

# The Influence of Delays in Real-Time Causal Learning

David A. Lagnado\* and Maarten Speekenbrink

*Department of Cognitive, Perceptual and Brain Sciences, University College London*

**Abstract:** The close relation between time and causality is undisputed, but there is a paucity of research on how people use temporal information to inform their causal judgments. Experiment 1 examined the effect of delay variability on causal judgments, and whether participants were sensitive to the presence of a hastener cue that reduced the delay between cause and effect without changing the contingency. The results showed that higher causal ratings were given to cause-effect pairs with less variable delays, but that conditions with an active hastener actually reduced participants' ratings of the causal cues. The latter finding can also be explained in terms of people's sensitivity to variability, because an undetected hastener leads to greater variability in experienced delays. Experiment 2 followed up previous research showing that people give higher causal ratings to cause-effect pairs with shorter delays. We examined whether this finding might be due to the greater probability of intervening events rather than the length of delay per se. The results supported the former conjecture: participants' causal ratings were influenced by the probability of intervening events in the cause-effect interval and not the mere length of delay. The findings from both experiments raise questions for current theories of causal learning.

**Keywords:** Causality judgments, temporal delays, contingency, delay variability, hasteners.

## INTRODUCTION

Causal knowledge is critical to the prediction, control and explanation of the world around us. There are numerous routes by which people acquire this knowledge, including covariation [1-4], interventions [5-7], temporal order [8-10], and prior knowledge [11, 12]. These different sources of information serve as noisy indicators of the underlying causal structure [13, 14]. No single source is an infallible guide, but in combination they can provide compelling evidence in favour of an underlying causal model. When confronted with a novel system, people seek to construct a generative causal model that allows them to mentally simulate and thus predict the system's actual and potential behaviour [15, 16]. This is often a hard inference problem, because the data the learner receives about the system can be noisy, incomplete and ambiguous. Hence the need to exploit multiple sources of evidence, under the assumption that regularities in the patterns of data generated by the system are determined by its causal structure.

In this paper we will focus on temporal information, and in particular how people use information about temporal delays between events to infer causal relations. Causal systems operate in real-time, and involve mechanisms and processes that unfold over time. Thus temporal information about the timing of events provides a natural source of knowledge about the underlying causal structure of a system. Temporal priority is a clear-cut example. Causes precede their effects, and therefore events that occur after a target

event cannot be causes of that event. People appear to be very sensitive to this constraint [9]. A more complex case is the temporal contiguity between events. The consensus view, ever since Hume [17], is that the more contiguous (closer in time) two events are, the more likely they are to be causally related. Previous research supports this view [18, 19] showing that when everything else is held constant (e.g., there is the same contingency between cause and effect), participants gave higher causal ratings to cause-effect pairs with shorter delays (and after a certain threshold delay, they no longer judged the pair as causally related). An important caveat is that in certain contexts, and with certain types of causal relation, people can tolerate much longer delays. One clear demonstration of this is when people have prior knowledge about the causal mechanism that explains the delays [20, 21].<sup>1</sup> For example, people expect a short delay between the click of a mouse and the appearance of an object on the computer screen, but tolerate a longer delay between the eating of seafood and a subsequent allergic reaction. This interplay between temporal cues and mechanism knowledge highlights the fact that the informational sources for causal relations are fallible, and that they can combine in a synergistic fashion.

## Variability of Delay

One dimension of temporal information that remains relatively unexplored is people's sensitivity to the *distribution* of time delays between putative cause and effect. This is an additional source of information about causality, because a near constant time delay between two event types is itself a

\*Address correspondence to this author at the Department of Cognitive, Perceptual and Brain Sciences, University College London, Gower Street, London WC1E 6BT, UK; Tel: +44 020 7679 5389; Fax: +44 020 7436 4276; E-mail: d.lagnado@ucl.ac.uk

<sup>1</sup>There are also situations where people might tolerate long delays without having explicit knowledge about the underlying causal mechanism (cf. taste-aversion learning in rats [25]).

suggestive cue that the events are causally related. For example, a precise and regular time interval between occurrences of X and Y seems unlikely under the hypothesis that the two events are entirely unrelated. A more plausible hypothesis is that they are causally related. This provides a natural explanation for the constant temporal delay between them. Such an inference is less compelling when the delay between X and Y is more variable. This is because the variability in delays itself calls out for some explanation, and one possibility is that there are alternative causes of the effect Y. The postulation of alternative causes is unnecessary in the constant delay case.

The inference that high variability in delays between putative cause and effect signals the presence of alternative causes is by no means infallible. There are various other possible reasons for the delay variability that need not undermine the causal status of the X-Y pair. One possibility is the action of *hasteners* – events that change the timing of an effect without altering the probability of the effect (see below for more details). The action of an unobserved hastener might account for the variability in delays between X and Y without detracting from the causal relation between them.

However, even if such possibilities are taken into account, the fact remains that a constant delay between cause and effect is better evidence (other things being equal) than a highly variable delay. This is because the latter case raises the possibility that alternative unknown causes are responsible, whereas the former requires no additional explanation.

It is important to note that information about delay variability is independent of: (i) the actual mean delay (since distributions can vary in both mean and variability, see Figs. 2 & 3); and (ii) the actual covariation between events (since two events can have identical contingency, and yet differ in the temporal distributions of their delays). This is not to say that the *perceived* contingency between events is similarly independent. Indeed it is possible that computations of contingency from a learner's perspective are influenced by the nature of delays between events [22]. This possibility will be discussed further in the General Discussion.

There has been very little research on how the distribution of temporal delays affects people's causal judgments, although there are suggestions in the literature that people are sensitive to the variability of delays. For example, Wasserman, Chatlosh and Neunaber [23] found that participants gave slightly higher causal ratings in a constant delay condition than in a variable delay condition, even though the mean delay was equivalent in both conditions. Their study used a free-operant paradigm (participants evaluated the effect of a button press on the illumination of a light), so it is possible that this study underestimated the effect of delay variability because of the powerful cue already provided by the participants' own interventions.

### Hasteners

Another unexplored issue is whether people are sensitive to factors that change the temporal distribution of delays without actually changing the probability of the effect. Such a factor is termed a *hastener* in the philosophy literature [24]. For example, buying 'speedy-boarding' on a cheap airline, allows you to board the plane quicker, but does not

alter whether or not you board the plane. Indeed there seem to be many real-world situations where the presence of a specific factor speeds up the time at which an effect occurs, without changing whether or not the effect occurs. Are people sensitive to this possibility? Would such features be categorized as causal, even though they only modulate the time at which the effect occurs, and do not change the contingency between cause and effect? To our knowledge these are unexplored questions.

Experiment 1 is designed to look at people's sensitivity to delay variability, and to the presence of hasteners. Neither question has received much attention in the psychological literature, and yet both might play a role in the use of temporal information to uncover causal structure. We predict that greater variability in the delays between cue-outcome pairs will lead to lower causal ratings. This is based on our hypothesis that higher delay variability will encourage participants to postulate alternative causes of the effect. It also follows from a more general supposition that people are sensitive to the uncertainty in sources of information, and attribute less weight to more unreliable (variable) sources.

### Delays and Intervening Events

The classic finding with respect to delays between putative causes and effects, as mentioned above, is that an increased delay tends to lower people's attributions of cause [18]. There are two main explanations for this finding, both of which have their roots in associative learning theory [25]. One account maintains that it is the duration of delay between events that directly affects the judged causal link between event pairs. This *intrinsic delay process* is readily explicated in associative terms [18]. According to associative theories, causal judgments are based on the formation of an associative link between mental representations of the cue and outcome, and the strength of this link dictates the strength of the explicit judgment. The greater the delay between presentations of cue and outcome, the weaker the associative link between their mental counterparts.

This is a simple and intuitive explanation, and can be given a rational justification under the assumption that closer events are more likely to be causally linked (other things being equal). An alternative explanation is that the longer the delays between putative cause and effect, the higher the probability that other causes might have occurred during that interval and produced the effect in question. This account can also be couched in associative terms. The intervening events will compete with the putative cause to predict the target effect, and thus reduce attributions to that cause alone [25]<sup>2</sup>.

This intervening events explanation is similar to our earlier speculation about the effect of delay variability on causal judgments, whereby higher delay variability is expected to lower causal ratings because it increases the feasibility of alternative possible causes of the effect. The disruptive potential of intervening events (even if they are unrelated to the actual effects) is magnified by the fact that they occur in between the actual cause and effect pairs. The qualitative temporal ordering (cause—intervening event—effect) pro-

<sup>2</sup>We thank Jose Perales for useful suggestions on this section, including the terms 'intrinsic delay process' and 'intervening-events' explanation.

vides weak evidence in favour of the intervening event as a cause of the effect [9]. The effect of delay on causal learning, and the viability of these two alternative explanations, will be investigated in Experiment 2.

### Overview of Experiment 1

Experiment 1 will explore the effect of delay variability on people's causal judgments. It will also examine whether people's judgments are sensitive to the presence of hasteners that reduce the delay between cause and effect without themselves changing the probability of the effect. To investigate these issues we use two novel causal learning tasks. In both tasks participants seek to identify the causes of a medical condition on the basis of real-time exposure to a virtual physiological system, with time varying presentations of putative causes (e.g., different types of bacteria) and effects (e.g., stomach cramps). In the learning phase participants watch a real-time animation of an endoscopic examination, and evaluate the causal efficacy of several types of bacteria/viruses (characterized by distinctive features such as feelers, tails and spots). To explore people's sensitivity to delay variability we have both high and low variability conditions. To explore their sensitivity to hasteners we have conditions where the hastener is active or inactive. People's causal judgments are assessed *via* explicit ratings of absolute causal strength, and through comparative ratings of causal contribution.

## EXPERIMENT 1

### Method

#### *Participants and Apparatus*

46 UCL students were paid £3 to participate in the experiment. Participants were run individually on computers using specially written software in C#.

#### *Design*

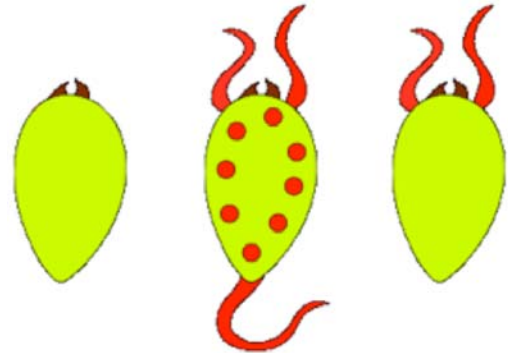
A 2x2 mixed design was used, with Hastener (active, inactive) as a between-subject factor and Variability (low, high) as a within-subject factor.

#### *Materials*

Participants were each presented with two causal learning tasks. These were formally identical, but differed in terms of the cover story and nature of the stimuli. Both tasks were couched in a medical setting, one involved detecting the effects of bacteria on stomach upsets, the other involved the effects of viruses on lung inflammation. In the bacteria task, participants had to infer the causal relation between different types of bacteria and a stomach cramp. The bacteria were depicted on the screen in a pictorial form. All bacteria had the same body and head shape, but differed in terms of the presence or absence of three features: tail, feelers, and spots on the back (see Fig. 1). Unknown to the participants, one of these features determined the probability of the effect (*cause*), another determined the time at which the effect occurred (*hastener*), and the third was inactive (*lure*). Which feature played which role was counterbalanced across participants.

There were eight different cue combinations (see Table 1), and each pattern was presented 10 times through the

course of the learning task, making a total of 80 presentations. The order of these was randomized for each participant.



**Fig. (1).** Examples of bacteria used in movie: left bacterium has no features present, middle bacterium has all three features present, right bacterium has only feelers present.

In the learning phase of the task, participants watched a short movie that represented the real-time pictures from an endoscope placed inside the stomach of a patient. Bacteria were presented at random locations against a constant background depicting the inside of the stomach. The time interval between the start of the experiment and the presentation of the first bacterium, and between subsequent appearances of bacteria, were determined by random draws from an exponential distribution with a mean of 6 seconds. An exponential distribution was used to ensure that the time intervals between bacteria were independent. Once a bacterium appeared on the screen, it persisted for two seconds before disappearing. The presence of the effect (a stomach cramp) was represented by a short red flash. The probability of the effect was determined by the presence or absence of the cause feature. When this feature was present, the probability of the effect was 0.8; when it was absent, the probability of the effect was 0.1. Thus, the presence of this feature greatly raised the probability of the effect, but there was still a small probability that the effect would occur in the absence of this feature (but only given the presence of a bacterium). The time delay between the appearance of the bacterium and the appearance of the effect was determined by whether the hastener was present (presence or absence of the hastener feature), and whether the hastener was active or inactive. This latter variable was a between-subject variable – half the participants always saw active hasteners, half the participants always saw inactive hasteners (i.e. equivalent to a second lure).

In the hastener active condition, when the hastener feature was present the mean time delay between cause (bacterium) and effect (stomach cramp) was 0.5 seconds; when it was absent, the mean time delay was 2 seconds. The active hastener reduced the time delay between bacteria and effect whether or not the cause feature was present. In the hastener inactive condition the mean delay between cause and effect was 2 seconds irrespective of the presence or absence of the hastener feature. For a summary of the probabilities and time delays see Table 1.

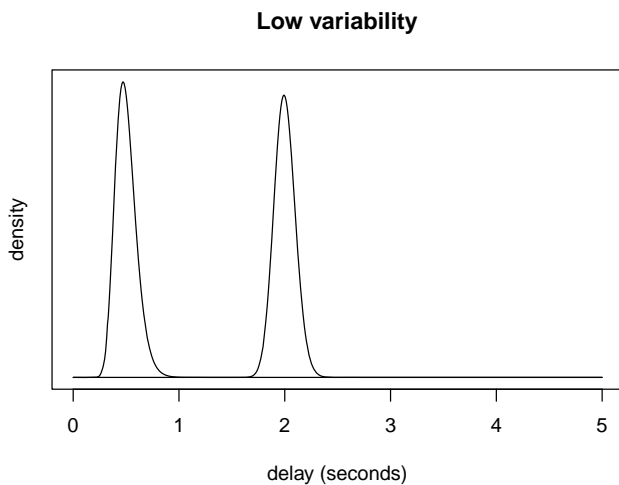
The variability of the delay between cause and effect was manipulated within-subjects, with two levels of variability:

**Table 1. Summary of Probabilities and Time Delays for the Eight Cue Combinations in the Learning Phase in Experiment 1**

Pattern	Cause	Hastener	Lure	P(Effect)	Hastener Active Mean Time Delay (Seconds)	Hastener Inactive Mean Time Delay (Seconds)
A	1	1	1	0.8	0.5	2
B	1	1	0	0.8	0.5	2
C	1	0	1	0.8	2	2
D	1	0	0	0.8	2	2
E	0	1	1	0.1	0.5	2
F	0	1	0	0.1	0.5	2
G	0	0	1	0.1	2	2
H	0	0	0	0.1	2	2

high and low. Thus, each participant experienced one task with low variability delays, and the other with high variability delays. In the low variability condition, the delays between cause and effect were randomly sampled from a lognormal distribution with a standard deviation of 0.1 seconds. In the high variability condition, the delays were randomly sampled from a lognormal distribution with a standard deviation of 1 second. Lognormal distributions were used to avoid the possibility of negative delays.

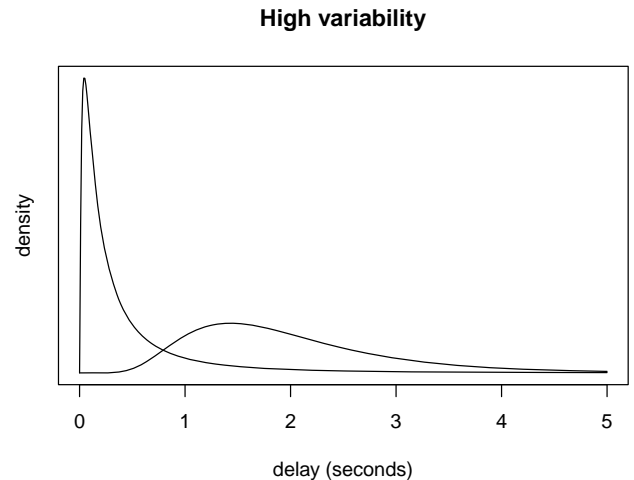
Figs. (2 and 3) show the distributions of delays for low and high variability conditions respectively. Note that in the active hastener condition the mean delay is 0.5 seconds when the hastener feature is present, and 2 seconds when it is absent. In the inactive hastener condition the mean delay is always 2 seconds.



**Fig. (2).** Distributions for delays between cause and effect in the low variability condition. Note that in the active hastener condition, mean delay is 0.5 secs when hastener is present, 2 secs when absent. In inactive hastener condition, mean delay is always 2 secs.

The virus task was formally equivalent to the bacteria task, but used a different cover story and visual stimuli. Participants had to infer the causal relation between different types of virus and lung inflammation. All viruses had the same circular body shape with three arms, but differed in

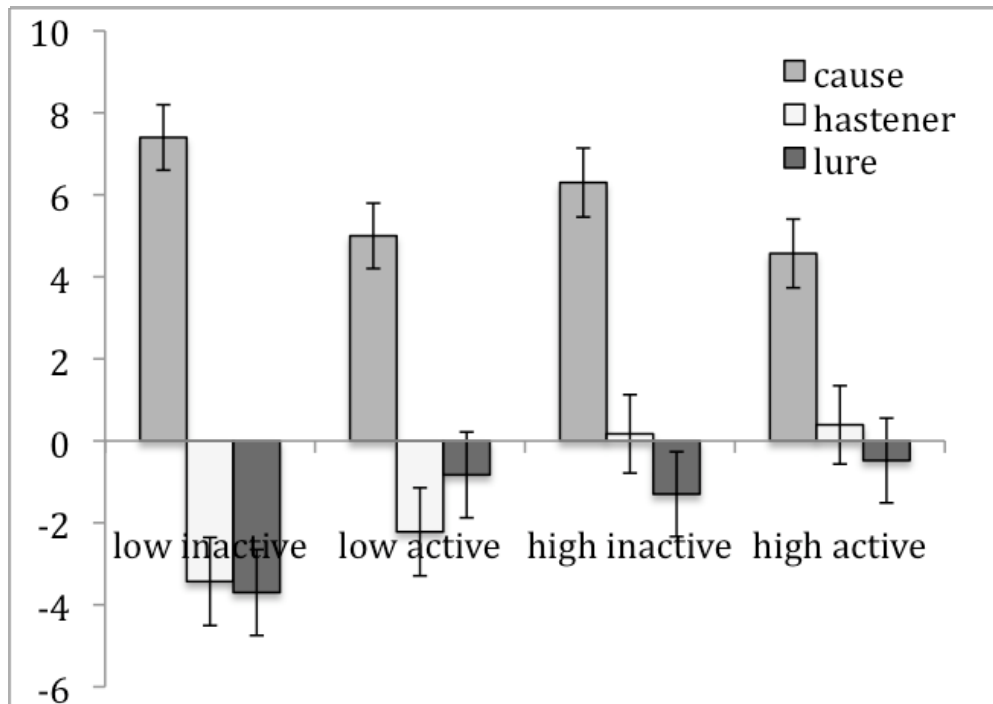
terms of the presence or absence of three features attached to these arms: a square, circle or triangle. The presence of lung inflammation was represented by a short green flash. In all other respects this task was identical to the bacteria task. The order of tasks was counterbalanced between participants.



**Fig. (3).** Distributions for delays between cause and effect in the high variability condition. Note that in the active hastener condition, mean delay is 0.5 secs when hastener is present, 2 secs when absent. In inactive hastener condition, mean delay is always 2 secs.

**Procedure**

Each participant completed two tasks. Before each task they were told that they were to take the role of a medical investigator assessing the effects of different bacteria (viruses) on the stomach (lungs). They were told that the bacteria (viruses) could differ with respect to the presence or absence of three features. They also performed a brief recognition task to check that they could correctly identify these different features. Each task consisted in a learning phase and a judgment phase. In the learning phase participants watched a movie that lasted for approximately 8 minutes. In the movie they viewed the real-time presentation of bacteria (viruses) and stomach cramps (lung inflammation). The former were represented by the visual images shown in Fig. (1), the latter by flashes on the screen.



**Fig. (4).** Mean absolute cause ratings ( $\pm$ SEM) for each cue type by delay variability (low, high) and hastener (active, inactive).

### Judgment Phase

At the end of the movie participants gave both absolute and comparative causal ratings. For the absolute ratings, participants were shown each of the three features separately, and rated the extent to which each feature caused the effect (stomach cramps or lung inflammation). Participants registered their ratings by moving a slider on a scale from +10 ('completely causes') to -10 ('completely prevents'), with 0 standing for 'neither prevents nor causes'. For the comparative ratings, all three features were presented together and participants were asked to divide up 100 points between each feature, according to its relative causal contribution to the effect (with higher numbers denoting greater causal impact).

## RESULTS

Overall there were no differences between the two tasks, so subsequent analyses ignore this factor.

### Absolute Causal Ratings

Fig. (4) presents the mean ratings for each cue type (cause, hastener, lure) according to delay variability (low vs. high) and whether the hastener was active or inactive. Inspection of Fig. (4) shows that participants rated the cause cue much higher than the other two cues in all conditions, and also suggests that the cause cue is rated higher when the hastener is inactive rather than active. It is also notable that the lure was given negative ratings in all conditions, and that the hastener was given negative ratings in the low variability conditions.

To explore these observations, a 2x2x3 ANOVA was conducted with Cue Type and Variability (low, high) as within-subject factors, and Hastener (active, inactive) as a

between-subject factor. There were main effects of Cue Type,  $F(2,88) = 52.65$ ,  $p < .001$ , and Variability,  $F(1,44) = 9.65$ ,  $p < .005$ , and significant interactions between Cue Type and Variability,  $F(2,88) = 4.39$ ,  $p < .05$ , and Cue Type and Hastener,  $F(2,88) = 3.10$ ,  $p = .05$ . There was no main effect of Hastener, and no other interactions.

The Cue Type by Variability interaction was explained by the hastener cue, which was rated negatively in the low variability condition (-2.83) but not in the high variability condition (.28),  $t(45) = 3.67$ ,  $p < .005$ . In contrast, the cause cue was rated more highly in the low variability condition (6.17) than in the high variability condition (5.43), although this difference was not significant,  $t(45) = 1.18$ , ns. The Cue Type by Hastener interaction was explained by the cause cue, which was rated higher when the hastener was inactive (6.85) than when it was active (4.79),  $t(44) = 2.16$ ,  $p < .05$ .

### Comparative Causal Ratings

Fig. (5) presents the mean ratings for two cue types (cause, hastener) according to delay variability and whether the hastener was active or inactive. There was no need to include ratings for the lure cue in the analysis, because these were completely constrained by the values for the other two cues (i.e., lure = 100-cause-hastener). Inspection of Fig. (5) shows that participants rated the cause cue much higher than the hastener cue in all conditions. It also suggests that the cause cue is rated higher when the hastener is inactive rather than active, and when variability is low rather than high.

A 2x2x2 ANOVA was conducted with Cue Type (cause, hastener) and Variability (low, high) as within-subject factors, and Hastener (active, inactive) as a between-subject factor. There was a main effect of Cue Type,  $F(1,44) = 90.10$ ,  $p < .001$ , a marginal effect of Hastener,  $F(1,44) = 3.24$ ,

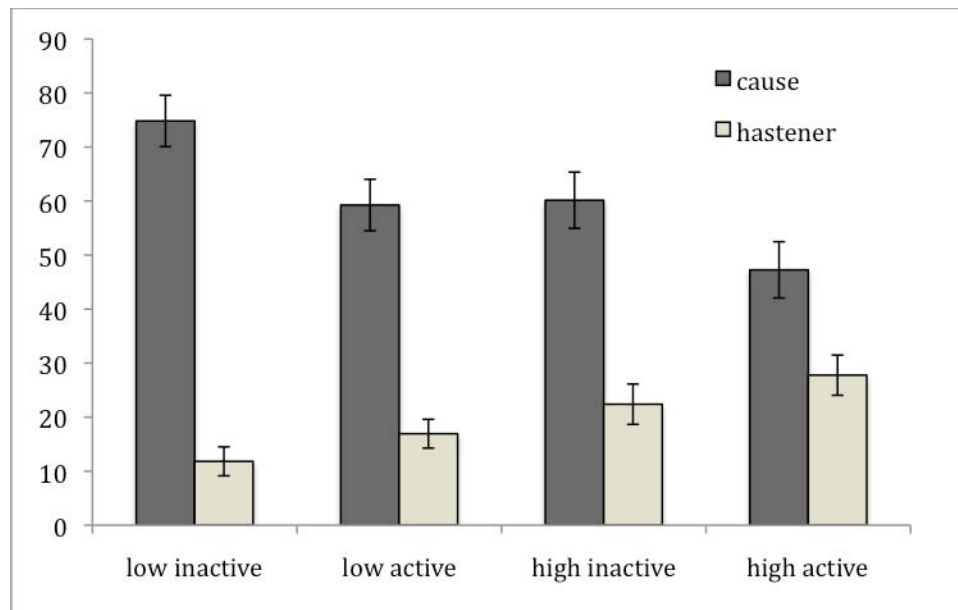


Fig. (5). Mean comparative cause ratings ( $\pm$ SEM) for each cue type by delay variability (low, high) and hastener (active, inactive).

$p = .079$ , but no effect of Variability,  $F(1,44) = .452$ , ns. There were significant interactions between Cue Type and Variability,  $F(1,44) = 18.53$ ,  $p < .001$ , and Cue Type and Hastener,  $F(1,44) = 5.17$ ,  $p < .05$ . There were no other interactions.

The Cue Type by Variability interaction was due to the fact that the cause cue was rated higher in the low variability condition (67.05) than the high variability condition (53.71),  $t(45) = 3.68$ ,  $p < .005$ , whereas the hastener cue was rated higher in the high variability condition (25.08) than the low variability condition (14.38),  $t(45) = 3.78$ ,  $p = .001$ . The Cue Type by Hastener interaction was explained by the cause cue, which was rated higher when the hastener was inactive (67.49) than when it was active (53.26),  $t(44) = 2.31$ ,  $p < .05$ .

## DISCUSSION

The absolute and comparative causal ratings pointed to similar conclusions. First, participants in all conditions managed to identify the unique feature that raised the probability of the effect, and they gave this feature higher causal ratings than the other two features. This fits with previous research showing that causal ratings are sensitive to the contingency between cause and effect [2, 4]. Second, the presence of an active hastener lowered the ratings given to the cause cue, both in conditions of high and low variability of delay. This is a notable finding, especially since the mean delay in the inactive hastener condition (2 secs) is longer than that in the active hastener condition (1.25secs). This seems to conflict with previous studies showing that people attribute greater causality with shorter delays between cause and effect [18]. However, as discussed below, this finding is potentially explained in terms of people's sensitivity to the variability of the delay between cause and effect. Third, the variability of time delay had an effect on causal ratings for both the cause cue and the hastener. For the comparative ratings, there was an intriguing interaction between the effects of variability of delay and the presence of the hastener. The cause cue was

rated as *more* causal when variability was low rather than high, but in contrast the hastener cue was rated as *less* causal when variability was low rather than high.<sup>3</sup> This pattern was also found for the absolute ratings, although only the hastener cue showed a significant effect of delay variability.

### Effect of Hastener

The difference in causal ratings for the cause, whereby ratings were lower when the hastener was active rather than inactive, is initially puzzling since the contingency in both conditions is the same, and the mean delay between cause and effect is actually lower in the active hastener condition. One plausible explanation for this effect is that in the active condition the cause-effect delays are more variable because they are effectively being sampled from two separate distributions (either mean = 0.5s when hastener is present or mean = 2s when hastener is absent). If participants are unaware that the hastener actually signals this switch in distribution, the active hastener condition will entail greater variability, and thus lower causal ratings for the cause.

### Effect of Variability

As predicted causal ratings for the cause were higher when the cause-effect delay was less variable. However, we also found that causal ratings for the hastener were *lower* when the cause-effect delay was less variable. This effect was not predicted, but might be because the low variability condition provided better conditions to uniquely identify the causal cue, and it was therefore easier to rule out the hastener feature as a cause.

<sup>3</sup>The reason for this discrepancy between absolute and comparative ratings is unclear. In a sense the comparative measures were the most appropriate in this context, because participants were choosing between alternative causal explanations. The absolute ratings were also potentially confusing because they allowed for negative ratings (i.e. a factor preventing an effect) whereas there was no mention of the possibility of preventative causes in the cover story. Thus it is possible that participants used this part of the scale to emphasize the difference between judged causes and judged non-causes. The comparative ratings provided a more straightforward way of doing this.

## EXPERIMENT 2

In the previous experiment, we showed that delay variability between cause and effect influenced people's causal judgments about the cause-effect relation. As mentioned in the introduction, previous studies [18, 19] have shown that causal attribution was susceptible to the length of delay between a potential cause and effect. In Experiment 2 we focus on the locus of this effect. In particular, we investigate whether it is the length of delay itself, or whether it is the occurrence of intervening events during that delay which drives the delay effect. Experiment 2 was designed such that the average delay between cause and effect, and the probability of intervening events between cause and effect, were varied independently.

## METHOD

### Participants and Apparatus

Twenty participants participated in the experiment for course credit or 3 pounds. Participants were recruited from the University College London subject pool. The mean age was 20.45 years ( $SD=2.96$ ). Participants were tested individually in sound dampened rooms using custom software written in C#.

### Design

The experiment had a 2x2 within subjects design, with average Delay (short, long) between cause and effect and the

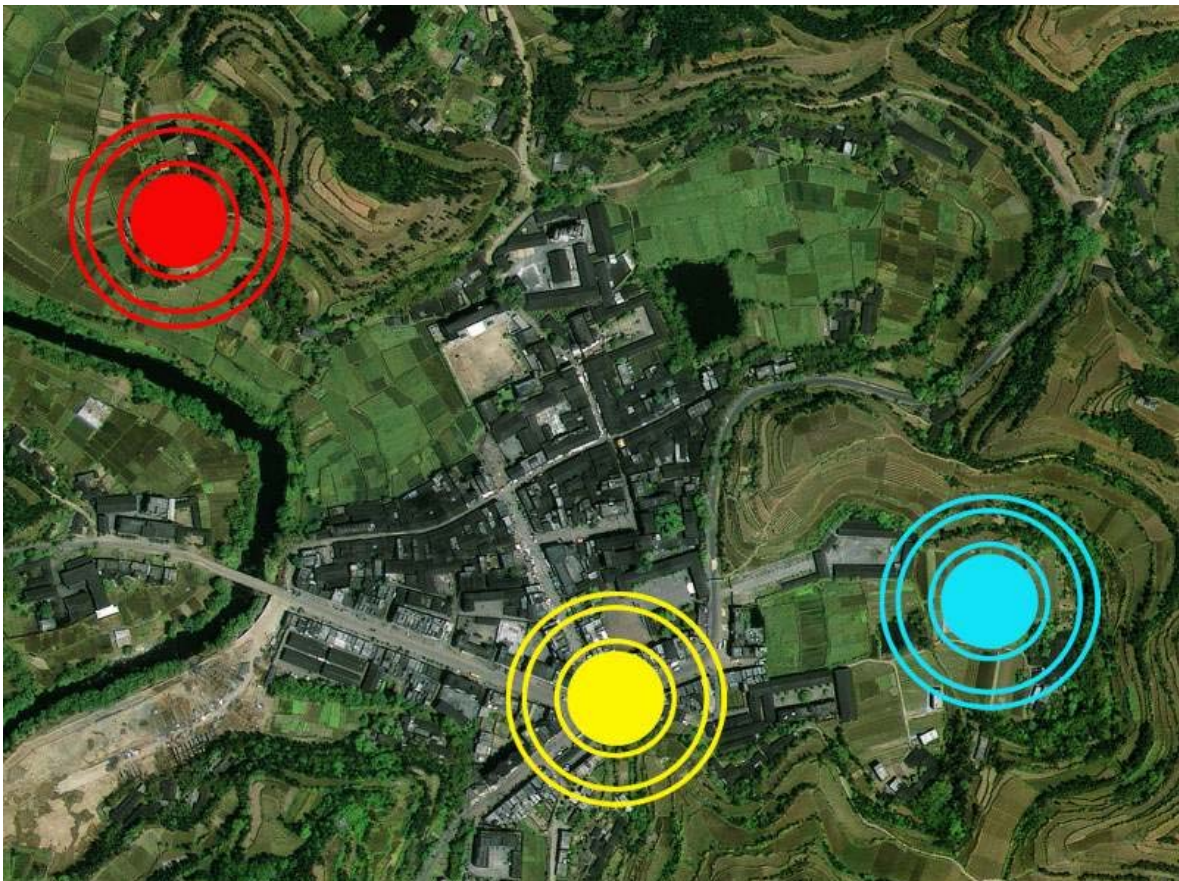
Probability of Intervening Events (PoIE: low, high) between cause and effect as experimental factors. The order of the resulting 4 conditions was counterbalanced across participants.

### Materials

Participants were each presented with four causal learning tasks. Each task consisted of a learning phase and a causal judgement phase. Participants were asked to take the role of a geologist, investigating the effects of different types of seismic waves on the occurrence of earthquakes. In each task, a different triad of wave types was studied in a particular geographical region. Unbeknownst to the participants, only one of the three types of wave (the *cause*) raised the occurrence of earthquakes, while the other two types (*lures*) had no effect.

Each task consisted of a learning and causal judgement phase. In the learning phase, participants watched a short movie that represented real-time observations of a particular geographical region. Seismic waves were presented at random locations on a static satellite image of the region (see Fig. 6). The waves were represented by circular shapes that blinked in alarm-like manner for 1 second and each wave type had a distinct colour. Earthquakes were represented by a bright flash on screen.

The cause wave had a probability of  $p = 0.8$  of resulting in an earthquake. If the wave was effective, the earthquake would occur after a random delay sampled from a lognormal



**Fig. (6).** Example of primary task stimuli. The circular images symbolise waves (i.e., cues). The different colours distinguish between 3 separate waves.

distribution. The mean of this distribution was 3 seconds in the short delay condition, and 6 seconds in the long delay condition. The standard deviation was 0.1 seconds in both conditions (the same value as the low variability conditions of Experiment 1). In addition to the earthquakes caused by the target cue (8 in total), there were also a total of 4 random occurrences of earthquakes, which served as a “background rate” of earthquakes.

Each wave type (Cause, Lure A, Lure B) was presented a total of 10 times during a task, with the order randomized for each participant. The time interval between waves was sampled randomly from an exponential distribution. In the low PoIE conditions, the mean of this distribution was chosen such that the probability of a wave occurring between the cause and effect was  $p = 0.35$ . In the high PoIE condition, the mean was chosen such that this probability was  $p = 0.65$ .

**Procedure**

Each participant completed four tasks. Before the first task, they were informed they were to take the role of a geologist investigating the causes of earthquakes. They were then shown the general appearance of the waves and earthquakes. Participants were informed that there would be 4 animations, each lasting at most 10 minutes.

At the start of each task, participants were told the types of wave they would be investigating, as well as the region in which the investigation took place. They were informed that sometimes waves may cause earthquakes and sometimes they may not and that an earthquake may take time to develop after the occurrence of a wave. In addition, they were reassured that some periods in which no waves would occur were normal. Participants were notified that there would be questions about the causal strength of each tremor at the end of each animation.

Each task consisted of a learning and causal judgment phase. In the learning phase, participants watched a movie that lasted between 1.28 and 9.25 minutes. The average length of each movie varied between conditions and was 1.43 in the short delay/high PoIE condition, 4.21 in the short

delay/low PoIE condition, 2.86 in the long delay/high PoIE condition and 8.41 minutes in the long delay/low PoIE condition. In each movie, they observed different types of seismic waves and earthquakes. The former were represented by the alarm-like circular shapes, and the latter by bright flashes. In each task, a different background image was used, to represent a different geographical region. Also, the seismic waves had different colours in each task. The order of regions, and the colour of the cause and lure waves, was randomized for each participant.

The judgment phase was similar to that of Experiment 1: for each type of wave, they were first asked to rate the extent to which it causes earthquakes, on a scale of 0 (does not cause the effect) to 10 (completely causes the effect). They were then asked for comparative ratings, in which they divided 100 points amongst the three types of cues.

**RESULTS**

**Absolute Ratings**

Fig. (7) depicts the mean ratings for each cue type.

The absolute ratings were analysed with a 2 (Delay) x 2 (PoIE) x 3 (Cue Type: cause, lure A, lure B) repeated measures ANOVA. This showed a main effect for Cue Type,  $F(2,38) = 40.21$ ,  $MSE = 8.57$ ,  $p < .001$ , which is due to the cause receiving a higher rating than either lure A,  $t(159) = 12.85$ ,  $p < .001$ , and lure B,  $t(159) = 11.78$ ,  $p < .001$ , while there was no difference between the lures,  $t(159) = 1.48$ ,  $p = .14$ . In addition, there was a main effect of PoIE,  $F(1,19) = 8.40$ ,  $MSE = 3.175$ ,  $p < .01$ , indicating overall higher ratings when the PoIE was high ( $M = 5.65$ ,  $SD = 3.17$ ) as compared to low ( $M = 5.16$ ,  $SD = 3.45$ ). The main effect of Delay was not significant,  $F(1,19) = .45$ ,  $MSE = 2.365$ ,  $p = .51$ . There were no significant interactions, although there was a trend towards a significant interaction between Delay and PoIE,  $F(2,38) = 3.47$ ,  $MSE = 5.55$ ,  $p = .078$ , possibly indicating that the effect of PoIE is larger for shorter delays (low PoIE:  $M=5.72$ ,  $SD=3.15$ ; high PoIE:  $M=5.13$ ,  $SD=3.52$ ) than for longer delays low PoIE:  $M=5.56$ ,  $SD=3.19$  ; high PoIE:  $M=5.19$ ,  $SD=3.40$ ).

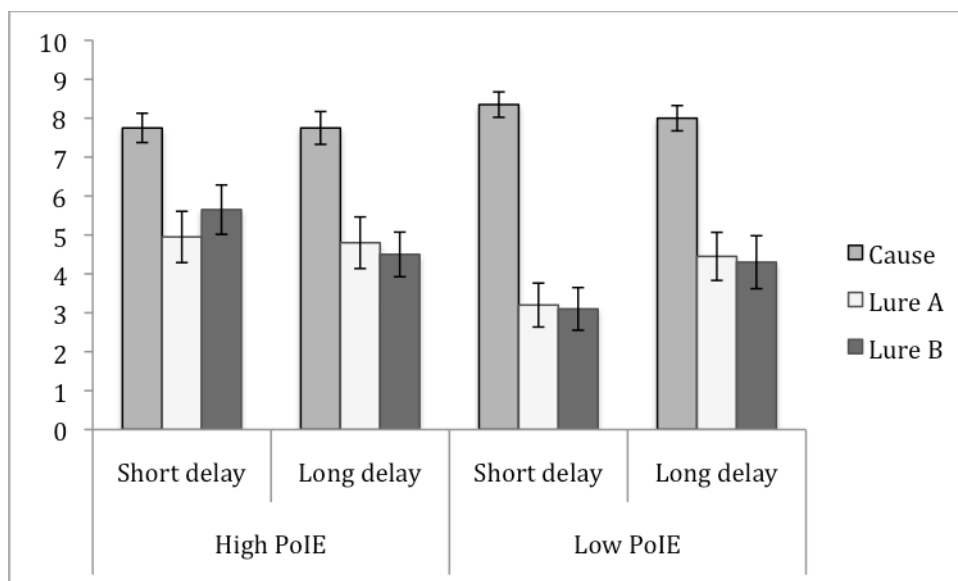


Fig. (7). Mean absolute ratings (±SEM) by delay and probability of intervening events (PoIE).



## Comparative Ratings

As in Experiment 1, we only analysed the ratings for the cause and one of the lures. The mean comparative ratings are shown in Fig. (8).

The comparative ratings were analysed with a 2 (Delay) x 2 (PoIE) x 2 (Cue Type: predictor, lure A) repeated measures ANOVA. The results of this analysis agreed largely with those for the absolute ratings. There was a significant main effect of Cue Type,  $F(1,19) = 101.34$ ,  $MSE=459.55$ ,  $p < .001$ , indicating a higher rating for the cause ( $M=61.33$ ,  $SD=24.87$ ) than the lure ( $M=17.89$ ,  $SD=16.46$ ). In addition, there was a significant interaction between Cue Type and PoIE,  $F(1,19) = 7.42$ ,  $MSE=474.18$ ,  $p = .013$ , which was due to the cause receiving a higher rating under a low PoIE ( $M=65.49$ ,  $SD=24.42$ ) compared to a high PoIE ( $M=57.18$ ,  $SD=24.77$ ),  $t(79) = 2.09$ ,  $p < .05$ , and the lure receiving a lower rating under a low PoIE ( $M=15.20$ ,  $SD=16.59$ ) compared to a high PoIE ( $M=20.58$ ,  $SD=15.98$ ), although this difference just failed significance,  $t(79) = 1.98$ ,  $p=.051$ . Other effects were not significant. In particular, there was no interaction between Delay and Cue Type,  $F(1,19) = .67$ ,  $MSE = 583.87$ ,  $p = .43$ .

## DISCUSSION

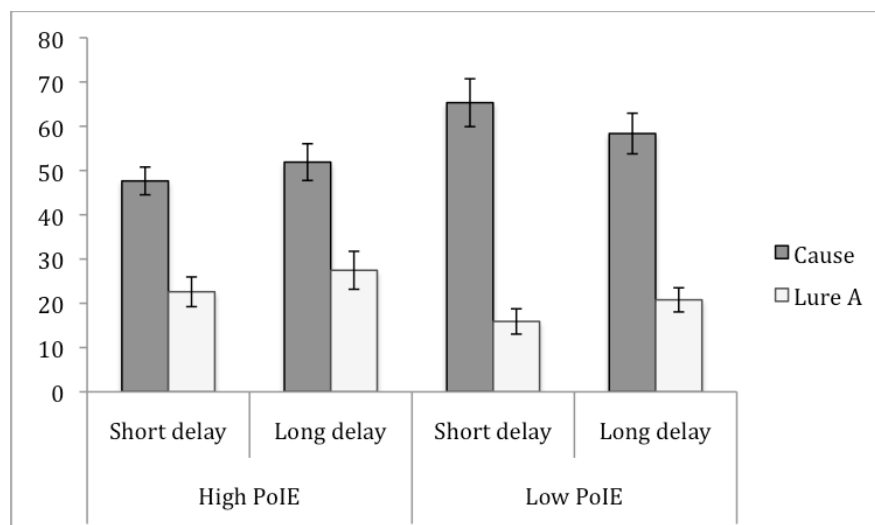
As in Experiment 1, the results of the absolute and comparative causal ratings were similar. As before, participants in all conditions managed to distinguish between a cause raising the probability of the effect and lures that had no effect. While participants were generally good at detecting the cause, their performance was clearly affected by the Probability of Intervening Events. When there was a high likelihood of another cue occurring between cause and effect, the cause was judged as less causally effective, while the lures were rated as more causally effective. In contrast, we found no effects of average delay. Hence, it is the intervening events that affect causal attribution, not the length of the delay between cause and effect. By definition, intervening events occur in between the actual cause and its effect, and because of this qualitative causal ordering they can be more easily confused for having a causal effect.

Experiment 2 used a different cover story (geological events) to Experiment 1 (physiological events) and it could be argued that the former would be more tolerant to temporal delays of the order of seconds. This suggests that it would be worthwhile replicating Experiment 2 using the medical cover story from Experiment 1. However, the results of such a replication would not affect the conclusions we aim to draw from the current studies. If the main finding was not replicated, i.e. there was a main effect of temporal delay not explained by intervening events, then this would point to the role of prior knowledge about mechanisms (e.g., geological vs. physiological) and their differing time frames of action. Our position is consistent with this conclusion. We maintain that people use multiple sources of information to infer causality, including prior knowledge, temporal information, and covariation [14], and that these can combine in synergistic ways.

## GENERAL DISCUSSION

The results from Experiment 1 show that people are sensitive to the variability in delay between cause and effect, assigning greater causal efficacy to less variable cause-effect pairs. This was illustrated in two ways. First, participants gave higher causal ratings to the cause in the low rather than the high variability condition. Second, participants gave lower causal ratings to the cause when the hastener was inactive rather than active. A straightforward interpretation of this finding is that participants experienced greater variability of cause-effect delays in the active condition (delays were sampled from two distributions, mean = 2s and mean = 0.5s) than in the inactive condition (delays were only sampled from one distribution, mean = 2s). This is particularly noteworthy because the overall mean delay is lower (=1.25s) in the active condition, and previous research suggests that shorter delays lead to higher causal ratings [18]. Overall, the pattern of results for both hastener and delay variability can be explained by people's sensitivity to the temporal distribution of delays, and in particular the variability in these distributions.

Experiment 2 followed up on the claim that people give higher causal ratings to cause-effect pairs with shorter delays



**Fig. (8).** Mean comparative ratings ( $\pm$ SEM) by delay and probability of intervening events (PoIE).

[18]. In particular, we explored whether this finding might be due to the probability of intervening events rather than the length of delay per se. The results supported the former conjecture. People's causal ratings were influenced by the probability of intervening events (in the cause-effect interval) and not the mere length of the delay. The greater the probability of intervening events, the more likely participants were to assign causality to these events rather than the actual cause. This effect is compounded by the fact that the qualitative ordering of events (cause—intervening event – effect) also supports this spurious inference. This fits with previous research by Lagnado and Sloman [9] showing that people readily use temporal ordering as a cue to causality, even when it conflicts with contingency information.

Taken together, these findings illustrate people's sensitivity to temporal distributions of cause-effect delays, but also suggest that this might be due to more general level inferences about the possibility of alternative causes of the target effect. In the case of the contrast between high and low delay variability, greater variability in delays might reduce confidence in the target cause-effect pair because of the possibility of alternative (unknown) causes that would explain this variability. In the case of the contrast between short and long delays, longer delays may typically lead to lower causal ratings because a greater number of intervening events are likely to be experienced. These intervening events by definition occur in between the actual cause and effect pairs, and in virtue of this qualitative temporal ordering are more likely to be attributed some causal efficacy.

It should be noted that delay variability could inform people's causal judgments in two ways. First, increased variability can suggest the action of alternative unknown causes, thus lowering one's confidence in the target causal relation. Second, increased variability might lower one's belief about the strength of the putative causal relation. These two possibilities can be conflated when a single causal judgment is taken [26], and our current studies cannot distinguish between them. However, we suspect that it is the former sense that is operating – that delay variability is affecting people's beliefs in the existence of a causal link rather than the strength of this link. Future studies, using additional response measures, could untangle these aspects.

Another issue that is underdetermined by the current data is whether participants in Experiment 1 were aware that the hastener (when active) modulated the time delay between cause and effect. This could be resolved in several ways. One approach, for example, would be to ask participants to make predictions about the likely time of an effect given the cause, and seeing whether their predictions were sensitive to the presence or absence of the hastener. Another approach would be to ask participants explicitly about the 'hastening' role of the hastener.

In addition, it would be informative to run an active hastener condition in which the hastener feature was not shown, but the participants still experienced the same distribution of time delays (i.e., sampled from two distributions, one with mean = 0.5s, the other mean = 2s). If participants persisted in attributing lower causal rating in this condition (as compared with the inactive hastener condition), then this would support

the claim that people were unaware of the predictive effect of the hastener (with regard to mean time delay), and lowered their causal ratings due to the increased variability in the cause-effect delays.

### Associative Explanations

What do the results of the current experiments tell us about the processes underlying causal judgment? As noted in the Introduction, associative theories can explain the sensitivity of people's judgments to intervening events in terms of cue competition between these events and the putative cause [25]. The greater the number or probability of intervening events, the more competition there will be to predict and thus explain the target outcome. This means that the degree of association between putative cause and target effect is likely to be reduced when the temporal delay between this pair is longer, purely due to the increased probability of intervening events.

While this finding thus admits of an associative account, it is also open to a cognitive inferential explanation. Thus, from a rational perspective, the increase in the probability of intervening events, and hence in alternative potential causes of the target effect, is bound to increase the learner's uncertainty about which event is the true cause. Indeed one might construe the associative process as an implementation of this rational-level inference [27]. Alternatively, one might seek a different kind of psychological mechanism to underpin the rational-level inference [28]. The current experimental results cannot rule on these issues.

Providing an associative explanation for the effect of delay variability is less clearcut. One approach would be to argue that the effect is due to the fact that a learner's computation of contingency is itself dependent on assumptions about temporal delays [22]. In particular, it could be argued that although Experiment 1 sought to keep the cause-effect contingency constant whilst independently manipulating delay variability, it is possible that participants computed different contingencies in the two delay variability conditions<sup>4</sup>.

To illustrate, consider the typical two-by-two contingency table that is assumed to underlie people's computations of contingency (see Table 2). The judged strength of a causal relation between cue and outcome is supposed to depend on the contingency  $\Delta P = P(\text{outcome}|\text{cue}) - P(\text{outcome}|\text{no cue}) = A/(A+B) - C/(C+D)$ . Thus, events counted in cells A and D increase the judged strength, and events in cells B and C decrease the judged strength. Typical causal learning experiments are trial-based, so it is relatively clear to the learner what constitutes an instance of each cell type. In real-time learning experiments, such as those conducted in this paper, there are no pre-designated trials for learners to compute contingency, and therefore any such computations are dependent on the time frame imposed by the learner.

For example, if participants use a fixed temporal window (starting at the occurrence of a putative cause) to categorize event types, and thus compute contingency, there are condi-

<sup>4</sup>We thank an anonymous reviewer for this point.

tions under which a low variability delay distribution will yield more event pairs that are categorised as instances of cell A, whereas a high variability distribution (with equivalent mean) will yield more event pairs that are counted as instances of cell B or C. This would lead to reduced causal judgments in the high variability context. This effect depends on the precise relation between the temporal window set by the learner and the mean (and shape) of the time delay distributions. Applied to Experiment 1, if participants set time windows that are greater than the mean delay, but not too much longer (i.e., <5sec in the 2 sec mean delay condition) then they would compute a lower contingency in the high variability condition.

**Table 2. Typical 2x2 Contingency Table in Causal Learning Analysis**

	Outcome Present	Outcome Absent
Cue present	A	B
Cue absent	C	D

Note. Contingency  $\Delta P = P(\text{outcome}|\text{cue}) - P(\text{outcome}|\text{no cue}) = A/(A+B) - C/(C+D)$ .

However, there are two reasons to suspect that this approach does not provide a satisfactory explanation for all the findings in Experiment 1. First, the temporal delays between events, even in the high variability condition, were still quite short (e.g., either mean = .5 or mean = 2 sec, SD = 1 sec) so it is less likely that learners imposed a threshold that was sufficiently short to lead to any substantial changes in contingency computations. Second, this approach cannot explain the effect of variability due to the switching between two distinct distributions (mean = .5 or mean = 2 sec) in the active hastener conditions. Whatever the threshold time, the probability that the effect occurs within the resulting time window will be at least as high in the active hastener condition as in the inactive hastener condition. Thus, if contingencies were computed by counting occurrences of the effect within a set time window, the contingency would be at least as high in the active hastener condition as in the inactive hastener condition. So the finding that the cause was rated lower in the active hastener condition cannot be explained from a pure contingency account.

## CONCLUSIONS

We have presented two novel empirical findings with respect to people's sensitivity to the variability in delay distributions, and the impact of intervening events in the interval between causes and effects. It is not clear whether both these findings can be accommodated within current associative theories of causal learning, or whether they fit better with alternative accounts that incorporate temporal properties more directly (e.g., the temporal coding hypothesis [29]; rate-based theory [30]). Whatever the process-level account, however, we would argue that people's sensitivity to temporal distributions of events and delays makes good sense from an inferential perspective. Given that the available evidence about causal systems is often noisy, incomplete and ambiguous, it pays to exploit whatever information there is. In a

dynamic world this will often consist in information about temporal distributions.

## REFERENCES

- [1] Cheng PW. From covariation to causation: a causal power theory. *Psychol Rev* 1997; 104: 367-405.
- [2] Dickinson A, Shanks DR, Evenden JL. Judgement of act-outcome contingency: The role of selective attribution. *Q J Exp Psychol* 1984; 36A: 29-50.
- [3] Shanks DR, Dickinson A. Associative accounts of causality judgment. In: GH Bower, Ed. *The psychology of learning and motivation: advances in research and theory* San Diego, CA: Academic Press 1987; Vol. 21: pp. 229-61.
- [4] Shanks DR. Judging covariation and causation. In: Koehler D, Harvey N, Eds. *Blackwell handbook of judgment and decision making*. Oxford, England: Blackwell 2004.
- [5] Gopnik A, Glymour C, Sobel DM, Schulz LE, Kushnir T, Danks D. A theory of causal learning in children: causal maps and Bayes nets. *Psychol Rev* 2004; 111: 1-31.
- [6] Lagnado DA, Sloman SA. The advantage of timely intervention. *J Exp Psychol Learn Mem Cogn* 2004; 30: 856-76.
- [7] Steyvers M, Tenenbaum JB, Wagenmakers EJ, Blum B. Inferring causal networks from observations and interventions. *Cogn Sci* 2003; 27: 453- 89.
- [8] Fernbach PM, Sloman SA. Causal learning with local computations. *J Exp Psychol Learn Mem Cogn* 2009.
- [9] Lagnado DA, Sloman SA. Time as a guide to cause. *J Exp Psychol Learn Mem Cogn* 2006; 32: 451-60.
- [10] White PA. Naive analysis of food web dynamics: a study of causal judgment about complex physical systems. *Cogn Sci* 2000; 24: 605-50.
- [11] Waldmann MR. Knowledge-based causal induction. In: Shanks DR, Holyoak KJ, Medin DL, Eds. *The psychology of learning and motivation* San Diego, CA: Academic Press 1996; Vol. 34: pp. 47-88.
- [12] Waldmann MR, Hagmayer Y, Blaisdell AP. Beyond the information given: Causal models in learning and reasoning. *Curr Dir Psychol Sci* 2006; 15(6): 307-11.
- [13] Einhorn HJ, Hogarth RM. Judging probable cause. *Psychol Bull* 1986; 99, 3-19.
- [14] Lagnado DA, Waldmann MR, Hagmayer Y, Sloman SA. Beyond covariation: Cues to causal structure. In: Gopnik A, Schultz L, Eds. *Causal learning: Psychology, philosophy, and computation*. New York: Oxford University Press 2007; pp. 154-72.
- [15] Pearl J. *Causality*. Cambridge, England: Cambridge University Press 2000.
- [16] Sloman SA. *Causal models; how people think about the world and its alternatives*. New York: Oxford University Press 2005.
- [17] Hume D. *A treatise of human nature*. 2<sup>nd</sup> ed. Selby-Bigge LA, Ed, revised by P. H. Niddich, Oxford: Clarendon Press (1739/1978).
- [18] Shanks DR, Pearson SM, Dickinson A. Temporal contiguity and the judgement of causality by human subjects. *Q J Exp Psychol* 1989; Section B, 41: 139-59.
- [19] Wasserman EA, Neunaber DJ. College students' responding to and rating of contingency relations: the role of temporal contiguity. *J Exp Anal Behav* 1986; 46: 15-35.
- [20] Buehner MJ, May J. Knowledge mediates the timeframe of covariation assessment in human causal induction. *Think Reason* 2002; 8: 269-95.
- [21] Buehner MJ, May J. Rethinking temporal contiguity and the judgment of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Q J Exp Psychol* 2003; 56A: 865-90.
- [22] Vallee-Tourangeau F, Murphy RA, Baker AG. Contiguity and the outcome density bias in action-outcome contingency judgements. *Q J Exp Psychol* 2005; 58B: 177-92.
- [23] Wasserman EA, Chatlosh DL, Neunaber DJ. Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learn Motiv* 1983; 14: 406-32.
- [24] Bennett J. Event causation: the counterfactual analysis. *Perspect Sci* 1987; 1: 367-86.
- [25] Dickinson A. *Contemporary animal learning theories*. Cambridge: Cambridge University Press 1980.

- [26] Griffiths TL, Tenenbaum JB. Structure and strength in causal induction. *Cogn Psychol* 2005; 51(4): 334-84.
- [27] Lagnado DA. A causal framework for learning and reasoning. *Behav Brain Sci* 2009; 32(2): 211-212.
- [28] Mitchell CJ, De Houwer J, Lovibond PF. The propositional nature of human associative learning. *Behav Brain Sci* 2009; 32: 183-98.
- [29] Acrediano F, Miller RR. Some constraints for models of timing: A temporal coding hypothesis perspective. *Learn Motiv* 2002; 33: 105-23.
- [30] Gallistel CR, Gibbon J. Time, rate, and conditioning. *Psychol Rev* 2000; 107: 289-344.

---

Received: October 01, 2009

Revised: December 16, 2009

Accepted: January 07, 2010

© Lagnado and Speekenbrink; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.