Causal Induction from Continuous Event Streams: Evidence for Delay-Induced Attribution Shifts

Marc J. Buehner[1] and Jon May[2]

Abstract

Contemporary theories of Human Causal Induction assume that causal knowledge is inferred from observable contingencies. While this assumption is well supported by empirical results, it fails to consider an important problem-solving aspect of causal induction in real time: In the absence of well structured learning trials, it is not clear whether the effect of interest occurred because of the cause under investigation, or on its own accord. Attributing the effect to either the cause of interest or alternative background causes is an important precursor to induction. We present a new paradigm based on the presentation of continuous event streams, and use it to test the Attribution-Shift Hypothesis (Shanks & Dickinson, 1987), according to which temporal delays sever the attributional link between cause and effect. Delays generally impaired attribution to the candidate, and increased attribution to the constant background of alternative causes. In line with earlier research (Buehner & May, 2002, 2003, 2004) prior knowledge and experience mediated this effect. Pre-exposure to a causally ineffective background context was found to facilitate the discovery of delayed causal relationships by reducing the tendency for attributional shifts to occur. However, longer exposure to a delayed causal relationship did not improve discovery. This complex pattern of results is problematic for associative learning theories, but supports the Attribution-Shift Hypothesis.

Keywords

causality, reasoning, time perception, continuity, induction, associative learning

[1]Cardiff University; [2]University of Plymouth

Causal Induction from Continuous Event Streams

How do humans and other intelligent systems learn that one thing causes another? The approach most commonly endorsed in cognitive science is that unobservable causal relations may be inferred from observable evidence, in the form of contingencies (Allan & Jenkins, 1980; Cheng, 1997; Dickinson, 2001; Shanks & Dickinson, 1987; Pearl, 2000; cf. Hume, 1739, 1888; Rescorla & Wagner, 1972). In experimental psychology this assumption is reflected in standard causal learning paradigms, which employ discrete learning trials or summary information, explicitly indicating whether or not a cause and effect have co-occurred on a particular occasion. Commonly, evidence pertaining to causal inference is classified according to contingency tables. Table 1 displays the simplest and most often used kind, one referring to a binary causal relation involving only one candidate cause *c* and one effect *e*, both of which only have two possible levels (present or absent). Entries in *cell A* include all trials where *c* and *e* occurred together, *cell B* refers to trials where *c* occurred, but *e* failed to occur, *cell C* refers to occasions where *e* occurred in the absence of *c*, and *cell D* contains occasions where neither *c* nor *e* were present. Different theoretical approaches to causal induction, ranging from associative learning and Pavlovian conditioning, to social psychological inference rules, from statistical and Bayesian decision models to computational theories of causal power, offer various solutions as to how such evidence could be interpreted with respect to the existence and strength of a causal relation between *c* and *e* (for an overview see Buehner & Cheng, 2005).

Table 1. A standard 2 x 2 contingency table.

		Effect *e*	
		present	absent
Candidate Cause *c*	present	A	B
	absent	C	D

The theoretical advances generated from this rich body of research notwithstanding, it may be grounded in an oversimplification of the learning and reasoning processes underlying causal cognition: Some everyday causes have immediate consequences, while others do not reveal their effects until later. In many cases the temporal structure of the causal relation is not immediately evident. Moreover, most causal induction tasks hardly ever present themselves as mere evidence-evaluation tasks. The world around us is a

continuous flux of events and is not carved up into neat learning trials. If the assumption that causal inference proceeds by evaluating covariational data is correct, then the ability to decide whether a particular event constitutes an instance of cell *A*, *B*, *C*, or *D* is an important precursor of causal inference. Before one can evaluate covariational data, one first has to obtain this data. Traditional experiments bypass this problem by presenting pre-processed covariational information in the form of discrete learning trials or summary tables. In this paper we present an experimental design which circumvents artifacts arising from conventional paradigms, and use it to test the Attribution-Shift Hypothesis originally suggested by Shanks & Dickinson (1989). Before we lay out a detailed specification of this hypothesis and its predictions, it is necessary to first review relevant evidence and theories pertaining to the role of temporal spacing in causal induction in general.

The Detrimental Effect of Delay on Causal Learning

The evidence most relevant to the role of temporal spacing in causal induction comes from studies inspired by animal learning. Using free-operant procedures, Shanks and colleagues (Shanks & Dickinson, 1987; Shanks, Pearson, & Dickinson, 1989) have studied the effect of response-outcome delays on human causal judgments. Participants had to judge how strongly pressing a key caused a triangle to light up on a computer screen; a high contingency (.75) was no longer detected if the delay exceeded two seconds. This finding has subsequently been replicated many times (Reed, 1992, 1999, 1996; Buehner & May, 2002, 2003, 2004). The explanation for the detrimental effect of cause-effect delays proffered by Shanks and Dickinson is derived from the principles of associative learning: "the size of the increment in associative strength accruing from a pairing decreases as the contiguity is degraded" (Shanks & Dickinson, 1987 p. 231). Indeed, associationism often cites David Hume as one of its main philosophical influences (e.g., Dickinson, 2001), and Hume himself (1739, 1888) noted that regular succession (i.e., contingency) and contiguity are the main cornerstones of causal inference.

Einhorn and Hogarth (1986) agreed with this position, but argued that in situations where cause-effect contiguity is low, (mechanical) knowledge of how a cause may bring about a delayed effect should bridge temporal gaps. They cited many anecdotal examples of how humans readily infer delayed causal relations. However, in the absence of such knowledge, delayed cause-effect pairings might go unnoticed, or serve as poor evidence for a causal relationship. A corollary of Einhorn and Hogarth's knowledge mediation hypothesis is that successful causal inference requires a good match between the reasoner's assumptions about the timeframe of the causal relation in question, and the time actually elapsing between cause and effect. If, for example, one expects the cause to produce its effect immediately, delayed pairings will not be seen as causal (Buehner & May, 2002; Buehner & McGregor, 2006).

Buehner and May's (2002) analysis of Einhorn and Hogarth's (1986) argument

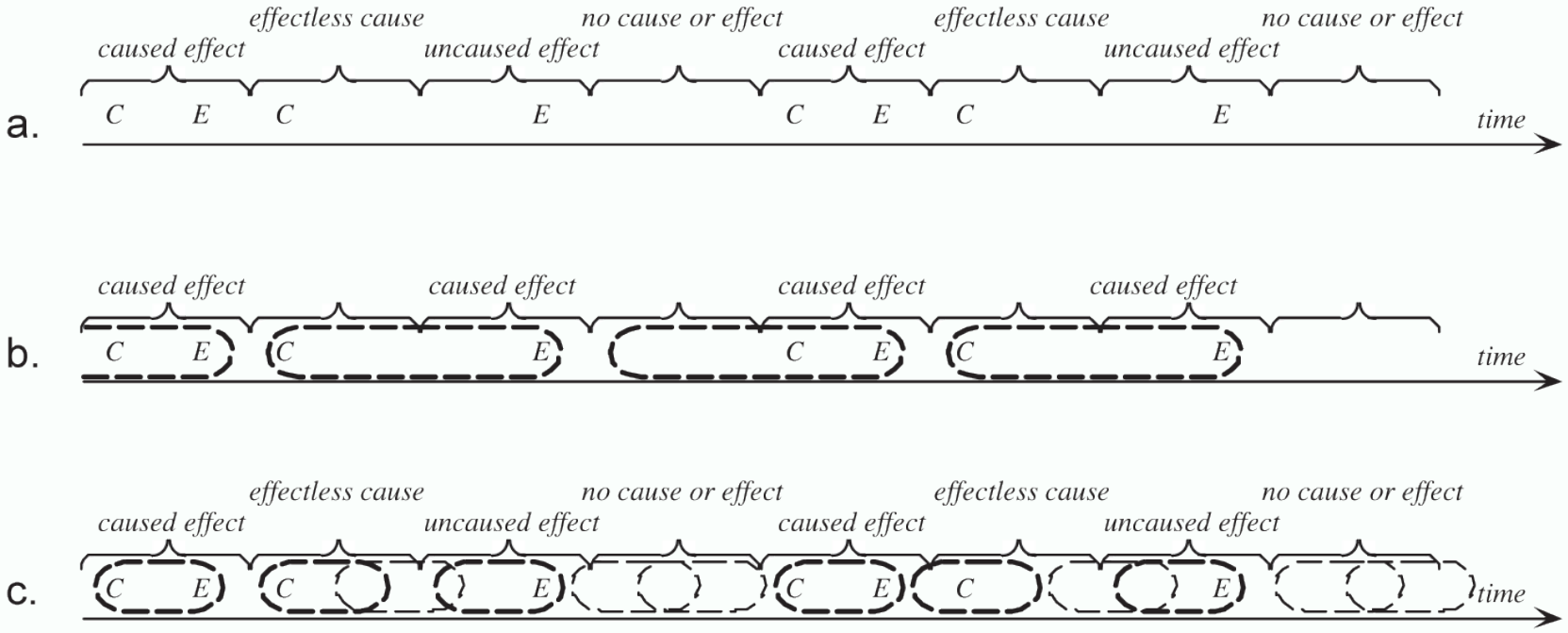
showed that there are two competing explanations for why causal estimates in studies such as Shanks et al.'s (1989) decreased as a function of the cause-effect delay. Shanks et al. noted that "subjects in judgment studies such as ours assume that the word 'causes' in the experimental instructions means 'causes immediately.' After all, they presumably have considerable experience of the immediacy of cause-effect relations in such electrical devices as computers" (1989, p. 155). Consequently, low causal ratings in delayed conditions could reflect a mismatch between participants' expectations and the evidence they perceived. In contrast, according to associationism, reductions in contiguity invariably bring about reductions in associative strength and thus decreased impressions of causal strength. A mismatch between expectations and experience falls outside the scope of simple associationism (see Miller & Barnet, 1993; and Savastano & Miller, 1998, for more refined associative models). Recent experimental work (Buehner & May, 2002, 2003, 2004) has shown that Shanks et al.'s finding might indeed have reflected a mismatch between participants' assumptions about the timeframe of the causal relation in question, and the available evidence. If participants in Buehner and May's studies were aware that the relation might imply a delay, action-outcome gaps of up to four seconds were no longer detrimental to causal learning.

The Attribution-Shift Hypothesis

Associative learning theory provides a *functional* analysis of the role of delay in causal learning: The longer the temporal gap between cause and effect, the weaker the increment in associative strength accruing from each pairing. The knowledge-mediation hypothesis offers a *rational*, *pragmatic* explanation of why successful causal inference is (sometimes) possible, even when the relation involves a delay. It is incomplete, however, as it does not provide a *functional* analysis of the detrimental effect of delay in the absence of mediating knowledge, or the lack of such effects when knowledge is present. Since knowledge-mediation falls outside the scope of associative learning theories, an alternative functional account is needed to flesh out the knowledge-mediation hypothesis.

One cognitive interpretation of the functional role of delay in causal induction is to propose that people adopt individual temporal thresholds, or windows, to decide when to attribute an effect to a preceding candidate cause, and when to attribute it to alternative causes. In other words, causes that produce delayed effects would not get credited with this evidence; instead, the evidence would be interpreted to indicate that the effect occurred on its own, or was produced by (unobserved) causes other than the one in question. This is illustrated by Figure 1, where a sequence of all four possible combinations of a cause C and an effect E (shown in 1a) can be interpreted as either evidence for four instances of completely contingent C-E combinations when a long temporal window is adopted (shown in 1b) or as evidence for a non-contingent sequence when a short temporal window is adopted (shown in 1c). In Figure 1c, an effect following the cause after a short delay falls

Figure 1. Parsing continuous event streams. The upper row (a) shows two instances of each class of cause-effect pairing. Despite the underlying trial structure, every event follows a cause, albeit at varying intervals. An observer could parse the event stream as shown by row (b), where a hypothetical constant temporal window has been superimposed at each event. In the lower row (c), a shorter temporal window provides the subjective impression of a cause-effect pairing as implemented by the trial structure, but there could be many more instances of windows containing neither cause nor effect, illustrated by the faint windows.



within the temporal window, and so is attributed to the cause, but effects following the cause after a longer delay fall outside the window, and so are not attributed to the cause, but to the constant background instead (and, as a corollary, the cause is seen as not having any effect). There are also many periods when the temporal window encompasses neither cause nor effect (the light dashed ovals in Figure 1c), in fact, more than represented by the structure shown in Figure 1a.

Expressed in terms of conditional probabilities, cause-effect delays might thus produce a shift in attribution away from P(e|c) toward P(e|¬c). Given that the cause-effect contingency $\Delta P = P(e|c) - P(e|\neg c)$ is a fundamental component in all current computational models of causal induction, such an attribution shift could readily explain why delays generally reduce causal ratings: temporal lags between cause and effect would effectively reduce the perceived contingency.

It is worthwhile at this point to briefly expand on the relation between conditional probabilities and causality in probabilistic accounts, and how these relate to associative learning theories. According to probabilistic accounts of causality, the reasoner has separate representations of the two conditional probabilities P(e|c) and P(e|¬c) and uses these to compute a third representation, namely that of causality. In associative terms, the equivalent to P(e|c) would be the association between the cause/background compound and the effect, and to P(e|¬c) it would be the association between the background alone and the effect; the representation of causality would be the (theoretical) associative strength of the cause alone, independent of the background. A hallmark feature of computational causal power accounts such as Cheng's (1997) is that they can represent causal strength as an unbound variable, that is, they can generalize outside of the training context (i.e., background), and make valid predictions in a novel context with different background probabilities, a feature which most associative models lack (see also Marcus, 1998)

In a simple associative account, the functional role of delays in causal learning is remarkably different to that of the Attribution Shift Hypothesis. Whereas the latter assumes that delayed pairings have the dual function of decreasing the subjective evidence for causal agency (with respect to the candidate cause), and at the same time increasing the subjective evidence of independent occurrence (due to unobservable causes), the latter argues that delayed pairings still accrue evidence for causal agency, but to a lesser extent than immediate pairings. More sophisticated associative models could employ "windows of associability," which in effect would allow them to simulate attribution shift-like behavior, or indeed have temporal gradients of associability so that delayed effects accrue strength for the background either if the delay exceeds a threshold, or according to a gradient discounting function.

The hypothesis that temporal contiguity has a direct influence on perceived contingency was first put forward by Shanks and Dickinson (1987) who also provided a preliminary test, again using an instrumental paradigm. As in their other studies, participants had

to evaluate how strongly pressing a key made a triangle flash on the computer screen. The novelty in this experiment, however, was that participants had to simultaneously judge the effectiveness of two keys in one experimental session, with the instruction to alternate pressing one and the other. In a control condition both keys A1 and A2 were subjected to an identical reinforcement schedule where P(e|c) was set to .75 and P(e|¬c) was .25 (how P(e|¬c) was operationalized is not specified, but presumably it was with reference to 1s periods), and presses on both keys, if reinforced, produced the effect immediately. The crucial experimental condition employed the same probabilities, but only one key, A3, produced the effect (O) immediately, while the other, A4, produced it after a 4 sec delay. As expected, participants rated A1 and A2 as very close to the actual ΔP, and provided considerably lower ratings for A4; A3 received the highest ratings, and ratings exceeded the actual ΔP (the report only contains descriptive analyses of the data). Shanks and Dickinson argued that:

> *If the effect of contiguity is mediated by a change in the perceived contingency, we should expect to have observed a decrement in the judgments not only for A4 but also for A3. According to this account, delaying the outcome for A4 does not just decrease the perceived P(O|A4) but also correspondingly increases P(O|¬A4). As these delayed outcomes were unlikely to have occurred in close association with A3, they should also have served to increment P(O|¬A3), thus reducing the perceived contingency for A3 as well as A4. (p. 235)*

Unfortunately, Shanks and Dickinson (1987) did not provide data about the distribution of responses and outcomes across the 240 sec sampling period, so their interpretation at best remains speculative. The fact that A3 attracted ratings that were higher, both compared to the control conditions A1 and A2, and to the actual ΔP, however, suggests that they were mistaken. A more plausible interpretation of their data is that at least some proportion of the delayed outcomes produced by pressing A4 were attributed to A3. In addition to decreasing P(O|A4), the delay appears to have also increased P(O|A3), which would have resulted in causal impressions higher than the programmed ΔP.

The observational paradigm we present below is better suited to test the hypothesis that cause-effect delays decrease causal ratings because effects are attributed to (a background of) alternative causes rather than the candidate cause in question. It allows strict control of the evidence available to participants, thus overcoming the need for post-hoc interpretations and speculations about what they might have perceived. Also, a more direct test of the Attribution Shift Hypothesis would involve direct probes of the conditional probabilities P(e|c) and P(e|¬c).

Predictions

Simple Associative learning theory (AL) and the Attribution Shift Hypothesis (ASH) make

a number of contrasting predictions with respect to how delays influence what is learnt, though, as we shall see, supplementing AL with a window or gradient of associability allows it to make nearly identical predictions. We shall try, nonetheless, to tease the two theoretical positions apart empirically. According to ASH, any delay-induced decrease in causal learning should be accompanied by decreases in subjective impressions of P(e|c), and corresponding increases in P(e|¬c). According to AL, causal induction does not involve computation and subsequent interpretation of conditional probabilities. This is not to say that AL cannot represent probability learning. According to Lagnado & Shanks (2002), for example, associative representations of causal strength can serve as direct substitutes for probability estimates, even though logically the two are not equivalent. While this could explain a delay-induced decrease of subjective P(e|c)—since causal ratings are lower, probability estimates would also be lower, it is not clear how a corresponding increase in P(e|¬c) would come about, unless, as noted above, one postulates a window or gradient of associability.

An extension from ASH suggests that sufficiently large shifts from P(e|c) to P(e|¬c) should produce negative subjective contingencies (even when the implemented contingency is positive). According to ASH, this would result in the impression that the cause prevents an otherwise occurring effect. This prediction does not readily follow from AL, which merely predicts weaker or no causal attribution in the case of delays. Again, however a window or gradient of associability would allow it to predict that pairings involving delays that exceed the threshold for associability would then accrue associative strength for the background. If the background context has accrued more strength than the candidate cause, the cause would likewise be interpreted to prevent the effect.

If ASH is correct in asserting that the function of delays is to shift attribution from *c* to alternative causes *a*, then any measures taken to reduce the credibility of *a* as a cause of *e* should reduce the detrimental effect of delay. One way how this could be achieved is to demonstrate that *a* fails to produce *e*. Once participants appreciate that *a* cannot be a cause of *e*, they might be less inclined to shift attribution away from *c* to *a*. An experimental implementation of this manipulation would involve Pre-exposure to the Unreinforced Background, a well established conditioning paradigm (e.g., Dickinson, Watt, & Varga, 1996). According to AL, such pre-exposure would likewise improve learning of delayed relations, through reducing the associability of the background context.

ASH and AL make dissociating predictions about what happens over the course of learning when the relation involves a delay. According to AL, delays reduce the size of the increment accruing from each pairing. This implies that evaluation of a delayed causal relation should improve over the course of training. More specifically, it should be possible to overcome, or at least alleviate the detrimental effect of delay by giving participants more experience with the relation. This prediction is unique to AL and does not follow at all from ASH: if delayed pairings serve as evidence for P(e|¬c), than experiencing more of these pairings will not have any beneficial effect.

General Method

In all three experiments, participants had to observe a continuous event stream in the form of a computerized representation of a tank driving across a military training range, and then had to evaluate whether sensor disks on the ground produced explosions when the tank rolled over them. The candidate cause thus was operationalized by the image of a tank rolling over the image of a disk, and the effect was an explosion happening on the tank. Depending on the experimental condition, explosions occurred either immediately, 2s or 5s after the tank rolled over the disk. In line with earlier studies on the effect of delay on human causal learning (Shanks et al., 1989; Buehner & May, 2002, 2003, 2004), instructions to participants explicitly mentioned that the effect sometimes might occur on its own, even though in fact this never happened during any of the experimental conditions. In our experiments, we did this by instructing participants that aircraft flying out of sight high above the training range may drop radar-guided bombs on the tank.

In all three experiments, participants viewed a continuous stream of events, which was not divided into clearly marked learning trials. Our strategy was to substitute a series of individual learning trials with short movie clips, where each clip can contain information about the presence or absence of the candidate cause and effect. We then spliced these movie clips together seamlessly to form one long movie. The end product viewed by participants thus did have an underlying trial structure, which was of course fully controlled by the experimenter, as illustrated in Figure 1a. However, participants were unaware of the underlying trial structure that generated the stream. As a consequence, the decision about whether or not a cause and effect formed a co-occurring pair constituted part of the inference process, as it must do outside the laboratory. Although from the experimenter's point of view, the underlying trial structure presented in Figure 1a constitutes no evidence for an association between cause and effect (all four types of trial occur twice, resulting in a contingency of 0), a participant could interpret this particular stream as showing that effects always follow causes, sometimes quickly but sometimes after a delay, and thus that there is a perfect association of cause and effect (Figure 1b).

Apparatus

All aspects of the experiment were administered via computer, programmed with Macromedia Director 7.0. The computer was programmed to deliver a continuously streaming film of a tank traversing a military training range. As in commercial animation, this appearance was created by keeping the tank stationary in the middle of the screen, but moving a looped landscape continuously from left to right behind the tank. The tank moved up and down over a range of ten pixels, but remained stationary with respect to the horizontal axis. On the bottom right corner of the tank was a light brown cloud of dust, again animated within a continuous loop.

Figure 2 displays filmstrips taken from the experiment. We encourage readers to view

example stimuli available from <http://www.cf.ac.uk/psych/home2/buehner/stimuli/> to get an impression of the continuous nature of the stimuli. Underlying the surface appearance of continuous motion was an underlying discrete trial structure. Each trial was built as follows. One second into each trial, an image of a disk placed on the ground entered the screen from the left. The software was programmed so that the disk would move smoothly along at the same speed as the landscape, thus creating the impression that the disk was stationary on the ground, and the tank was moving toward it. The disk moved behind the tank 1.67s after appearing at the left of the screen (subjectively, the appearance was that the tank rolled over the disk). The disk was no longer visible after meeting the tank. If the

Figure 2. A film-strip of screenshots from the Experiments. The tank remains in the centre of the screen, moving up and down slightly, while the landscape scrolls past from left to right (note the positions of the clouds and trees in successive frames). When a disk appears (frame 2), it moves at the same speed as the landscape. The impression is of a camera panning to follow a tank moving from right to left across stationary disks. After the disk has met the tank, an explosion can occur (frames 3 to 5), with a puff of smoke and flame also moving along with the background.

trial was one without an effect, the tank met the disk, and nothing happened. If the trial was of a type where an effect occurred, it did so either immediately on the tank's impact with the disk (0s), or after a delay of 2s or 5s. The effect consisted of a 1.5s long animation of an explosion, superimposed over, and partially occluding the tank. All trials lasted 10.17s, regardless of whether there was an explosion or not, and whether it occurred immediately or after a delay (i.e., trials ended 7.5s after the disappearance of the disk on no effect and immediate effect trials, 5.5s after the onset of an explosion on 2s delay trials, and 2.5s after the onset of an explosion on 5s delay trials). The length of trials (equal across all conditions of all experiments) of 10.17s thus consisted of 1s (disk appears) + 1.67s (disk disappears under tank) + 5s (maximum delay on effect present trials) + 1.5s (duration of explosion, if present) + 1s (inter-trial interval). Trials followed each other immediately and seamlessly, although, of course, the trial boundary was not apparent to the participant at all, and as far as they were concerned, the tank kept on traveling across the landscape.

Procedure and Cover Stories

For all three experiments, participants read instructions asking them to imagine that they were military officers sent on a training course to develop their observational skills in the context of anti-tank defense. They were told that they would observe a tank traversing

a military training range, and that aircraft flying high above and out of sight might drop radar-controlled bombs on the tank. It would not be possible to see the aircraft, nor any bombs; the only thing they could observe would be an explosion near the tank. Instructions further asked participants to concentrate on some special devices that had been placed on the range. In Experiment 1, participants were asked to evaluate whether the devices produce or prevent explosions near the tank, or whether they are unrelated to explosions. In Experiments 2 and 3, participants were asked only to evaluate whether the devices produce explosions. Participants were told to carefully observe the tank traversing the training range, and to decide whether and how strongly the devices produced (Experiments 2 and 3), or produced or prevented (Experiment 1) explosions. Full instructions are included in Appendix A.

Participants in all three experiments then viewed a demonstration to familiarize themselves with the experimental materials. The demo consisted of three trials: one "empty" trial in which the tank traversed the field for10.17s; one in which the disk was present, but no effect occurred; and one in which the effect occurred in the absence of the disk. Participants were of course unaware of this underlying trial structure; from their perspective they saw the tank roll across the landscape for 30.5s (3 x 10.17s), with a disk appearing after 11.67s, the tank rolling over this disk after 13.34s, and 9.67s after this, at 23s, an explosion occurring. Note that the missing trial from this practice sequence is the one in which a disk and an explosion were present: this was not used here because the delay between the two was varied in the main experiment, and so we did not want to create any expectations about its duration (the interval between the disk meeting the tank and the explosion in the demonstration was almost twice as long as the longest experimental delay). After viewing the demonstration, participants had the opportunity to ask clarifying questions, and then proceeded to the experimental movies. Each experimental movie consisted of a number of short clips determined by the condition of the relevant experiment (see specific Design sections), arranged in random order and spliced together seamlessly.

After each movie participants had to rate whether and how strongly the devices produced (Experiments 2 and 3), or produced or prevented (Experiment 1) explosions; Experiment 1 employed a scale from –100 to 100, where –100 meant that rolling over the devices prevented explosions every time, and 100 meant that rolling over them produced explosions every time; 0 meant that the devices had no relation to whether or not explosions happen. Experiments 2 and 3 used a scale from 0 to 100 with the two same labels as these points had in Experiment 1. Participants then had to answer two questions aimed at probing their subjective impression of the conditional probabilities P(e|c) and P(e|¬c): "Overall, how likely were explosions when the tank rolled over the devices?" and "Overall, how likely were explosions without the tank rolling over the devices?" Participants had to enter ratings on a scale from 0 to 100 where 0 meant that "explosions *never* happened

when a device was present/absent," 50 meant that "explosions *sometimes* happened when a device was present/absent," and 100 meant that "explosions *always* happened when a device was present/absent." Once participants provided all ratings for a condition, they immediately proceeded to the next experimental condition, that is, the next movie. Overall, each experiment lasted about 20-25 minutes.

On the definition of conditional probabilities in a continuous paradigm

Figure 1 illustrates that the definition and operationalization of conditional probabilities is non-trivial in a continuous paradigm. Without discretely marked learning trials, it is not self-evident whether a given effect can be attributed to a preceding candidate cause or to the background of alternative causes. From the experimenter's point of view, P(e|c) of course still has a definite value, since the continuous flow of events is determined by a probability schedule; for each occurrence of *c*, the schedule dictates the probability with which *e* will follow after the programmed delay. The participant's evaluation of the evidence depends crucially on the temporal window he or she is assuming to be operating for the causal relation in question. Whereas the hypothetically programmed structure in Figure 1a is intended to yield P(e|c) = .5, the subjective value of this conditional probability is .5 only if a short temporal window is assumed (Figure 1c); a long window (Figure 1b) yields a value of 1.0 for the same evidence (this is assuming that all effects occurring in the window are counted. Some temporal assumptions are of course more restrictive, in that immediate effects would not be attributed to the cause if a delayed mechanism is expected, cf. Einhorn & Hogarth, 1987).

The operationalization of P(e|¬c) on the other hand is difficult even from the experimenter's perspective. While the schedule clearly controls the delivery of *c* (and thus allows controlled delivery of instances of *e* associated with *c*, i.e., rendering P(e|c) defineable), no equivalent control exists for ¬*c* events. In the absence of discrete trials, every granular unit of time that is not taken up by *c* itself (or *c* plus the temporal window attached to it) constitutes ¬*c*. It follows that ¬*c* is thus not quantifiable and therefore also cannot serve as the basis for the conditional probability P(e|¬c). Some readers might be tempted to argue that operationalization of P(e|¬c) is straightforward due to the discrete trial structure that is underlying our paradigm. It would indeed be possible to include trials where *c* is absent, and *e* could be programmed to occur on these trials according to the same probabilistic rules that determined the programmed value of P(e|c). In fact, we even included one such trial in the demonstration phase that preceded each experiment (i.e., no disk, but an explosion).

However, a moment's reflection reveals that doing that would dramatically violate one of the fundamental assumptions underlying all contemporary models of causal inference from contingency information, namely that the composite of alternative causes *a* (other than candidate cause *c)* remains constant, irrespective of the presence of *c* (see Cheng,

1997). Assuming that *a* causes *e* to the same extent, both when *c* is present and when it is absent, allows one to use P(e|¬c) as a direct estimate of P(e|a). Knowing P(e|a) is essential for causal inference, in order to inform us how much of P(e|c) needs to be discounted as being caused by *a* (in the ΔP model this discounting is achieved simply by subtracting the estimate of P(e|a) from P(e|c). Cheng (1997) provides a more normative form of discounting). Keep in mind that in a continuous paradigm, every moment of time that is not occupied by *c* (plus its window) by definition is ¬*c*. Attempting to provide information about P(e|¬c) by including¬*c* trials thus is not only a wasted effort, as there are already plenty, in fact *countless* instances of ¬*c* even if there is not a single ¬*c* trial; including such trials would effectively provide two different values for P(e|a), one that is relevant for the ¬*c* trials, and one that applies for the trials during which *c* does occur (i.e., the periods of time during such trials where *c* plus the attached window is not present; this latter value would be 0). In other words, including ¬*c* trials would render the evidence uninterpretable.

In a continuous paradigm, the way to supply information about P(e|¬c) is to schedule instances of *e* according to a set (random) rate, and thus independent of and not related to *c,* throughout the experiment. One could for instance program the apparatus to deliver an explosion with a probability of .1 every 2s, irrespective of whether or not or when *c* has occurred. We would than have information about the rate with which *e* occurs in the absence of *c*, but it is not clear how this information could be combined with the conditional probability P(e|c) to yield causal estimates (cf. Gallistel, 2002). It is, however, possible to provide information that P(e|¬c) = 0. To do so, one simply would never have any occurrence of *e* that is not associated with *c*. Crucially, one would not even need to include "empty" ¬*c* trials (although one could without doing any harm), since every unit of ¬*c* time counts as evidence toward P(e|¬c). Throughout the experiment there are extended periods of time where *e* could happen due to alternative causes, but in fact does not (for instance 8.5s for each unreinforced trial, 7s for each trial with immediate reinforcement). There is no reason for participants to believe that on trials where *e* happens due to *c*, there could not be a second instance of *e*, due to *a*. Most of the studies investigating the influence of delay on causal learning have used P(e|¬c) = 0 (Shanks et al., 1989; Buehner & May, 2002, 2003, 2004), probably to avoid making the task too difficult for participants. We have chosen to keep with this tradition. In order to set P(e|¬c) = 0, we simply kept the background rate of *e* at zero. We did not include any "empty" ¬*c* trials.

Experiment 1

The first experiment tests whether our new paradigm replicates the basic findings established in research using free-operant paradigms. Shanks and Dickinson (1987) have shown that both cause-effect contingency and contiguity influence impressions of causal strength. Buehner and May (2002, 2003, 2004) demonstrated that the impact of cause-

effect contiguity is mediated by higher-level knowledge about the timeframe of the causal mechanism in question, so that delayed contingencies could be inferred if they were expected by participants, although any experience of an immediate contingency blocked this mediation. Experiment 1 employs variations in contingency, contiguity, and knowledge of mechanism to see whether the same complex pattern of results also emerges from a continuous paradigm. In order to directly test ASH, we collected subjective probability estimates in addition to causal ratings. Here the aim is to examine the relationship between participants' estimates of P(e|c) and P(e|¬c). To avoid ambiguity, in this and all remaining experiments the effect never occurs due to alternative causes other than *c*, so any non-zero estimates of P(e|¬c) reflect misattribution, rather than overestimation.

An interesting corollary of Shanks & Dickinson's (1987) attribution shift hypothesis is that (unexpected) cause-effect delays might render an objectively positive contingency subjectively negative. Consider a design where programmed P(e|c) = .5, and P(e|¬c) = 0. If cause-effect delays result in a full shift of attribution from P(e|c) toward P(e|¬c), subjective P(e|c) will be 0 and subjective P(e|¬c) will be .5, resulting in an overall subjective notion of ΔP to be negative. This effect does not rely on a complete attribution shift, but would be obtained as long as more than half the instances contributing toward P(e|c) get subjectively attributed to P(e|¬c). To test this corollary, we included the possibility of negative causal ratings, although all conditions employed positive contingencies.

Method

Participants

Twenty-three (16 female, 7 male; median age: 27 years) volunteers were recruited via the University of Sheffield volunteers' e-mail distribution list and received £4 for their efforts.

Apparatus, Procedure, and Cover Story

The apparatus and procedure were as described in the General Method section. Knowledge of Mechanism was manipulated between subjects with two different Cover Stories. Both sets of instructions provided the same general information as outlined in the General Method section. More specifically, they stated that participants had to observe whether devices placed on the training range harm or protect the tank when it rolls over them; alternatively, the devices could also have no function at all and thus be unrelated to explosions occurring near the tank. Instructions stated that the protective function of the devices would be to emit a jamming radio signal that interferes with the radar-tracking of the bombs that may be dropped by out-of-sight high-flying aircraft. The nature of the harmful function varied between cover stories. For the *Mine* group, instructions stated that the devices could be landmines that explode as soon as the tank rolls over them; for the

Sensor group, instructions stated that the devices could be sensors that triggered a radar-controlled missile to be launched from several miles away to pursue the tank and explode on it. Instructions in both groups stated that regardless of which function the devices had, they would always look the same. Participants were asked to carefully observe the tank and to decide whether and how strongly the devices produced or prevented explosions.

Design

A combination of the within-subject factors *Contingency*, with the levels .5 and .75, and *Delay*, with the levels 0s, 2s, and 5s produced six experimental conditions per participant. The factor *Cover Story*, with the levels "mine" and "sensor," was varied between participants. There were 16 learning trials per condition, and the candidate cause (disk) was present on all of them. Variations in *Contingency* were achieved by varying the total amount of trials with a scheduled explosion (i.e., 8 when *Contingency* was .5 and 12 when *Contingency* was .75). Variations in *Delay* were achieved by setting the cause-effect delay accordingly. The order of trials within each condition was randomized for each participant, and trials were presented seamlessly, effectively preventing participants from noticing the underlying trial structure. Based on earlier observations that prior experience of immediate or delayed causal relations in a within-subject design influences the parsing of subsequent delayed relations (Buehner & May, 2002, 2003), we thought it prudent to counterbalance whether participants experienced a delayed or immediate condition as their first condition. To this end, we created two counterbalanced *Order* conditions, 0s-first and 5s-first. Participants in the 0s-first group first saw the condition with *Contingency* = .5 and *Delay* = 0s, whereas participants in the 5s-first group first saw the condition with *Contingency* = .5 and *Delay* = 5s. The order of the remaining five conditions was randomized, so neither *Delay* nor *Contingency* were blocked.

Results and Discussion

Causal Ratings

An overall ANOVA (Significance level .05, with effect size η^2 computed as a ratio of SS-effect to SS-Total) on the causal ratings showed main effects of *Contingency* $F(1,19) = 4.41$, $\eta^2 = .03$, and *Delay* $F(2,38) = 7.48$, $\eta^2 = .09$, but also of *Order* $F(1,19) = 5.03$, $\eta^2 = .03$, and an interaction of these three factors $F(2,28) = 3.58$, $\eta^2 = .04$. There was also a marginal effect of *Cover Story* $F(1,19) = 4.20$, $\eta^2 = .03$, $p = 0.055$, and an interaction of *Order* and *Cover Story* $F(1,19) = 6.01$, $\eta^2 = .04$.

Overall, causal ratings were higher for the .75 than for the .50 contingency conditions, and ratings declined as the cause-effect delay increased. In three conditions, participants actually provided slightly *negative* ratings (of between -.09 and -.15), indicating that the disk was *preventing* explosions rather than causing them: all three of these were condi-

Figure 3. Deviations of mean causal ratings from observed contingencies in Experiment 1 broken down by Cover Story, Delay, and Delay Condition Order. Error bars indicate standard errors. Participants who read the Sensor cover story, and who then experienced the 5 second delay condition first, consistently gave near normative ratings of causal strength, whereas other participants only rated the zero second delay condition as normatively causal.

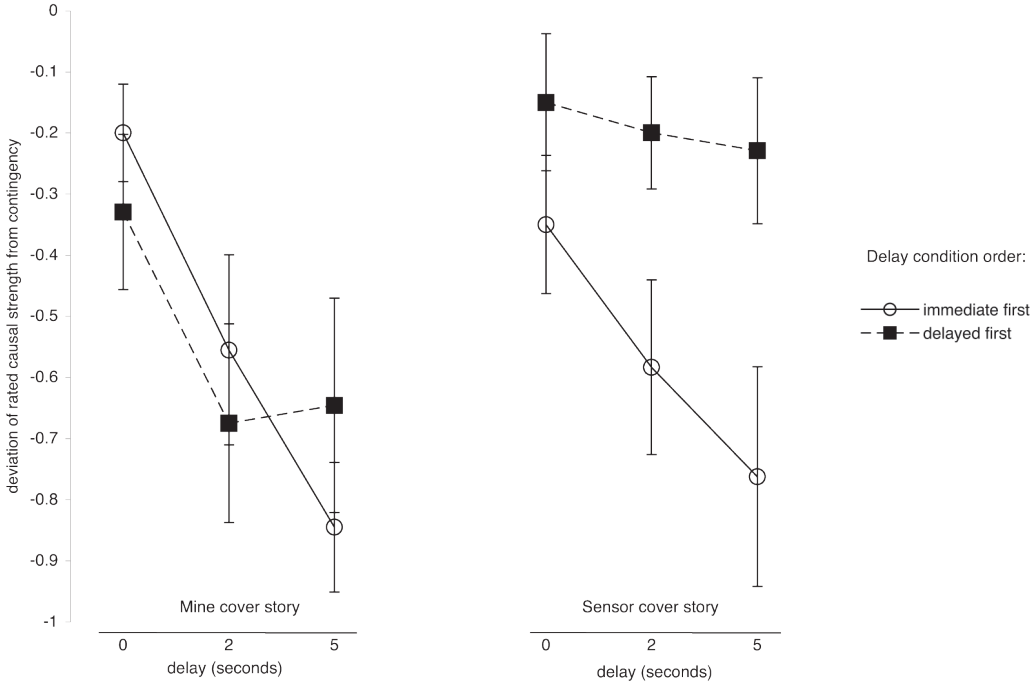
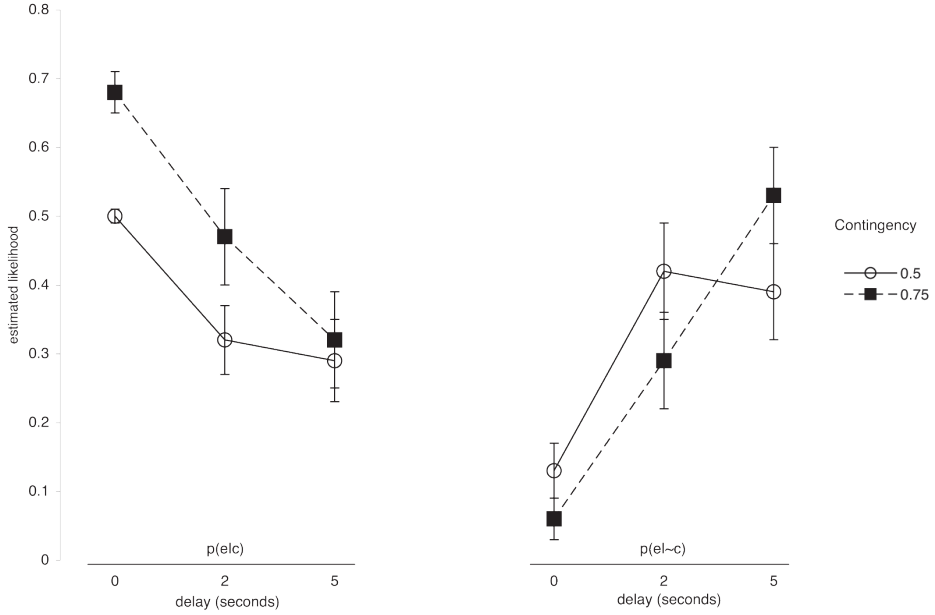


Figure 4. Mean estimates of conditional probabilities in Experiment 1 broken down by Delay and Contingency. Error bars indicate standard errors. The two estimates are inversely related, r = –0.66, and as delay increases, P(e|c) declines and P(e|¬c) correspondingly increases.



tions involving the immediate Mine cover story and a cause-effect delay (the contingency/delay combinations of 0.50/2s, 0.50/5s and 0.75/5s). Figure 3 illustrates the deviations of causal ratings from the experienced contingency (to allow the data to be collapsed over the two contingencies, which would produce different normative ratings). The exception to the trend of declining causal ratings with increasing delay is the group of participants who were instructed about the delayed nature of the causal mechanism and worked on a delayed condition first: they provided consistently high (and near normative) causal ratings for all conditions throughout the experiment and were largely unaffected by delays. Dividing the sample by *Order,* further ANOVAs showed only a main effect of *Delay* $F(2, 18) = 8.79$, $\eta^2 = .19$, for the participants who experienced the zero second delay first, but only a main effect of *Cover Story* $F(1,10) = 37.2$, $\eta^2 = .13$, for those who experienced the five second delay first.

Probability Estimates

Figure 4 displays participants' subjective estimates of the conditional probabilities P(e|c) and P(e|¬c). An overall ANOVA on estimates of P(e|c) revealed main effects of *Contingency*, $F(1,19) = 10.4$, $\eta^2 = .04$, *Delay*, $F(2,38) = 18.7$, $\eta^2 = .19$, and a *Contingency* x *Delay* x *Order* interaction, $F(2,38) = 6.94$, $\eta^2 = .05$. An ANOVA on estimates of P(e|¬c) revealed only a main effect of *Delay*, $F(2,38) = 19.6$, $\eta^2 = .22$, and a marginal interaction of *Contingency* x *Delay* $F(2,38) = 2.87$, $\eta^2 = .03$, $p = 0.069$. Estimates of the two conditional probabilities are clearly inversely related, and overall correlate $r = -0.66$. P(e|c) and the causal ratings are also closely related, $r = 0.61$, but P(e|¬c) and causal ratings are less strongly related, $r = -0.37$. The main effects of *Delay* on estimates of both P(e|c) and P(e|¬c) suggest that cause-effect delays did serve to decrease subjective impressions of contingency by reducing the perceived P(e|c) and at the same time increasing impressions of P(e|¬c), as originally hypothesized by Shanks and Dickinson (1987).

The absence of any effects associated with *Cover Story* stands in contrast to the results from the causal ratings. We would have expected the effects of *Delay* to be modulated by Cover Story. The causal ratings show that participants who were led to expect a delayed causal mechanism and who then experienced a delayed condition first did bridge the temporal gap between cause and effect and rated delayed conditions as highly causal; it is somewhat puzzling then why this same subgroup of participants, like all other participants, decreased their estimates of P(e|c) and increased their estimates of P(e|¬c) as the delay increased. One explanation could be that participants interpreted our question "Overall, how likely were explosions when the tank rolled over the devices?" as referring to the absolute temporal position of the explosions, rather than as to referring to whether or not explosions were causally associated with the tank rolling over the device. From such a perceptual perspective, it is of course rational to indicate that in the delayed conditions, explosions tended to happen when the tank was not rolling over the device, even if one

believed that the tank rolling over the device actually produced the later explosion. In retrospect, this question was ambiguously worded.

Summary

Overall, our observation-based paradigm replicated earlier findings: reductions of contingency and contiguity impair causal judgments, and extended these findings beyond free-operant methods. The data from Experiment 1 support the attribution-shift hypothesis in two very distinct ways. First, estimates of P(e|c) and P(e|¬c) decreased and increased, respectively, as the cause-effect delay increased. Second, causal ratings generally declined below zero with a 5s delay, suggesting that participants had perceived a negative contingency, i.e., P(e|c) < P(e|¬c), despite the fact that the programmed contingency in fact was always positive. As mentioned in the prediction section, simple AL models would not predict this pattern, but models supplied with temporal gradients or windows could.

Our manipulation of cover story significantly modulated the detrimental influence of delay on causal ratings, but only in participants who experienced a delayed condition at the beginning of the experiment. If instructions about a delayed causal mechanism were directly followed by evidence of immediate cause-effect pairings, in contrast, the instructions were disregarded and no longer had any beneficial effects throughout the experiment. Moreover, participants in the *Sensor* group, who expected a delayed relation, still rated the immediate 0s condition as highly causal. This result, taken together with our earlier findings about how knowledge mediates the timeframe of covariation assessment in causal learning from free-operant paradigms (Buehner & May, 2002, 2003, 2004) suggests that delay assumptions induced via cover stories may not be plausible enough to be maintained across an entire experiment when clashing with evidence of immediate relations, at least in a computer-based design (See Buehner & McGregor, 2006, for a successful implementation of delay assumptions with a realistic physical apparatus). The subsequent experiments thus avoid manipulations of cover story and focused exclusively on previous experience.

Experiment 2

The probability estimates from Experiment 1 suggest that one way in which delays could interfere with optimal causal attribution is that people fail to attribute effects to causes that happened some time ago. Rather, people might attribute delayed effects to the context or background of alternative causes, as originally suggested by Shanks and Dickinson (1987). Both in Shanks and Dickinson's original experiment and in our Experiment 1, there was ample rationale for participants to do so. After all, the instructions explicitly mentioned a constant unobservable alternative cause that might also produce the effect.

One corollary from this attribution shift hypothesis is that discrediting the ability

of alternative causes to produce the effect should improve assessment of delayed causal relations to some degree. If participants think that the effect does not happen on its own, they might be more willing to accept a delayed relation between candidate cause and effect, compared to situations where alternative causes are easy to conjure. Support for this hypothesis from the animal learning literature comes from Dickinson, Watt, & Varga (1996), who showed that giving rats prior exposure to an un-reinforced background context similar to the subsequent experimental situation improved detection of delayed contingencies. Rats in their Experiment 1 had to learn that pressing a lever delivered a food reward after a certain delay, and responded more on this delayed schedule, if, prior to the experimental sessions, they had spent time in the experimental chamber without the opportunity to press the lever. A critical element for the success of this pre-exposure was that no reinforcements were delivered during this period. According to Dickinson et al., pre-exposure to the un-reinforced background served to minimize context conditioning. While the non-contiguous reinforcement in the experimental sessions might have established an association between the constant background context and reinforcement, the un-reinforced pre-exposure periods effectively lowered such associations.

Expressed in cognitive terms, pre-exposure serves to point out that the effect does not happen on its own or in association with any attributes of the experimental situation other than the manipulated attribute. This should in turn increase the tendency to attribute delayed deliveries of the effect to an earlier occurrence of the cause (as opposed to a constant background of alternative causes). In Experiment 2, we tested the capacity of pre-exposure to improve human causal learning by adding a one-minute segment of the tank traversing the landscape, without rolling over any disks, and with no explosions occurring.

Method

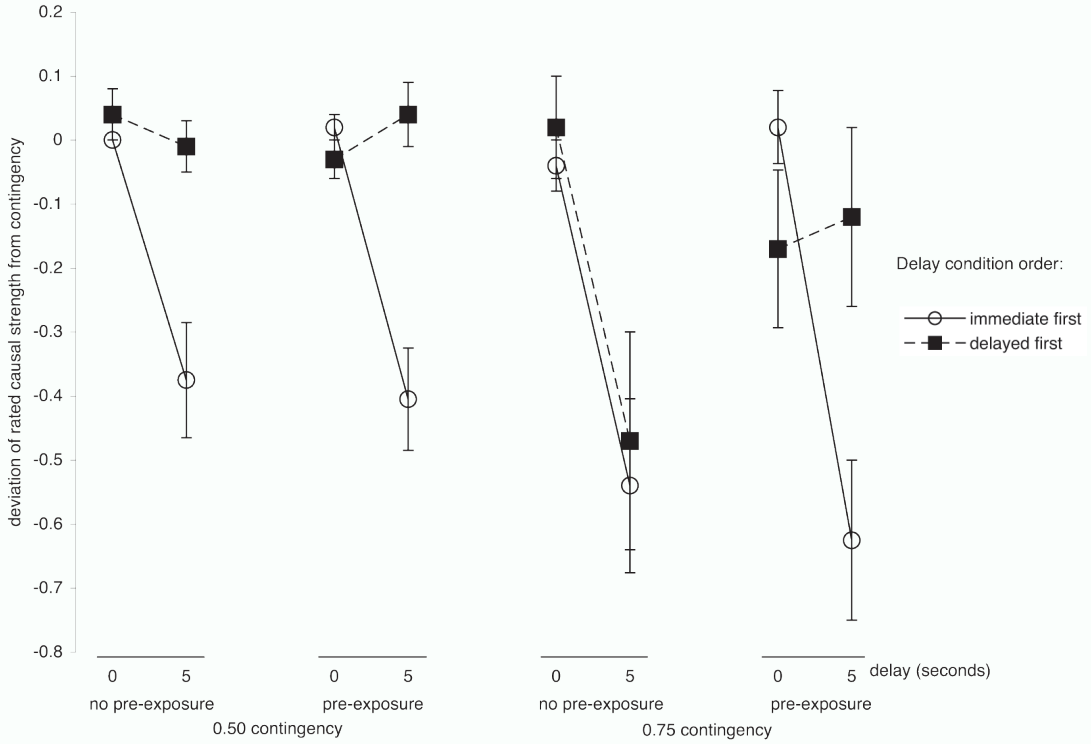
Participants

Twenty-three (19 female, 4 male) volunteers were recruited via the University of Sheffield volunteers e-mail distribution list and received £4 for their efforts.

Apparatus and Procedure

The apparatus and procedure were identical to Experiment 1, except that the instructions (common across all participants) did not discuss the nature of the sensors (preventive, neutral, or generative) or the timeframe associated with the sensors. Participants were still informed about the possibility of distant aircraft dropping smart bombs onto the tank. The instructions stated, however, that participants should concentrate on whether "special kinds of sensor" placed on the training range produced explosions when the tank rolled over them. One further change involved the nature of the rating scale, which now only

Figure 5. Deviations of mean causal rat[illegible]m observed contingencies from Experiment 2 broken down by Pre-Exposure, Contingency, Delay, and Delay Condition Order. Error bars indicate standard errors.



ranged from 0 to 100, thus no longer enabling participants to provide preventive ratings. See Appendix for full Instructions.

Design

The factors *Delay* (0s, 5s) and *Contingency* (.5, .75) varied within participants, *Pre-exposure* (yes, no) varied between participants. As in Experiment 1, we counterbalanced whether participants were exposed to a condition with 0s or 5s delay. To this end, the first condition always contained a .5 contingency, but whether it was paired with 0s or 5s delay was counterbalanced; the remaining three conditions were then presented in random order. As in Experiment 1, each condition comprised 16 learning "trials," and the number of "reinforced trials" depended on the contingency implemented in that condition. Participants in the Pre-exposure group experienced a one-minute segment of the tank traversing the landscape (no disks, no explosions) immediately before their first experimental condition. This pre-exposure phase was seamlessly integrated into the first condition, and was not mentioned to participants in any way. This condition thus lasted 222.72s, while the others lasted 162.72s.

Results and Discussion

Causal Ratings

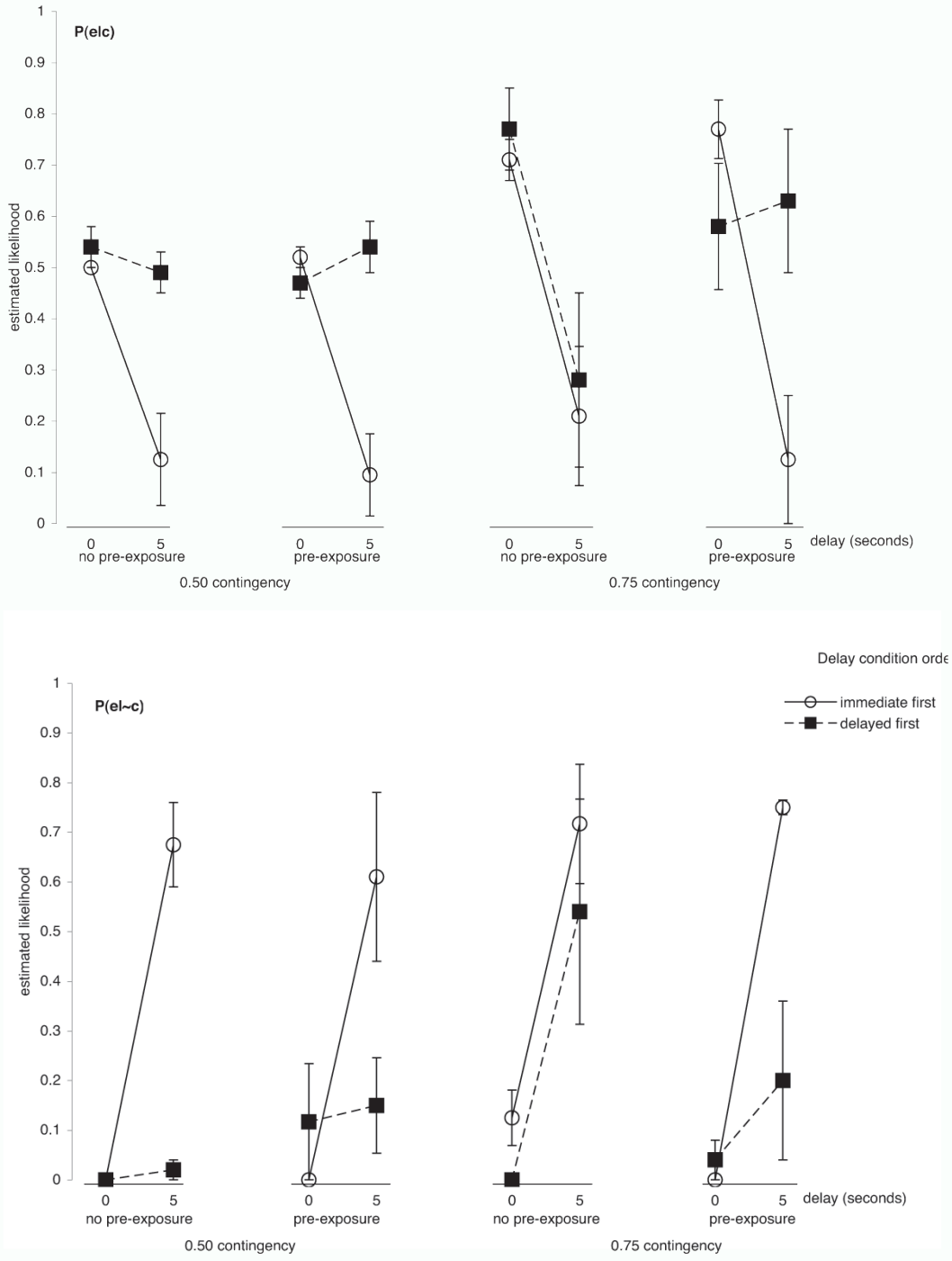
As in Experiment 1, an overall ANOVA indicated a main effect of the counterbalancing factor *Order* $F(1,19) = 4.95$, $\eta^2 = .04$, and several interactions. Dividing the sample according to *Order*, further ANOVAs indicated that for those who experienced an immediate relationship in the first condition, there was a main effect of *Delay* $F(1,10) = 44.9$, $\eta^2 = .56$, and an interaction of *Delay x Contingency* $F(1,10) = 26.6$, $\eta^2 = .06$; while for those who experienced a delayed condition first there was a marginal effect of *Delay* $F(1,9) = 3.66$, $\eta^2 = .07$, $p = 0.088$ and an interaction of *Delay x Pre-exposure* $F(1,9) = 5.14$, $\eta^2 = .10$. Figure 5 shows that participants who experienced an immediate condition first always rated the delayed conditions as less normative than the immediate conditions, and rated the immediate conditions near normatively, irrespective of whether they had been pre-exposed to a causally inactive background at the beginning of the experiment. In contrast, participants who saw a delayed condition as the first condition were susceptible to the pre-exposure manipulation: delayed-first participants who experienced 60s of pre-exposure to the background context without causes or effects rated the 5s delay just as highly as the 0s delay, and in both conditions were only a little below normative ratings. Delayed-first participants who did not receive pre-exposure, on the other hand, were negatively affected by cause-effect delays and rated contingencies paired with a 5s delay less normatively than the corresponding immediate (0s) contingencies.

An exception to this pattern is the rating of the .50 contingency/5s delay condition, which was rated normatively (a mean causal rating of .49) by participants who were in the delayed-first order and who received no pre-exposure. This pattern also emerged in the probability estimates (see Figure 6). Keep in mind that for these participants, this conditions would have been the very first condition they worked on. Buehner and May (2002, 2003) have shown that delayed contingencies are rated higher when they are presented at the beginning of an experiment, before participants have had experience of immediate contingencies, compared to when they are presented later in the experiment, after participants have experienced an immediate contingency. The same effect seems to have been at work in this experiment.

Probability Estimates

Figure 6 displays participants' mean probability estimates. The overall ANOVAs for the two probability estimates also indicated strong effects of *Order* (for $P(e|c)$, $F(1,19) = 12.4$, $\eta^2 = .07$; for $P(e|\neg c)$, $F(1,19) = 12.5$, $\eta^2 = .09$) and interactions, and so the sample was again split according to *Order* for further analysis. Those who had received an immediate condition first showed effects of *Delay* ($P(e|c)$, $F(1,10) = 67.4$, $\eta^2 = .59$; $P(e|\neg c)$, $F(1,10) = 52.5$, $\eta^2 = .57$) and *Contingency* ($P(e|c)$, $F(1,10) = 7.06$, $\eta^2 = .05$); those who had received a delayed

Figure 6. Mean estimates of conditional probabilities in Experiment 2 broken down by Pre-exposure, Contingency, Delay, and Delay Condition Order. Error bars indicate standard errors.



condition first showed an effect of *Delay* (P(e|¬c), $F(1,9) = 7.17$, $\eta^2 = .10$), an interaction of *Delay x Pre-exposure* (P(e|c), $F(1,9) = 5.47$, $\eta^2 = .11$), and a marginal interaction of *Delay x Contingency* (P(e|¬c), $F(1,9) = 4.57$, $\eta^2 = .07$, $p = 0.061$).

Overall, as in Experiment 1, P(e|c) and P(e|¬c) were inversely related $r = -0.79$ and P(e|c) and causal ratings were highly related $r = 0.89$; while P(e|¬c) and causal ratings were now also strongly inversely related $r = -0.77$. Again, cause-effect delays have, in general, simultaneously lowered participants' impressions of P(e|c) and increased their perceptions of P(e|¬c). The exception to this pattern are participants who experienced a delayed condition first and who had been pre-exposed to the background context, and those who experienced the .50 contigency/5s delay without pre-exposure. For these participants, impressions of P(e|c) did not decrease, and ratings of P(e|¬c) did not increase with cause-effect delays.

Summary

Pre-exposure to the un-reinforced background decreased participants' tendency to attribute delayed effects to a background of alternative causes. This effect of pre-exposure is apparent in its modulation of the impact of delay on both causal ratings and probability estimates. Similarly to Experiment 1, its influence interacted with the counterbalancing, such that it was only effective in the participants who worked on a delayed condition first. If the pre-exposure phase was immediately followed by a contiguous, no-delay condition, the advantage was lost. One interpretation of this counterbalancing effect is that experience of an immediate cause-effect contingency created such a strong impression of contiguous causality that subsequent delayed relations in contrast appeared non-causal. The knowledge gained from pre-exposure (i.e., that the background is incapable of producing the effect) was thus lost or overridden by subsequent experience of immediate relations. The effect of pre-exposure is commensurate with both AL and ASH.

Experiment 3

Associative accounts of human causal reasoning fundamentally challenge the notion that estimated causal strength is grounded on a computational process involving the conditional probabilities P(e|c) and P(e|¬c) (e.g., see Baker, Murphy, & Vallée-Tourangeau, 1996; Shanks, Lopez, Darby, & Dickinson, 1996). Instead, estimates of causal strength are updated every time the candidate cause is present. Importantly, such models "impose a low memory load on the organism because experience is stored as a small number of associative strengths. (…) [and] information about past events is lost in the computation. In other words, these models do not have episodic memory" (Baker et al., 1996, p. 1). Following Shanks and Dickinson's (1987) analysis, associative theories of causal induction should be able to account for the influence of cause-effect delays without recruiting an attributional shift from P(e|c) to P(e|¬c).

One way in which associative learning theory can explain detrimental effects of delay is by simply allowing that delayed effects produce weaker increments of associative strength on any given learning trial: "the size of the increment in associative strength accruing from a pairing decreases as the contiguity is degraded" (Shanks & Dickinson, 1987, p. 231). This assumption implies that detrimental effects of cause-effect delay can be overcome by prolonged training (Dickinson, 2001). A large number of weak increments or a small number of larger increments might produce similar judgments. If the effect of delay is limited to producing weaker increments on a trial-by-trial basis, one would even expect that judgments from delayed and contiguous schedules be identical at asymptote, reflecting the maximum strength the stimulus can support. However, most theorists would argue that delays have a detrimental effect overall, above and beyond weakening the trial-by-trial increments, resulting in a lower level of asymptote (Ralph Miller, personal communication April 2003). Be that as it may, an associative analysis of causal judgment predicts that estimates of delayed causal relations should improve to some extent, as the amount of learning trials increases.

A knowledge-mediation account of human causal induction, on the other hand, would not predict improvements as the amount of experience with the schedule increases. Because the pattern of conditional probabilities does not change as learning progresses, it would be difficult for this account to explain any improvements. Knowledge-based theories could account for improvements based on prolonged learning only by drawing on reasoning mechanisms external to the causal induction process itself. Prolonged exposure to the causal induction problem might, for instance, increase the chance of sudden insight into the delayed nature of the causal mechanism. If improvements were due to such processes, then causal ratings should improve as a step-function of learning trials, to reflect the sudden nature of the insight. Associative learning theory, in contrast, would predict gradual improvements, typically a negatively accelerated learning curve.

In Experiment 3 we tested whether prolonged experience with delayed causal relations directly improves causal ratings by manipulating the number of learning trials participants saw before they had to make a judgment. We departed from the common strategy, of studying acquisition functions by probing judgments at various points throughout one learning phase (e.g., Shanks, 1985, 1987), and instead manipulated the number of learning trials within participants. We chose this method for reasons of ecological validity. When asking participants for multiple judgments of the same relation across time, one implies that the relation changes over time (why else would one ask participants to update their beliefs?). We were only interested in how the overall assessment of delayed relations changed as a function of learning trials.

Method

Participants

Twenty-four (11 female, 13 male) volunteers were recruited via the University of Sheffield volunteers e-mail distribution list and received £4 for their efforts.

Design, Apparatus, and Procedure

The apparatus and general procedure were adapted from Experiment 2, with the pre-exposure factor removed from the design and an additional manipulation of the number of learning trials per condition. The design employed the factors *Delay* (0s, 5s), and *Length* (16, 24, 32 learning trials), varied within participants, and *Contingency* (.5, .75) varied between participants. The total length of each condition thus was 162.72s, 244.08s, and 325.44s for conditions comprising 16, 24, and 24 "trials," respectively. As in the previous experiments, we counterbalanced whether participants worked on a immediate or delayed contingency as their first condition. Toward this end, we let all participants begin with a condition comprising 16 learning trials, but counterbalanced whether it was associated with 0s or 5s delay (to create a between subjects factor *Order*). The remaining five conditions were presented in random order.

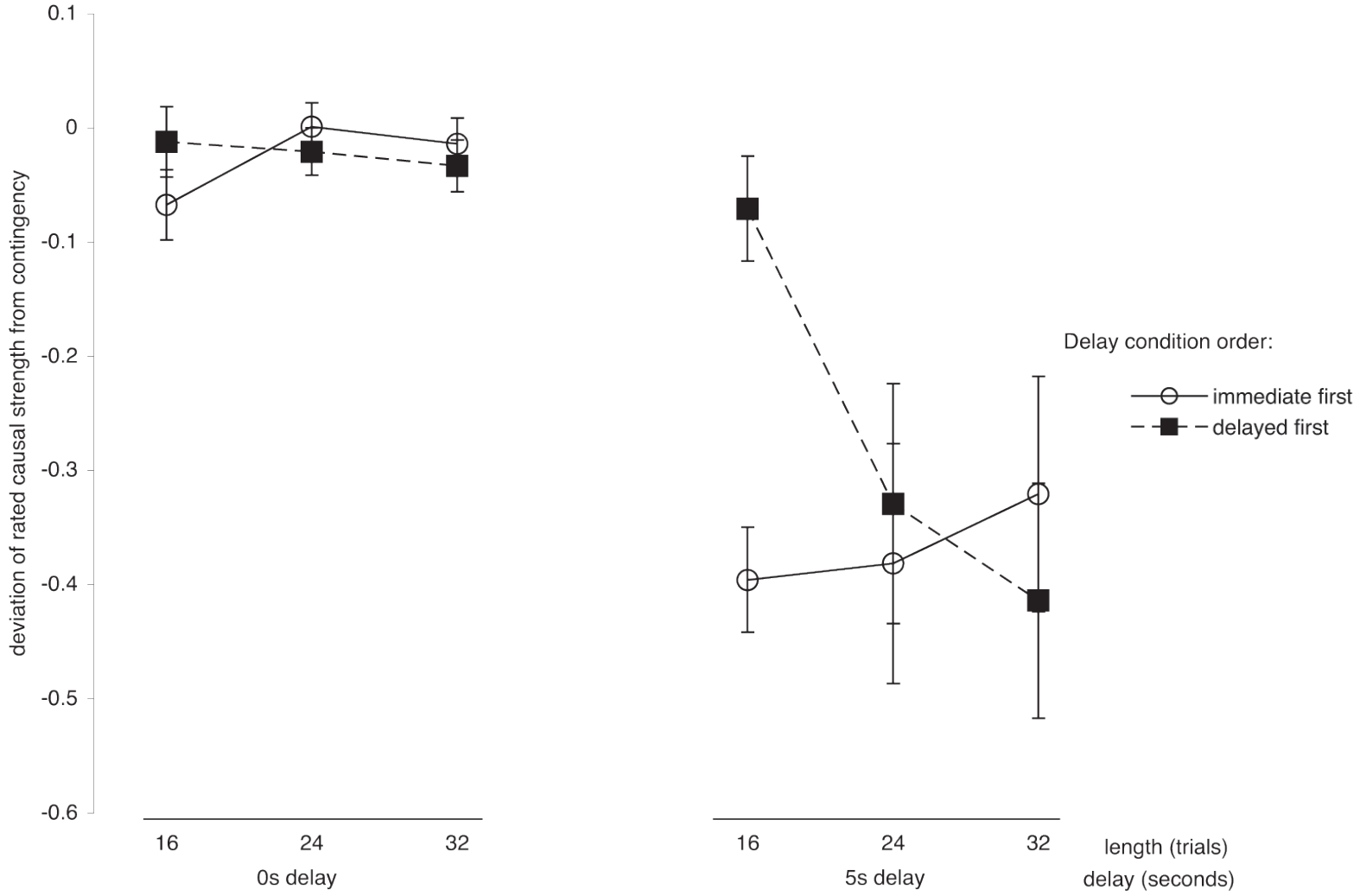
Results and Discussion

As in Experiment 2, the P(e|c) estimates were almost identical to the causal ratings, and overall they correlated $r = 0.82$, with P(e|c) and P(e|¬c) being inversely related $r = -0.83$. Causal ratings and P(e|¬c) correlated $r = -0.64$. In light of this, only analyses of the causal ratings will be reported. An ANOVA was conducted, with the between subjects factors of *Contingency* and *Order* and the within subjects factors of *Delay* and *Length*.

The ANOVA revealed main effects of *Contingency*, $F(1,20) = 6.44$, $\eta^2 = .06$ and *Delay*, $F(1,20) = 25.4$, $\eta^2 = .26$, and interactions between *Length* x *Order*, $F(2,20) = 8.92$, $\eta^2 = .03$, *Length* x *Delay*, $F(2,20) = 3.85$, $\eta^2 = .02$, and *Length* x *Delay* x *Order*, $F(2,20) = 3.76$, $\eta^2 = .02$. The *Contingency* effect reflected higher ratings given to the .75 contingency condition (mean = .52) than to the .50 contingency condition (mean = .38), and the *Delay* effect reflected higher ratings given to the 0s condition (mean = .60) than the 5s condition (mean = .31).

Because of the interactions of the within participant factors with *Order,* the sample was split and separate ANOVAs conducted on each level of this factor, as in the previous two experiments. Both groups showed effects of *Delay* (0s first: $F(1,10) = 12.0$, $\eta^2 = .30$; 5s first: $F(1,10) = 15.4$, $\eta^2 = .22$), and the 5s first group also showed a main effect of *Length* $F(2,20) = 8.45$, $\eta^2 = .09$ and an interaction of *Delay x Length* $F(2,20) = 6.96$, $\eta^2 = .07$. *Contingency* had a marginal effect in the 0s first group $F(1,10) = 4.32$, $\eta^2 = .09$, $p = 0.064$, and there was a marginal interaction of *Contingency x Delay* $F(1,10) = 4.19$, $\eta^2 = .06$, $p = 0.068$ in the 5s first group.

Figure 7. Deviations of mean causal ratings from observed contingencies from Experiment 3 broken down by Delay and Delay Condition Order as a function of learning trials. Error bars indicate standard errors.



These results are illustrated in Figure 7. The 0s delay conditions are rated by all participants with causal strengths remarkably close to the actual contingencies, regardless of the number of trials used to build the movie clip used for the condition. The 5s delay conditions are not rated as normatively: only the condition with 16 trials, when rated by those who were experiencing it as their first condition in the experiment, received ratings approximating to the actual contingency. In subsequent conditions, with 24 or 32 trials, these same participants rated the 5s delay as less causal, which is the opposite pattern to that predicted by associative learning theory.

We investigated the impact of *Length* on evaluations of delayed causal relations in more detail with a set of planned comparisons. We compared causal ratings from the 16 and 32 trials conditions involving a 5s delay separately for each combination of *Contingency* and *Order*. There was no difference in causal ratings after 16 and 32 trials of the .75 contingency in participants who worked in the .75/0s condition first, $t(5) = .89$. However, ratings of the .75 contingency from participants who worked in the .75/5s condition first did differ significantly between 16 and 32 trials, $t(5) = 3.71$, with the 16 trial condition ($M = 60.83$) receiving higher ratings that the 32 trial condition ($M = 15$). Analogously, there was no difference in causal ratings after 16 and 32 trials of the .50 contingency in participants who worked in the .50/0s condition first, $t(5) = 1.94$; the difference in participants who worked on the .50/5s condition first was in the opposite direction to that predicted by an associative learning account, with 16 trials ($M = 50$) producing higher ratings than 32 trials ($M = 27.17$), but this fell short of significance, $t(5) = 2.25$, $p = .07$.

The differences between few and many learning trials in participants who first worked on a delayed contingency is consistent with results from Experiments 1 and 2, and our earlier work (Buehner & May, 2002, 2003), in that delayed conditions receive comparatively high ratings when presented at the beginning of an experiment, but lower ratings when presented later on, after participants have seen immediate conditions. Keep in mind that we implemented the counterbalancing of delay versus immediate first by fixing the first condition to always comprise 16 trials, and varying whether this condition was paired with 0s or 5s delay. Participants who worked on a delayed condition first thus always began with the 16 trial/5s condition. Because this condition would not have been preceded by immediate conditions, it would be expected to produce comparatively higher causal ratings than subsequent conditions (of greater length) involving the same delay. The effect of *Length* in Delay–first participants, while opposite to the pattern predicted by associative learning theory, thus need not be especially problematic for it. It may well be a methodological artefact. It remains the case, however, that the better learning of delayed relations with prolonged training was not apparent for Immediate–first participants, either.

Another interpretation of the results allows a reconciliation with the associationism prediction: prolonged exposure could indeed gradually increase the cause-effect association, but in cases where the cause-effect gap is too large (i.e., the effect falls outside

the "window of associability," see also General Discussion), these increases accrue for the background instead. In other words, rather than serving to increase causal ratings for the candidate, prolonged learning might paradoxically increase causal ratings for the background. A growing conviction that the cause happens "on its own" is then accompanied by decreasing ratings of the causal strength of the candidate cause. However, our results did not show the general negative effect of *Length* that this interpretation suggests. Apart from the Delay–first 5s conditions, no condition shows any change in ratings of causal strength as the amount of exposure increases.

Overall, then, Experiment 3 fails to provide evidence in support of the notion that prolonged exposure to delayed causal relations improves their assessment. This result casts doubt on the associative learning interpretation of causal inference that stipulates that cause-effect delays result in weaker trial-by-trial increments compared to contiguous pairings. If this were the case, we should have found at least some form of improvement over learning trials. Instead, it appears that participants overall judged delayed contingencies as less causal than contiguous cause-effect pairings, and this impression persisted irrespective of the amount of experience with the relation. Thus, at least in humans, factors other than sheer amount of repeated exposure guide the discovery of delayed causal relations.

General Discussion

The Attribution Shift Hypothesis Reconsidered

We began this paper by explaining that one reason why disruptions in temporal contiguity might impair causal learning could be that delayed effects erroneously get attributed to a background of alternative causes, instead of to the candidate cause. In a probabilistic framework of causality, delays thus degrade subjective impressions of contingency though a decreased impression of P(e|c) combined with a simultaneous increase in subjective P(e|¬c). This notion, originally suggested (and simultaneously contested) by Shanks and Dickinson (1987), carries strong intuitive appeal and makes perfect sense from a computational perspective. If one assumes that a causal learner strives for explanation, each occurrence of *e* needs to be explained. If, for whatever reasons, *c* fails to get credit for an occurrence, it follows that some alternative or competing cause *a* will be credited instead.

One way to contest the attribution shift hypothesis in the face of evidence that shows impaired causal learning from delayed contingencies, is to maintain that delayed instances of *e* fall into an attributional black hole. In other words, one would have to claim that delayed *e*s simply go unnoticed. Alternatively, one might argue that they are noticed, but that reasoners do not feel compelled to explain what caused them. A more principled suggestion was made by Shanks and Dickinson (1987). Their argument was that all cause-effect pairings, whether immediate or delayed, accrue causal strength for the cause-effect association, albeit very small ones in the case of delays. Because the associationist frame-

work they endorsed has no room for concepts such as conditional probabilities (unless they are interpreted as associative weights, i.e., treated as equivalent to causal strength, see e.g., Lagnado & Shanks, 2002), the notion that causal judgment is a function of subjective conditional probabilities is irreconcilable with this approach. Instead, causal judgments from delayed contingencies decrease not due to an attributional shift, but due to weaker increments per trial.

The combination from Experiments 1-3 in this paper, however, strongly supports the notion that delayed cause-effect pairings serve as evidence that the background causes the effect. Support comes in direct and indirect ways. Firstly, estimates of conditional probabilities P(e|c) and P(e|¬c) were sensitive to disruptions in temporal contiguity as predicted by the ASH across all experiments. Secondly, Experiment 1 showed that delays can render a programmed positive contingency subjectively negative, suggesting that participants thought that P(e|c) < P(e|¬c), an interpretation that is supported directly by participants' estimates of these probabilities. Thirdly, Experiment 2 showed that pre-exposure to a causally inactive background facilitates learning of delayed contingencies. The probability estimates collected in this experiment suggest that the effect of pre-exposure was indeed to discredit the ability of alternative causes *a* to produce *e*, in other words to prevent an otherwise occurring shift from P(e|c) to P(e|¬c). Whereas most groups of participants in this experiment displayed the same delay-induced attributional shift, the subgroup of participants with successful pre-exposure was immune to it, suggesting that these participants evaluated the delayed contingencies correctly, because they did not erroneously attribute delayed effects to *a*. Fourthly, Experiment 3 clearly showed that, contrary to associationist principles, delays did not simply weaken the increment in associative strength accruing from each delayed pairing. If this had been the case, then we would have observed at least some improvement over prolonged training.

In sum, in the absence of compelling reasons to do otherwise, delayed effects get attributed to causes other than an objectively covarying prior event. We will next discuss other possible functions or conceptualizations of cause-effect delays.

The "window of associability" and temporal gradients

A common concept within the associative learning framework is the notion of a "window of associability." This means that reinforcers can only update associative strengths of events that fall within a certain time window preceding the reinforcer. If an event falls outside this window (i.e., if the cause-effect delay is too long), then no reinforcement takes place at all (and presumably the context gains associative strength). A more refined version of this is to assume that the window is not fixed, but instead forms a gradient, such that the associability tails off according to some discounting function. The window of associability appears conceptually similar to ASH—after all it likewise suggests that delays impair causal learning not necessarily through weaker increments, but by undermining the CS's gain of associative strength while simultaneously incrementing strength for the context.

Under this assumption, the 5s delay we imposed in our experiments might generally have placed the tank rolling over the disk outside the window of associability for explosions. Some manipulations, such as the instructions in Experiment 1 could have served to increase the size of the window, allowing associations between causes and effects to be formed for a subgroup of participants. Other manipulations, such as exposing participants to prolonged training, on the other hand, would not have beneficial effects under such an account. If the cause is outside the window of associability then it simply would not gain strength at all, not even a weak increment.

The similarity between the window of associability and ASH breaks down when one considers the results from Experiment 2, however. There is no way that pre-exposure could serve to increase the window of associability; its role in associative terms is clearly specified as reducing context conditioning (cf. Dickinson et al., 1996). Presumably, a prerequisite of countering the impact of cause-effect delays with pre-exposure is that the candidate cause is within the window of associability of the effect in the first place. Otherwise, pre-exposure might well decrease the amount of context conditioning, but the candidate cause would still be outside of the window, and hence no associations could be formed. Thus, we have to accept that the delayed relations in Experiment 2 involved causes that were within the window of associability. Because the delayed relations in Experiments 1 & 3 employed the same delay, 5s, as in Experiment 2, they must likewise have been within the window of associability. Thus, the function of the cover story in Experiment 1 could not have been to increase the window of associability—the data from Experiment 2 force one to accept that for our experiments this window is at least 5s long, so cover stories were not needed to extend it. Instead, it appears more likely that their function was to guide attribution.

Furthermore, the failure of prolonged training to increase causal ratings in the delayed conditions cannot be explained by a too narrow window of associability. While ASH can account for the results from all experiments, the window of associability cannot.

The temporal coding hypothesis and rate estimation theory

For a long time, influential associative learning theories have embraced the concept of trial-based learning (e.g., Rescorla & Wagner, 1972; Pearce & Hall, 1980; Pearce, 1987). Consequently, the role of time in the formation of associations has been somewhat overlooked. Notable exceptions to this trend are Miller's temporal coding hypothesis (see for example Miller & Barnet, 1993; Savastano & Miller, 1998) and Gallistel & Gibbon's (2000) rate estimation theory (RET). Miller strongly challenges the notion that the role of temporal contiguity is confined to influence the overall strength (or size of increments) in associative strength. Instead, he argues, time is an elementary component of the association. In a series of experiments, he demonstrated that animals acquire and retain information, not only about the strength of a CS-US association, but also about the temporal characteristics of this relation.

Gallistel and Gibbon's (2000) model goes one step further and argues that associations in the original sense envisioned by Pavlov do not exist. Rather, they argue, organisms keep track of the timing of events and compute rates of occurrence relative to certain salient stimuli (e.g., US occurrence in the presence vs. absence of the CS). These rates are then compared and decision thresholds govern whether or not (and when) a conditioned response occurs. While Miller's temporal coding hypothesis postulates that temporal information is an additional component acquired with associations, Gallistel and Gibbon argue that temporal information is all that is acquired.

It is beyond the scope of this paper to provide a detailed review of these theories, but we can at least ask whether and how well these important theoretical developments can account for our pattern of results in human causal reasoning. The temporal coding hypothesis could be used to explain the instructional effects we found in Experiment 1 and elsewhere (Buehner & May, 2002, 2003, 2004). It would be difficult for standard associative learning theory to explain how the exact same physical stimulus combination (involving delayed cause-effect pairings) is seen as causal under one set of instructions, but not another. However, if temporal information were an integral part of the association, it could account for instructional effects: the nature of the instructions, and the mention of a delayed or immediate causal mechanism, could have served to provide in memory an existing association and concomitant timeframe for this association. It is less clear, however, that the temporal coding hypothesis would make predictions different to standard associative learning theory about pre-exposure (Experiment 2) and prolonged learning (Experiment 3).

Rate-estimation theory, on the other hand, could probably not account for the instructional effects of Experiment 1. It can, however, explain the beneficial effects of pre-exposure to the un-reinforced background. Providing participants with additional time in the un-reinforced background is conceptually identical to increasing the inter-trial interval (ITI). According to RET, "increasing the average interval between trials increases the rate of acquisition" (Gallistel & Gibbon, 2000 p. 298). This can explain why participants in the pre-exposure condition of Experiment 2 have reached an acquisition criterion in the delayed condition, whereas the No pre-exposure participants have not. Gallistel and Gibbon's analysis of delay and trace conditioning explains why delayed procedures usually result in retarded acquisition: these procedures usually increase the CS-US delay but do not increase the ITI accordingly. Gallistel and Gibbon show that an appropriate adjustment of the ITI removes the detrimental effect of delay. This is commonly referred to as time scale invariance. Because our procedure kept the ITI constant across conditions, RET, like standard associative learning theory, would predict delays to result in retarded acquisition. The results from Experiment 3 show no evidence of retardation, however. Delays simply lowered causal ratings irrespective of the amount of learning. Retardation,

in contrast, would imply that the detrimental effect of delay could be (at least to some extent) overcome by prolonged training.

One could argue that the 5s delay in our experiments simply lowered the overall strength of a possible association, in other words, effectively lowered the asymptote. The data from the 0s conditions of Experiment 3 suggest that participants had already reached asymptote after 16 trials. If a 5s delay depressed the asymptote, participants might have plausibly already reached (the lower) asymptote in the 5s conditions after 16 trials and any additional learning trials would not have had any scope for improvement. While this notion is possible in principle, it is hard to reconcile with the results from Experiment 2. In Experiment 2, all aspects of the experimental context were identical to Experiment 3, so there is no reason why a delayed sensor-explosion association should support different levels of strength in one rather than another condition. Yet, the pre-exposure participants in Experiment 2 provided causal ratings that were much higher than those from the no pre-exposure group, or those from participants in Experiment 3. If delays effectively lower the absolute strength of a possible association, and ratings in Experiment 3 reflect the maximum strength possible under this delayed schedule, we would not have observed increased ratings in Experiment 2. In sum, while RET might be able to account for the results in Experiment 2, its predictions with respect to Experiment 3 (which are identical to those of standard associative learning theory) could not be supported, and the effects we found in Experiment 1 fall outside of its scope.

Knowledge-based Computation of Causal Power from Real-Time Data

We began this paper by criticizing contemporary accounts of causal learning, because they assume that these probabilities are readily available to the reasoner. Our critique does not discredit the approach altogether. We could show, however, that assessing the value of these probabilities is a fundamental prerequisite for the reasoning process. Once values for these probabilities have been obtained (either via direct and simplified information, as in the case of standard causal inference tasks, or via event parsing and segmentation of a continuous stream of events into causal and non-causal episodes, as in most real-world causal induction tasks, and in the experiments reported here), they can be fed into the causal reasoning engine. Our results have confirmed that degradations in cause-effect contiguity (such as delays) result in non-normative event parsing, in that delayed effects are not perceived as co-occurring with their causes, producing an incorrect assessment of the conditional probabilities, and also degraded causal ratings.

A new class of models for causal learning that has recently been advanced is that of causal structure learning models (e.g., Griffiths & Tenenbaum, 2005). According to these models, the reasoner compares the likelihood that the observed data is generated from a system that entails a cause-effect link to that of a system that contains no such link. Crucially, the comparison takes into account all possible strengths that the purported

causal relation might have. The measure that is achieved, causal support, thus reflects a degree of belief in the existence of a causal relation. Naturally, the stronger the relation is, the stronger the belief would be, and evidence suggesting that the effect occurs on its own will weaken the belief. In this respect, these models capture similar ideas as those of probabilistic causal learning. They go beyond these models, however, in that they also predict that belief increases as a function of sample size—a feature that is notoriously lacking in probabilistic models. Another advantage of these models is that they are not limited to contingency data but can instead also accept rates as input, and, thereby could at least in principle account of effects related to cause-effect timing, though a detailed computational specification of delay effects is at present not implemented in these models. Note that allowing the models to accept rates also solves the problem of defining $P(e|\neg c)$ in continuous time, which we discussed in the introduction.

A persistent feature across all our experiments was the order effect we found. In general, delayed relations were more positively evaluated when they formed the first evidence to be evaluated by the reasoner; delayed relations which had been preceded by immediate ones created lower causal impressions. Furthermore, manipulations aimed at overcoming the detrimental effects of delay (Instructions in Experiment 1, Pre-exposure in Experiment 2) were successful only, if they were followed by a demonstration of a delayed relation at the beginning of the experiment; evidence of immediate relations at the beginning of the experiment destroyed the impact of these manipulations. Similar order effects have been reported before (Buehner & May, 2002, 2003), and most likely reflect the artificiality of a computer-based scenario simulating real-world causal learning. Instructions, however convincing they may seem, can only ever be just that, and participants will always know that what they are observing is a computer-generated sequence of events. On a computer, it is always possible that the simulated effect follows the simulated cause immediately (there is no physical causal mechanisms that takes time to unfold), and thus it is only rational for learners to pick up on that. Furthermore, once they have experienced immediacy, delayed pairings will appear less appealing, regardless of previous instructions or manipulations. Interestingly, structure learning models might—at least conceptually—account for such order effects: it is plausible to assume that a delayed regularity provides stronger evidence for a causal relation in the absence of any other evidence than when it is evaluated in light of previous experience of immediate regularities. Again, though, a detailed computational specification of order effects is at present lacking in these models.

Our results have also shown that higher level knowledge about a delayed mechanism can correct contiguity biases. Hagmayer and Waldmann (2002) found that top-down knowledge in the form of temporal assumptions drives the segmentation strategies used to aggregate covariational data from unstructured lists. We have extended their results into a real-time scenario. That causal inference employs top-down components is by now widely accepted (e.g., see Waldmann, 1996; Buehner & Cheng, 2005). Ultimately, of

course, all (prior) knowledge is based on experience, and the burning question is how it is acquired in the first case, without initial top-down support. In the problems examined in this paper, the question is how hypotheses about the timeframe of causal relations can be extracted from continuous event streams. Temporal hypotheses are of course not the only example of causality related top-down knowledge that must have roots in a bottom-up process. Another example concerns the postulation of causal mechanism, recruited to help distinguish between genuine and spurious causes (e.g., although Russell's rooster crows reliably every morning before sunrise, we refrain from crediting it with causality because we know of no mechanism by which a bird could influence planetary motions). Lien & Cheng (2000) demonstrated that reasoners acquire (from statistical information) abstract categories of what is and is not causal, and use these categories to classify novel covariations as genuine or spurious. These abstract categories can be interpreted as an early form of knowledge of causal mechanism that allow subsequent hypothesis testing and strive for mechanistic explanation. The important lesson from Lien & Cheng for our purposes here is that they could show that top-down components (knowledge of mechanism) influencing an essentially bottom-up process (causal learning from statistical data) can themselves be acquired by this very bottom-up process.

We speculate that top-down components guiding the parsing of events into causal and non-causal episodes may likewise be acquired by the very bottom-up process that they ultimately end up influencing. Experiment 2 successfully explored this hypothesis with an idea inspired by associative learning theory. Pre-exposure to a causally ineffective background is most likely not the only way for participants to acquire temporal assumptions. Future research might investigate the same problem from different angles, and apply ideas from areas such as artificial grammar (Reber, 1969) or statistical (Fiser & Aslin, 2002) learning. Fiser and Aslin, for instance, have shown that people derive higher-order temporal structures from passive observation of very brief continuous event sequences, and can use these statistics for subsequent categorization tasks. Their results suggest that temporal contiguity should also play a role in statistical learning. For the human reasoner faced with the task of inferring relationships between events as they unfold in time: "What's to come is still unsure: in delay there lies no plenty" (*Twelfth Night, II:3)*.

Appendix

Instructions for Experiment 1

Imagine you are a military officer sent on a training course aimed at developing your skills in observing what happens on a battlefield. You already know that a lot of things are happening at once during combat, and it can be difficult to keep track of how effective and well functioning one's equipment is.

The context in which you will be assessing the effectiveness of different kinds of equipment is anti-tank defence.

You will observe a tank traversing a military training range. Aircraft flying over the range may drop radar-controlled smart-bombs on the tank. These aircraft are very high above the training range, and you cannot actually see them, nor whether or when they've launched a smart-bomb to pursue the tank. All you can see is when an explosion happens near the tank.

Your concern is with some special devices that have been placed on the training range.

These devices can have two different functions:

They can either be harmful for the tank, or they can protect the tank.

Section that was presented to the Mine group

In the harmful case, the devices are **landmines**.

As soon as the tank rolls over them, they explode.

Section that was presented to the Sensor group

In the harmful case, the devices are **trigger-sensors**.

As soon as the tank rolls over them, they trigger a rocket to be launched from several miles away to pursue and explode on the tank.

Both sets of Instructions continued:

In the protective case, the devices are **jamming transmitters**.

As soon as the tank rolls over them, they emit a radio signal that interferes with the radar-tracking of the smart-bombs that the aircraft drop on the tank (i.e., they prevent bombs from exploding).

Alternatively, the devices may be malfunctioning and have no effect whatsoever on whether or not there are explosions.

Alternatively, the devices may be malfunctioning and have no effect whatsoever on whether or not there are explosions.

Regardless of which function the devices have, they always look the same.

It is your task in this training program to carefully observe the tank traversing the training range and to decide whether and how strongly the devices produce or prevent explosions when the tank rolls over them.

Instructions for Experiments 2 and 3

Imagine you are a military officer sent on a training course aimed at developing your skills

in observing what happens on a battlefield. You already know that a lot of things are happening at once during combat, and it can be difficult to keep track of how effective and well functioning one's equipment is.

The context in which you will be assessing the effectiveness of different kinds of equipment is anti-tank defense.

You will observe a tank traversing a military training range. Aircraft flying over the range may drop radar-controlled smart-bombs on the tank. These aircraft are very high above the training range, and you cannot actually see them, nor whether or when they've launched a smart-bomb to pursue the tank. All you can see is when an explosion happens near the tank.

Your concern is with special kinds of sensors that have been placed on the training range. It is your task in this training program to carefully observe the tank traversing the training range and to decide whether and how strongly the sensors produce explosions when the tank rolls over them.

References

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of response alternatives. *Canadian Journal of Psychology, 34(1)*, 1-11.

Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). *Associative and normative models of causal induction: Reacting to versus understanding cause.* In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), Causal Learning (Vol. 34, pp. 1- 45). San Diego, CA: Academic Press.

Baker, A. G., Vallee Tourangeau, F., & Murphy, R. A. (2000). Asymptotic judgment of cause in a relative validity paradigm. *Memory and Cognition, 28(3)*, 466-479.

Buehner M. J., & Cheng, P. W. (2005). *Causal Learning.* In R. Morrison & K. J. Holyoak (Eds.) Cambridge Handbook of Thinking and Reasoning (pp 143-168).

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology:Learning, Memory, and Cognition, 29(6)*, 1119-1140.

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*, 865-890.

Buehner, M. J., & May, J. (2004). Abolishing the Effect of Reinforcement Delay on Human Causal Learning. *Quarterly Journal of Experimental Psychology, 57B(2),* 179-191.

Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Toward a general timeframe bias in causal induction. *Thinking and Reasoning, 12(4)*, 353-378.

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

Dickinson, A. (2001). Causal Learning: An associative analysis. *Quarterly Journal of Experimental Psychology, 54B(1)*, 3-25.

Dickinson, A., Watt, A., & Varga, Z. I. (1996). Context conditioning and free-operant acquisition under delayed reinforcement. *Quarterly Journal of Experimental Psycholog, 49B(2)*, 97-110.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99(1)*, 3-19.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(3)*, 458-467.

Gallistel, C. R. (2002). *Frequency, contingency and the information processing theory of conditioning.* In P. Sedlmeier & T. Betsch (Eds.), Frequency processing and cognition. Oxford, UK: Oxford University Press.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107(2)*, 289-344.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51(4)*, 285-386.

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

Hume, D. (1739/1888). *A treatise of human nature*. In L. A. Selby-Bigge (Ed.), Hume's treatise of human nature. Oxford, UK: Clarendon Press.

Lagnado, D. A., & Shanks, D. R. (2002). Probability judgment in hierarchical learning: a conflict between predictiveness and coherence. *Cognition, 83*, 81-112.

Lien Y. W. & Cheng, P.W. (2000). Distinguishing genuine from spurious causes. A coherence hypothesis. *Cognitive Psychology, 40(2)*, 87-137.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107(1)*, 195-212.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology, 37(3)*, 243-282.

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

Miller, R. R., & Barnet, R. C. (1993). The role of time in elementary associations. *Current Directions in Psychological Science, 2(4)*, 106-111.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review, 94(1)*, 61-73.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87(6)*, 532-552.

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

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*.

Reed, P. (1992). Effect of a Signaled Delay Between an Action and Outcome On Human Judgment of Causality. *Quarterly Journal of Experimental Psychology, 44B(2)*, 81-100.

Reed, P. (1996). No evidence for blocking in human judgments of causality by stimuli presented during an outcome delay. *Learning and Motivation, 27(3)*, 317-333.

Reed, P. (1999). Role of a stimulus filling an action-outcome delay in human judgments of causal effectiveness. *Journal of Experimental Psychology-Animal Behavior Processes, 25(1)*, 92-102.

Rescorla, R. A., & Wagner, A. R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*. In A. H. Black & W. F. Prokasy (Eds.), Classical Conditioning II: Current theory and research (pp. 64-99). New Yowk: Appleton-Century Crofts.

Savastano, H. I., & Miller, R. R. (1998). Time as content in Pavlovian conditioning. *Behavioural Processes, 44(2)*, 147-162.

Shanks, D. R. (1985). Continuous Monitoring of Human Contingency Judgment Across Trials. *Memory & Cognition, 13(2)*, 158-167.

Shanks, D. R. (1987). Acquisition Functions in Contingency Judgment. *Learning and Motivation, 18(2)*, 147-166.

Shanks, D. R., & Dickinson, A. (1987). *Associative Accounts of Causality Judgment.* In G. H. Bower (Ed.), Psychology of Learning and Motivation-Advances in Research and Theory (Vol. 21, pp. 229-261). San Diego, CA: Academic Press.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (Eds.). (1996). *The psychology of learning and motivation (Vol.34):Causal Learning.* San Diego, CA: Academic Press.

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). *Distinguishing associative and probabilistic contrast theories of human contingency judgment.* In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), Causal Learning (The Psychology of Learning & Motivation, Vol. 34, pp. 265-311). San Diego, CA: Academic Press.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal Contiguity and the Judgment of Causality By Human Subjects. *Quarterly Journal of Experimental Psychology, 41B(2)*, 139-159.

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

Acknowledgments

We thank Mark Haselgrove for helpful discussions of associative learning concepts, and Lorraine Allen and Steve Lindsay for comments on an earlier version of this paper. This research has been funded by ESRC grant R000223692, and was conducted while both authors held appointments at the University of Sheffield, England.

Paper submitted on March 4, 2009.

The final version accepted on August 20, 2009.