively they should be applied in different learning contexts. Rather, our primary goal was to examine how people learn causal relations when repeatedly intervening or observing over time, an important issue that previous research has largely neglected even though such scenarios may dominate everyday experience.

Experiments 1–3 investigated repeated *interventions* over time. In Experiments 1 and 2, we tested whether people use the temporal strategy at all. In Experiment 3, we tested whether people appropriately switch between the temporal and non-temporal strategies based on background knowledge about the dependence vs. independence of events.

In Experiments 4–7, we examined whether people use the proposed temporal strategy for causal structure learning with *observed* events. Experiment 4 tested whether people can learn the direction of causal relationships among two variables and Experiments 5 and 6 tested whether people can learn three variable structures. Experiment 7 tested whether being primed to think about exogenous influences facilitates the ability to learn the direction of causal relationships from observations.

2. Experiment 1a

Experiment 1 tested whether people use the temporal strategy (compare states of variables before and after an intervention), or whether they treat the sequence of events as temporally independent. For five different causal structures, we created two sets of stimuli with different orders. One order was created to uphold the stability assumptions and another to violate them. Any difference in the inferences between the stable and unstable conditions suggests use of the temporal strategy.

2.1. Method

2.1.1. Participants

Twenty students from Yale University participated, either for payment at $10 per hour or for partial fulfillment of an introductory psychology course requirement.

2.1.2. Stimuli and design

There were ten sets of stimuli produced by crossing 5 causal structures [chain $X \rightarrow Y \rightarrow Z$; common cause $X \leftarrow Y \rightarrow Z$; common effect $X \rightarrow Y \leftarrow Z$; one link $X \rightarrow Y$, Z; no links (no relationships between X, Y, Z)] $\times$ 2 trial orders (stable vs. unstable). For a given causal structure, the stable and unstable conditions had exactly the same 24 trials; the only difference was the order of the trials (see Table 4). Each of the three variables was manipulated and turned "on" four times and "off" four times; interventions always changed the value of the variable from the previous trial.

The 24 trials were created from joint probability distributions of static causal structures for which the base rate for exogenous variables was .5 and causal relations were deterministic. For the common-effect scenario $X \rightarrow Y \leftarrow Z$, Y was on if either X or Z was on (inclusive disjunction). The 24 trials adhered to the Markov condition. For example, for $X \rightarrow Y \rightarrow Z$, out of the four times when Y was set to 1 and the

Table 4
Summaries of the 10 conditions in Experiment 1.

Trial	$X \rightarrow Y \rightarrow Z$ Chain						$X \leftarrow Y \rightarrow Z$ Common cause						$X \rightarrow Y \leftarrow Z$ Common effect						$X \rightarrow Y\ Z$ One link						$X\ Y\ Z$ No link					
	Stable			Unstable			Stable			Unstable			Stable			Unstable			Stable			Unstable			Stable			Unstable		
	X	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*
1	**1**	1	1	1	1	**1**	1	**1**	1	**1**	1	1	**1**	1	0	1	**1**	1	**1**	1	0	1	1	**1**	**1**	0	0	**1**	1	1
2	**0**	0	0	0	0	**0**	0	**0**	0	**0**	0	0	**0**	0	0	0	**0**	0	**0**	0	0	0	0	**0**	1	**1**	0	**0**	0	0
3	0	**1**	1	0	0	**1**	1	**1**	1	1	1	**1**	0	1	**1**	1	1	**1**	0	0	**1**	1	**1**	1	1	1	**1**	1	**1**	1
4	0	**0**	0	0	**1**	1	**0**	1	1	0	0	**0**	0	0	**0**	0	0	**0**	**1**	1	1	0	**0**	0	**0**	1	1	0	**0**	0
5	0	0	**1**	**1**	1	1	**1**	1	1	1	**1**	1	0	**1**	0	**1**	1	1	**0**	0	1	**1**	1	1	0	**0**	1	1	1	**1**
6	0	0	**0**	**0**	0	0	1	1	**0**	0	**0**	0	0	**0**	0	**0**	0	0	0	0	**0**	**0**	0	0	0	0	**0**	0	0	**0**
7	**1**	1	1	1	**1**	1	1	1	**1**	0	0	**1**	0	1	**1**	0	**1**	1	0	0	**1**	**1**	1	1	0	0	**1**	1	**1**	0
8	1	**0**	0	0	**0**	0	0	**0**	0	1	1	**0**	0	**0**	1	1	**0**	0	0	**1**	1	**0**	0	0	**1**	0	1	0	**0**	1
9	1	**1**	1	0	0	**1**	**1**	0	0	0	0	**1**	0	**1**	1	0	**1**	0	0	**0**	1	**1**	1	0	**0**	0	1	**1**	1	0
10	1	1	**0**	0	**1**	1	**0**	0	0	1	1	**0**	**1**	1	1	1	**0**	1	**1**	1	1	0	0	**1**	0	**1**	1	**0**	0	1
11	1	1	**1**	**1**	1	1	0	0	**1**	0	**0**	0	**0**	1	1	1	**1**	0	**0**	0	1	1	1	**0**	0	1	**0**	1	1	**0**
12	**0**	0	0	**0**	0	0	0	0	**0**	1	**1**	1	**1**	1	1	0	**0**	1	0	0	**0**	0	0	**1**	0	**0**	0	0	0	**1**
13	**1**	1	1	**1**	1	1	1	**1**	1	**0**	0	0	**0**	1	1	**1**	1	0	0	**1**	0	1	1	**0**	0	**1**	0	**1**	0	0
14	**0**	0	0	1	1	**0**	0	**0**	0	**1**	1	1	0	0	**0**	**0**	1	1	0	**0**	0	**0**	0	1	0	1	**1**	**0**	1	1
15	0	0	**1**	1	**0**	0	0	0	**1**	0	**0**	0	**1**	1	0	**1**	1	0	**1**	1	0	**1**	1	0	**1**	1	1	1	0	**0**
16	0	0	**0**	**0**	0	0	0	0	**0**	1	**1**	1	1	**0**	0	**0**	1	1	1	1	**1**	**0**	0	1	1	1	**0**	0	1	**1**
17	0	**1**	1	1	**1**	1	**1**	0	0	**0**	1	1	1	**1**	0	1	1	**0**	1	**0**	1	1	**1**	0	**0**	1	0	1	**0**	0
18	0	**0**	0	0	**0**	0	**0**	0	0	**1**	0	0	1	1	**1**	0	1	**1**	1	**1**	1	0	**0**	1	**1**	1	0	0	**1**	1
19	**1**	1	1	**1**	1	1	1	**1**	1	**0**	1	1	1	1	**0**	1	1	**0**	1	1	**0**	0	**1**	0	1	**0**	0	**1**	0	1
20	1	**0**	0	1	1	**0**	1	1	**0**	**1**	0	0	1	1	**1**	0	1	**1**	1	1	**1**	1	**0**	1	1	0	**1**	**0**	1	0
21	1	**1**	1	1	**0**	0	1	1	**1**	1	**1**	1	1	**0**	1	**1**	1	1	1	1	**0**	0	**1**	1	1	**1**	1	1	**0**	1
22	1	1	**0**	**0**	0	0	**0**	1	1	0	**0**	0	1	**1**	1	**0**	0	0	1	**0**	0	1	**0**	0	1	**0**	1	0	**1**	0
23	1	1	**1**	1	1	**1**	**1**	1	1	1	1	**1**	1	1	**0**	1	1	**1**	1	**1**	0	1	1	**1**	1	0	**0**	1	0	**1**
24	**0**	0	0	0	0	**0**	0	**0**	0	0	0	**0**	**0**	0	0	0	0	**0**	**0**	0	0	0	0	**0**	**0**	0	0	0	1	**0**

Note: Boldface represents an intervention. 1 represents that a light bulb is on, and 0 off.

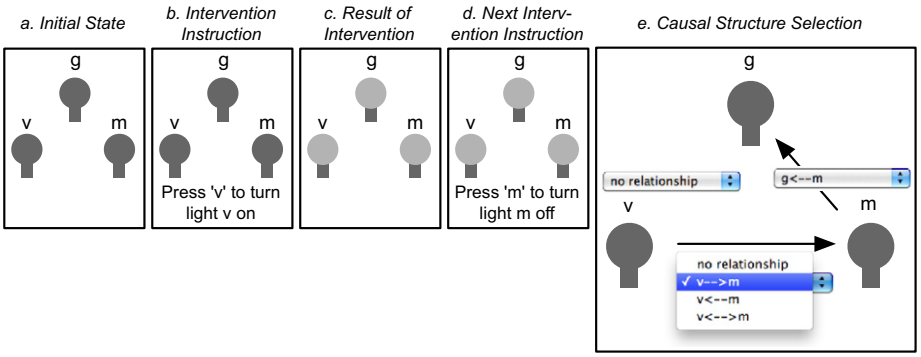


Fig. 2. Sample screenshots from Experiment 1.

four times that Y was set to 0, half the time $x = 1$ and half the time $x = 0$. Because the 24 trials were determined by joint probability distributions of static causal structures, if people use atemporal strategies, they would infer the "correct" model for both the stable and unstable conditions.

According to the temporal strategy, the two orders imply very different causal structures. Consider the stable chain $X \rightarrow Y \rightarrow Z$ condition in Table 4. Whenever X was manipulated, Y and Z also changed (e.g. the transition from Trials 1 to 2). Whenever Y was manipulated, Z also changed, but X stayed the same (e.g., Trials 2–3). When Z was manipulated, X and Y stayed the same (e.g., Trials 4–5). In contrast, unstable conditions were designed to systematically violate the stability heuristics and thus suggest the presence of links that did not exist according to the non-temporal strategy. Consider the unstable chain condition in Table 4. On Trial 2, Z was changed from 1 to 0, and X and Y also changed to 0, suggesting that Z caused X and Y. Additionally, the existence of real links was obscured (caveat in Heuristic C). For example, on Trial 5, X was changed from 0 to 1, but Y was already at 1. In sum, the unstable condition suggested different links from the stable condition, and did not identify one causal structure that consistently explained all the transitions between trials.

2.1.3. Procedures

Participants first read a cover story explaining that they would be presented with three light bulbs, would be given instructions to turn on or off specific lights, and that they were supposed learn how each light affects the others. Participants initially worked through an example scenario, exactly the same as the following 10 test scenarios except with only two light bulbs. Participants worked with all 10 scenarios in a Latin square design grouped by causal model such that the scenarios alternated between stable and unstable conditions.

During each scenario, participants were instructed to make 24 interventions to turn on or off specific lights. Initially, all three bulbs were off (e.g., Fig. 2a). Then participants were instructed to intervene to turn on or off a specific bulb by pressing a key (e.g., Fig. 2b), at which point the result of the intervention was displayed (e.g., Fig. 2c). Two seconds later, instructions appeared for the next intervention (e.g., Fig. 2d).

At the end of the 24 intervention trials within a scenario, participants chose the causal structure reflecting their belief of how the bulbs influenced one another (see Fig. 2e). Participants selected arrows indicating the direction of the causal relationships between the three light bulbs. For each pair of bulbs (e.g., X and Y), participants chose between "no relationship; neither light influences the other", "$X \rightarrow Y$; X influences Y", "$X \leftarrow Y$; Y influences X", or "$X \leftrightarrow Y$; X and Y both influence each other." Participants did not receive feedback of the accuracy of their causal model.

Across the 10 scenarios, the letter-names for the three light bulbs were changed, and the positions of causal light bulbs were changed (e.g., the bottom left light bulb was sometimes a cause, and sometimes an effect, and sometimes had no relation with the other lights). Similar counterbalancing was used for all future experiments.

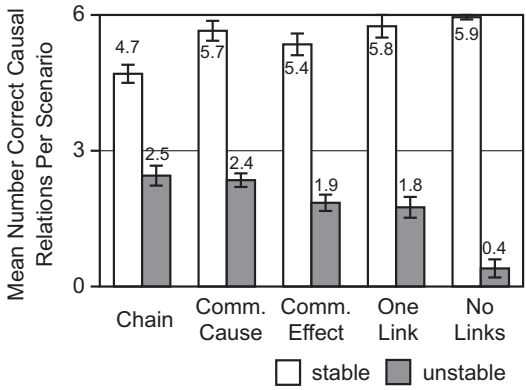


Fig. 3. Mean (std. error) number of correct causal relations inferred in Experiment 1a.

2.2. Results and discussion

The accuracy of causal structure inferences was assessed in the following way. (This system is used for all experiments except 4 and 5.) For each pair of variables, *X* and *Y*, *X* can cause *Y* or not, and *Y* can cause *X* or not. Thus for each pair of variables, participants had the possibility of identifying zero, one, or two correct causal relations. Suppose that *X* causes *Y*. If participants chose $X \rightarrow Y$ they would receive a score of 2. If participants chose $X \leftrightarrow Y$ they would receive a score of 1; they correctly identified $X \rightarrow Y$ but also falsely inferred $X \leftarrow Y$. If participants chose no relationship between *X* and *Y*, they would receive a score of 1; they correctly inferred that *Y* does not cause *X*, but they failed to infer $X \rightarrow Y$. Finally, if they chose $X \leftarrow Y$, they would receive a score of 0; they falsely inferred $X \leftarrow Y$ and failed to infer $X \rightarrow Y$. Across the three variables in a given scenario, participants had the possibility of identifying zero to six correct causal relations; chance performance is 3.

For all of the five causal structures, participants identified more correct causal relations in the stable than unstable conditions $ts(19) > 8.32$, $ps < .01$ (Fig. 3), suggesting that they used the temporal strategy for learning causal structures. If they had used the atemporal strategy, they would have inferred the "correct" relations in both the stable and unstable conditions.

There were two trends in participants' errors. First, in the stable chain condition $X \rightarrow Y \rightarrow Z$, participants frequently inferred the direct link $X \rightarrow Z$. When *X* was intervened upon, the states of *Y* and *Z* frequently changed (see Fernbach & Sloman, 2009). Second, in the unstable conditions, participants frequently endorsed links that do not exist according to the atemporal strategy. Consider the no links unstable condition as an illustration. In regards to the atemporal strategy, regardless of whether a light was turned on or off, the other two lights had a 50% chance of being on or off. In regards to the temporal strategy, when a light was turned on or off, the other two lights sometimes changed state (e.g., Table 4, no links, unstable, Trials 1–2). Such changes after interventions often led participants to endorse all six causal relationships in the unstable conditions. In fact, in all the unstable conditions, accuracy was below chance responding of 3, all $ts(19) > 2.46$, $ps < .03$. In sum, the results strongly suggest that participants used the temporal strategy.

3. Experiment 1b

Causal learning experiments typically randomize the order of trials to eliminate any influence of order. How would this affect causal structure learning? It is possible that people will simply learn that the trials are independent and use an atemporal strategy appropriate for static causal structures.

However, if people use the temporal strategy, then random orders would lead them to infer causal links that do not exist according to atemporal theories. Previous work on the gambler's fallacy suggests that people interpret truly random events that are independent and identically distributed as being influenced by prior events. For example, when observing a sequence of coin flips, after seeing

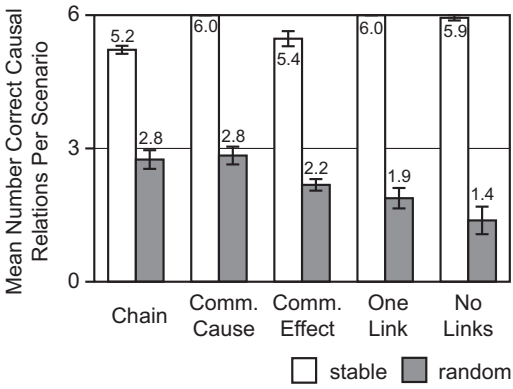


Fig. 4. Mean (std. error) number of correct causal relations inferred in Experiment 1b.

a string of heads, people predict that the next coin flip is "due to" land on tails (e.g., see Oskarsson, Van Boven, McClelland, and Hastie (2009) for a review). In this vein, people may irresistibly interpret the events as connected over time and use the temporal strategy.

3.1. Method

3.1.1. Participants

There were 32 participants from the same population.

3.1.2. Stimuli and design

The stimuli, design, and procedures were the same as Experiment 1a except that unstable conditions were replaced with random conditions, which contained the exact same 24 trials as the stable conditions but in a pseudo-random order. They were pseudo-random because, just as in Experiment 1a, a variable could not be intervened upon to stay in the same state as in the previous trial.

3.2. Results and discussion

The results appear in Fig. 4. Overall, participants inferred more correct causal relations in the stable conditions than in the random conditions, all $ts(31) > 12.30$, all $ps < .01$, suggesting use of the temporal strategy. If participants had realized that in the random conditions the trials were independent and used an atemporal strategy appropriate for static causal structures, then they should have been at least above chance for learning causal structures in the random condition. However, for the common effect, one-link, and no-links causal structures, participants were significantly below chance in the random conditions, all $ts(31) > 5.18$, all $ps < .01$. For the random chain condition, $t(31) = 1.19$, $p = .24$, and random common cause condition, $t(31) < 1$, participants were not significantly different from chance.

Unlike Experiment 1a, participants worked with blocks of stable and random scenarios. We thought that perhaps participants might have learned to use the temporal strategy in the stable block, and continued to use the temporal strategy in the random block. However, a between-subjects analysis looking only at the block that participants saw first revealed all the same effects as the within-subjects analysis.

In sum, participants did not appear to use the atemporal strategy when the data trials were in fact temporally independent. This finding might be taken as a caution that even in standard causal learning experiments with randomized orders, people do not necessarily treat the trials as statistically independent.

One benefit of the approach used in Experiment 1 is that the 24 data trials were created by static causal structures; if participants had used the atemporal strategy, they would have inferred what we

called the "correct" causal structure in both the stable and unstable conditions. However, a limitation to Experiment 1 was that we only tested select, hand-made trial orders.

4. Experiment 2

In Experiment 2, we tested many different stable trial orders by using the computer to simulate dynamic causal structures with stable states. For the unstable conditions, we switched the order of two key trials from the stable conditions. This manipulation accomplished a number of goals. First, it tested whether the use of the temporal strategy generalizes to other orders that were not hand-crafted. Second, this approach allowed us to hold constant the order of the interventions, which could have potentially influenced the ease of learning. For example, in Experiment 1, the stable-chain condition started off with the following sequence of interventions [$x = 1$, $x = 0$, $y = 1\ldots$], but the unstable chain condition started off with a different set of interventions [$z = 1$, $z = 0$, $z = 1\ldots$]. Third, this approach allowed more precision to identify how specific changes in the trial orders produced specific inferences, and we created a simple model to predict participants' inferences.

4.1. Method

4.1.1. Participants

There were 20 participants from the same population as Experiment 1.

4.1.2. Stimuli and design

There were ten conditions created by crossing the 5 causal structures × 2 trial orders (stable vs. unstable). For each of the 10 conditions, a different set of data was created for each participant.

The stable stimuli (see Table 5) were created by simulating a dynamic causal structure in which the states of variables remain stable. For a given causal structure (e.g., common cause), a sequence of interventions was randomly chosen such that there were 2 "on" interventions and 2 "off" interventions per variable, producing a total of 12 trials. After each intervention was chosen, the values of the other two variables were deduced using the principles of stability; see Fig. 1 and Table 2 in the introduction. Essentially, states of variables remained stable over time unless they were intervened upon or one of their causes was intervened upon.

The data for the unstable conditions were exactly the same as the stable conditions except that two critical trials that had the same intervention but different outcomes were swapped. For example,

Table 5
Summary of one participant's stimuli for Experiment 2.

	X→Y→Z Chain						X←Y→Z Common Cause						X→Y←Z Common Effect						X→Y Z One Link						X Y Z No Links					
	stable			unstable			stable			unstable			stable			unstable			stable			unstable			stable			unstable		
Trial	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**X**	**Y**	**Z**
1	0	0	**1**	0	0	**1**	0	0	**1**	0	0	**1**	**1**	1	0	**1**	1	1	0	**1**	0	0	**1**	0	0	**1**	0	0	**1**	0
2	**1**	1	1	**1**	1	1	**1**	0	1	**1**	0	0	1	**0**	0	1	**0**	0	0	1	**1**	0	1	**1**	0	0	**0**	0	0	**0**
3	1	**0**	0	1	**0**	0	**0**	0	1	**0**	0	1	1	1	**1**	1	1	**1**	**1**	1	1	**1**	1	0	**1**	0	0	**1**	0	0
4	1	**1**	1	0	**1**	1	1	**1**	1	1	**1**	1	1	**0**	1	1	**0**	1	1	**0**	1	1	**0**	1	1	0	**1**	0	0	**1**
5	1	1	**0**	1	1	**0**	0	**0**	0	0	**0**	0	1	**1**	1	1	**1**	1	**0**	0	1	**0**	0	1	1	0	**0**	1	0	**0**
6	**0**	0	0	**0**	0	0	0	0	**1**	0	0	**1**	1	1	**0**	1	1	**0**	0	**1**	1	0	**1**	1	**0**	0	0	**0**	0	0
7	**1**	1	1	**1**	1	1	0	0	**0**	0	0	**0**	1	1	**1**	1	1	**1**	0	1	**0**	0	1	**0**	**1**	0	0	**1**	0	0
8	1	1	**0**	1	1	**0**	**1**	0	0	**1**	0	1	**0**	1	1	**0**	1	1	0	**0**	0	0	**0**	0	**0**	0	0	**0**	0	0
9	**0**	0	0	**0**	0	0	**0**	0	0	**0**	0	0	**1**	1	1	**1**	1	0	**1**	1	0	**1**	1	1	0	0	**1**	1	0	**1**
10	0	0	**1**	0	0	**1**	1	**1**	1	1	**1**	1	**0**	1	1	**0**	1	1	**0**	0	0	**0**	0	0	0	0	**0**	0	0	**0**
11	0	**1**	1	1	**1**	1	1	1	**0**	1	1	**0**	0	0	**0**	0	0	**0**	0	0	**1**	0	0	**1**	0	**1**	0	0	**1**	0
12	0	**0**	0	0	**0**	0	0	**0**	0	0	**0**	0	0	**1**	0	0	**1**	0	0	0	**0**	0	0	**0**	0	**0**	0	0	**0**	0

Note: This is a sample set of data seen by one participant. Boldface represents an intervention. 1 and 0 represent the presence and absence of hormones, respectively. Arrows show the two trials that were switched from the stable to unstable conditions.

consider Table 5, common cause, unstable condition. Trials 2 and 8 were swapped from the stable condition. During the transitions from Trials 1 to 3, when X was turned on and off, Z went off and on, which violates the stability heuristics and suggests that X influences Z.

Because the 12 data trials were chosen by simulating a dynamic causal structure as it underwent 12 interventions, these 12 trials need not conform to the Markov principle as explained for static causal structures. For example, in the $X \rightarrow Y \rightarrow Z$ condition in Table 5, after the two interventions setting $z = 0$, the state of X was 1 because it happed to be 1 before the intervention on Z. Additionally, after the two interventions setting $z = 1$, the state of X was 0. This negative dependence occurred merely due to the particular sequence of interventions that were randomly chosen for this one participant. Because the data do not necessarily conform to the Markov principle, if participants use the atemporal strategy they would not necessarily infer the "correct" causal structure. Most importantly, however, if people use atemporal strategies they would infer the same causal structures in the stable and unstable conditions. In contrast, if people use the temporal strategy they would be more likely to infer the "correct" causal structure, the dynamic causal structure that actually generated the data, in the stable condition, and would be misled by the two rearranged trials in the unstable condition.

4.1.3. Procedure and design

Participants first read a cover study story asking them to pretend that they were assistants in a biology lab studying hormones in amoebas. They would "produce" or "suppress" hormones by injecting chemicals into the amoebas and "learn how each hormone affects the others." They were told that the "hormones work immediately... without any perceivable delay." "+" and "−" signs denoted the presence and absence of the hormones.

Participants worked with an example scenario and then 20 test scenarios, two sets of 10 scenarios [5 causal structures × 2 trial orders (stable vs. unstable)]. Within each set of 10 scenarios, the scenarios were grouped by causal structure such that the scenarios alternated between stable and unstable conditions. Additionally, the order of the five causal structures was counterbalanced in a Latin-square design. Participants received different sets of data for the first and second sets of 10 scenarios. The procedures, flow, and dependent variables within each scenario were the same as in Experiment 1.

4.2. Results and discussion

The first and second blocks of 10 scenarios produced the same pattern of results, so we collapsed across the two blocks. For all of the five causal structures, participants identified more correct causal relations in the stable than unstable conditions $ts(19) > 7.25$, $ps < .01$ (Fig. 5), suggesting that participants used the temporal strategy.

One difference between the results of Experiments 1 and 2 is that participants inferred correct causal relations above chance in the unstable conditions in Experiment 2 because only two trials were

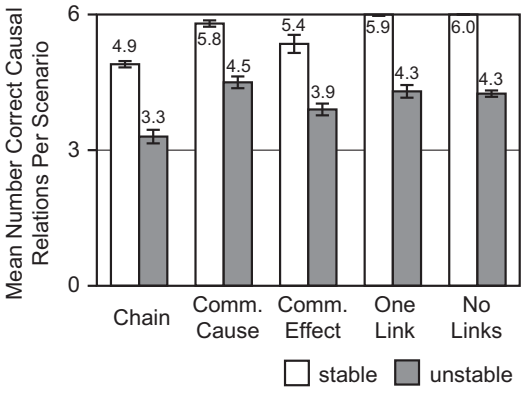


Fig. 5. Mean (std. error) number of correct causal relations inferred in Experiment 2.

rearranged. Yet, this small manipulation still produced large differences between the stable and unstable conditions, attesting to the robustness of the effect.

We created a simple model of the temporal strategy based on the three stability heuristics explained in the Introduction. If a variable Y ever changed after an intervention on a variable X, then the model inferred that X caused Y. If Y always stayed the same after an intervention on X, then the model inferred that X did not influence Y. This simple model correctly predicted 92.5% of all the causal relations that participants inferred among all three variables. The model predicted participants' inferences equally well in the stable ($M = 92\%$, $SD = 6\%$) and unstable ($M = 93\%$, $SD = 9\%$) conditions, $t(19) < 1$. In sum, this simple temporal strategy robustly explains when participants infer that causal links do and do not exist.

5. Experiment 3

The purpose of Experiment 3 was to determine how flexibly people switch between the temporal and atemporal strategies given background knowledge about whether they are repeatedly intervening on one entity (i.e. dependent observations) or are intervening only once on each entity and do not know the state of the variables before the intervention (i.e. independent observations). Previous studies with a similar manipulation have successfully moderated whether people treat the trials as dependent or independent (Rottman & Ahn, 2009). To test which strategy people use, we crossed the cover story manipulation with the stable vs. unstable manipulation from Experiment 1.

One likely outcome is that participants would use the temporal strategy for the dependent scenario (like Experiments 1 and 2) but switch to using a non-temporal strategy for the independent events scenario. In this case, there would be a difference in accuracy of inferred causal relations for the stable vs. unstable conditions for the dependent scenario, but there would be no difference between the stable vs. unstable conditions in the independent scenario.

Alternatively, if people continue to use the temporal strategy in the independent scenario, there may still be some difference in accuracy between the stable vs. unstable conditions.

5.1. Method

5.1.1. Participants

There were 16 participants from the same population.

5.1.2. Design and procedures

Participants worked with eight scenarios created by crossing the number of amoebas cover story (one vs. many) $\times$ trial order (stable vs. unstable) $\times$ causal structure (common cause, $X \leftarrow Y \rightarrow Z$ vs. one link $X \rightarrow Y$; Z). The design was entirely within subjects. The eight scenarios were ordered in a Latin square such that each scenario appeared first for some participants, and the scenarios were grouped by number of amoebas. The data trials were the same as in Experiment 1, so the following paragraphs focus on the number of amoebas manipulation.

The one-amoeba condition, analogous to a within-subjects design, emphasized the dependent nature of the data. Participants were told: "In each of the following scenarios you will observe hormones within one amoeba. You will first intervene to produce or suppress one hormone. Then you will observe whether the hormones are present or absent." Additionally, a picture of one amoeba was present for the entire scenario to emphasize the repeated interventions on a single entity. The transition from one trial to the next is shown in Fig. 6a–c.

The many-amoebas condition emphasized the independent nature of the data. Participants were told: "In each of the following scenarios you will observe hormones within 24 different amoebas. You will first get a new amoeba, and then intervene to produce or suppress one hormone. Finally you will observe whether the hormones are present or absent." The transition from trial to trial proceeded in the following way. Two seconds after the results of a given intervention were displayed (e.g., Fig. 6d), participants were instructed to "Press the spacebar to get the next amoeba" (e.g., Fig. 6e). When the spacebar was pressed, a picture of a new amoeba with a different shape, appeared (e.g.

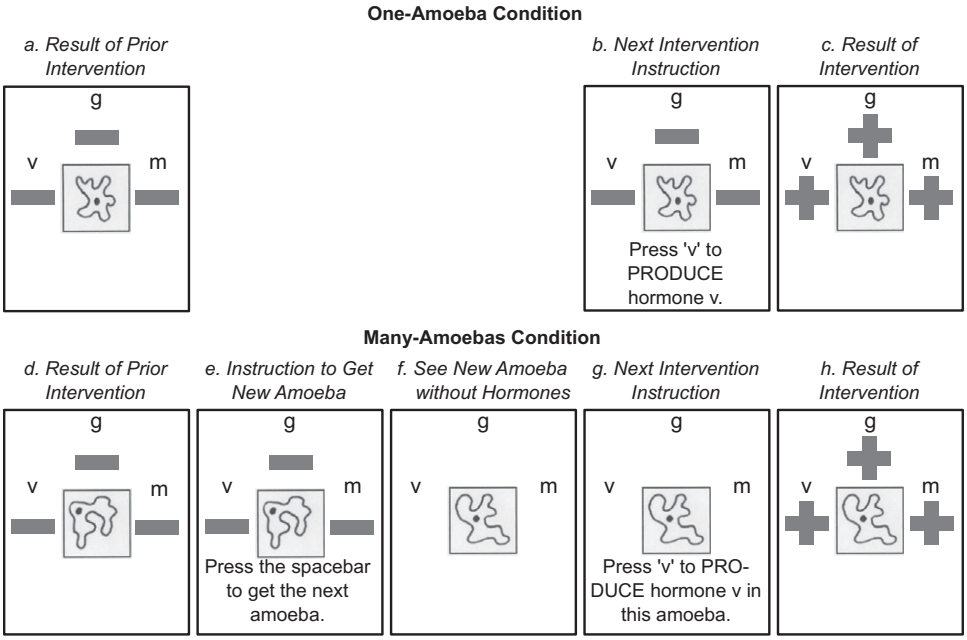


Fig. 6. Sample screenshots from Experiment 3.

Fig. 6f). Simultaneously, the states of the hormones in the previous amoeba ("+" and "−" marks) disappeared to clarify that the states of the hormones were unknown before the intervention and to make it perceptually difficult to track the changes of the hormones over time. Two seconds later, the prompt for the next intervention appeared (e.g., Fig. 6g). When the intervention key was pressed, the states of the hormones appeared for the current amoeba (e.g., Fig. 6h).

After each scenario, participants selected the causal relationships between the hormones.

5.2. Results and discussion

A 2 number of amoebas (one vs. many) × 2 trial order (stable vs. unstable) × 2 causal structure (common cause vs. one link) repeated-measures ANOVA was performed. There was a main effect of trial order; participants correctly identified more causal relationships in the temporally stable than unstable orders, $F(1, 15) = 45.28, p < .01, \eta_p^2 = .75$ (Fig. 7). There was not a significant main effect of causal structure nor of the number of amoebas, Fs < 1.[1]

Did participants switch from the temporal strategy in the one amoeba condition to the atemporal strategy in the many amoeba condition? This was tested with an interaction between number of amoebas and trial order. Though there was a large difference between the stable and unstable orders for the one-amoeba condition, there was a smaller difference between the stable vs. unstable, many-amoebas conditions, suggesting less use of the temporal strategy, $F(1, 15) = 12.61, p < .01, \eta_p^2 = .46$. However, there was still a significant difference between the stable and unstable, many-amoebas conditions, $t(15) = 3.59$, $p < .01$, suggesting that participants still used the temporal strategy. Furthermore, participants did not simply transfer the temporal strategy from the one-amoeba condition; they were more accurate in the stable than unstable many-amoebas conditions even before experiencing the one-amoeba scenarios, $t(7) = 3.21$, $p = .02$.

[1] Out of all the other two-way and three-way interactions, the only notable finding was that there was a larger difference between the stable vs. unstable conditions for the common cause structure than the one-link structure, $F(1,15) = 4.03$, $p = .06$, $\eta_p^2 = .21$.

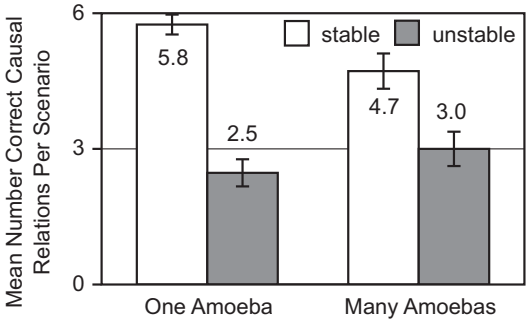


Fig. 7. Mean (std. error) number of correct causal relations inferred in Experiment 3.

The 24 data trials were created from a joint probability table for static causal structures, so if participants had used the atemporal strategy they should have been able to identify the correct causal relations, or they should have at least been above chance in both the stable and unstable conditions. However, participants were not above chance in the many amoeba, unstable condition, $t(15) < 1$.

Additionally, participants identified fewer correct causal relations in the many-amoeba than one-amoeba, stable condition, $t(15) = 2.57$, $p = .02$. This decrease suggests less use of the temporal strategy for the many amoebas condition, but not necessarily an increase in the use of an atemporal strategy. Participants may have not known an alternative strategy to use in the many amoeba condition. Alternatively, this finding could simply have resulted from the temporal strategy being more difficult to perform in the many-amoebas condition because participants would have had to remember the states of the prior hormones (Fig. 6e) to compare to the current hormones (Fig. 6h).

In sum, even when participants were given a cover story suggesting that trials were *in*dependent, participants still treated the trials as *dependent*; they compared the hormone levels in the amoeba in the current trial to those in a different amoeba in the previous trial.

6. Experiment 4

Experiments 4–7 investigated whether people use the temporal strategy when *observing* variables over time. In the introduction we explained how it is possible to learn the direction of causal relations from observations if states of variables are stable, which is impossible according to atemporal strategies for static causal structures. In Experiment 4, we tested whether people use the temporal strategy to learn the direction of a causal relation between two variables.

6.1. Method

6.1.1. Participants

At a main pedestrian crossroad on campus, participants were recruited for a 5 min psychology experiment that paid $2. Thirty-six people participated; they were mainly Yale students, but also included visitors to campus and nearby residents; demographics were not collected. Participants were tested on a laptop computer at the same location.

6.1.2. Stimuli

The datasets for the two conditions are presented in Table 6. The same sixteen trials were used in the two conditions, and the overall contingency, assessed with *phi*, a binary measure of correlation, was 0.5.

The sixteen trials were presented in different orders in the two conditions. In the directional condition, Y sometimes changed on its own (e.g., Trials 1–5). According to the temporal theory, these transitions suggest that an exogenous event changed the state of Y, but because X remained stable, these transitions suggest that Y does not influence X. At other times, both X and Y changed simultaneously

Table 6
Summary of stimuli for Experiment 4.

Trial	$X \rightarrow Y$ directional		Not directional	
	X	*Y*	*X*	*Y*
1	1	**1**	1	1
2	1	**0**	1	1
3	1	**1**	0	0
4	1	**0**	**1**	0
5	1	**1**	**0**	0
6	**0**	0	0	**1**
7	**1**	1	0	**0**
8	**0**	0	1	1
9	0	**1**	1	1
10	0	**0**	0	0
11	0	**1**	0	**1**
12	0	**0**	0	**0**
13	**1**	1	**1**	0
14	**0**	0	**0**	0
15	**1**	1	1	1
16	**0**	0	1	1

Note: 1 represents a "positive mood" and 0 represents a "negative mood". Boldface represents hypothetical exogenous influences on the bolded variable that participants may *infer* to explain the changes from trial to trial. Bold numbers were not denoted in any way for participants.

(e.g., Trials 6–8), suggesting some causal relationship between *X* and *Y*. Overall, because $X \leftarrow Y$ is not plausible, one might infer $X \rightarrow Y$.

In the non-directional condition, *X* and *Y* changed independently of the other equally often (e.g., Trials 3–5, or Trials 5–7), which does not clearly identify one causal direction over the other. Additionally, sometimes both *X* and *Y* changed simultaneously (e.g., Trials 2–3). If people believe that there is some causal relationship between the two but cannot determine the direction, they might infer that both *X* and *Y* influence each other.

6.1.3. Procedures

Participants read the following cover story:

"One interesting social phenomenon is that one person's mood may spread to other people. If you are in a good mood one day, you might put your roommate into a good mood, and vice versa.

In [Fig. 8], you see two roommates. Bill is in a positive mood (+), and Tim is in a negative mood (−).

Please pretend that you are a psychologist studying the moods of roommates. You are trying to figure out if one person's mood influences the other. You might discover that Bill's mood influences Tim's mood, or Tim's mood influences Bill's mood, or both, or that neither influences the other.

In the following scenario, you will observe Bill and Tim's moods over a period of 16 consecutive days.

Please remember that moods influence one another on the same day. For example, if Bill gets into a negative mood on Monday, and if Bill's mood influences Tim's mood, then Tim will also be in a negative mood on Monday."

These instructions were intended to accomplish a number of goals. First, we thought that it would be intuitive for moods to stay stable across a period of days. Second, we hoped that participants would be able to consider possible exogenous influences on people's moods (e.g., getting a good or bad score on a test), which we hypothesized would facilitate the temporal strategy. Finally, we included the instruction that "moods influence one another on the same day" to ensure that we were not testing anything about temporal delay. The stimuli also did not have any pairs of trials in which one person's mood changed on one trial, and then the other person's mood changed on the next trial.

Fig. 8. Stimuli in Experiment 4. *Note*: + represents a positive mood and − represents a negative mood.

Next, participants worked with both the directional and non-directional scenarios in a counterbalanced order. In the directional scenario, the people were named Nick and John, and in the non-directional scenario, the people were named Amy and Lisa (they are simply labeled as *X* and *Y* in Table 6). In each scenario, participants were shown a sequence of 16 screens (e.g., Fig. 8) representing 16 consecutive days. After each screen appeared, there was a delay of 2 s, and then participants were prompted to "Press the spacebar to see the next day," at which point the people's moods on the next day appeared. At the end of the 16 days, participants chose one of the following four options:

"No Relationship means that neither person's mood influences the other.
$X \rightarrow Y$ means that Person *X*'s mood influences Person *Y*'s mood.
$X \leftarrow Y$ means that Person *Y*'s mood influences Person *X*'s mood.
$X \leftrightarrow Y$ means that both people's moods influence each other."

6.2. Results and discussion

Fig. 9 presents the proportion of participants who chose each of the four response options in each condition. A McNemar test on the proportions of two related samples revealed that participants inferred $X \rightarrow Y$ more in the directional than non-directional condition, $p < .01$. As can be seen in Fig. 9, most of the participants who inferred $X \rightarrow Y$ in the directional condition inferred $X \leftrightarrow Y$ in the non-directional condition. Looking only at people who inferred either $X \rightarrow Y$ or $X \leftarrow Y$ in the directional condition, a binomial test confirmed that participants were more likely to infer $X \rightarrow Y$ instead of $X \leftarrow Y$, $p < .01$. In sum, people readily learned the direction of causal relationships from observing stable states over time.

7. Experiment 5

Experiment 5 tested whether people can use the temporal strategy to distinguish a chain, $X \rightarrow Y \rightarrow Z$, *from a common cause structure,* $X \leftarrow Y \rightarrow Z$, when observing stable variables. According to non-temporal strategies, these structures cannot be distinguished because they are in the same Markov class; all three variables are correlated, and *X* and *Z* are conditionally independent given the state of *Y*.

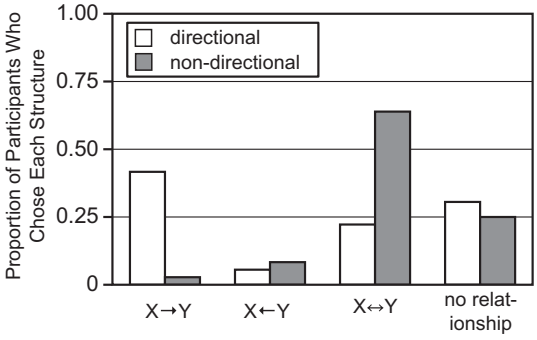


Fig. 9. Proportion of participants who chose each graph per condition in Experiment 4.

Table 7
Summary of stimuli for Experiment 5.

Trial	$X \rightarrow Y \rightarrow Z$ Chain			$X \leftarrow Y \rightarrow Z$ Common cause		
	X	Y	Z	X	Y	Z
1	0	0	0	0	0	0
2	0	0	**1**	0	0	**1**
3	0	0	**0**	0	0	**0**
4	0	**1**	1	**1**	0	0
5	0	**0**	0	**0**	0	0
6	**1**	1	1	1	**1**	1
7	**0**	0	0	1	1	**0**
8	**1**	1	1	1	1	**1**
9	1	1	**0**	**0**	1	1
10	1	1	**1**	**1**	1	1
11	1	**0**	0	0	**0**	0
12	1	**1**	1	1	**1**	1
13	1	1	**0**	**0**	1	1
14	1	1	**1**	0	1	**0**
15	**0**	0	0	**1**	1	0
16	0	0	**1**	1	1	**1**
17	0	0	**0**	0	**0**	0
18	0	**1**	1	**1**	0	0
19	0	1	**0**	1	0	**1**
20	0	**1**	1	**0**	0	1
21	0	**0**	0	0	0	**0**
22	0	0	**1**	1	**1**	1
23	0	0	**0**	**0**	1	1
24	**1**	1	1	**1**	1	1
25	1	**0**	0	0	**0**	0
26	1	0	**1**	**1**	0	0
27	1	0	**0**	**0**	0	0
28	1	**1**	1	0	0	**1**
29	1	1	**0**	0	0	**0**
30	1	1	**1**	1	**1**	1
31	**0**	0	0	1	1	**0**
32	**1**	1	1	1	1	**1**

Note: 1 represents a "positive mood" and 0 represents a "negative mood". Boldface represents hypothetical exogenous influences on the bolded variable that participants may *infer* to explain the changes from trial to trial. Bold numbers were not denoted in any way for participants.

7.1. Method

7.1.1. Participants

There were 28 participants from the same population and tested in the same location as in Experiment 4.

7.1.2. Stimuli

Both conditions had the same set of 32 trials with different orders (Table 7). The 32 trials were determined with the following parameters: $P(C = 1) = .5$, $P(E = 1|C = 1) = .75$, and $P(E = 1|C = 0) = .25$, where C represents a cause and E represents an effect.

In the chain condition $X \rightarrow Y \rightarrow Z$, the trials were ordered such that sometimes Z changed state by itself, sometimes Y and Z changed state together and sometimes X, Y, and Z all changed state together. If participants use the temporal strategy, they might attribute Z changing state by itself (e.g., Trials 2–3) to an exogenous influence on Z, and also infer that Z does not influence X or Y. They might interpret Y and Z changing together (e.g., Trials 4–5) as evidence of an exogenous influence on Y (after all Z does not appear to influence Y), which further influences Z. Finally, they might interpret all three

variables changing state together (e.g., Trials 7–8) as evidence of an exogenous influence on X (the above transitions suggest that neither Y nor Z influences X), which influences Y and Z.

In the common cause condition $X \leftarrow Y \rightarrow Z$, sometimes X and Z changed state by themselves and sometimes all three variables changed state together. If participants use the temporal strategy, they would interpret X and Z changing state by themselves (e.g., Trials 2–3, 4–5) as evidence of an exogenous influence on X or Z, respectively, and also infer that neither X nor Z influences any of the other two variables. Furthermore, they might interpret all three variables changing state simultaneously (e.g., Trials 11–12) as evidence of an exogenous influence on Y, which influences X and Z (because X and Z seem not to influence Y).

7.1.3. Procedure

Participants first read the emotion transmission cover story from Experiment 4, modified to contain three friends. Participants were also told at the beginning about chains (''one person's mood influences a second person's mood, which in turn influences a third person's mood'') and common causes (''there is one main person whose mood influences both other people'') and that their goal was to determine if a chain or common cause best describes these three friends.

To encourage participants to think of moods as stable, participants also read: ''People can also stay in good moods or bad moods for a period of days.'' Finally, participants were also instructed, ''On each day, please consider possible events that influenced people's moods. For example, Allison may have performed well on a test, which put her into a good mood, and spread to Bill. Or, Chris may have done poorly on a test, but his bad mood did not spread to Allison or Bill.'' The reason for this instruction was to encourage them to think about unobserved events that may have manipulated the people's moods, which we believed would facilitate using the temporal strategy (see Section 9).

After reading the instructions, participants worked with both scenarios in a counter-balanced order. From the beginning of each condition, participants were instructed that at the end of the scenario they would choose whether the scenario is best described by the chain (Person $X \rightarrow$ Person $Y \rightarrow$ Person Z) or common cause (Person $X \leftarrow$ Person $Y \rightarrow$ Person Z). The two options were visible during the entire scenario and after the 32 trials participants chose one of the two causal structures.

7.2. Results

Seventy-one percent (20 out of 28) of the participants chose the chain structure in the chain condition, which is significantly above chance according to a binomial test, $p = .01$. Seventy-five percent (21 out of 28) of the participants chose the common cause structure in the common cause condition, $p < .01$. A McNemar test revealed that these two response patterns were significantly different, $p < .01$. In sum, participants were able to distinguish the chain and common cause structures by repeated observations over time even though the two conditions contained the same set of trials and even though the chain and common cause structures are Markov equivalent.

In Experiment 5, we chose the 32 trials based on a joint probability distribution determined by static causal structures. However, a limitation to Experiments 4 and 5 was that we only tested select, hand-made trial orders. We address this limitation in Experiment 6.

8. Experiment 6

In Experiment 6, we sought to investigate whether people use the temporal strategy for learning other structures (chain, common cause, common effect, one link, and no links) from a variety of computer-generated datasets. To accomplish these goals, we compared a ''stable'' vs. ''random'' order of data within each of the five causal structures. The stable conditions were created by simulating dynamic causal structures with stable states. The random conditions were created by randomizing the order of the trials from the stable conditions. If participants use the temporal strategy, they would more likely be able to infer the ''correct'' causal structure that was used to generate the trials in the stable than random condition.

8.1. Method

8.1.1. Participants

Twenty-four students from Yale University participated, either for payment at $10 per hour or for partial fulfillment of an introductory psychology course requirement.

8.1.2. Stimuli

For each of the five causal models, 20 versions of the stable and random orders were created (see Table 8 for an example seen by one participant); Participants 21–24 worked with the same sets of stimuli as Participants 1–4.

The stimuli for the stable conditions were created in the following way. For a given scenario, the computer simulated the temporal functioning of a dynamic causal model with stable states as it underwent 24 interventions; 4 interventions per variable to set its state to 1, and 4 interventions per variable to set its state to 0. To generate each trial, a computer chose an intervention that would change the state of one of the three variables. Once an intervention was chosen, the computer determined the states of the other variables based on the stability heuristics; when a cause was intervened upon, the effects changed, but when an effect was intervened upon, the other variables stayed the same. Even though interventions were used to generate the stimuli, participants did not know which variable was intervened upon; they merely observed the variables over time.

There was one other criterion for choosing interventions; a causal variable was only manipulated if its effect variable was not already at ceiling or floor. For example, for the one link condition [$X \rightarrow Y, Z$], X would only be set to 1 if Y was previously set at 0. Avoiding ceiling and floor effects was intended to ease the causal structure learning process and provide the most sensitive test of whether people use this temporal strategy. For example, a transition from ($x = 0$, $y = 1$) to ($x = 1$, $y = 1$) is ambiguous; X could influence Y but Y was already at ceiling, or it could be that X and Y are unrelated.

Below we describe the characteristic types of transitions that participants observed in the *stable* conditions. In the common effect condition $X \rightarrow Y \leftarrow Z$, sometimes Y changed by itself, sometimes X and Y changed together, and sometimes Z and Y changed together. (Additionally, sometimes X and Z changed independently because ceiling and floor effects could not always be avoided.) In the one link condition [$X \rightarrow Y, Z$], sometimes Y and Z changed individually, and sometimes X and Y changed together. In the no links condition, each variable always changed by itself. The characteristic transitions for the chain and common cause were described in Experiment 5.

The random conditions comprised the exact same 24 trials as in the dependent conditions, except the order of the trials was randomized. This randomization resulted in violations of the stability heuristics. For example, on Trial 7 in the random chain $X \rightarrow Y \rightarrow Z$ condition in Table 8, X went into a bad mood while Y and Z stayed in a good mood, which suggests that X does not influence Y or Z.

Because the stimuli were created by simulating dynamic causal structures, they do not necessarily uphold the Markov condition (see Section 4), so atemporal strategies would not necessarily learn the "correct" causal relations. Most importantly, however, atemporal strategies would infer the same causal structure in the stable and random conditions. Any difference in the inferences between the two conditions would suggest that participants used the temporal strategy.

8.1.3. Design and procedure

The same cover story about mood transmission was used. Participants were told that their goal was to "figure out which people's moods influence which other people's moods." One additional change was made in the cover story. Because we did not want the cover story about "friends" to bias participants to only infer positive causal relationships, participants were told that people could be "frenemies" and "be in a good mood because another person is in a bad mood."

After reading through the instructions, participants worked with one example scenario and then 10 test scenarios; 5 causal structures $\times$ 2 orders (stable vs. random). The scenarios were grouped into the stable vs. random scenarios; half the participants received the stable scenarios first and the other half received the random scenarios first. The scenarios within each block were randomized in a Latin square design such that each scenario sometimes appeared first.

Table 8
Summary of one participant's stimuli for Experiment 6.

Trial	$X \rightarrow Y \rightarrow Z$ Chain						$X \leftarrow Y \rightarrow Z$ Common cause						$X \rightarrow Y \leftarrow Z$ Common effect						$X \rightarrow Y\ Z$ One link						$X\ Y\ Z$ No link					
	Stable			Random			Stable			Random			Stable			Random			Stable			Random			Stable			Random		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	0	0	**1**	1	1	1	0	0	**1**	0	0	1	0	1	**1**	0	0	0	0	0	**1**	1	1	0	**1**	0	0	1	0	0
2	0	0	**0**	0	0	0	**1**	0	1	1	0	1	0	0	**0**	1	1	0	0	0	**0**	1	1	1	1	**1**	0	0	1	1
3	**1**	1	1	0	0	1	1	0	**0**	0	0	0	0	1	**1**	0	0	1	0	0	**1**	0	0	0	1	**0**	0	1	1	0
4	1	1	**0**	0	0	0	**0**	0	0	1	0	1	0	**0**	1	0	0	0	**1**	1	1	0	0	1	1	**1**	0	1	0	0
5	1	1	**1**	0	1	1	0	0	**1**	1	1	1	0	**1**	1	0	1	1	**0**	0	1	1	1	1	1	1	**1**	0	0	1
6	1	**0**	0	1	1	1	**1**	0	1	0	0	0	0	**0**	1	0	0	1	0	0	**0**	0	0	0	1	1	**0**	1	1	0
7	1	**1**	1	0	1	1	**0**	0	1	1	0	1	**1**	1	1	0	0	0	0	0	**1**	1	1	0	1	1	**1**	1	1	1
8	**0**	0	0	1	1	1	**1**	0	1	1	1	1	1	1	**0**	1	1	0	0	0	**0**	0	0	0	1	1	**0**	1	0	0
9	0	**1**	1	1	1	1	1	0	**0**	1	1	1	1	**0**	0	0	0	0	**1**	1	0	1	0	0	1	1	**1**	1	1	0
10	0	1	**0**	0	0	0	1	0	**1**	0	0	1	1	**1**	0	0	1	0	1	1	**1**	0	0	1	1	**0**	1	0	1	0
11	0	1	**1**	0	1	0	**0**	0	1	1	0	0	1	1	**1**	0	0	0	1	1	**0**	0	0	0	**0**	0	1	1	0	0
12	0	**0**	0	0	0	0	**1**	0	1	0	0	1	1	1	**0**	1	1	1	1	**0**	0	0	0	0	0	0	**0**	1	1	0
13	0	**1**	1	0	0	0	**0**	0	1	1	0	1	**0**	0	0	0	1	1	1	**1**	0	0	0	0	**1**	0	0	0	0	0
14	0	1	**0**	0	1	1	0	0	**0**	0	0	0	0	1	**1**	1	1	1	**0**	0	0	1	1	0	**0**	0	0	0	0	0
15	0	1	**1**	0	0	0	1	**1**	1	1	1	1	0	0	**0**	1	1	0	0	**1**	0	1	1	0	**1**	0	0	0	0	0
16	0	**0**	0	0	1	1	1	1	**0**	0	0	1	**1**	1	0	0	1	1	0	**0**	0	0	0	0	**0**	0	0	1	0	1
17	**1**	1	1	1	1	0	1	1	**1**	1	1	0	**0**	0	0	1	1	0	0	**1**	0	1	1	0	0	**1**	0	0	0	1
18	**0**	0	0	0	1	0	0	**0**	0	0	0	1	**1**	1	0	1	1	0	0	**0**	0	0	1	0	**1**	1	0	0	0	0
19	0	**1**	1	0	0	0	1	**1**	1	1	1	1	**0**	0	0	0	1	1	0	**1**	0	0	1	0	1	**0**	0	1	1	1
20	0	**0**	0	1	1	1	0	**0**	0	1	0	1	**1**	1	0	0	0	0	0	**0**	0	0	0	0	**0**	0	0	1	1	0
21	**1**	1	1	0	1	1	1	**1**	1	0	0	0	1	**0**	0	1	1	0	**1**	1	0	0	0	0	0	0	**1**	1	1	1
22	**0**	0	0	1	1	1	0	**0**	0	0	0	0	1	**1**	0	1	1	0	**0**	0	0	0	0	1	0	**1**	1	0	1	0
23	**1**	1	1	1	0	0	1	**1**	1	0	0	0	**0**	0	0	1	0	0	**1**	1	0	0	0	1	0	1	**0**	1	0	0
24	**0**	0	0	0	0	0	0	**0**	0	1	0	0	0	**1**	0	1	0	0	**0**	0	0	0	1	0	0	**0**	0	0	0	0

Note: 1 represents a "positive mood" and 0 represents a "negative mood". Boldface represents hypothetical exogenous influences on the bolded variable that participants may *infer* to explain the changes from trial to trial. Bold numbers were not denoted in any way for participants. No bold numbers were included for random conditions because there is not a consistent interpretation of hypothesized exogenous influences that could explain the changes in the states of the variables.

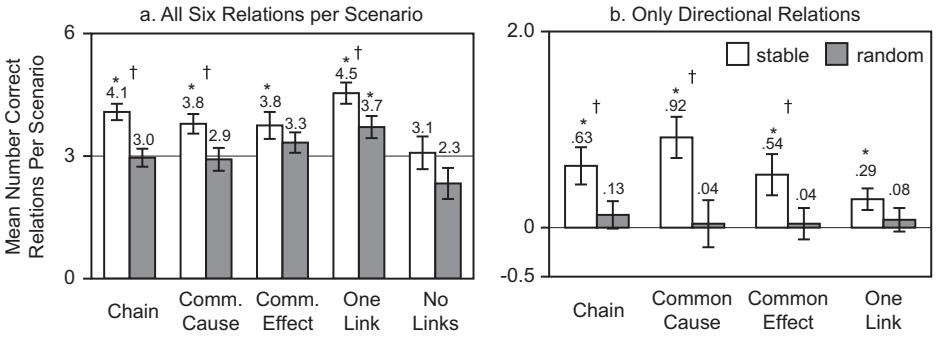


Fig. 10. Mean (std. error) number of correct causal relations inferred in Experiment 6. *Note*: *Mean is significantly greater than chance. †Means are significantly different.

The flow of each scenario was the same as Experiments 4 and 5. At the end of each scenario, participants chose which people's moods they thought influenced which other people's moods using the same procedure in Experiments 1–3; participants could choose any combination of the six potential causal links between the three people.

8.2. Results and discussion

The basic patterns of effects held regardless of whether participants worked with the stable or random conditions first, so we collapsed across this factor. We first report the total number of correct links endorsed out of the six links for a given causal structure, the same measure used in Experiments 1–3 (see Fig. 10a). Participants chose more correct links in the stable than random conditions for the chain, $t(23) = 3.87$, $p < .01$, common cause, $t(23) = 2.26$, $p = .03$, and one link structures, $t(23) = 2.50$, $p = .02$. There was not a significant difference for the common effect structure, $t(23) = 1.02$, $p = .32$, and the no links structure, $t(23) = 1.74$, $p = .10$.[2]

A second set of analyses was run only looking at pairs of variables that had a direct causal relationship (Fig. 10b). For example, for the common effect condition $X \rightarrow Y \leftarrow Z$, there is no direct link between X and Z, so only the X–Y pair and the Y–Z pair were assessed. The reason for this analysis was to ensure that the differences between the stable and random conditions are not just due to differences in the links that participants inferred between pairs of variables without direct causal relations. Each link in a causal structure was coded as 1 if a participant endorsed the correct direction, 0 if a participant either endorsed both directions or endorsed neither direction, and -1 if a participant endorsed the opposite direction. Thus, scores for the chain, common cause, and common effect structures could range from +2, if a participant endorsed both of the correct links and no other links, to -2, a participant inferred the exact opposite direction for the two links. For the one link condition, scores could range from +1 to -1. Zero reflects chance responding or an inability to infer a causal direction.

Overall, summing across the four causal structures, participants were more likely to identify the correct causal links in the stable condition ($M = 2.38$, $SD = 2.28$) than the random condition ($M = 0.29$, $SD = 1.80$), $t(23) = 3.52$, $p < .01$. Participants in the stable condition consistently inferred the direction of links as predicted by the temporal strategy $t(23) = 5.10$, $p < .01$. Participants in the random condition were ambivalent about the direction of the causal links and not significantly different from chance, $t(23) < 1$.

[2] Furthermore, participants were above chance in choosing correct links in all the stable conditions, $ts(23) > 2.26$, $ps < .04$, except no links, $t < 1$. Out of the random conditions, the only condition above chance was the one link structure, $t(23) = 2.66$, $p = .01$. Participants were not different from chance for the random chain, common cause, and common effect conditions, $ts(23) < 1.32$, $ps > .20$. Participants were marginally below chance for the no links random condition, $t(23) = 1.74$, $p = .10$.

The stable and random conditions within each causal structure were compared using non-parametric Wilcoxon Signed-Ranked tests because of the limited range of values (Fig. 10b). Participants inferred the correct directions of causal links more in the stable than random conditions for the chain, $p = .04$, common cause, $p = .02$, and common effect structures, $p = .04$. There was not a significant difference for the one link structure $p = .24$ (but see Sections 6 and 9).[3]

One interesting finding was that there was no significant difference between the stable vs. random, no link conditions. Surprisingly, participants did infer some links in the stable conditions, perhaps possibly due to delay cues. For example, in the no-links stable condition in Table 8, *Y* changed in Trial 4 and *Z* changed in Trial 5, potentially suggesting that *Y* influenced *Z*.

In sum, participants were for the most part able to learn directional causal structures even in an unrestricted learning paradigm and with a wide variety of orders of data.

9. Experiment 7

Experiment 7 tested one of our core assertions of how people are able to learn directional causal structures from observations over time, that when people observe a change in a variable, they infer that some exogenous influence must have produced the change (see also Table 3 in the Introduction). Inferring exogenous influences is critical for the following reason. Suppose one is observing *X* and *Y* over time. Transitions when *X* and *Y* change simultaneously is evidence that some causal relationship exists between the two variables, but this does not identify which variable is the cause. However, if people have some reason to believe that external influences on *Y* do not get transferred to *X*, then when *X* and *Y* change simultaneously, they might infer that there was an external influence on *X* that transferred to *Y*.

In Experiment 7, we primed half the participants to think about exogenous influences. These participants first worked with five scenarios in which they had the opportunity to intervene on the variables (their own interventions were exogenous influences on the causal structures), and later worked with five observation test scenarios. The other group of participants worked with two sets of five observation scenarios; the second half of the five observation scenarios was the control condition. We compared the last five observational scenarios across groups to test whether the participants who were primed to think about external influences (interventions) later inferred directional causal relationships according to the temporal strategy more than participants who had worked only with observational scenarios.

9.1. Method

9.1.1. Participants

There were 20 participants from the same population.

9.1.2. Design

The experiment was a 2 priming (intervention then observation [Int–Obs] vs. observation then observation [Obs–Obs]) $\times$ 5 causal structures design. The results of the second phases of the two conditions were compared; the data collected in the first half of the experiment were not analyzed.

9.1.3. Stimuli

The stimuli for the intervention condition was exactly the same as the stable conditions in Experiment 1 (see Table 4).

The observation conditions had 32 trials (see Table 9) which were chosen by making joint probability tables with the following parameters for which *C* represents a cause and *E* represents an effect: $P(C = 1) = .5$, $P(E = 1|C = 1) = .75$, $P(E = 1|C = 0) = .25$. For the common effect scenario, the two causes combined in a roughly noisy-or parameterization to produce the effect; the effect was present with

[3] Furthermore, One-Sample Wilcoxon Signed-Ranked tests reveal that for all four of these stable conditions, participants inferred the directional links above chance, $ps \leqslant 0.02$. For the random conditions, participants were not different from chance, $ps \geqslant 0.48$.

Table 9
Summary of observational stimuli in Experiment 7.

Trial	$X \rightarrow Y \rightarrow Z$ Chain			$X \leftarrow Y \rightarrow Z$ Common cause			$X \rightarrow Y \leftarrow Z$ Common effect			$X \rightarrow Y\ Z$ One link			$X\ Y\ Z$ 0 Link		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	**1**	0	0	**1**	**1**	1	0	0	**1**	0	0	0	**1**
3	0	0	**0**	0	0	**0**	**0**	0	0	0	**0**	0	0	0	**0**
4	0	**1**	1	**1**	0	0	0	1	**1**	0	**1**	0	0	0	**1**
5	0	**0**	0	**0**	0	0	0	0	**0**	0	**0**	0	0	0	**0**
6	**1**	1	1	1	**1**	1	**1**	1	0	**1**	1	0	0	0	**1**
7	**0**	0	0	1	1	**0**	1	1	**1**	**0**	0	0	0	0	**0**
8	**1**	1	1	1	1	**1**	**0**	1	1	**1**	1	0	0	0	**1**
9	1	1	**0**	**0**	1	1	0	0	**0**	**0**	0	0	**1**	0	1
10	1	1	**1**	**1**	1	1	0	**1**	0	**1**	1	0	1	**1**	1
11	1	**0**	0	0	**0**	0	0	**0**	0	**0**	0	0	1	**0**	1
12	1	**1**	1	1	**1**	1	0	1	**1**	1	**1**	0	1	**1**	1
13	1	1	**0**	**0**	1	1	**1**	1	1	1	**0**	0	1	**0**	1
14	1	1	**1**	0	1	**0**	**0**	1	1	1	**1**	0	1	**1**	1
15	**0**	0	0	**1**	1	0	0	**0**	1	1	**0**	0	1	**0**	1
16	0	0	**1**	1	1	**1**	0	**1**	1	1	**1**	0	1	**1**	1
17	0	0	**0**	0	**0**	0	0	**0**	1	1	1	**1**	**0**	1	1
18	0	**1**	1	**1**	0	0	**1**	1	1	1	**0**	1	0	1	**0**
19	0	1	**0**	1	0	**1**	**0**	1	1	1	**1**	1	0	1	**1**
20	0	**1**	1	**0**	0	1	0	0	**0**	1	**0**	1	0	1	**0**
21	0	**0**	0	0	0	**0**	**1**	1	0	1	**1**	1	0	1	**1**
22	0	0	**1**	1	**1**	1	1	1	**1**	**0**	0	1	0	1	**0**
23	0	0	**0**	**0**	1	1	1	**0**	1	**1**	1	1	0	1	**1**
24	**1**	1	1	**1**	1	1	1	**1**	1	**0**	0	1	0	1	**0**
25	1	**0**	0	0	**0**	0	1	1	**0**	**1**	1	1	**1**	1	0
26	1	0	**1**	**1**	0	0	1	1	**1**	**0**	0	1	1	**0**	0
27	1	0	**0**	**0**	0	0	1	0	**0**	**1**	1	1	1	**1**	0
28	1	**1**	1	0	0	**1**	1	**1**	0	**0**	0	1	1	**0**	0
29	1	1	**0**	0	0	**0**	1	**0**	0	0	**1**	1	1	**1**	0
30	1	1	**1**	1	**1**	1	1	1	**1**	0	**0**	1	1	**0**	0
31	**0**	0	0	1	1	**0**	1	1	**0**	0	**1**	1	1	**1**	0
32	**1**	1	1	1	1	**1**	**0**	0	0	0	**0**	1	1	**0**	0

Note: 1 represents that a hormone was present, and 0 absent. Boldface represents hypothetical exogenous influences on the bolded variable that participants may *infer* to explain the changes from trial to trial. Bold numbers were not denoted in any way for participants.

a probability of 1/8 if neither cause was present, 3/4 if one cause was present, and 7/8 if both causes were present. Most critically, the trials were ordered to uphold the stability heuristics. (See Sections 7 and 8 for a more detailed explanation of the characteristic transitions of the different causal structures.)

9.1.4. Procedures

Participants worked with the cover story about three hormones in an amoeba similar to Experiment 2. Participants were first randomly assigned to the "Intervention then Observation" (Int–Obs) or the "Observation then Observation" (Obs–Obs) condition. (See Table 10 for the procedural flows of these two conditions.)

For the Intervention phase in the Int–Obs condition, participants were told that they would "first intervene to produce or suppress one hormone by injecting chemicals into the amoeba,..." "observe whether the hormones are present or absent,..." and afterwards "intervene on the hormones again." The purpose of the intervention phase was simply to get participants thinking about the idea of an intervention and how interventions can discriminate causal structures. Participants were not told in any way that they should use any of the knowledge or experience from the intervention phase during the observation phase. Within and across phases, the hormones that were causes and effects changed

Table 10
Procedural flow of Experiment 7.

Intervention then Observation Condition	Observation then Observation Condition
• Intervention cover story	• Observation cover story
• 2-Variable intervention practice scenario	• 2-Variable observation practice scenario
• 5 Intervention scenarios	• 5 Observation scenarios
• Observation cover story	• 5 Observation scenarios
• 2-Variable observation practice scenario	
• 5 Observation scenarios	

location on the screen. In sum, there was no way for participants to directly transfer knowledge about how specific hormones interact from the intervention phase to the observation phase.

For the observation phases, participants were told that they would "observe hormones within one amoeba over a period of 32 h" and that they were essentially "fast-forwarding through a time-lapse video, stopping at each hour to check the status of the hormones." This cover story was designed so that participants understood that they were working with the states of hormones over time within one amoeba. The procedures for the observation blocks were essentially the same as in Experiments 4–6.

9.2. Results and discussion

The main question is whether participants in the observation phase of the Int–Obs condition used the temporal strategy to infer causal structures more than the participants in the second half of the Obs–Obs condition. Out of all six links (Fig. 11a), participants in the Int–Obs condition inferred more correct links than participants in the Obs–Obs condition for all the causal structures, all $ts(18) > 2.53$, all $ps \leqslant 02$. All five of the Int–Obs conditions were above chance, $ts(9) > 3.99$, $ps < .01$. However, out of the Obs–Obs conditions, only the one-link condition was above chance (3), $t(9) = 4.99$, $p < .01$; all the rest of the Obs–Obs conditions were not significantly different than chance, $ts(9) < 1.8$, $ps \geqslant 0.1$.

The low performance in the Obs–Obs conditions was not unexpected. In Experiments 4–6, the cover story about emotion transmission was used so that participants could easily think about unobserved causes that impacted people's emotions. However, unobserved factors that produce changes in hormones over time are not necessarily so obvious and may be the reason for the worse performance in this experiment. This explanation fits with the point of the current experiment that thinking about unobserved factors responsible for changes in observed variables facilitates the ability to learn causal directions.

As in Experiment 6, we ran another analysis only of pairs of variables with direct links (Fig. 11b). For each of the four causal structures, Mann–Whitney–Wilcoxon tests for independent samples re-

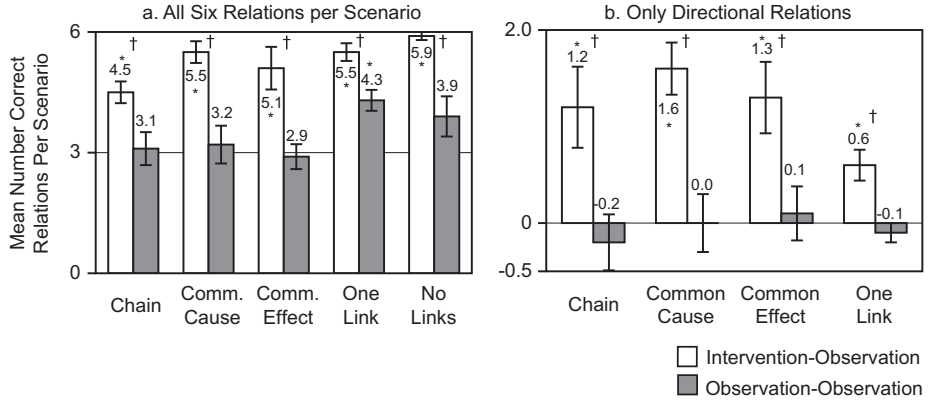


Fig. 11. Mean (std. error) number of correct causal relations inferred in Experiment 7. *Note*: *Mean is significantly greater than chance. †Means are significantly different.

vealed that participants were more likely to infer the direction of the causal relations predicted by the temporal strategy in the Int–Obs condition than the Obs–Obs condition, all $ps \leqslant 0.04$. One-sample Wilcoxon Signed-Ranked tests revealed that participants inferred the direction of the causal relations above chance for all the Int–Obs conditions, all $ps < 0.02$, but participants in the Obs–Obs conditions did not, all $ps > .31$.

In sum, participants in the Int–Obs condition frequently endorsed the directional causal links suggested by the temporal strategy. However, participants in the Obs–Obs condition often did not show consensus in inferring a causal direction. This experiment demonstrates that thinking about unobserved factors responsible for producing changes in a causal system facilitates learning a causal structure from repeated observations over time.

10. General discussion

Regardless of whether one is reasoning about emotions, stock prices, or the weather, the states of variables are typically influenced by prior states. Though this observation is simple and intuitive, most theories of causal learning have focused on cases with independent trials and have ignored this important issue. The current experiments demonstrate that the temporal stability of variables allows people to infer complex causal structures that they otherwise have difficulty learning.

Experiments 1–3 examined how people learn causal structures from repeated *interventions*. If participants intervened upon X and Y simultaneously changed, they inferred that X caused Y. Such a strategy assumes that variables typically stay stable and do not coincidentally change at the exact moment of an intervention. Experiments 1 and 2 demonstrated that people learn causal structures quickly and reliably when variables are stable but not unstable. Experiment 3 demonstrated that this temporal strategy seems to be a default; people make inferences about changes of variables from trial to trial even if the cover story suggests that consecutive trials represent completely independent entities.

Experiments 4–7 examined how people learn causal structures from *observing* states over time. Existing theories suggest that "Markov equivalent" causal structures (e.g., $X \rightarrow Y$ vs. $X \leftarrow Y$; $X \rightarrow Y \rightarrow Z$ vs. $X \leftarrow Y \rightarrow Z$) cannot be distinguished unless one variable occurs before the other; "correlation does not imply causation." However, in Experiments 4–6, we found that people distinguish a wide range of structures when observing states over time. People assumed that effects may change state due to an exogenous factor without their observed causes changing simultaneously, but when causes change state they usually produce a change in their effects. Experiment 7 demonstrated that thinking about possible exogenous influences on the observed variables facilitates this temporal strategy.

10.1. Scope and compatibility with other theories

Most of this paper has focused on distinguishing the proposed learning strategy from existing theories. Though there are important differences, they are often compatible with other strategies and causal structure learning may often be over-determined (Lagnado et al., 2007).

In a sense, Experiments 1–3 can be viewed as an elaboration of Fernbach and Sloman's (2009) proposal that if X is intervened upon, correlated changes in Y will be viewed as caused by X. However, Fernbach and Sloman (2009; Experiment 1) only tested the influence of this strategy on inferring mediator variables in chains of punctate events. In scenarios with stable states as investigated here, this strategy has much more dramatic influences on a wide variety of structures. The current experiments are also compatible with theories suggesting that temporal delay is a cue to learning causal structures (e.g., Hagmayer & Waldmann, 2002; Lagnado & Sloman, 2004, 2006). Our experiments used scenarios when there is no delay. However, in the real world, both cues may be present simultaneously. For example, after intervening on X, there might be a delay, then Y might change state and stay in the new state.

Mayrhofer and Waldmann (2011) found that people often assume that causes are sufficient to produce their effects, which allows them to infer the direction of causal relationships from pure observations. This heuristic has similarities to our proposal that when a cause changes it will usually

produce a change in its effect. However, there are also some important differences. Whereas the heuristic proposed here is about temporal change over time, Mayerhofer and Waldmann's is about punctate variables. In fact, in our stimuli (e.g., Experiment 4), causes were not sufficient. Another difference is that our stimuli often involved two alternative states (positive mood vs. negative mood) rather than presence vs. absence, and it is unclear how a sufficiency heuristic would apply to two alternative states.

In Experiment 3, we found that people seem to use the temporal strategy even when the trials were framed as independent. However, Steyvers et al. (2003) have shown that people can use the independent-events strategy to learn causal structures. An aspect of their paradigm likely encouraged understanding that the trials were independent. The variables were categorical with many possible levels, so from trial to trial, it would have been very unlikely for the variables to stay stable. With binary variables, it may be harder to learn that trials are independent because from trial to trial, each variable has a good chance of staying the same. Furthermore, the gambler's fallacy suggests that people tend to think that sequential trials are statistically dependent even when they are independent (see Oskarsson et al. (2009) for a review). In sum, features of the scenario may bias people's ability to understand trials as being independent.

Though we have shown that people used the temporal strategy in a number of experiments, they did not always use it. In Experiment 7, participants only used the temporal strategy after they had first gone through an intervention phase priming them to think about possible exogenous factors that produce changes in the observed variables. Experiment 7 used more abstract stimuli involving hormones in amoebas rather than emotions in people (Experiments 4–6). Participants may have had a harder time reasoning about exogenous factors responsible for the changes in the hormones or they may have had certain prior beliefs about the stability of emotions but not hormones. The temporal strategy may depend upon having certain types of background knowledge about the scenario.

10.2. Open questions and limitations

There are several open questions with regard to how broadly people use the strategies proposed here. First, can people use temporal strategies to learn probabilistic causal relationships? By probabilistic, we mean that when a cause changes its effects *sometimes* change. For interventions, probabilistic causal relationships likely will not pose a challenge so long as variables are stable over time. However, when observing variables over time with probabilistic relations, the temporal strategy might not always lead to useful inferences. Consider $X \rightarrow Y$ with a weak relationship and suppose that there are few exogenous influences that change the state of Y. In such instances it is possible that X would change more frequently than Y and the temporal strategy might even lead to inferring the wrong causal direction. However, people may use this heuristic if they assume that causes are generally strong (Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008).

Second, in the current manuscript we only investigated cases in which variables stayed stable until the next intervention. However, there are many other types of temporal relationships. For example, after going on vacation, a person's anxiety is likely to be lowered temporarily; the tables in the introduction oversimplify this point. Continuous variables also allow for the possibility of other sorts of dynamic environments, such as variables that slowly fluctuate or variables that decay at a constant rate (e.g., Hagmayer et al., 2010). It is likely that many of these higher order parameters (e.g., whether variables are temporally dependent, if so, what are the temporal dynamics of the system) would both guide how causal relations are learned and could also be learned themselves (Griffiths & Tenenbaum, 2009).

Third, in some of these studies we collected informal debriefings. The main finding was that participants often found the tasks much harder in the unstable, random, and many-entity conditions. Some participants were able to verbalize strategies consistent with those proposed here. For example, one participant wrote, "Person 1 . . . seemed to have fairly stable moods . . . . However, when Person 1's moods switched, Person 2's switched along with him. This would lead one to believe that Person 1's moods impact Person 2's more." However, it would be desirable in future studies to use other verbal measures such as think-aloud protocols for convergent evidence of use of temporal strategies.

10.3. Implications for prior research

In many experiments with punctate events (e.g., Gopnik et al.'s puppet machine, 2004), both the atemporal and temporal strategies predict the same learning. For example, imagine a scenario in which whenever one activates *X*, *Y* becomes temporarily activated. Both strategies suggest that *X* influences *Y*. In terms of the atemporal strategy, when *X* is manipulated, *Y* is active, and when *X* is not manipulated, *Y* is inactive. In terms of the temporal strategy, *before X* manipulated, *Y* is inactive, and *after X* is manipulated, *Y* becomes active. For this reason, in the current experiments we used stable states and changed the orders of the trials, in which case the two strategies make different predictions. The results of the current studies demonstrate a clear need for a more systematic investigation of the strategies that people use to learn causal structures in various scenarios.

10.4. Implications for models of causal learning

Future research should develop models of causal structure learning over time. The temporal models we have developed differ fundamentally from most existing atemporal models of causal learning because our models make inferences based on transitions between trials. The fundamental unit of inference is not a punctate trial, but how variables change over time.

In the Appendix, we present a heuristic model for learning causal direction from observations. Though this heuristic model does correlate with participants' inferences, the modest correlations suggest that it does not capture all the reasoning going on. We have also developed a rational model to learn causal direction from observations (see Rottman & Keil, 2011). However, the rational model is considerably more complex than the heuristic model, and did not explain participants' inferences better than the heuristic model, so we do not present it here.

10.5. Conclusion

The bulk of previous research on causal learning has investigated how people learn causal relationships from atemporal, independent events. Yet, in the real world, variables often are stable for periods of time, and assuming that most objects have stable dispositions may be an adaptive strategy to help prune down a huge set of possible causes (e.g. Elga, 2007). Together, these experiments demonstrate that there is rich structure in how events unfold over time, and people readily identify the structure in these temporal patterns. Given the importance of causal reasoning for prediction and action, and given that we experience the world temporally, the abilities demonstrated here may reflect a fundamental way in which we develop representations of the causal structure of the world. For learning causal relationships, temporal reasoning may be the norm rather than the exception.

Acknowledgments

This research was supported by an NSF Graduate Research Fellowship (Rottman) and NIH Grant R37 HD023922 (Keil). The authors thank Eric Taylor for helpful comments.

Appendix A. Heuristic model of learning causal direction from observations over time

The input of the model are the sequential pairs of trials, "transitions," observed by participants. The model produces two scores for each pair of variables that reflect how likely it is that one variable causes the other. For a scenario with only two variables, *X* and *Y*, the model would produce a score for the two links $X \rightarrow Y$ and $X \leftarrow Y$. These scores have no limits; the higher the score the more likely that the link exists.

For each transition, the model runs the following function. If both variables changed (e.g., $[x = 0, y = 0]$ at t_0 to $[x = 1, y = 1]$ at t_1), then the model adds 1 to the scores for both links because this is evidence of some causal relationship. If one variable changed (e.g., $[x = 0, y = 0]$ at t_0 to $[x = 0, y = 1]$ at t_1) then the model subtracts 1 from the structure in which the variable that changed is the cause, in this

case $X \leftarrow Y$. The reason is that when Y changed to 1, X failed to change, which suggests that Y does not influence X. Finally, there is one exception to the above rule. If one variable was already at ceiling or floor when the other changed, then the scores are left unchanged. For example, if the variables change from $[x = 0, y = 1]$ at t_0 to $[x = 0, y = 0]$ at t_1, when Y changed to 0, X was already at 0, so this trial is not evidence against $X \leftarrow Y$. This model can be generalized for larger numbers of variables by considering each pair of variables at each transition.

A.1. Experiment 4

In the directional condition, the model gave $X \rightarrow Y$ a higher score (7) than $X \leftarrow Y$ (3). In the non-directional condition, both $X \rightarrow Y$ and $X \leftarrow Y$ had a score of 2.

A.2. Experiment 5

For the chain $X \rightarrow Y \rightarrow Z$, the model gave a higher score for $X \rightarrow Y$ (7) than $X \leftarrow Y$ (3) and a higher score for $Y \rightarrow Z$ (15) than $Y \leftarrow Z$ (7). The model also gave a higher score for $X \rightarrow Z$ (7) than $X \leftarrow Z$ (−5). For the common cause condition $X \leftarrow Y \rightarrow Z$, the model gave higher scores for $X \leftarrow Y$ (7) than $X \rightarrow Y$ (1) and a higher score for $Y \rightarrow Z$ (7) than $Y \leftarrow Z$ (1). The model also had a somewhat higher score for $X \leftarrow Z$ (3) than $X \rightarrow Z$ (−1).

A.3. Experiments 6

Experiment 6 tested whether people could learn five different three-variable structures by comparing conditions with either a stable or random order, and each participant worked with unique sets of data. Thus, Experiment 6 presented the best opportunity to test the model under a variety of orders. Both participants' responses and the output of the model comprised six numbers corresponding to the likelihood of each of the six causal relationships (participants responses were binary whereas the model produced continuous predictions). For each participant and each condition [5 causal structures × 2 (stable vs. random)], we calculated the Fisher-transformed correlation between the participant's six responses and the likelihood of the six links inferred by the model. (For participants who inferred that all links were present or all links were absent for a given condition, no correlation can be calculated because there is zero variance. In such cases a correlation of zero was used.) These Fisher-transformed correlation coefficients were used for inferential statistics. The inversely transformed averages of these correlations are presented in Fig. A1.

In the stable conditions, the average correlations between participants' causal structure inferences and the model were significantly above zero for four of the five conditions, $ts(23) \geqslant 2.28$, $ps \leqslant .03$. As discussed in Experiment 6, participants did not appear to use the temporal strategy in the stable

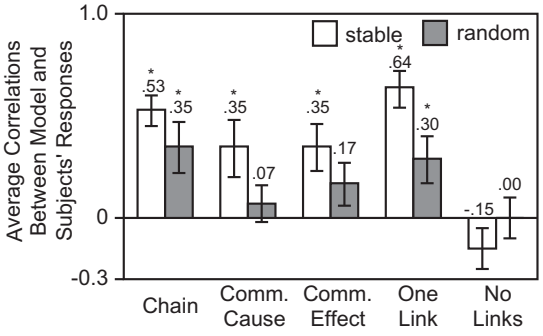


Fig. A1. Average correlations (std. error) between model simulations and participants' responses in Experiment 6. *Note*: *Correlation is significantly greater than 0, $p < .05$.

no-links condition, and there was not a significant correlation between participants' inferences and the model, $t(23) = 1.40$, $p > .10$.

In the random conditions, participants' inferences were significantly correlated with the model for the chain, $t(23) = 2.58$, $p = .02$, and one link, $t(23) = 2.44$, $p = .02$. The other three conditions were not, $ps > .10$. Participants' inferences in the random conditions appeared to be less correlated with the model than those in the stable conditions. It would make sense that participants would use the temporal strategy less in the random conditions because the temporal patterns in the data did not consistently identify one causal structure. To test this possibility, we ran a 5 (causal structure) $\times$ 2 (stable vs. random) repeated-measures ANOVA, which revealed a main effect of stability, $F(1,23) = 8.87, p < .01, \eta_p^2 = .28$. Follow-up tests revealed that when comparing the stable vs. random conditions within a causal structure, only the one-link condition was significantly different, $t(23) = 2.56$, $p = .02$. There was also a main effect of causal structure in the ANOVA, $F(4,92) = 9.44$, $p < .01, \eta_p^2 = .29$; participants' inferences were closer to those of the model in for some conditions than others. There was no interaction between stability and causal structure.

In sum, these correlations suggest that participants may have used a strategy similar to the proposed model in some conditions. However, the modest correlations suggest that participants may have also been influenced by other strategies, particularly in the random and no link conditions.

We do not discuss simulations for Experiment 7 because the conditions are qualitatively the same as Experiments 5 and 6.

References

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*(3), 211–241.

Dienes, Z., & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 848–862.

Dienes, Z., & Fahey, R. (1998). The role of implicit memory in controlling a dynamic system. *Quarterly Journal of Experimental Psychology Section A, 51*(3), 593–614.

Elga, A. (2007). Isolation and folk physics. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited* (pp. 106–119). Oxford: Oxford UP.

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

Friedman, N., Murphy, K., & Russell, S. (1998). Learning the structure of dynamic probabilistic networks. In *Conference on uncertainty in artificial intelligence*. Madison, WI: Morgan Kaufmann.

Glymour, C. N., & Cooper, G. F. (1999). *Computation, causation, and discovery*. Menlo Park, CA: AAAI Press.

Gonzalez, C., Vanyukov, P., & Martin, M. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior, 21*(2), 273–286.

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), 3–32.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*(4), 661–716.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., & Lagnado, D. (2010). Spontaneous causal learning while controlling a dynamic system. *The Open Psychology Journal, 3*, 145–162.

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.). *Causal learning: Psychology, philosophy, and computation* (pp. 86–100). Oxford: Oxford UP.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory and Cognition, 30*(7), 1128–1137.

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), 451–460.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.). *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford, UK: Oxford UP.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*(4), 955–984.

Lunt, P. K. (1988). The perceived causal structure of academic failure. *British Journal of Social Psychology, 27*, 171–179.

Lunt, P. K. (1989). The perceived causal structure of unemployment. In F. Olander & K. Grunert (Eds.), *Understanding economic behaviour* (pp. 107–120). Amsterdam: Kluwer.

Lunt, P. K. (1991). The perceived causal structure of loneliness. *Journal of Personality and Social Psychology, 61*(1), 26–34.

Mayrhofer, R., & Waldmann, M. R. (2011). Heuristics in covariation-based induction of causal models: Sufficiency and necessity priors. In C. H. Carlson & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 3110–3115). TX: Austin.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory and cognition, 37*(3), 249–264.

Neapolitan, R. E. (2004). *Learning Bayesian networks*. Upper Saddle River, NJ: Pearson Prentice Hall.

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological bulletin, 135*(2), 262–285. doi:10.1037/a0014821.

Osman, M. (2010). *Controlling uncertainty: Decision making and learning in complex worlds*. Chichester, UK: Wiley-Blackwell.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge UP.

Rottman, B. M., & Ahn, W. K. (2009). Causal learning about tolerance and sensitization. *Psychonomic Bulletin and Review, 16*(6), 1043–1049.

Rottman, B. M., & Keil, F. C. (2011). Learning causal direction from repeated observations over time. *Proceedings of the 32st Annual Conference of the Cognitive Science Society*.

Rottman, B. M., & Keil, F. C. (2010). Connecting causal events: Learning causal structures through repeated interventions over time. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2825–2830.

Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory and Cognition, 34*(2), 411–419.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: MIT Press.

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

White, P. A. (2006). How well is causal structure inferred from cooccurrence information? *European Journal of Cognitive Psychology, 18*(3), 454–480.