

Structure Awareness in Action-Outcome Learning Eradicates the Detrimental Effect of Reinforcement Delays

W. James Greville (GrevilleWJ2@Cardiff.ac.uk), Adam Cassar (AdCas43@Goolglemail.com)
Mark Johansen (JohansenM@Cardiff.ac.uk), Marc J. Buehner (BuehnerM@Cardiff.ac.uk)

School of Psychology, Cardiff University,
Tower Building, Park Place, Cardiff, CF10 3AT, Wales, UK

Abstract

Many studies of Action-Outcome Learning have demonstrated that reinforcement delays exert a detrimental influence on learning performance. Different theoretical perspectives offer varying explanations for this effect. A rational perspective suggests that as long as action-outcome pairings can be clearly recognized, delays should not interfere with the inductive process. Here we tested this idea by manipulating whether action-outcome contingencies were clearly identifiable as such by providing structural information in real time. In the absence of such information, we replicated the familiar detrimental effects of delay. Providing structural markers, and thus allowing easy identification of action-outcome pairings, eradicated this effect. Importantly, two additional experiments indicate that these results cannot be attributed to alternative explanations involving outcome salience or better awareness of timing. We conclude that when the environment allows Action-Outcome Learning to be conceptualized as a contingency learning task, learners are capable of covariation computation and immune to variations of response-outcome timing.

Keywords: Causality, Contiguity, Reinforcement, Structure, Computation

Introduction

The detrimental effect of a cause-effect delay on the learning of a causal relation is well established. However, the precise reason for this effect is still the subject of some debate. While it seems fairly intuitive that delayed causal relations might be more difficult to detect, and judged as weaker, compared to more immediate relations, this raises the question of how we ever manage to infer delayed causal relations of more than a few seconds. Yet we manage to do so routinely in day-to-day life. At the same time, laboratory experiments using basic stimuli have demonstrated that delayed causal relations of more than a few seconds could not be distinguished from non-contingent alternatives (Shanks, Pearson, & Dickinson, 1989). It therefore follows that in real-world causal induction, some other tangible source of information must be brought to bear that enables us to correctly identify delayed causal relations.

There have been a plethora of studies investigating the ability of humans to judge event contingencies (e.g. Shanks, 1987; Wasserman, Chatłosh, & Neunaber, 1983). A long-standing paradigm is the instrumental free-operant procedure (FOP), whereby participants evaluate the effectiveness of their responding (for instance pressing a key on a keyboard) in producing an outcome (such as a flash or a tone). These experiments are typically

programmed with an invisible underlying trial structure, whereby the condition timeline is divided into several temporal segments. If a response is made during a particular segment, then an outcome will be scheduled to occur (with a certain probability) at the end of that segment. A key consideration that is often overlooked in such experimental designs is whether this trial structure is apparent. This may play a critical role in the mediation of empirical cues such as delay.

Several potential explanations for the effect of delay have been offered stemming from different theoretical motivations. Traditional associative accounts argue that causal induction is simply an extension of associative learning, and is as a consequence governed by the same principles as other forms of learning such as Pavlovian and instrumental conditioning. This perspective adopts the Humean assertion that temporal contiguity is necessary for learning to occur. Degradation of this contiguity leads to weaker increments of associative strength and thus universally attenuates learning.

Cognitive perspectives on causal learning, on the other hand, tend to focus on event contingencies. Most proponents of this view agree that the sensory input available to us, in the form of presence or absence of events, is computed to provide an assessment of the covariation between candidate causes and effects. In the simplest terms, the possible event combinations are as follows: Both cause and effect occur (c, e), the cause occurs without the effect ($c, \sim e$), the effect occurs without the cause absent ($\sim c, e$), and neither cause nor effect occur ($\sim c, \sim e$). These event frequencies are often represented in a 2×2 contingency matrix, and form the basis for many different computational models of learning (see, e.g., Hammond & Payne, 1983). Provided that this information can be discerned from the available evidence, contiguity is not essential.

The role of contiguity from this perspective is instead limited to determining whether or not events are classed as contingent. Longer intervals increase the likelihood of intervening events to occur between action and outcome, which compete for explanatory strength and place greater demands on processing and memory resources. Accordingly, where there is some temporal separation between cause and effect, the crucial decision revolves around deciding whether this constitutes a case of c, e , or separate cases of $c, \sim e$ and $\sim c, e$. The greater the delay, the more likely the latter becomes, and the effect will not be attributed to the cause. This is therefore known as the *attribution shift* hypothesis (Buehner & May, 2009).

Experiments by Buehner & May showed that by appealing to higher level knowledge, the detrimental effect of delay can be modulated (2002) and abolished completely (2004). Participants were presented with action-outcome learning tasks in different thematic scenarios. By manipulating the context using cover-stories, a delay between cause and effect was made to seem plausible by providing explicit information regarding the expected timeframe of the causal mechanism. In a scenario where participants evaluated the effectiveness of pressing a switch on the illumination of a lightbulb, one group of participants were told that the bulb was an ordinary bulb that should light up right away, while another group of participants was instructed that the bulb was an energy-saving bulb that lights up after a delay. For this latter group there was no decline in ratings with delay; delayed and immediate causal relations were judged as equally effective.

These findings were consistent with the knowledge-mediation hypothesis (Einhorn & Hogarth, 1986): reasoners have pre-existing ideas about specific mechanisms by which causes produce their effects, which in turn enables flexible interpretation of incoming evidence, including appraisal of delayed causal relations. However, a problem with this approach is circularity: if causal learning is governed by top-down assumptions regarding causal mechanisms, where does this knowledge come from in the first place?

Perhaps some causal knowledge is innate. Stimulus selectivity in rats (Garcia & Koelling, 1966) demonstrates that animals indeed have pre-existing conceptions about the types of stimuli that can elicit particular physiological reactions. It is therefore not unreasonable to suggest that animals (including humans) may likewise have some prior expectation about certain potential mechanisms, which may well include non-contiguous causal relations. Nevertheless, it seems appropriate to search for other means by which the connection between a proximal candidate cause and a distal effect may be bridged. Are there cues that can mitigate the impact of delay without recourse to knowledge of mechanism?

Our goal here was to create a paradigm by which the underlying trial structure could be made evident without appealing to prior knowledge, or manipulation of expectations using cover stories or thematic contexts. Instead, we aimed to convey this information using stimuli that are directly observable in the learning environment and thus demonstrate that empirical cues can be used to infer delayed causal relations without any prior cognitive bases. This was achieved by using a brief auditory tone to signal the end of each trial. This tone occurred regardless of whether an effect occurred or not, and if an effect was scheduled it occurred simultaneously with the tone. The tone thus marked the point at which an effect could potentially occur.

Our hypothesis represents a convergence of two traditionally opposing perspectives on causal learning. In accordance with associative learning theory, we predict a decline in causal ratings as delay increases and no additional

information is provided. However, when the tone is introduced, providing markers that effectively reveal the delineation into trials, then contiguity becomes unimportant. The task will reduce to a simple contingency judgment, and we should see no delay-induced decline in ratings, as predicted by a computational account of causal learning.

Experiment 1

Method

Participants

33 undergraduate students from Cardiff University were recruited via an online participation panel. Participants included both males and females, with a modal age of 19 years. Either course credit or £3 payment was awarded for completion of the experiment. One participant failed to make any responses during two of the experimental conditions and thus was dropped from the analysis.

Design

The factors *trial length* (2s/5s) and *trial structure* (*apparent* vs. *not*) combined to produce four experimental conditions. Previous studies have found manipulation of trial length as an effective determinant of action-outcome delay, thus 2s and 5s were classed as *short delay* and *long delay* conditions respectively. For the *apparent* conditions, the end of each trial was signaled by an auditory tone, with the commencement of the next trial coinciding with tone offset. Meanwhile no additional cues were provided for the *not apparent* condition. Effectively, each trial ran seamlessly into the next, with no markers delineating one trial from the next (other than the occurrence of an effect). All participants experienced all four conditions, providing a 2x2 within-subjects design. The conditions were blocked such that the two *apparent* conditions were always presented one after the other, likewise for the two *not apparent* conditions. The order of which *apparent* or *not apparent* condition was presented first, or whether the *apparent* or *not apparent* block was presented first, was counterbalanced. At the end of each condition, participants were presented with the following question: "Please enter a rating from 100 to -100 to indicate the effect you think the button had on the triangle's behavior. 0 means it had no effect, +100 means it always made it light up, and -100 means it always prevented it from lighting up." The rating provided by participants constituted the dependent measure.

Apparatus, Materials & Procedure:

The experiment was conducted on an Apple "Mac Mini" computer running Microsoft Windows XP and Python 2.4.1, with a 17" LCD display, with standard headphones used to deliver the auditory stimulus. The stimuli consisted of an outline of an equilateral triangle and an image of a red circular button situated directly beneath it. Participants were free to click on this button with the mouse at any point. On doing so, the button stimulus 'depressed' for 500ms.

An effect constituted the triangle 'lighting up' (the transparent background became bright yellow and a 'glow' effect appeared around the triangle border) for 500ms. The

occurrence of the effect was determined probabilistically. If a response was made during the trial, $P(e|c)$ was 0.7; if no response was made, $P(e|\sim c)$ was 0.2. Only the first response in each trial altered the probability from 0.2 to 0.7, with subsequent responses having no influence.

For the apparent conditions, at the end of each trial, an auditory tone of 1000Hz was played for 500ms. This tone signaled the end of the trial, with the next trial beginning on termination of the tone. If an effect was scheduled, it occurred at this point of the trial to coincide precisely with the tone. For not apparent conditions, an equivalent 500ms was added to the end of each trial and the effect (if scheduled) occurred during this period. This ensured identical trial lengths and reinforcement delays between apparent and not apparent conditions. Each condition comprised 60 consecutive trials; total condition lengths were thus 150s and 330s for 2s and 5s conditions.

Participants were instructed to determine to what extent pressing the button caused or prevented the triangle from lighting up. Apparent conditions included the following additional instructions: "Each problem is divided into a series of trials. The end of each trial is marked by a beep. The triangle can only light up once per trial, and if it does so, it will light up at the end of the trial (i.e. to coincide with the beep)."

Results & Discussion

Causal Ratings

All analyses adopted a significance level of 0.05. One participant failed to make any responses during two of the experimental conditions and thus was dropped from the analysis altogether. One additional data point which was more than two standard deviations from the mean was also removed from the analysis for causal ratings. Figure 1 shows that ratings fell sharply in the *not apparent* conditions as trial length (and resultant action-outcome delay) was increased. However, a corresponding decline is not seen for the apparent conditions; there appears to be no difference between 2s and 5s. This suggests that the provision of trial structure information nullified the deleterious impact of delay.

A 2x2 within-subjects ANOVA corroborated these impressions, finding significant main effects of delay ($F_{(1,31)} = 7.276$), trial structure ($F_{(1,31)} = 4.322$), and a significant delay x structure interaction ($F_{(1,31)} = 4.719$). This supports the original hypothesis. However we must exercise caution in the interpretation of these results. Because we employed a free-operant paradigm, it is possible that participants' response behavior differed between conditions, resulting in different objective response-outcome contingencies (cf. Buehner & May, 2003). If any such differences occurred were between the *apparent* and *not apparent* conditions, then manipulation of trial structure would be confounded with contingency and our results compromised. In addition, because participants were free to respond at any given time, there is no guarantee that increasing trial length will produce a concomitant increase in the action-outcome delay. A

participant could respond at any point during the trial and therefore it is perfectly possible that contiguous cause-effect pairings will be experienced in both the 2s and 5s conditions. A closer inspection of the behavioral data is therefore warranted.

Behavioral Data

Response rate was calculated as the total number of responses, both reinforced and unreinforced, produced by participants across the entire duration of the condition and including all responses made during each trial. Mean action-outcome interval was calculated as the time between the first response in a given trial and the subsequent effect (if one occurred). If the response was unreinforced then this was not included in the calculation.

An analysis of behavioral data using 2x2 within-subjects ANOVAs on response rate and action-outcome interval revealed that, as expected, action-outcome intervals were significantly longer for trials of 5s length than for 2s ($F_{(1,32)} = 84.942$) confirming that controlling trial length was effective in manipulating reinforcement delay. We also found an effect of trial length on response rate ($F_{(1,32)} = 28.437$), which replicates earlier findings (e.g. Buehner & May, 2003). The important comparisons, however, were those involving trial structure. Specifically, if action-outcome intervals were significantly shorter for *apparent* than *not apparent* conditions, then our case for structural insight would be weakened by a mediation through experienced delay. Likewise, differences in response rate would entail different objective contingencies experienced across these conditions.

However, there was no significant main effect of trial structure on either response rate ($F_{(1,32)} = 0.814$) or action-outcome interval ($F_{(1,32)} = 1.495$); neither was there significant interaction between trial length and trial structure for response rate ($F_{(1,32)} = 0.026$) or action-outcome interval ($F_{(1,32)} = 0.033$). We can thus have confidence that our results concerning causal ratings are purely driven by structure information, and are not mediated by behavioral differences.

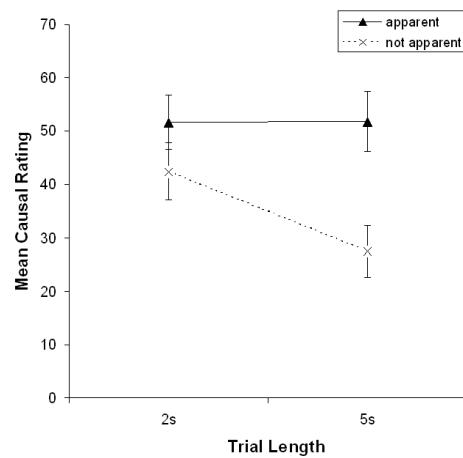


Figure 1: Mean causal ratings for Experiment 1. Error bars show standard error.

This finding suggests that causal learning in real time can, under certain conditions, be approached as a contingency learning task. When trial structure is apparent, contingency information can easily be discerned, and events accurately assigned to the cells of the contingency matrix. Under such circumstances, delays do not interfere with learning, as predicted by contingency-based or covariational models. Indeed in this case, judgments closely matched actual ΔP . Reinforcement delays thus are only detrimental to causal learning when they introduce ambiguity concerning response-outcome pairings.

It is important to note that our structural manipulation presented a tone simultaneously with the outcome. The tone therefore cannot act as a signal, bridging the temporal gap (Reed, 1992). There are however some other potential alternative explanations that must be ruled out.

Experiment 2A

Research in classical conditioning has demonstrated that increasing outcome salience increases the associative strength gained on each successive trial (e.g. Rescorla & Wagner, 1972). It could be argued that the tones marking the end of trials in the apparent conditions served to increase the saliency of the outcome, which coincided with them. If the causal learning process is subject to this property of associative learning, it might be responsible for alleviating the effect of delay. It could therefore be that our results are in fact driven by salience rather than structural insight.

To explore the effect of outcome salience, we modified the original paradigm such that under one set of conditions, outcome salience was increased, but without providing trial structure information. Accordingly, in one set of conditions, the triangle flash was accompanied by the same auditory tone used to provide structural markers in Experiment 1, adding to the salience of the outcome. The crucial distinction between this and the first experiment was that that here, the tone did not sound on occasions where there was no outcome, and thus did not convey trial structure information.

Method

Participants

32 participants, recruited as those in Experiment 1, completed the experiment to receive either £3 payment or course credit.

Design

Trial Length was either 2s or 5s as in the previous experiment, and Salience was either standard (no tone) or enhanced (tone present). Accordingly this gave four conditions which were presented in a blocked counterbalanced design as in the previous experiment.

Apparatus, materials & procedure

As before, except that in the *enhanced* conditions, the outcome was accompanied by the auditory tone, and participants received the following extra instructions: "When the triangle flashes, it will be accompanied by a tone." The *standard* conditions were identical to the *not*

apparent conditions in the previous experiment. This and the following experiment were conducted in a small computer lab using Windows XP machines, and testing multiple participants at once. Partitions between machines and use of headphones ensured that each participant could focus exclusively on their own task.

Results & Discussion

Causal Ratings

Two data points which were more than two standard deviations from the mean were removed from the analysis. Figure 2 shows that causal ratings declined as trial length increased from 2s to 5s for both standard and enhanced conditions. There also appeared to be a slight positive influence of enhanced outcome salience. Most importantly there appeared to be little difference between 5s-salient and 5s-standard conditions, suggesting that increasing outcome salience alone cannot replicate the observed effect from Experiment 1.

A 2x2 within-subjects ANOVA found the expected significant main effect of delay ($F_{(1,30)} = 5.634$). There was no significant effect of salience ($F_{(1,30)} = 1.705$) nor was the salience x delay interaction significant ($F_{(1,30)} = 0.036$).

Behavioral Data

We found the expected main effects of delay on both response rate ($F_{(1,32)} = 33.512$) and action-outcome interval ($F_{(1,32)} = 355.372$). The effect of salience was non-significant on both response rate ($F_{(1,32)} = 0.199$) and action-outcome interval ($F_{(1,32)} = 1.361$). The interaction between salience and delay was non-significant for both response rate ($F_{(1,32)} = 1.779$) and action-outcome interval ($F_{(1,32)} = 1.643$).

The non-effect of increasing outcome salience was not wholly anticipated; the literature suggests that this manipulation might have enhanced learning (although such a trend, albeit non-significant, was seen). More importantly however, the decline from 2s and 5s remains for the salient condition, while there is no real difference between the 5s-

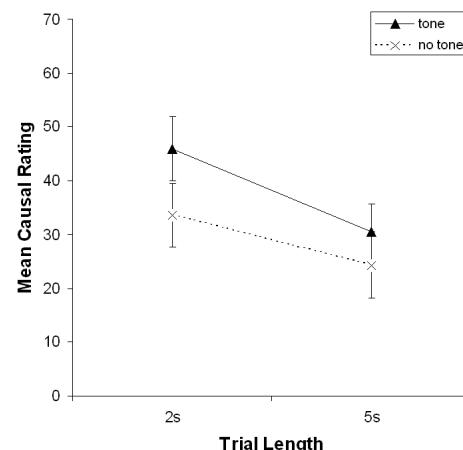


Figure 2: Mean causal ratings for Experiment 2A. Error bars show standard error.

salient and the 5s-standard conditions. We can thus rule out increased outcome salience as an alternative explanation for the effects observed in Experiment 1. We turn next to examine another potential confound, the presence of an auditory pulse.

Experiment 2B

In Experiment 1, the tone sounding at regular intervals (at the end of each trial) effectively produced a metronomic pulse that might have influenced participants' perception of time. In Experiment 1, the meter of this auditory pulse changed in line with trial length, such that there was either a relatively quick pulse occurring every 2s, or a slower pulse every 5s. Importantly, although trial length was different, there was one beat per trial in each case, so the marking of the passage of time was consistent for both conditions. Thus there was an imposed degree of perceptual similarity which could have accounted for the lack of difference between 2s and 5s when the tone was present.

To test this alternative explanation, we modified Experiment 1 in a fairly simple manner that would retain the auditory pulse, without necessarily providing information regarding trial structure. The tone was thus moved so that it did not occur at the end of each trial, thus demarcating one trial from the next, but rather occurred midway through each trial. Each tone was separated by the exact same interval, thus still providing a regular pulse, but was now no longer contiguous with the end (or beginning) of each trial, and therefore did not convey (useful) trial structure information.

Method

Participants

34 psychology students from Cardiff University received either £3 payment or course credit for participation.

Design

As for the previous experiments, four experimental conditions were produced by combining the factors *Trial Length* (2s or 5s) and *Pulse* (present/not present) and presented in a blocked counterbalanced design.

Apparatus, Materials & Procedure:

The apparatus, location and procedure was identical to the previous experiment, except that participants in the pulse conditions received the following extra instructions: "You will hear a tone sounding at regular intervals. This is a pulse to help you keep track of time."

Results & Discussion

Causal Ratings

Figure 3 suggests that the auditory pulse did not alleviate the effect of delay. Interestingly however, it does seem that the presence of the pulse did improve judgments of causality across the board; both for 2s and 5s, although it did not noticeably improve judgments at 5s relative to 2s. This general effect could be due to a slowing down of the internal pacemaker by the auditory pulse. Studies have provided evidence that human time perception is determined

by a temporal oscillator, the frequency of which can be altered by interference from an imposed rhythm (Treisman, Faulkner, Naish & Brogan, 1990). Slowing the frequency means time seems to pass more quickly and the subjective duration of intervals is shortened. As a result, the perceived delay between cause and effect could have been decreased by the presence of the auditory pulse.

A within-subjects ANOVA found significant main effects of both pulse ($F_{(1,31)} = 4.413$) and delay ($F_{(1,31)} = 5.523$), but importantly no significant pulse x delay interaction ($F_{(1,31)} = 0.988$). These results suggest that the effect in Experiment 1 is not attributable to the presence of the auditory pulse alone, and is due to our manipulation of trial structure information. However, one has to be cautious in the interpretation of a null result. While the interaction indeed falls considerably short of significance, one can notice from Figure 3 that the slope from 2s to 5s for the *pulse* condition is less steep than that for the *no pulse* condition. One might therefore suggest that a more powerful experiment may also have elicited a significant interaction.

These slight concerns can be alleviated by an inspection of the behavioral data. The bisection of the trial by the tone had the potential to induce a change in the behavior of participants. Some significance may have been attached to the tone, for instance being perceived as marking the start of the trial, or a point at which they should respond. Participants may therefore only have responded at or after the tone, and by doing so, effectively cutting the trial in half, and significantly reducing the action-outcome delay. This would account for the increase of 5s relative to 2s – if trial length is indeed truncated in this fashion, it will have been shortened by approximately 2.5s compared to 1s.

Behavioral Data

An analysis of the behavioral data reflected these suspicions. While the main effect of pulse on response rate was non-significant ($F_{(1,34)} = 2.760$), there was indeed a significant effect of pulse on action-outcome interval ($F_{(1,34)} = 25.983$). Mean action-outcome intervals where no pulse

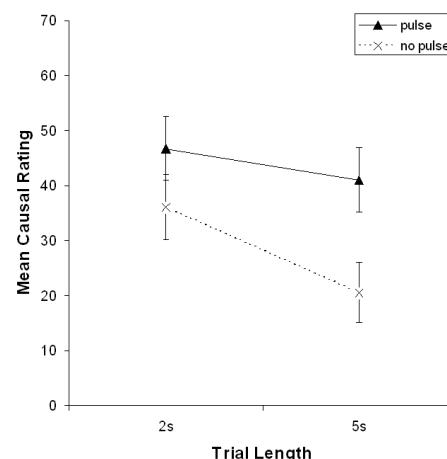


Figure 3: Mean causal ratings for Experiment 2B. Error bars show standard error.

was present were 1.38s and 3.46s at trial lengths of 2s and 5s respectively; with the inclusion of the pulse, these were shortened to 1.11s and 2.78s. This can explain both the main effect of pulse, through the overall reduction in delay, and also the smaller decline in ratings from 2s to 5s (when the pulse was present) as there is a smaller discrepancy in delay. Consistent with the previous experiments, we also found the expected main effects of trial length on both response rate ($F_{(1,34)} = 13.819$) and action-outcome interval ($F_{(1,34)} = 546.072$). The interaction between pulse and trial length was significant for action-outcome interval ($F_{(1,34)} = 5.477$) but not for response rate ($F_{(1,34)} = 1.054$).

We can therefore be confident in our assessment that auditory pulse is not the determinant of the interaction observed in Experiment 1; when trial structure was present, 5s conditions received significantly higher ratings than when it was not, despite a lack of difference in actual action-outcome interval. If this effect were driven by the pulse, then in the present experiment, coupled with the behavioral shift, a significant interaction should have been even more likely, yet was not obtained.

General Discussion

This paper has demonstrated that by providing structural information in a real-time causal judgment task, the detrimental effect of temporal separation between action and outcome can be abolished. When cause and effect pairings are clearly delineated, the learning process appears to reduce to a simple contingency assessment which is unaffected by delay. Two follow-up studies ruled out potential alternative explanations for this effect, thus we can have confidence in the validity of the trial-structure manipulation.

These findings are consistent with a rational perspective on causal induction and could be regarded as a step towards overcoming the problem of circularity that hampers the causal mechanism view. It may well be that in the absence of clear structural information, other sources of knowledge (such as expectations based on previously acquired mechanistic beliefs) serve to divide the event stream into meaningful patterns of event co-occurrence. Importantly, we have shown that such beliefs are not necessary when structural information is apparent in the input, and furthermore, that such information serves to overcome the well-established detrimental effects of reinforcement delay.

References

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

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgement of causality: effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology*, 56A(5), 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., & May, J. (2009). Causal Induction from Continuous Event Streams: Evidence for Delay-Induced Attribution Shifts. *Journal of Problem Solving*, 2(2).

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

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4(3), 123-124.

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.

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 York: Appleton-Century Crofts.

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.

Treisman, M., Faulkner, A., Naish, P. L., & Brogan, D. (1990). The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, 19(6), 705-743.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of Causal Relations in Humans. *Learning and Motivation*, 14, 406-432.