Running head: TIME AS GUIDE TO CAUSE

Time as a guide to cause

David A. Lagnado

University College London

Steven A. Sloman

Brown University

Address for correspondence:

David A. Lagnado

Department of Psychology

University College London

Gower Street London WC1E 6BT, UK

d.lagnado@ucl.ac.uk

Telephone: +44 (0) 20 7679 5389

Fax: +44 (0) 20 7436 4276

Abstract

How do people learn causal structure? In two studies we investigated the interplay between temporal order, intervention and covariational cues. In Study 1 temporal order overrode covariation information, leading to spurious causal inferences when the temporal cues were misleading. In Study 2 both temporal order and intervention contributed to accurate causal inference, well beyond that achievable through covariational data alone. Together the studies show that people use both temporal order and interventional cues to infer causal structure, and that these cues dominate the available statistical information. We endorse a hypothesis-driven account of learning, whereby people use cues such as temporal order to generate initial models, and then test these models against the incoming covariational data.

KEYWORDS: Causal learning, temporal order, intervention, covariation

Introduction

Inferring what causes what is notoriously difficult both in principle (Hume, 1748) and in practice (McKim & Turner, 1997). Does drug usage cause crime or does crime cause drug usage? Or perhaps something else – maybe something quite complicated – causes both? One cue to help us infer causal structure comes from the order in which events occur over time. Causes precede their effects, so an event that comes later cannot be the cause of an earlier event. The relation between drug usage and crime is clarified by the discovery that for female offenders drug usage typically precedes criminal activity (Johnson, 2004). This reduces the likelihood that involvement in crime is the cause of drug use.

But temporal order can mislead. Lightning does not cause thunder, roosters do not cause the sun to rise, and the petrol gauge showing empty does not cause the car to stop, despite the temporal precedence of the first event in each case. Consider the recent health scare in the UK about the link between the MMR jab and autism. Many parents were convinced that the vaccine had caused their child to develop autism because of the onset of behavioral symptoms soon after the jab was given. This mistaken inference had potentially harmful consequences, reducing the use of the vaccine in the general population and adding to the risk of childhood disease.

More generally, there are two main ways that temporal order can misinform us about causal structure. First, two events might be causally

linked, but our assumptions about their temporal order might mislead us about the direction of this link. Second, two events might be spuriously correlated due to a common cause, as in the example of lightning and thunder. In this case the temporal delay experienced between the two events can lead to the erroneous belief that the earlier event causes the later one, when there is in fact no causal relation between the two.

Previous research

Despite its potential to both inform and mislead our causal inferences, the role of temporal order has been assumed rather than investigated in the recent psychological literature. Earlier work did examine the spatiotemporal conditions that promote causal inference (Michotte, 1946; Shultz, 1982; Waldmann & Holyoak, 1992), but current studies tend to presuppose the causal structure under investigation, and explore how people quantify the parameters of this structure (Griffiths & Tenenbaum, in press). With respect to temporality, research has focused on the influence of time delay on judgments of causal strength, a major finding being that as the delay between two events increases, judgments of causal strength decrease (Shanks, Pearson & Dickinson, 1989), unless there is good reason to expect a delay (Buehner & May, 2002, 2003; Hagmayer & Waldmann, 2002). However, in these learning paradigms events are pre-sorted as potential causes or effects (e.g., button presses and lights going on; shell firings and explosions), so the main causal relations are already presumed. The participant's task is to quantify the

strengths of these relations rather than to infer whether or not the relations exist. This leaves us with the question of how people use temporal order to infer the underlying causal structure that links events. This is the central question addressed in this paper.

Covariation information

Along with the temporal order of events, people are often exposed to covariational information. Repeated observations of a causal system can reveal statistical relations between events and in turn these statistical relations can point to a causal relation. Current learning models focus on how people translate this covariational data into judgments of causality (Cheng, 1997; Shanks, 2004). However, covariation alone will rarely suffice for inferring a unique causal structure (see Spirtes, Glymour & Scheines, 1993). For example, a strong correlation between cholesterol level and heart disease does not by itself reveal whether cholesterol levels cause heart disease or vice-versa, or whether both are the result of a common cause.

Covariational data is more informative when combined with additional assumptions about the system under study (Waldmann, 1996). For example, if one assumes that effects never occur due to hidden causes[1] (and cannot occur spontaneously), then covariation data might be sufficient to infer causal direction. This is because the assumption of no hidden causes implies that an event that occurs alone cannot be an effect-only variable, so it must be a cause variable (and also possibly an effect variable too). For example,

suppose that A and B are correlated, and there are no other hidden causes. If one sometimes observes A alone then one can infer that it must be a cause of B, not an effect. This simplifying strategy would also work if hidden causes were assumed to be very rare (rather than impossible).

Covariational information is also more discriminating when combined with information about temporal order. Recall the example of a correlation between cholesterol and heart disease. If high cholesterol turns out to *precede* heart disease then the possibility that the latter causes the former can be ruled out. Normatively, joint knowledge of covariational and temporal information seems critical in many types of causal induction. How people make such judgments remains an open question. They might use both cues or they might focus on one or the other. The trade-off between temporal and covariational information will be examined in Study 1.

Intervention

Another fundamental route to causal knowledge is to intervene on the system under study. By manipulating certain variables and observing their effect (or lack of effect) on other variables, one can learn about their underlying causal relations. This is fundamental to the experimental method in science, as well as to the informal experiments that we conduct everyday. One critical advantage of intervention is that it can discriminate between causal structures that are difficult or impossible to distinguish by observation alone. Intervention provides this inferential advantage by disconnecting

intervened-on variables from their causes (Pearl, 2000; Spirtes, Glymour & Scheines, 1993).

However, interventions are typically confounded with a temporal order cue (Lagnado & Sloman, 2004): an intervention takes place prior to any of its effects, so an implicit temporal ordering is set up by the act of intervention itself. Does interventional learning benefit people due to this implicit temporal cue or for some other reason? This question will be pursued in Study 2.

Temporal order vs. covariation

Study 1 pits temporal order cues against covariational cues and measures people's judgments of causal structure. We constructed a learning environment in which people could use both temporal and statistical cues to induce causal structure. The paradigm was inspired by viruses (electronic or otherwise) whose temporal order of transmission is not necessarily reflected by the order in which they manifest. This is because there can be variability in the time of transmission of a virus from host to host, as well as variability in the time it takes for an infection to reveal itself.

Suppose that your computer crashes due to an email virus. Twenty minutes later your office-mate's computer also crashes (see Figure 1). A natural inference is that your computer transmitted the virus to your office-mate's computer (model 1). But this is not the only possibility. Perhaps you both received the virus from a common source, but it arrived at your

computer first (model 2). Another possibility is that your office-mate received the virus first, and then transmitted it to you, but it took longer to corrupt their computer (model 3).

Clearly the temporal order in which the email virus manifests itself (by crashing the computer) does not guarantee an inference about the order in which the computers were infected, nor about who infected who. More generally, the causal order in which events occur cannot simply be read off from the temporal order in which events occur (or appear to occur). Temporal order is often a reliable cue, but it is also fallible.

Other cues can help narrow down the possibilities. In the email virus example, the covariations amongst the presence or absence of viruses on the three computers can help to discriminate among the models. This is greatly facilitated if one assumes that there is only one external input to the network (computer C) and no other hidden causes of the virus. For example, if repeated observations show that sometimes computer B is infected without computer A being infected one can rule out model 1. Further, if computer A is sometimes infected without computer B being infected, one can also rule out model 3 and thus conclude that model 2 is correct.[2] In this case the temporal ordering of events is ambiguous between several possible causal models, but the covariational information, in combination with certain assumptions about the absence of hidden causes, can be used to discriminate amongst them.

The key question addressed in Study 1 is how temporal order and covariation are integrated to infer causal structure. In particular, we test the hypothesis that temporal order can override sparse covariational information and lead to spurious causal inferences. To test this implication, the covariational information in the learning environment in Study 1 was held constant, but the temporal order was manipulated.

STUDY 1

Participants sent test messages to a small computer network in order to figure out which connections were working. They completed four problems each with the same underlying causal structure (see Figure 2), but with information presented in various temporal orders. In each case the covariational information seen by the participants, together with the assumption that email messages could not appear unless they had been sent by the participants themselves (e.g., no hidden causes), was sufficient to uncover the unique causal structure. However, the validity of the temporal order information was varied between problems. In some cases it matched the underlying structure; in other cases it suggested alternative structures. Participants were made fully aware that the temporal information was unreliable.

Method

Participants and apparatus

Twenty-four students from University College London were paid £7 (about $12) to participate in the experiment. They were tested individually on a PC.

Materials

The same probabilistic model was used to generate network responses for all four problems (Figure 2). Each working connection had a 0.8 probability of passing a message from one computer to another, and messages could not appear on a computer unless they had been passed by a connected computer. The probabilities of each possible pattern of messages are shown in Table 1.

In each problem the computer network was presented on the screen in a fixed spatial configuration (see Figure 2). The source computer (A) was always placed at the bottom, but the locations of the other three computers (B, C and D) were systematically varied across problems. The computers in each network were labelled with a different set of three-letter names. These sets of labels were rotated across problems for each participant.

Procedure

Participants were instructed that they would be presented with four similar inference problems. In each problem their task was to test a faulty computer network in order to establish which connections were working. To do this they had to send test messages to one computer (labelled 'A' in Figure 2) and then see which other computers received the message. They were told

that some connections work 80% of the time and some do not work at all, and that a working connection from computer X to computer Y implied nothing about whether there was a working connection from Y to X. They were also instructed that messages could not reach a computer unless they had been sent by the participants themselves (e.g., there were no hidden causes of the email messages).

Participants were also informed that there would be time delays in the appearance of the messages on the computer screens. They were told that these delays could be either due to variability in the time it takes for a message to be transmitted from computer to computer or in the time it takes for a message, once received, to be displayed on a computer monitor. The implication of this latter possibility was emphasized: "This means that it is possible for a message to be transmitted by a computer before it is displayed on its own screen (in the same way that you may pass on a virus before it becomes active on your own computer)".

Each problem consisted of a learning and a testing phase. In the learning phase participants sent 100 test messages to the network, one at a time. To ensure that participants were engaged in the task, on every fifth test they were asked to predict whether a specified computer would receive the message. These questions were rotated across computers.

There were four temporal order conditions. In each condition information about whether a computer had received a test message was

displayed in a different temporal order (see Table 2). In condition 1 the information for all computers was displayed simultaneously, so there were no temporal cues. In condition 4 the temporal ordering matched the actual network structure, whereas in conditions 2 and 3 it suggested different causal structures. Temporal order was created by inserting one second delays.

Each problem was followed by an identical test phase. Participants were asked a set of 10 questions. First, for each of nine possible connections they were asked whether they thought that the connection was working (and were reminded that working connections still only worked 80% of the time). They responded "yes" or "no". Second, they were asked an inferential question: "Suppose you had to send an important message to computer C. Would it be better to send it from B or D?" After answering this question participants proceeded to the next problem.

Results and Discussion

Structural questions. The proportions of links endorsed by participants in each time condition are shown in Table 3. An ANOVA with *time condition* and *link* as within-participant factors revealed a main effect of link, $F(8,184) = 13.5$, $p < 0.001$, no main effect of time condition, $F(3, 69) < 1$, and a significant interaction between time and link, $F(24, 552) = 7.72$, $p < 0.001$.

The relation between link choices and time condition are illustrated in Figure 3. For each time condition a summary model was constructed from the

links that were endorsed by more than 50% of participants. Inspection of these models shows use of both temporal and covariational information, with the former dominating the latter. Thus, even though all four problems had identical structure and therefore generated the same covariational data, link choices were heavily influenced by the temporal ordering. This is most apparent when the temporal orderings conflicted with the underlying structure (conditions 2 and 3). In both conditions participants inserted links that were implied by the temporal cues but not by the patterns of covariation (e.g., link D→C in problem 2, and links D→C, A→D, C→B in problem 3).

Inferences. The proportion of participants choosing computer B (the correct answer) was 75% for time condition 1, 62.5% for time condition 2, 20.8% for time condition 3, and 75% for time condition 4. A within-participant ANOVA revealed a significant effect of time condition, $F(3,69) = 7.95$, $p < 0.001$. Paired comparisons showed that condition 3 was significantly lower than the other three conditions (condition 3 vs. 1, $t(23)= 4.03$, $p < 0.001$; condition 3 vs. 2, $t(23)= 3.12$, $p < 0.01$; condition 3 vs. 4, $t(23)= 4.51$, $p < 0.001$), but no other differences were significant.

A more revealing analysis combined responses in conditions 2 and 3 into a 'time delay' category (where messages on computer C were preceded by messages on computer D), and compared this with a 'simultaneous' category that combined conditions 1 and 4 (where the messages on C and D appeared simultaneously). Mean correct responses in the 'simultaneous'

category (75%) were significantly higher than those in the 'time delay' category (41.7%), $t(94) = 3.48$, $p = 0.001$.

These results reflect the strong influence of temporal ordering on participants' judgments. Those that experienced a systematic delay between the appearance of messages on computers D and C were more likely to use D (rather than B) to send a message to C. This mimics the pattern of link selections. Most participants (89%) who chose computer D also erroneously endorsed a connection from D to C, whereas only 58 per cent who did not choose computer D did.

STUDY 2

Recent work in psychology suggests that both adults and children can learn causal structure through the appropriate use of interventions (Gopnik et al., 2004; Lagnado & Sloman, 2002, 2004; Sobel, 2003; Steyvers et al., 2003). As noted in the introduction, what is special about intervention, as opposed to mere observation, is that it can modify the structure of the system under study. Suitably chosen interventions thus allow one to distinguish between causal models that are 'observationally' equivalent (Pearl, 2000; Spirtes, Glymour & Scheines, 1993).

To illustrate, suppose you know that listening to country music and suicide rates are correlated across numerous metropolitan areas (Stack & Gundlach, 1992), and you want to determine whether listening to country music causes suicide rates, suicide rates causes listening to country music, or

both result from a common cause. In the absence of any prior assumptions or knowledge, the covariational data alone are insufficient to answer this question. However, suitable interventions can determine the correct model. For example, if you intervene on the amount of airtime devoted to country music, and the suicide rate changes, you can infer a causal link from country music to suicide. This is because your intervention disconnects the level of country music on the radio from its usual causes, and thus eliminates any confounding variables that may affect both levels of country music and suicide rates.

However, there are several factors aside from the special kind of information that intervention affords that might drive its advantage over observation. First, interventions and temporal order are typically confounded (both in laboratory experiments and the real world). An intervened-on variable must change its value *before* changes in its effects. Even if changes appear to be simultaneous, people can infer that changes other than their interventions are effects, not causes. In this way people can benefit from intervention even if they fail to change their causal models appropriately to represent their interventions (Lagnado & Sloman, 2004).

To return to the example above, suppose one uses the simple heuristic that any changes that occur after one's intervention are effects of the intervened-on variable. This also permits one to infer that listening to country music causes suicide rates from the fact that suicide rates change *after* the

airtime devoted to country music is changed. However, in this case it is achieved without explicit representation of the intervention or the potential structural modifications that it entails. In this way an intervener can benefit from the fact that an intervention decouples the intervened-on variable from its other causes without being aware of this fact. It is enough that the intervener is using some heuristic based on temporal order. This confound between intervention and temporal order needs to be teased apart before we conclude that people are rational experimentalists.

Another potential difference between intervention and observation lies in the distribution of information that people can receive about a system. While observers may typically receive a representative sampling of the system's autonomous behavior, the information that interveners receive depends on what interventions they make. This is not just because their interventions can modify the system, but also because their choices modulate the frequencies of the data that they receive. For example, if all your interventions are directed at one specific cause of an event you will receive little information about alternative causes of that same event. Lagnado and Sloman (2004) ruled out selective information as a determinant of learning success in their experiments, but it is important to examine the pattern of interventions that interveners make, to establish whether this can affect the difference between intervention and observation.

Third, interventional learning may promote more directed hypothesis-testing. Thus someone who repeatedly intervenes on a system is in a better position to test their own hypotheses than someone who merely observes the system. So far there is mixed evidence as to whether the opportunity to hypothesis test enhances interventional learning. Sobel and Kushnir (2003) report that it does; Lagnado and Sloman (2004) and Osman and Heyes (2005) found that it does not.

Intervention vs. temporal order

Lagnado and Sloman (2004) showed that temporal order cues improved causal learning, irrespective of whether learners were intervening or just observing. However, generally low levels of performance made it difficult to quantify the separate effects of temporal order cues and intervention. The low levels of performance observed in these studies can be attributed in part to the possibility of unknown hidden causes. These made recovery of the correct causal models more difficult in both the intervention and observation conditions (see Lagnado and Sloman, 2004, for a detailed discussion). To boost learning performance in the current study, and thereby permit a more robust analysis of the relation between temporal order and intervention, participants were given causal problems with no hidden causes.

Overview of Study 2

Participants either manipulated or observed on-screen switches (see Figure 4) in order to figure out the causal connections between these

components. They completed six problems each with a different causal model (see Table 4, column 2). Participants were divided into three groups: those who could freely intervene on the causal system, those who simply observed the system's behavior, and those who observed the results of another person's interventions (yoked to the active interveners). Within each group participants were presented with information about the switches' values in two temporal orders, either consistent with, or opposite to, the underlying causal structure.

Method

Participants and apparatus

Seventy-two students from Brown University were each paid $10 to participate in the experiment. They were tested individually on a PC.

Materials

There were six learning tasks. In each task probabilistic data were generated according to a specific causal model. Each causal model was made up from several identical components connected by causal links (see Table 4). All causal links were probabilistic and of the same strength. If component A was linked to component B, then there was an 80% chance that activation of A would lead to activation of B and no chance that B would activate if A did not (so there were no hidden causes).

Each component had an internal state (activated /not activated), a slider indicator of this state (on/off), and a binary switch (on/off) used to

activate this internal state. Components could either be activated directly – by someone clicking on the binary switch -- or indirectly -- through the activation of a linked component. Slider indicators simply registered whether or not their components were activated, and could not be used to activate components (see Figure 4). The components (but not the links) of a causal system were displayed on the screen in spatial configurations that did not give any clues as to the underlying causal structure.

Procedure

All participants were told that they would be presented with six short learning tasks. In each task they had to figure out how the components in a system were causally connected. They were warned that the links were probabilistic, and thus not guaranteed to work on every occasion. They were not explicitly told that there were no hidden causes that might activate the components, although this would have been a natural assumption to make on the basis of the instructions.

Participants were given a fixed number of trials for each task (see Table 5, column 3), and the task order was counterbalanced. They divided into three groups: active intervention, yoked intervention and observation. Those in the active intervention group were told that they would be able to intervene on the system by switching on any of the components (one at a time) and seeing which other components were activated. At the start of each trial all the components were switched off, and participants could choose just one

component to intervene on (switch on). After this intervention, the activation statuses of all the components were displayed via the slider indicators. This was done in one of two temporal orders (see below). See Figure 4 for an example of the screen appearance during an intervention trial.

Participants in the yoked intervention group were told that they would be watching another agent trying to learn about the system. They were in fact yoked to the performance of individuals from the active intervention condition. On each trial they were shown which component the agent had intervened on, and the slider indicator values for all the components. Again these were displayed in one of two temporal orders.

Participants in the observation group were told that they would be watching the autonomous behavior of the system. On each trial they observed the slider values for each of the components. These was generated trial-by-trial by randomly activating one of the source components (those that had no links feeding into them from other components) and letting this permeate through the system according to the probabilistic links. Participants were given no clues as to which component had been externally activated, and which were activated by a linked component.

The temporal order of the display of the components' states (i.e., their slider indicators) was manipulated (see Table 4). In the consistent time condition the slider values were displayed in a temporal order that matched the causal model. For example, for the $A \rightarrow B$ model, the slider value on

component *A* was displayed before component *B*. In the inconsistent time condition the reverse ordering was used. Both conditions used 0.5 second delays throughout. The temporal ordering (consistent or inconsistent) was systematically varied within-participants; it was fixed within a particular task, but counterbalanced across tasks.

All participants were forewarned that there would be temporal delays in the display of slider values, and that these could not be used as a reliable guide to causal order. They were not informed, however, about the systematic nature of these time delays. After completing each task participants were asked about the causal links between components. They were asked about the presence or absence of a causal link for all possible links in the model as in Study 1.

Results and Discussion

The mean correct model choices[3] by group (intervention, yoked intervention, observation) and time condition (consistent, inconsistent) are shown in Figure 5. A mixed ANOVA revealed a main effect of group, $F(2, 69) = 25.96$, $p < 0.001$, a main effect of time condition, $F(1, 69) = 20.88$, $p < 0.001$, and an interaction between group and time, $F(2, 69) = 4.00$, $p < 0.05$. Paired comparisons showed that in both the yoked intervention (79.2% vs. 52.8%) and observation condition (54.2% vs. 16.7%) correct model choices were significantly higher when the temporal order was consistent rather than inconsistent ($t(23) = 2.84$, $p < 0.01$ and $t(23) = 3.87$, $p < 0.01$, respectively).

However, in the active intervention condition showed no significant difference between temporal conditions (77.8% vs. 73.6%, t(23) = 0.65, *ns.*). So, while the expected drop in performance with inconsistent temporal order was evident in the yoked intervention and observation groups, it was not in the active intervention group.

Individual link analysis

An alternative way to analyze the response data is in terms of individual links (as in Study 1). This allows us to look at the effect of temporal order on link choice in both the time consistent and time inconsistent conditions. For each participant, and across all six models, we computed the proportion of links that were chosen *with* or *against* the experienced temporal order. All links between variables were included in the analysis except those that linked variables that appeared simultaneously (see Table 4).

The summary results for each of the three learning conditions are shown in Figure 6. A mixed ANOVA was conducted with time direction (with, against) as a within-participant factor, and group (intervention, yoked intervention, observation) as a between-participant factor. It revealed a main effect of time direction, $F(1,69) = 33.43$, $p < 0.001$, a main effect of group, $F(2,69) = 7.43$, $p < 0.01$, and a time direction by group interaction, $F(2,69) = 11.27$, $p < 0.001$. Paired comparisons showed that links that fit the experienced time direction predominated in the observation condition (76.3% vs. 23.7%, $t(23) = 6.83$, $p < 0.001$) and yoked intervention condition (56.7% vs. 43.3%, t(23)

= 2.11, $p < 0.05$). There was no significant difference in the active intervention condition (52.3% vs. 47.7%, $t(23) = 0.81$, *ns.*). These results reinforce the conclusions drawn from the analysis of correct models: time direction has a strong effect on link choices in the observation condition, somewhat less in the yoked condition, and none in the intervention condition.

Choices between Markov equivalent models

As noted above, one of the main limitations of purely covariational data is their failure to discriminate between observationally (Markov) equivalent models. This meant that unless observers made the assumption that there were no hidden causes the best that they could do was to infer a class of Markov equivalent models.

Did experienced time order affect which Markov equivalent model observers selected? To address this question we looked at all those model choices that corresponded either to the actual model or one of its Markov equivalents (we term these choices *Markov correct*). We then compared how many of these Markov correct choices went *with* or *against* the experienced temporal order (both in the time consistent and time reversed conditions). For example, in problem 1 the model that fitted the experienced time direction was A→B in the time consistent condition and B→A in the time reversed condition.

Across all six problems 41% of the total model choices (59 out of 144) were Markov correct and fitted the experienced time order, whereas only

14.5% (10 out of 144) were Markov correct and against the experienced time order. This difference was highly significant ($t(23) = 5.78$, $p < 0.001$), and shows that experienced temporal order exerts a strong effect on choices between Markov equivalent models.

Pattern of interventions

The distribution of interventions made by the active interveners (and also seen by the yoked interveners) were analyzed separately for each of the six models. Separate ANOVAs were conducted for each problem, with time (consistent, inconsistent) and node intervened-on as within-participant factors. These revealed no effects of time (for all models $Fs < 1$), but main effects of node for five of the models (model 1, $F(1,22) = 4.87$, $p < 0.05$; model 3, $F(2,44) = 11.21$, $p < 0.01$; model 4, $F(2,44) = 14.58$, $p < 0.01$; model 5, $F(3,66) = 11.30$, $p < 0.01$; model 6, $F(3,66) = 18.52$, $p < 0.01$) and a marginal effect of node for model 2 ($F(2,44) = 2.98$, $p = 0.06$).

Table 5 shows the overall distributions of interventions for each model (collapsed across time consistent and inconsistent conditions). Paired comparisons revealed that for all the problems the effect-only node was chosen significantly less than the other nodes: model 1, $t(23) = 2.23$, $p < 0.05$; model 2, $t(23) = 2.23$, $p < 0.05$; model 3, $t(23) = 4.61$, $p < 0.01$; model 4, $t(23) = 4.03$, $p < 0.01$; model 5, $t(23) = 6.79$, $p < 0.01$; model 6, $t(23) = 8.26$, $p < 0.01$.

By symmetry these comparisons also showed that the root nodes in models 1, 2 and 4 were chosen significantly more than the other nodes.

Additional paired comparisons showed that this did not hold for the chain model (3), where there was no difference between the proportion of interventions on root nodes (A) and intermediate nodes (B), $t(23) = 0.30$, *ns*. However, in both model 5 and model 6 there was a significant preference for the root nodes ($t(23) = 2.33$, $p < 0.05$ and $t(23) = 3.14$, $p < 0.01$, respectively).

In sum, the analyses of the distributions of interventions revealed: (1) no difference in the pattern of interventions between time consistent and time inconsistent conditions; (2) a lower proportion of interventions on the effect-only nodes; (3) a preference for the root nodes except in model 3.

The tendency not to intervene on effect-only nodes is perfectly reasonable as these interventions convey the least amount of information about causal structure (Scheines, Easterday, & Danks, in press; Steyvers et al., 2003). The tendency to intervene on root nodes in models 1, 2 and 4 is likewise reasonable, as is the lack of a preference for the root node in model 3. However, the marked preference for the root node in model 6 is not optimal, and is reflected in the lower number of correct choices for that model (see Table 6).

The distributions of interventions also suggest that interveners do not see a particularly unrepresentative sampling of trial types as compared to observers. This would only occur if the interveners had omitted to intervene on the root nodes of a model. In fact the above analysis shows that the opposite is true, that they have some preference for root nodes. This implies

that the key difference between interveners and observers, in terms of the kind of information that they receive, hinges on the structural changes due to intervention rather than simply different frequencies of trial types.

Study 2B

Why were active interveners unaffected by temporal order? One possible reason emerged in post-experimental questioning: interveners might have overcome the inconsistent temporal order cue by first figuring out that in some problems the temporal ordering was reversed and then computing the correct model by mentally reversing the time order in which the slider information had been displayed. This would have been particularly easy in the two-variable problem.

A follow-up study (2B) was designed to test this idea. The inconsistent temporal order was made uninformative by making the order of events random on each trial with the constraint that the veridical causal order was never used. Apart from this difference the method and procedure were identical to the previous intervention condition. Twenty-four new participants from the same population as in Study 2 completed all six problems in an intervention condition with temporal order (either consistent or randomized) as a within-participant factor. The results for this group are also shown in Figure 5 (labelled intervention 2), and their choices for the individual problems are shown in Table 6.

When participants received information in the consistent temporal order they performed significantly better than when they received it in the randomized order (76.4% vs. 52.8%, $t(23) = 2.99$, $p < 0.01$). Indeed the level of performance for interveners in the randomized time order condition was no different to that achieved by observers in the time consistent condition in Study 2 (52.8% vs. 54%, $t(46) = 0.13$, *ns*). Inspection of Figure 5 shows parallel shifts from time consistent to time inconsistent conditions for intervention 2 and observation. This suggests that intervention and time order provide separate additive cues to causal structure.

GENERAL DISCUSSION

These studies investigated how people use temporal order to infer causal structure. Study 1 found that temporal order overrode covariation information, leading to spurious inferences when temporal cues were misleading. Study 2 found that both temporal order and intervention contributed to accurate causal inference, well beyond that achievable through covariational information alone. However, when interveners received information in a randomized order (Study 2B) they performed no better than observers who received information in a consistent order.

Taken together these findings show that temporal order and intervention afford separate cues to causal structure, and work best when they combine rather than conflict. This explains the efficacy of interventions

in most everyday contexts, where temporal order supplements the evidence provided by intervention.

Hypothesis-driven vs. data-driven learning

Two broad theoretical approaches to human causal learning can be distinguished. Data-driven theories assume that the learner builds up causal knowledge by extracting statistical relations from the correlational data they are exposed to (Glymour, 2003; Shanks, 2004). As noted in the introduction, however, there are both theoretical and practical reasons why this cannot be the full story. The studies in this paper reinforce this conclusion: people struggle to infer causal structure when exposed to covariational information alone. This held true even when the induction problem was simplified by ensuring that there were no unknown hidden causes of the observed patterns of covariation.

In contrast, hypothesis-driven approaches (e.g., Steyvers et al., 2003; Tenenbaum & Griffiths, 2003; Waldmann, 1996; Waldmann & Hagmayer, 2001) maintain that people use prior assumptions to guide the learning process. One elaboration of this idea argues that learners use a variety of non-statistical cues (temporal order, intervention, prior knowledge) to construct their initial causal models, and these models are then confirmed or revised in the light of subsequent data (Lagnado, Waldmann, Hagmayer, & Sloman, in press).

Our current findings are readily interpretable within the hypothesis-driven framework. Participants in both studies allowed the temporal order cue -- which was apparent on the very first trial -- to dictate the kinds of models they constructed and tested. And the effects of this cue persisted even when the covariational data were inconsistent with it.

To illustrate, consider the causal judgments in Study 1. When the data were presented in temporal order ABDC participants incorrectly endorsed the link from D to C (in line with the fact that D preceded C), but correctly endorsed the link from A to C (Figure 3, condition 2). A plausible explanation for this pattern of judgments is that participants first used the temporal ordering to hypothesize an initial model (A→B→D→C). This model was then confirmed by most of the test trials (e.g., patterns 1, 3, 4 and 5, accounting for 87% of the trials). However, occasionally they saw a test pattern that contradicted this model (pattern 2: A, B, C, ~D). To accommodate this new evidence, they added a link from B to C, but did not remove the redundant link from D to C, because this still fit the temporal ordering.

Similarly, when the data were presented in temporal order ADCB (condition 3) participants were likely to have constructed an initial model that matched the temporal order (A→D→C→B). This too would have been confirmed by the majority of trials (patterns 1 and 5, accounting for 71% of trials). In the face of the anomalous patterns (2, 3, 4) about 60% of participants

inserted the A→B link. This model would account for 87% of the data patterns (all except pattern 2).

The advantage for interventional learning found in Study 2 also supports a hypothesis-driven account. Interventions will be most effective if they can target specific hypothetical models to test. At the very least someone must conjecture that the variable that they choose to intervene on is a potential cause. Our analyses of the distribution of interventions in Study 2 confirm this view. There were significantly less interventions made on effect-only nodes than on other nodes. And this makes good sense, because they cause nothing else to happen in the system and are least informative about the correct causal structure. Other studies have also found that for simple models people are close to optimal in their choice of interventions (Scheines et al., in press; Steyvers et al., 2003). This would be unlikely if they did not formulate hypothetical models in advance.

Using cues in proportion to their reliability

A more general perspective on the current findings is that people are using cues such as temporal order and covariation in proportion to the reliability (or variability) of these cues. On this view the dominance of temporal over covariational information in controlling people's judgments was due to the greater reliability of the temporal order cue. Indeed in Study 1, and the time consistent conditions in Study 2, the temporal order cue had no variability. In contrast, the probabilistic nature of the models in both studies

ensured that the covariation information was variable. This view would predict that if the reliability of the temporal cue in Study 1 was reduced, for example, by adding noise to the temporal delays, then people would place less weight on this cue.

The idea that people combine cues in proportion to their reliability receives independent support from Griffiths (2005), who showed that participants' confidence in the existence of a causal relation was proportional to the variability of the covariational information they were presented. A key idea there, as here, is that people use covariation information to corroborate their prior hypotheses about causal links.

The benefits of intervention

There are various reasons why interventions can enhance learning of causal structure (Lagnado & Sloman, 2004). The main focus of recent research has been on the special information it provides due to the fact that interventions modify the structure of the system under investigation (Gopnik et al., 2004; Steyvers et al., 2003). The additive effect of intervention in Study 2 confirms the importance of this aspect in human learning. However, there are several other factors that typically accompany interventions and that can also contribute to its efficacy.

A learner who can make repeated interventions can benefit from the fact that they can more readily test their own hypotheses. Although Lagnado and Sloman (2004) and Osman and Heyes (2005) found no difference between

'active' and 'yoked' interveners, Sobel and Kushnir (2003) reported an advantage for the former. Study 2 sheds some light on this question. Recall that in the time consistent condition active interveners and yoked interveners performed equally well, but in the time-reversed condition the former outperformed the latter. This suggests that under favorable learning conditions (e.g., with an informative time cue) the freedom to choose one's own interventions adds nothing to the benefits of interventional information. However, under more demanding conditions (e.g., time reversed), the ability to choose one's own interventions, and perhaps make the intervened-on variable more salient, does help learners.

The main question addressed in Study 2 was the interplay between temporal order and intervention. Previous experimental studies often confound these factors, so it is possible that people simply use a temporal cue to reap the benefits of intervention. Our results show that there is an effect of intervention beyond that gained through the temporal order cue, but also show that temporal order accounts for a fair proportion of the advantage.

This has important implications for our understanding of intervention, which is at the heart of causality. Despite all the continuing excitement about interventional over observational learning, it looks like some of the benefit might just have to do with the temporal cue and attentional benefits. This implies that any temporal cue might help people learn, as long as it is

consistent with the to-be-learned structure. This may explain the appeal of flow diagrams.

Conclusions

In sum, our studies show that people use both temporal order and interventional cues to infer causal structure, and that these cues dominate the available statistical information. They also support a hypothesis-driven account of learning, whereby people use cues such as temporal order to generate initial models, and then test these models against the incoming covariational data.

This allows us to make good inferences on the basis of slender data, but can also expose us to certain causal illusions. As our studies show, the latter are especially prevalent when the temporal order of events conspires to mislead us. And this holds in everyday life. Compelling as the inference might be, the fact that you just yelled at the television set does not give you credit for your team's game-winning goal.

References

Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking and Reasoning*, 8(4), 269-295.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgment of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology*, 56A(5), 865-890.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2), 367-405.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Sciences*, *7*, 43-48.

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, 1–31.

Griffiths, T. L. (2005). Causes, coincidences, and theories. Unpublished doctoral thesis.

Griffiths, T. L., & Tenenbaum, J. B. (in press). Elemental causal induction. *Cognitive Psychology*.

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

Hume, D. (1748). *An enquiry concerning human understanding.* Oxford, England: Clarendon.

Johnson, H. (2004). *Drugs and crime: a study of incarcerated female offenders.* Research and public policy series, No. 63. Canberra: Australian Institute of Criminology.

Lagnado, D. A., & Sloman, S. A. (2002). Learning causal structure. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society,* (pp.560-565). Mahwah, NJ: Erlbaum.

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

Lagnado, D.A., Waldmann, M. R., Hagmayer, Y., & Sloman, S.A. (in press). Beyond covariation: Cues to causal structure. To appear in Gopnik, A., & Schultz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation.* New York: Oxford University Press.

McKim, V., & Turner, S. (1997). *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences.* Notre Dame: University of Notre Dame Press,

Michotte, A. E. (1946/1963). *The perception of causality.* London, England: Methuen & Co.

Osman, M., & Heyes, C. (2005) Practice doesn't always make perfect: Goal induced decrements in the accuracy of action- and observation-based

problem solving. *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*.

Pearl, J. (2000). *Causality.* Cambridge, England: Cambridge University Press.

Shanks, D. R. (2004). Judging covariation and causation. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making*. Oxford, England: Blackwell.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgment of causality by human subjects. *Quarterly Journal of Experimental Psychology, Section B*, 41(2), 139-159.

Scheines, R., Easterday, M., & Danks, D. (in press). Teaching the normative theory of causal reasoning. To appear in Gopnik, A., & Schultz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation.* New York: Oxford University Press.

Shultz, T. R. (1982). Rules of Causal Attribution. *Monographs of the Society for Research in Child Development*, 47(1).

Sobel, D. M. (2003). Watch it, do it, or watch it done. Manuscript submitted for publication.

Sobel, D. M., & Kushnir, T. (2003). Interventions do not solely benefit causal learning. *Proceedings of the Twenty-fifth Annual Meeting of the Cognitive Science Society,* (pp. 1100-1105). Mahwah, NJ: Erlbaum.

Spirtes, P., Glymour, C., & Schienes, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Stack, S., & Gundlach, J. (1992). The effect of country music on suicide. *Social Forces, 71,* 211-218.

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

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems,* 15, 35-42. Cambridge, MA: MIT Press.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 34, pp. 47–88). San Diego, CA: Academic Press.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121,* 222-236.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*, 27–58.

Footnotes

1. Strictly speaking this amounts to the assumption that there are no hidden causes other than exogenous inputs to the system, and that these inputs only affect root variables.

2. More generally, the three models imply different patterns of *conditional* dependence. For example, only model 2 implies that A and B are probabilistically independent conditional on C.

3. The correct model for each problem was the causal model that was actually used to generate the data. Of course observers are at a disadvantage with respect to interveners here because covariational data alone is insufficient to identify a unique model (except for the model in problem 2). It seems likely that observers assumed that there were no additional hidden causes (i.e., no switches activated purely spontaneously), in which case unique identification is possible. The later analyses in terms of individual link choices and Markov equivalent models, however, shows that temporal order is the driving influence in participant's judgments regardless of which models are counted as correct.

Table 1. Probabilities of each possible pattern of messages in Study 1.

Pattern	Computers with message present	Probability
1	*ABCD*	0.512
2	*ABC*	0.128
3	*ABD*	0.128
4	*AB*	0.032
5	*A*	0.200

Table 2.

Temporal order of display of information (one second delays) for the four conditions in Study 1.

	Time steps			
Condition	t_1	t_2	t_3	t_4
1	*A* *B* or *C* or *D*			
2	*A*	*B*	*D*	*C*
3	*A*	*D*	*C*	*B*
4	*A*	*B*	*C* or *D*	

Note: Inclusive 'or' is used throughout. For example, in condition 4 at time step 3, either *C* or *D*, or both can occur.

Table 3.

Proportion of participants affirming causal relation for each relation in each time order. True causal relations are shown in bold.

	Time Order			
	1	2	3	4
Causal relation	Simultaneous	*ABDC*	*ADCB*	*AB[CD]*
AB	**1.00***	**0.96***	**0.58**	**0.92***
AC	0.38	0.17	0.88*	0.29
AD	0.33	0.13	0.54	0.33
BD	**0.75***	**0.79***	**0.21**	**0.79***
DB	0.63	0.38	0.79*	0.50
BC	**0.75***	**0.96***	**0.38**	**0.88***
CB	0.50	0.46	0.50	0.46
DC	0.25	0.83*	0.71*	0.21
CD	0.29	0.21	0.33	0.29

* $p < 0.05$ greater than 50%

Table 4.

Temporal order of display of information (with 0.5 second delays between time steps) for the two time conditions in Study 2.

Task	Model	Time Consistent				Time Reversed			
		t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4
1	$A \rightarrow B$	*A*	*B*			*B*	*A*		
2	$A \rightarrow C \leftarrow B$	*A* or *B*	*C*			*C*	*A* or *B*		
3	$A \rightarrow B \rightarrow C$	*A*	*B*	*C*		*C*	*B*	*A*	
4	$B \leftarrow A \rightarrow C$	*A*	*B* or *C*			*B* or *C*	*A*		
5	$A \rightarrow C \rightarrow D$ $\nearrow$ B	*A* or *B*	*C*	*D*		*D*	*C*	*A* or *B*	
6	$A \rightarrow B \rightarrow C \rightarrow D$	*A*	*B*	*C*	*D*	*D*	*C*	*B*	*A*

Note: Inclusive 'or' is used throughout. For example, in the time consistent condition of task 2, at time step 1 either *A* or *B*, or both can occur.

Table 5.

Distribution of interventions for each model in the learning phase of Study 2.

Task	Model	Tests	% Intervention			
			A	*B*	*C*	*D*
1	$A \rightarrow B$	10	54	46		
2	$A \rightarrow C \leftarrow B$	20	37	34	30	
3	$A \rightarrow B \rightarrow C$	20	37	36	27	
4	$B \leftarrow A \rightarrow C$	20	45	27	28	
5	$A \rightarrow C \rightarrow D$ $\nearrow$ B	30	28	31	25	16
6	$A \rightarrow B \rightarrow C \rightarrow D$	30	33	25	25	17

Table 6.

Percentage of correct choices for each model in Study 2 and 2B.

		Study 2						Study 2B	
		% Correct choices							
		Time Consistent			Time Reversed			Time Consistent	Time Randomized
Task	Model	1	2	3	1	2	3		
1	$A \rightarrow B$	92	75	83	83	67	17	100	67
2	$A \rightarrow C \leftarrow B$	83	83	75	92	67	25	100	67
3	$A \rightarrow B \rightarrow C$	75	92	42	67	58	0	83	58
4	$B \leftarrow A \rightarrow C$	100	92	58	83	75	25	100	58
5	$A \rightarrow C \rightarrow D$ $\nearrow$ B	75	75	25	50	17	8	50	33
6	$A \rightarrow B \rightarrow C \rightarrow D$	42	58	42	67	33	0	42	33

Figure captions

Figure 1. Three possible models when the virus appears on computer A before computer B. Model 1 is the natural inference on the basis of temporal order. Model 2 has a common cause C that infects both A and B separately. Model 3 has a structure directly opposite to order in which the viruses appear.

Figure 2. Structure of computer networks for all conditions in Study 1.

Figure 3. Summary model choices for the four temporal order conditions in Study 1. Note that only links endorsed by > 50% participants are shown, and the thickness of the arrows corresponds to the percentage of participants selecting that link (thickest link = 100%).

Figure 4. Screen in learning phase for problem 2. Top panel shows screen prior to intervention. Bottom panel shows screen after participant has switched on A, C has switched on, but B has remained switched off. In the temporally consistent condition the slider on A would have registered before the slider on C. In the temporally inconsistent condition the slider on C would have registered first.

Figure 5. Percent correct model choices in Study 2 showing effects of intervention and temporal order. Note that for intervention2 the temporal order in the inconsistent condition was randomized rather than reversed.

Figure 6. Proportion of links chosen by time direction for four of the six models in Study 2.

Figure 1

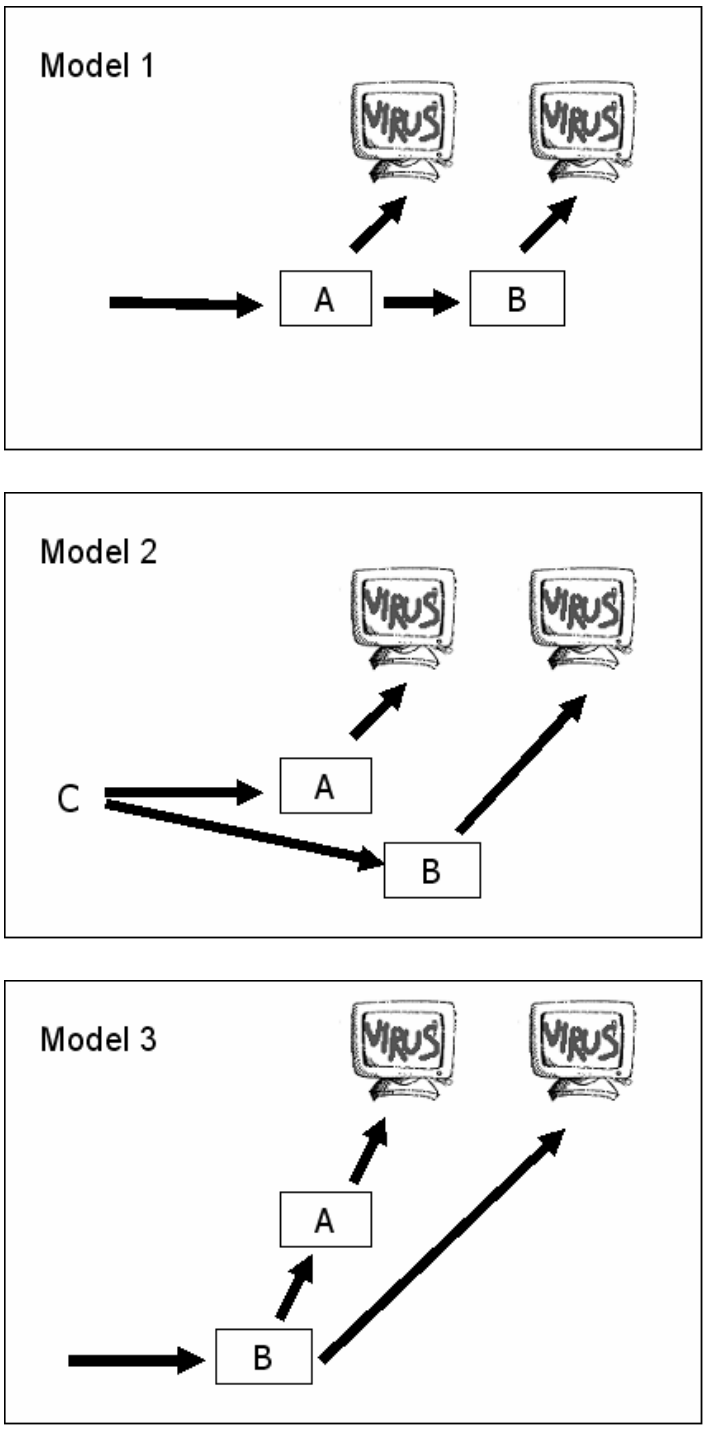

Model 1
VIRUS
VIRUS
A
B
Model 2
VIRUS
VIRUS
C
A
B
Model 3
VIRUS
VIRUS
A
B

Figure 2

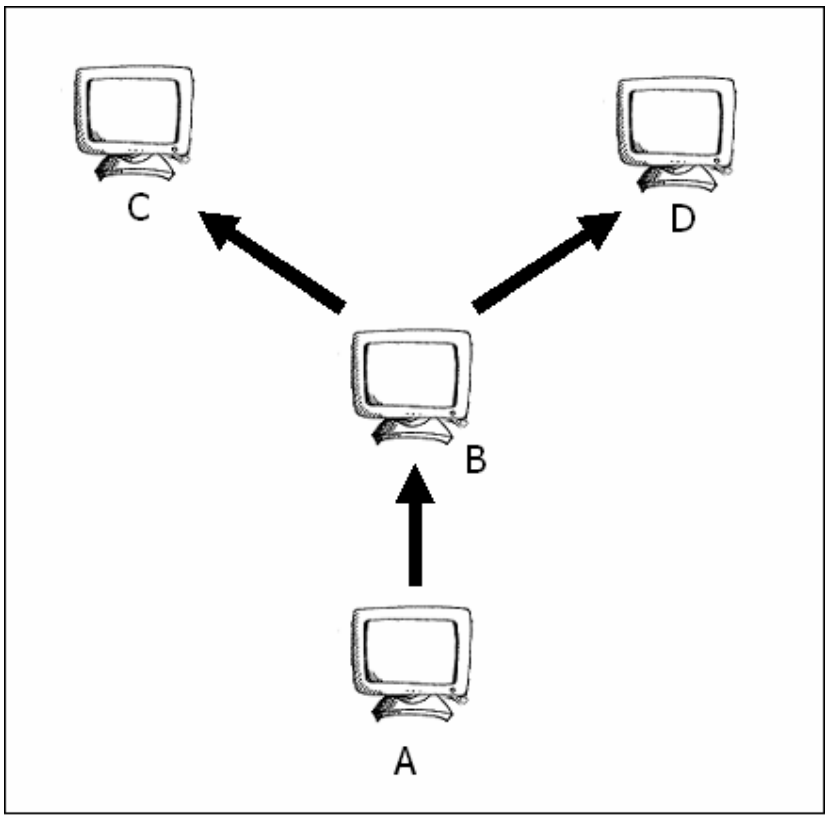


Figure 3

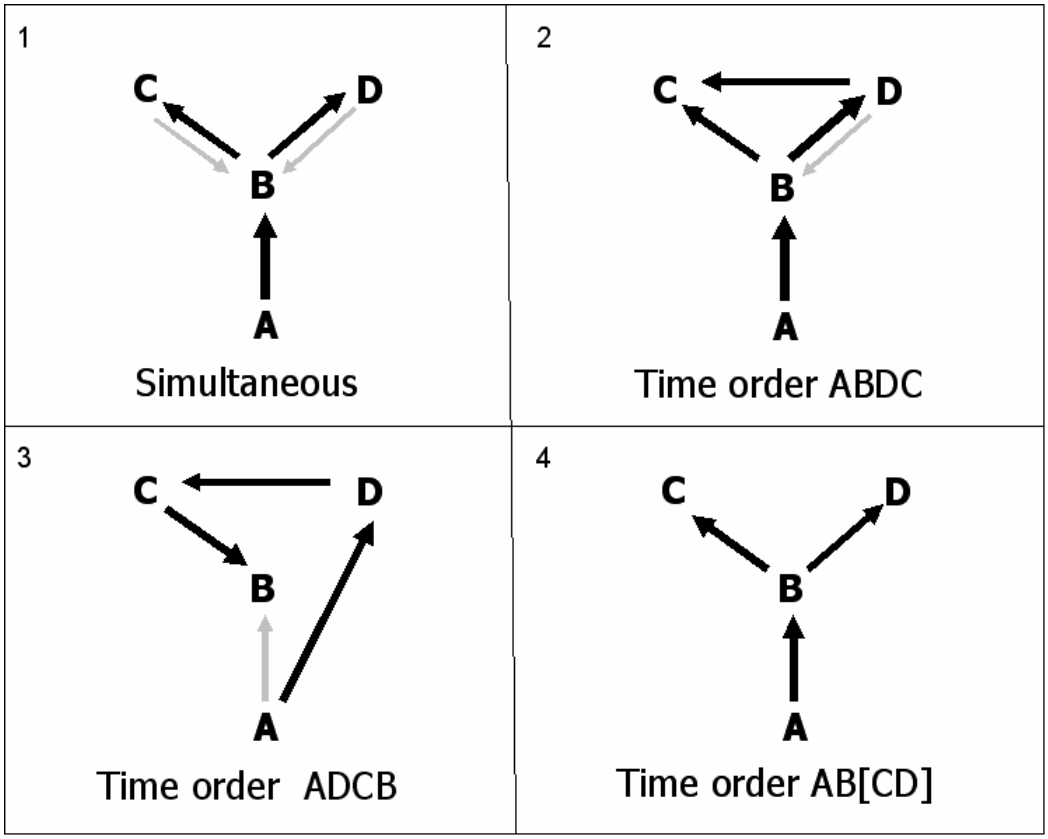

1
C
D
B
A
Simultaneous
2
C
D
B
A
Time order ABDC
3
C
D
B
A
Time order ADCB
4
C
D
B
A
Time order AB[CD]

Figure 4

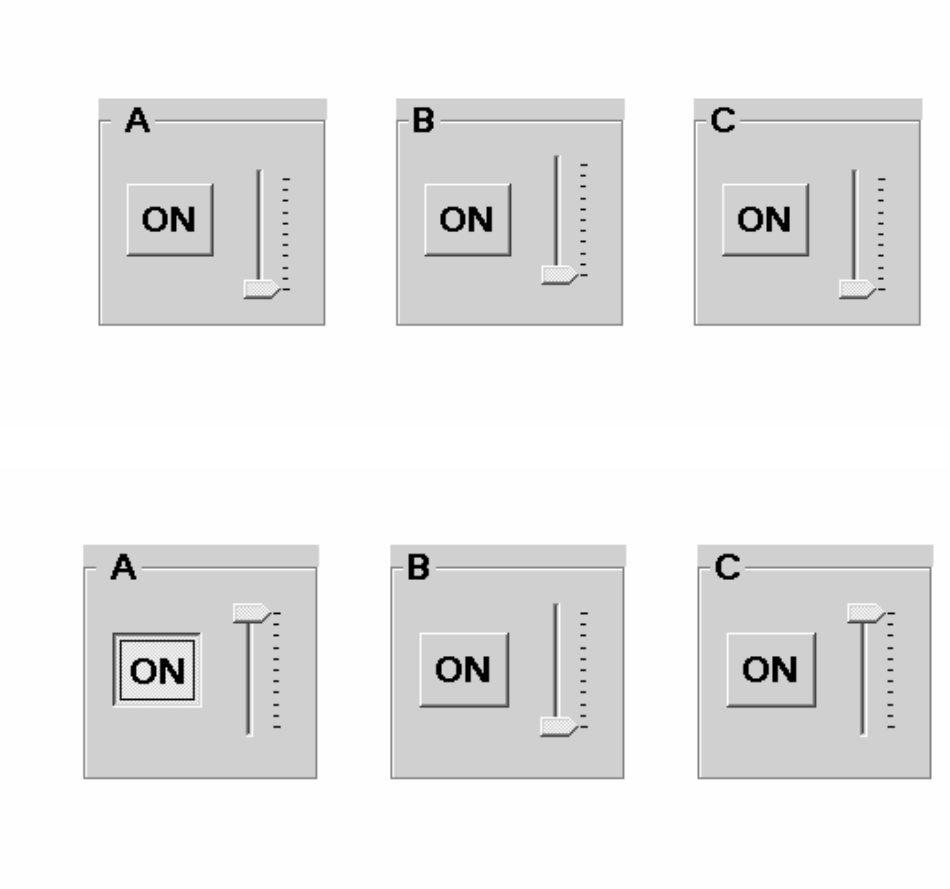


Figure 5

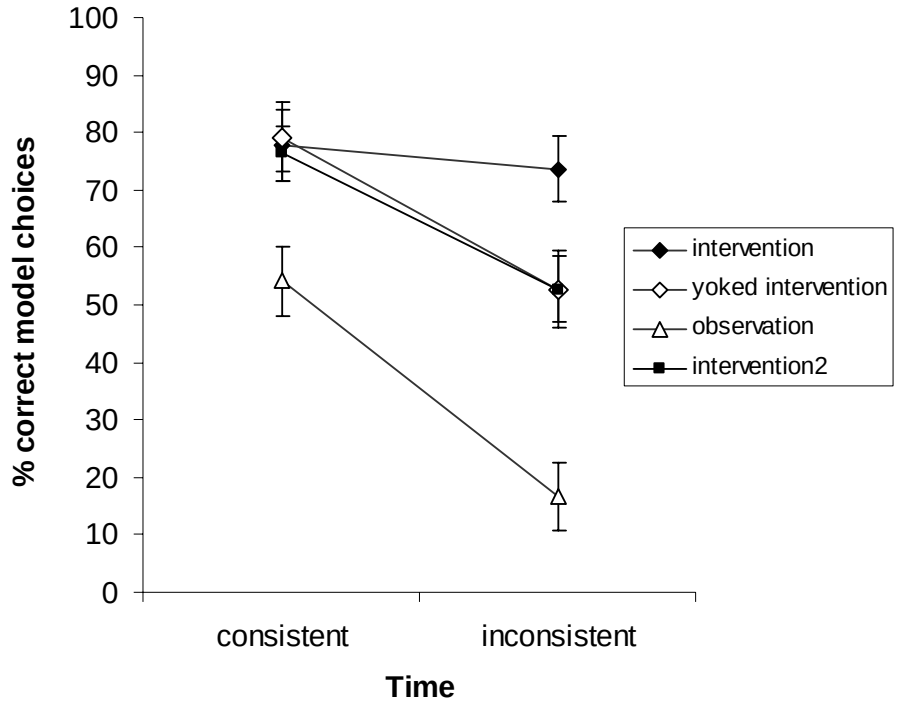
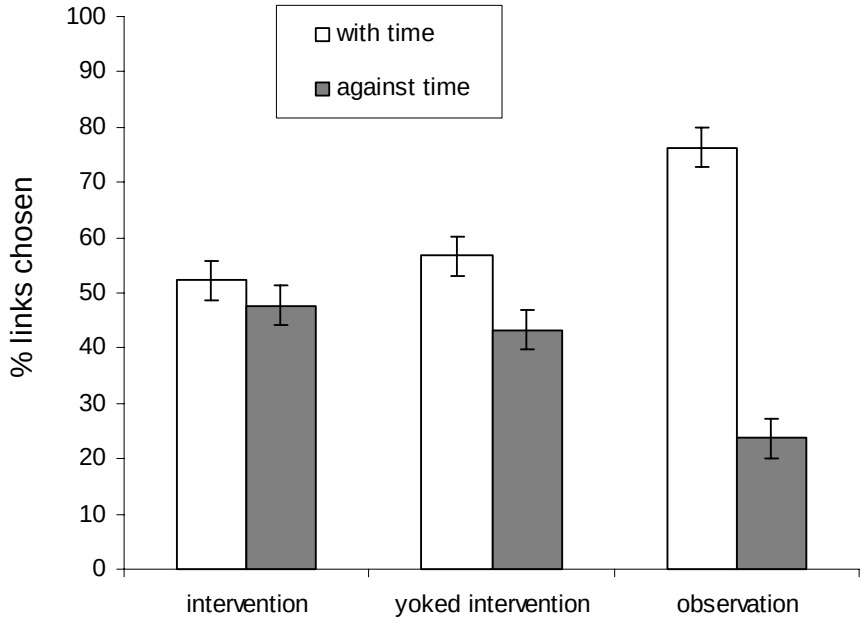

100
90
80
70
60
50
40
30
20
10
0
% correct model choices
consistent
inconsistent
Time
intervention
yoked intervention
observation
intervention2

Figure 6


100
90
80
70
60
50
40
30
20
10
0
% links chosen
with time
against time
intervention
yoked intervention
observation