Structural awareness mitigates the effect of delay in human causal learning

W. James Greville · Adam A. Cassar ·
Mark K. Johansen · Marc J. Buehner

Published online: 12 April 2013
© Psychonomic Society, Inc. 2013

Abstract Many studies have demonstrated that reinforcement delays exert a detrimental influence on human judgments of causality. In a free-operant procedure, the trial structure is usually only implicit, and delays are typically manipulated via trial duration, with longer trials tending to produce both longer experienced delays and also lower objective contingencies. If, however, a learner can become aware of this trial structure, this may mitigate the effects of delay on causal judgments. Here we tested this “structural-awareness” hypothesis by manipulating whether response–outcome contingencies were clearly identifiable as such, providing structural information in real time using an auditory tone to delineate consecutive trials. A first experiment demonstrated that providing cues to indicate trial structure, but without an explicit indication of their meaning, significantly increased the accuracy of causal judgments in the presence of delays. This effect was not mediated by changes in response frequency or timing, and a second experiment demonstrated that it cannot be attributed to the alternative explanation of enhanced outcome salience. In a third experiment, making trial structure explicit and unambiguous, by telling participants that the tones indicated trial structure, completely abolished the effect of delays. We concluded that, with sufficient information, a continuous stream of causes and effects can be perceived as a series of discrete trials, the contingency nature of the input may be exploited, and the effects of delay may be eliminated. These results have important implications for human contingency learning and in the characterization of temporal influences on causal inference.

Keywords Causality · Contiguity · Reinforcement delay · Trial structure · Free-operant procedure · Associative learning · Decision making · Reasoning · Judgment

Introduction

Causal learning is a core cognitive competency that enables us to impose structure on the world and to intervene on the environment to achieve desired outcomes. The principles underlying causal learning are still debated (Dickinson, 2001b; Griffiths & Tenenbaum, 2005, 2009; Holyoak & Cheng, 2011). In most cases, a causal relationship between one event and another cannot be directly perceived. Rather, the connection must be inferred by detecting patterns in the occurrence of these events. Most contemporary theories of causal learning acknowledge three crucial cues to causality, first described by David Hume (1739/1888): temporal order (i.e., causes must precede their effects), contingency (the causes must reliably and repeatedly produce their effects), and contiguity (the causes and effects must occur closely together in time).

Temporal order is almost unanimously accepted as a necessity for causal learning. Most researchers also agree that in order for two events to be classified, respectively, as cause and effect, some form of statistical dependence (i.e., contingency) of the latter on the former is necessary. Broadly, the stronger the contingency between cause and effect, the stronger the inferred relationship between them. In the case of two binary variables, at any given point there are four possibilities: Both the cause and the effect may be either present or absent, which can be represented using a 2×2 contingency table. Table 1 illustrates the four possible outcomes relevant to a simple binary causal relation, where the cause c is either present or absent ($\neg c$), and the effect e likewise is also either present or absent ($\neg e$). The key debate is how exactly this information is used to obtain a metric of causality. One of the most well-known and longstanding

---

W. J. Greville (✉)
College of Medicine, Swansea University, Swansea, UK
e-mail: w.j.s.greville@swansea.ac.uk

A. A. Cassar
Cardiff University, Cardiff, UK

M. K. Johansen · M. J. Buehner
School of Psychology, Cardiff University, Cardiff, UK

Table 1 Standard 2×2 contingency matrix showing the four possible combinations of cause and effect occurrence and nonoccurrence

		Effect <i>e</i>	
		Present	Absent
Cause <i>c</i>	Present	A $e c$	B $\neg e c$
	Absent	C $e \neg c$	D $\neg e \neg c$

models is the ΔP statistic (Jenkins & Ward, 1965), which calculates contingency using the A, B, C, and D cells from the 2×2 matrix as: $A/(A + B) - C/(C + D) = P(e|c) - P(e|\neg c)$. Though more recent models have been developed that more accurately reflect human judgments than ΔP (Cheng, 1997; Griffiths & Tenenbaum, 2005), this metric can provide a useful estimate of causality in many cases.

Similarly, it is generally agreed that the stronger the contiguity between two events, the stronger the impression of causality (but see Buehner & McGregor, 2006), and that a lack of contiguity—that is, a delay between cause and effect—has a detrimental effect on causal learning: Causal judgments tend to decline as the temporal interval separating cause and effect increases. For instance, Shanks, Pearson, and Dickinson (1989) investigated the effect of delay on causal learning using a computer-based task in which human participants were required to rate the effectiveness of pressing a key to make a triangle light up on the computer screen. In their experiments, a triangle lit up with 75 % probability when the spacebar was pressed, and the interval between response and outcome was varied between conditions. Shanks et al. found that as the delay increased, participants' causal ratings decreased correspondingly. Indeed, participants were unable to distinguish conditions involving delays of 4 s or longer from noncontingent control conditions.

Research has demonstrated that the influence of both statistical (contingency) and temporal (contiguity) information can be considerably affected by the mode of information presentation or the presence of additional information (e.g., Buehner & May, 2002, 2003, 2004). Building on such research, we present a new “structural-awareness” hypothesis, which posits that when reasoners become aware that a real-time causal induction task has an underlying contingency structure, they make use of this structure to connect temporally separated causes and effects. More specifically, the structural-awareness hypothesis suggests that detrimental effects of cause–effect delay can be overcome when the continuous time stream affords segmentation of events into the four constituent units of a contingency table and learners

are aware of this underlying structure. Effectively, this reduces causal induction in real time to a simple contingency-learning task.

Theoretical perspectives on delay

While most contemporary theories of causal learning have recognized the general importance of contiguity, different theoretical perspectives suggest different mechanisms for the detrimental impact of delays on learning. According to an associative perspective (Allan, 1993; Baker, Murphy, & Vallée-Tourangeau, 1996; Dickinson, 2001a), causal learning is simply a reflection of the extent to which an association between cause and effect has been learned. Thus, learning a causal relationship is equivalent to learning the relationship between a conditioned stimulus (CS) and an unconditioned stimulus (US) in Pavlovian conditioning, or a response and an outcome in instrumental conditioning. The detrimental impact of delay on learning is well established (Grice, 1948; Wolfe, 1921), with degradations of contiguity leading to weaker increments of associative strength between stimuli. For instance, trace conditioning (in which a discrete interval separates the CS and US) is generally less effective than delay conditioning (in which the US immediately follows the CS), and the longer the delay, the weaker the association (Solomon & Groccia-Ellison, 1996). A notable exception is the “Garcia effect,” in which conditioned taste aversion can be established with delays of considerable length (Garcia, Ervin, & Koelling, 1966). As a general principle, though, delays are considered to impede the formation of associations, and are therefore assumed to attenuate human judgments of causality via the same associative processes.

Cognitive perspectives (Ahn, Kalish, Medin, & Gelman, 1995; Einhorn & Hogarth, 1986) adopt a different interpretation: The contiguity between cause and effect is seen to simply make the causal relation easier to detect and to learn. Put differently, delay places greater demands on cognitive resources, as events must be held in memory for longer, and also increases the likelihood that other events will occur during the delay, and thus compete for explanatory strength. A lack of contiguity thus introduces uncertainty as to whether a given effect was generated by the cause in question or by some other mechanism. What then becomes crucial, if there is temporal separation between cause and effect, is whether this imparts an impression of a single event of c and e , or of two separate events, c and $\neg e$ followed by $\neg c$ and e . The greater the delay, the more likely the latter interpretation becomes, and the effect will not be attributed to the cause (Buehner & May, 2009).

Yet, in day-to-day life, humans routinely demonstrate the capacity to connect causes with their effects over a broad

range of delays, sometimes of days, weeks, or months. While delayed causal relations might be more difficult to detect and might be judged as weaker, as compared to more immediate relations, the results of studies such as Shanks et al. (1989) have raised questions as to how we ever manage to infer delayed causal relations of more than a few seconds. Regardless of which theory of causal learning one subscribes to, it therefore follows that in real-world causal induction, some other source of information must enable us to correctly identify delayed causal relations. Buehner and May demonstrated that the effect of delay may be reduced (Buehner & May, 2002), or even eliminated completely (Buehner & May, 2004), by invoking high-level knowledge. Specifically, a “cover story” describing a context for the experienced events was provided prior to the experiment, whereby the delay between cause and effect was made either plausible (a grenade being launched or an energy-saving light bulb being switched on) or implausible (an ordinary light bulb being switched on). Participants in delay-plausible conditions showed a marked reduction in the extent to which delays negatively impacted their causal judgments.

Here, we considered another means by which learners could overcome the detrimental effects of delay in causal learning: Rather than appealing to top-down theories of causal mechanism or power, we asked whether, in certain circumstances, structural properties of the learning environment itself could be exploited to facilitate causal learning. More specifically, we investigated whether revealing the underlying trial structure frequently used in causal-learning experiments (and representative of many real-world causal-learning problems) would help reasoners recognize the contingency nature of the learning problem. The proposed structural-awareness hypothesis predicts that enabling learners to recognize the contingency structure of a learning problem unfolding in real time would allow them to discount temporal information and instead shift their focus to the statistical regularity between cause and effect.

Trial structure in the free-operant procedure

The experiments of Shanks et al. (1989) and Buehner and May (2002, 2003, 2004) were based on the instrumental free-operant procedure (FOP), a paradigm that has dominated research in causal learning over the past 40 years. In a standard FOP, participants evaluate the effectiveness of pressing a key in producing an outcome (such as a flash or a tone). These experiments are typically programmed with an invisible underlying structure, whereby the condition time line is segmented into trials of a fixed duration, also referred to as the *sampling interval* (Chatlosh, Neunaber, & Wasserman, 1985; Hammond, 1980; Reed, 1992, 1999;

Wasserman, Chatlosh, & Neunaber, 1983; Wasserman & Neunaber, 1986). If a response is made during a given trial, an outcome is scheduled to occur (with a certain probability) at the end of that trial, creating a response–outcome contingency. The delay separating response and outcome may be potentially influenced by adjusting the length of each trial, with longer trials tending to produce longer delays. If, for instance, a response is made at the beginning of a trial, then response–outcome intervals will increase in line with the trial length.

However, the FOP places no restrictions on whether or when participants may respond in the sampling interval. Consequently, the response–outcome interval on any given trial cannot be precisely controlled by the experimenter, but is instead dependent on when the participant chooses to respond. There is, therefore, no guarantee that longer trials will produce a concomitant increase in the response–outcome delay actually experienced. Furthermore, participants are free to respond more than once per trial, but typically only the first response is subjected to the reinforcement schedule. As a result, longer trials in an FOP may reduce the actual contingency experienced by the participant (Buehner & May, 2003).

Taking this into consideration, a key feature of such experimental designs that is often overlooked or unreported is whether the underlying trial structure is apparent to the participant. This may play a critical role in the interpretation of the statistical and temporal relations between cause and effect. For instance, if the structure is clearly delineated, this may prompt participants to modify their response behavior and respond no more than once per trial, preserving the objective contingency. The effect of delays may likewise be assuaged in similar fashion. An apparent trial structure may prompt participants to adjust the timing of the response to just prior to the end of the trial, analogously to the timing of the conditioned response in animals experiencing fixed-interval reinforcement schedules (Gallistel & Gibbon, 2000; Gibbon, 1977). Delaying responses in this manner would lead to a reduction in the experienced response–outcome interval, thus preserving the associative strength between response and outcome.

At a higher level, if participants become aware of the trial structure, this understanding may enable them to note whether or not they made *any* response, and then observe whether an outcome occurred at the trial end, thus focusing on the true underlying contingency. Furthermore, if participants become aware that an outcome can only occur at the end of the trial, effectively the actual response–outcome delay is irrelevant; regardless of whether a response is made early or late, the outcome is anchored to the same point. Participants thus may be able to ignore the delay and focus only on whether or not an outcome was delivered at the end

of the trial (and whether they made a response), and thus to make a purely contingency-based estimate of causality while ignoring the lack of contiguity. Thus, low-level cues denoting trial structure may serve to invoke higher-level reasoning processes.

In sum, trial structure information may mitigate the effect of delays through changes in the frequency or timing of responses, through a higher-level awareness of trial structure, or through a combination of the two. Our goal was to address this issue of trial structure. Far from being merely an artifact of an experimental paradigm, such trial structures, in which the occurrence of an outcome is governed according to a defined temporal arrangement, are abundant in everyday life. A prosaic example is the undergraduate admissions cycle: Applications can be submitted at any point up to a published deadline, and the outcome is announced on a prespecified date. Exactly where in the cycle candidates submit an application has no bearing on the likelihood of them securing a place, nor would multiple submissions to the same institution increase the odds. Applicants can submit their application just before the deadline, and so experience no delay between action and outcome. Alternatively, an understanding of the structure means that applicants appreciate that it is irrelevant when in the cycle they submit—the decision will always be announced on the same day. We aimed to determine whether awareness of such trial structures can influence subsequent causal judgments, and if so, whether such an influence can be achieved through changes in response timing and frequency, or top-down structural awareness, or both.

Our experiments

To this end, we created a paradigm in which the underlying trial structure was revealed. In Experiment 1, we manipulated the conventional FOP by using a brief auditory tone to mark the end of every trial (regardless of whether or not an effect occurred). However, we did not explicitly tell participants about this structural information or encourage them to use it in Experiment 1. In Experiment 2, we evaluated the possibility that the trial markers could serve merely to highlight the occurrence of outcomes, via increasing their salience. Experiment 3 then built upon Experiment 1, with the addition of explicit information about the trial markers.

Experiment 1

The aim of this experiment was to contrast the effects of apparent and not-apparent trial structures. We adopted an FOP similar to that of Shanks et al. (1989), manipulating delay by using trials of different lengths—specifically, 2 s (short) or 5 s (long). The critical manipulation was whether

markers indicating trial structure were present or not. For conditions with trial markers, a tone was played at the end of each trial in order to delineate one trial from the next.

As we discussed above, this manipulation could have several possible effects: Firstly, the provision of trial structure information might lead participants to adapt their behavior by responding only once and just prior to the end of the trial, so that response timing might preserve both contiguity and contingency, despite differences in trial length. Alternatively, or additionally, trial markers might instill an understanding of trial structure: If structural insight revealed the contingency between response and outcome, this might simply render delays and multiple responses inconsequential, enabling participants to overcome the effects of increased trial length without requiring a modification of response-timing behavior.

Method

Participants A group of 39 undergraduate students were recruited via an online participation panel hosted by the School of Psychology at Cardiff University. The median and modal ages of the participants were both 19 years, and course credit was awarded for participation.

Design We combined two levels of the factor Trial Length (2 s and 5 s, classed as the short-delay and long-delay conditions, respectively) with two levels of the factor Trial Markers (present and absent) to produce four experimental conditions. For the markers-present 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 cues were provided for the markers-absent conditions, and each trial ran seamlessly into the next, with nothing delineating the boundary (other than the occurrence of an effect on some trials). All participants experienced all four conditions, for a 2×2 within-subjects design. The conditions were blocked such that the two markers-present conditions were always presented together, and likewise for the two markers-absent conditions. Across participants, we counterbalanced whether the markers-present or the markers-absent conditions came first and, within each block, whether the conditions with short or long trials came first, for eight unique condition orders. 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 ratings provided by participants constituted the dependent measure.

Apparatus, materials, and procedure The experiment was conducted in a small computer lab, using Python version 2.4.1 on Windows PCs with 19-in. LCD widescreen displays. Standard headphones were used to deliver the auditory stimuli. The participants were tested in small groups, with partitions between individual workstations and the use of headphones ensuring that each participant could focus exclusively on his or her own task. Participants used the mouse to click on the button, and the keyboard to type in ratings. The experiment took approximately 20 min to complete.

The stimuli consisted of the outline of an equilateral triangle with an image of a red circular button situated directly beneath. Participants were free to click on this button with the mouse at any point. On doing so, the button stimulus “depressed” for 500 ms. An effect constituted the triangle “lighting up” for 500 ms. The occurrence of the effect was determined probabilistically: If a response was made during the trial, $P(e | c)$ was .7; if no response was made, $P(e | \neg c)$ was .2. Multiple responses on a given trial had no cumulative effect.

For the markers-present conditions, an auditory tone of 1000 Hz was played for 500 ms at the end of each trial (i.e., after every 2- or 5-s interval, depending on the condition). This 500-ms period did not count as part of the trial, but was *in addition* to the 2 or 5 s that had already elapsed. The tone thus 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 in the trial to coincide precisely with the tone. In order to ensure that the trials were of consistent length for both the markers-present and markers-absent conditions, the same additional 500 ms was added to the end of the trials in which no tone sounded, with the effect (if scheduled) again occurring during this period. Each condition comprised 60 consecutive trials; the total condition lengths were thus 150 and 330 s, respectively, for the 2-s and 5-s trial lengths. Figure 1 provides an illustration of the structure of the experiment and of the four possible types of trial (i.e., A, B, C, and D in the 2×2 contingency matrix in Table 1), in which a response was either made or not made, and an outcome either occurred or did not. The distinction between the markers-present and markers-absent conditions is indicated by the presence of the tone that delineates one trial from the next.

The general instructions for all participants in this series of experiments are provided in the [Appendix](#). For the present experiment, there was no difference in the instructions between blocks; participants in conditions in which the marker was present were not explicitly informed of its purpose, nor were they informed in advance that the markers would be present.

Results and discussion

All of the analyses for this and the following experiments adopted a significance level of .05. Three participants who provided causal ratings more than two standard deviations

from the mean were classed as outliers and excluded from all subsequent analyses. For the analyses of response frequency and timing, in addition to the participants already excluded on the basis of their causal ratings, those who returned data more than two standard deviations from the mean on a particular measure were also excluded from the analysis of that measure.

Causal ratings A preliminary analysis included the counterbalancing factor Block Order (markers present vs. absent first) as a between-subjects variable. Critically, we wanted to examine whether participants who first experienced the markers-present conditions experienced carryover effects into the subsequent block of markers-absent conditions, and in particular whether this would negate the effect of longer trials. Thus, we were primarily interested in the interaction between order and trial markers and in the three-way interaction between order, trial markers, and trial length. However, no significant interactions were found between either the former, $F(1, 34) = 2.064$, $MSE = 883.503$, $p = .160$, or the latter, $F(1, 34) = 0.391$, $MSE = 846.128$, $p = .536$, set of factors; hence, the remainder of the analysis focused on the within-subjects effects.

Figure 2 indicates that overall ratings were higher for the conditions including trial markers than for those that did not. Both conditions show that increasing trial length produced a reduction in causal ratings, but the decline appears less steep in the conditions with trial markers, suggesting that trial markers may have partly alleviated the effect of delay. A 2×2 within-subjects ANOVA confirmed significant main effects of trial markers, $F(1, 35) = 4.488$, $MSE = 910.355$, $\eta_p^2 = .114$, and trial length, $F(1, 35) = 5.372$, $MSE = 962.707$, $\eta_p^2 = .133$. However, the interaction between these two factors was not significant, $F(1, 35) = 0.971$, $MSE = 831.407$, $p = .331$.

The implication of these results is that the provision of trial markers facilitated participants’ ability to connect cause and effect, with their judgments corresponding more closely to the programmed contingencies when markers were provided. However, longer trials still had a negative influence on causal judgments that was not completely eliminated by providing these markers. Before moving on to consider further implications of these results, we must first examine response-timing patterns to determine whether manipulating trial length was indeed a direct determinant of experienced response–outcome delays, and whether the provision of trial markers affected the timing and/or frequency of participants’ responses.

Response frequency and timing Mean response–outcome intervals were calculated as the time between the last response in a given trial and the subsequent outcome (if one occurred). Unreinforced trials were not included in the calculations. Response frequency, meanwhile, was calculated as the total number of both reinforced *and* unreinforced

Fig. 1 Representative diagram of the markers-present (top) and markers-absent (bottom) conditions in Experiment 1. Each kind of trial is represented by a segment of the horizontal time line between two dashed vertical lines, where the pointing fingers indicate the occurrence of the response/cause, the triangles indicate the occurrence of the

outcome/effect, and the musical notes indicate the occurrence of the trial-marker tones. The four types of trials corresponding to the possible combinations of occurrence and nonoccurrence of responses/causes and outcomes/effects are shown at the bottom, where A, B, C, and D represent individual cells from the contingency matrix in Table 1

responses produced by participants across the entire duration of the condition.

Table 2 reports the mean response–outcome intervals and mean total responses for all three experiments. For the present experiment, both intervals and response totals were elevated for long as compared to short trials, as expected. We found little discernible difference between the conditions with and without markers when trials were short; however, in conditions with longer trials, response frequency was somewhat lower, and response–outcome intervals were longer, when markers were present than when they were absent.

An analysis of response timing using 2×2 within-subjects ANOVAs confirmed that response–outcome intervals were

significantly longer for trials 5 s in length than for the 2-s trials, $F(1, 32) = 215.824$, $MSE = 0.353$, $\eta_p^2 = .871$, demonstrating that controlling trial length was effective in manipulating experienced delay. The effect of trial length on total responses was also significant, $F(1, 32) = 14.172$, $MSE = 5,532.221$, $\eta_p^2 = .307$, which replicates previous findings (e.g., Buehner & May, 2003). The important comparisons, however, were those involving trial markers—specifically, to elucidate whether the effects of structural information involve an elicited change in response patterns or are due solely to a higher-level understanding of structure. Trial markers did not exert a significant effect on total responses, $F(1, 32) = 2.609$, $MSE = 1,007.408$, $p = .116$, nor did the interaction between trial markers and trial length reach significance, $F(1, 32) = 3.079$, $MSE = 1,610.471$, $p = .089$. However, a significant main effect of markers on response–outcome interval was evident, $F(1, 32) = 4.184$, $MSE = 0.262$, $\eta_p^2 = .116$, as well as a significant interaction between markers and trial length, $F(1, 32) = 5.970$, $MSE = 0.276$, $\eta_p^2 = .155$. Specifically, as Table 2 shows, experienced intervals for the conditions with 5-s trials were longer when markers were present.

These findings suggest that the provision of trial markers did indeed affect participants' behavior in terms of response timing and frequency. Response frequency was lower in the long-trial condition when markers were present than when they were absent, albeit not significantly so. More importantly, the significant effect of trial markers on response–outcome intervals was in the opposite direction from what might be predicted if structural information had affected response timing in a way that would facilitate causal judgment; it seems that participants tended to respond earlier when markers were present, thus creating longer response–outcome delays. A possible explanation for this finding is

Fig. 2 Mean causal ratings in Experiment 1 as a function of trial length for conditions with trial markers either present or absent. Error bars show standard errors

Table 2 Mean total responses and experienced response–outcome intervals in each condition for all three experiments

		Total Responses		Response–Outcome Interval (s)	
		Trial Length 2 s	Trial Length 5 s	Trial Length 2 s	Trial Length 5 s
Experiment 1	Markers present	38.849 (25.471)	75.333 (73.583)	1.257 (0.343)	2.999 (0.794)
	Markers absent	35.515 (16.192)	96.515 (104.067)	1.297 (0.235)	2.595 (0.619)
Experiment 2	Enhanced salience	28.500 (16.017)	54.000 (22.639)	1.348 (0.191)	2.682 (0.467)
	Standard salience	31.864 (18.859)	80.136 (53.082)	1.307 (0.193)	2.790 (0.431)
Experiment 3	Apparent	43.608 (22.215)	76.143 (53.137)	1.135 (0.194)	2.966 (0.634)
	Not apparent	34.786 (47.324)	74.321 (47.324)	1.166 (0.210)	2.854 (0.498)

Standard deviations are given in parentheses.

that participants took the trial markers as a signal to respond, bearing in mind that no prior information had been given about the meaning of the tone. If participants tended to respond fairly quickly after the tone, this would account for the longer response–outcome delays, relative to conditions without markers, in which participants would respond at any given time. However, this behavioral change did not negatively impact causal ratings, suggesting that structural information invokes higher-level processes. The effect of the trial markers on causal ratings was therefore not confounded with changes in response timing, but rather persisted in spite of this effect.

Our preliminary conclusion therefore was that trial markers served to make the trial structure apparent to participants, and that this information was then subsequently utilized to assist causal learning. Clearly, the provision of trial structure was insufficient to completely negate the effect of delay, as indicated by the absence of an interaction between trial markers and trial length (though we did find a trend in the expected direction). We attempted to address this point in Experiment 3, by providing explicit information about the purpose of the trial markers, but the immediate implication was that an identifiable trial structure enhanced causal attribution, consistent with the structural-awareness hypothesis. There remained, however, another potential explanation as to how the presence of the markers might have enhanced causal ratings without modifying overt behavior, which we addressed in the next experiment.

Experiment 2

It is important to emphasize that the tone used to delineate trials always occurred simultaneously with the outcome (on trials in which the outcome occurred), rather than preceding it. The tone therefore could not act as a signal for the outcome (Reed, 1992), thus ruling out a potential associative explanation for the observed effect—that of bridging the temporal gap with a second CS. However,

since the tone always coincided precisely with the outcome of the triangle lighting up, it could be argued that this additional stimulus increased the salience of the outcome in conditions with trial markers. Enhancing outcome salience increases the associative strength gained on each successive trial (e.g., Rescorla & Wagner, 1972), so if the process of causal induction is subject to this property of associative learning, the boost in associative strength might be responsible for enhancing causal attribution in those conditions including the tone. It could therefore be that the effect of providing trial markers in Experiment 1 was in fact driven by enhanced outcome salience rather than structural insight.

To address this question, we modified the original paradigm, such that outcome salience was increased, but without providing (additional) 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. The crucial distinction between this and the first experiment was that here, the tone did not sound on occasions in which there was no outcome, and thus did not convey additional trial structure information.

If enhanced outcome salience was responsible for the results of Experiment 1, here we would anticipate higher ratings in conditions in which the outcome was accompanied by the tone than in conditions with no tones. If, on the other hand, the results of Experiment 1 were attributable to an awareness of trial structure, since the tone no longer served to delineate all trials, it should have little influence on judgments in the present experiment.

Method

Participants A group of 33 undergraduate students from Cardiff University, with an average age of 19 years, were recruited via an online participation panel, receiving either £3 payment or course credit in return for participation.

Design The trial length was either 2 or 5 s, and salience was either standard (tone absent) or enhanced (tone present). This created four conditions, which were presented in a blocked, counterbalanced design, as in Experiment 1.

Apparatus, materials, and procedure The experiment was conducted in the same location using the same apparatus as in Experiment 1. The paradigm was a straightforward adaptation of the previous study, with the standard-salience conditions being identical to the markers-absent conditions in Experiment 1. In the enhanced-salience conditions (illustrated in Fig. 3), the outcome was accompanied by the tone, which was not delivered at any other point. Participants in the enhanced-salience conditions received the following extra instruction at the start of the experiment: “When the triangle flashes, it will be accompanied by a tone.”

Results and discussion

One participant failed to comply with the instructions, and one further participant received incorrect stimuli due to computer malfunction; both participants were removed from the analysis. Three additional participants, who provided ratings more than two standard deviations from the mean, were excluded as outliers from all subsequent analyses.

Causal ratings Neither the Block Order $\times$ Salience interaction, $F(1, 26) = 0.128$, $MSE = 747.910$, $p = .724$, nor the three-way interaction between order, salience, and trial length, $F(1, 26) = 2.960$, $MSE = 534.365$, $p = .097$, reached significance. The subsequent analyses therefore did not include block order as a factor.

Figure 4 shows the mean causal ratings for each condition. Ratings evidently declined as trial length increased from 2 to 5 s for both the standard and enhanced conditions. There appears to be little difference between the conditions with and without salience enhanced by the tone, although the enhanced-salience conditions did attract slightly higher ratings. Analysis revealed a significant main effect of trial length on the causal ratings, $F(1, 27) = 8.900$, $MSE = 1,017.361$, $\eta_p^2 = .234$, but no main effect of salience, $F(1, 27) = 1.133$, $MSE = 723.742$, $p = .297$, and no significant interaction between trial

Fig. 4 Mean causal ratings in Experiment 2 as a function of trial length for conditions with either enhanced outcome salience (tone) or standard outcome salience (no tone). Error bars show standard errors

length and salience, $F(1, 27) = 0.118$, $MSE = 572.154$, $p = .734$. Enhancing outcome salience with a tone therefore did not significantly improve learning about the causal relationship.

Response frequency and timing The same exclusion criteria for outliers were applied as in the previous experiment, and the results for Experiment 2 are shown in the middle of Table 2. We found the expected main effects of trial length on both total responses, $F(1, 21) = 41.236$, $MSE = 725.903$, $\eta_p^2 = .663$, and response–outcome interval, $F(1, 24) = 333.269$, $MSE = 0.149$, $\eta_p^2 = .933$. The effect of salience was not significant for response–outcome interval, $F(1, 24) = 0.276$, $MSE = 0.098$, $p = .604$, nor was the Salience $\times$ Trial Length interaction significant, $F(1, 24) = 1.470$, $MSE = 0.094$, $p = .237$. However, a significant influence did emerge of salience on total responses, $F(1, 21) = 6.862$, $MSE = 697.518$, $\eta_p^2 = .246$, qualified by a significant interaction between salience and trial length, $F(1, 21) = 6.696$, $MSE = 425.998$, $\eta_p^2 = .242$. Specifically, fewer responses were made when the tone was not present, and particularly so for the long-delay conditions.

The significant effect of the tone on total responses warrants further consideration. Since response frequencies were lower when the tone was present, this meant fewer responses

Fig. 3 Representative diagram of the enhanced-salience conditions in Experiment 2. The standard-salience conditions were identical to the markers-absent conditions in Experiment 1

per outcome and, in other words, a higher objective response–outcome contingency when the tone was present. This might be expected to have a facilitatory effect on causal judgments. Furthermore, since the effect was greatest in the long-delay condition, this should have elevated ratings further in this condition. In other words, the effect of the tone on total responses should have induced causal ratings to be more like those obtained in Experiment 1. However this was not the case: Simply providing the tone in conjunction with the outcome was not sufficient to increase causal ratings.

To summarize, adding the tone in conjunction with the outcome had no significant influence on participants' causal ratings. We found an inhibitory effect of the tone on response frequency, which would have increased objective contingency, yet this still did not enhance judgments of causality. The key conclusion, then, is that enhanced outcome salience is unlikely to be the explanation for the facilitatory influence of trial markers on causal judgments in the preceding experiment. Furthermore, the fact that the additional presence of the tone decreased total responses (thus increasing objective contingency) but did not affect causal ratings suggests that the nonsignificant trend in response frequency in Experiment 1 is similarly unlikely to have influenced causal ratings. This further supports the argument that the main effect of trial markers was due to structural insight rather than changes in response frequency or timing.

Experiment 3

Experiment 1 demonstrated that markers delineating trial structure can facilitate causal learning, while Experiment 2 showed that enhanced outcome salience was unlikely to be the explanation for this result. There remained, however, a key question: If the trial markers did indeed endow participants with the ability to identify structure, and thus to connect delayed effects with the responses that caused them, why was the effect of delay not completely abolished in conditions with trial markers? Though Fig. 2 suggests a partial amelioration of delay effects, this was not statistically significant. One potential reason is that participants were not given any explicit expectancy beforehand in the instructions that the trial markers would be present, and were not informed of the markers' purpose. As a result, it may have taken some time before participants came to realize the significance of the tone and that it indicated structure; indeed, some participants may never have come to this realization. A possible follow-up to this study would be to extend the learning time spent by participants on this task, to see whether the effects of providing trial markers would become more pronounced. However, a simpler alternative, used here, was to inform participants in advance of the presence and purpose of the tones marking the structure.

This eliminated the need for participants to infer the significance of the tone and helped reduce the task to simple contingency estimation from the outset.

Method

Participants A group of 34 undergraduate students with an average age of 20 years were recruited via an online panel, with partial course credit awarded for participation.

Design A 2×2 factor design was again employed, with the same levels of Trial Length (2 and 5 s) combining with Trial Structure (apparent and not apparent), providing four conditions.

Apparatus, materials, and procedure The experiment was conducted on an Apple “Mac Mini” computer running Microsoft Windows XP and Python 2.4.1, with a 17-in. LCD visual display and standard headphones used to deliver the auditory stimulus.

Perceptually, each condition was identical to the four conditions in Experiment 1. The key difference was that, prior to the apparent conditions (when markers were present), participants received the following additional information:

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).

Participants were thus notified in advance that each condition was divided into a series of trials, and that these trials would be denoted by auditory markers; hence, the trial structure was apparent from the outset.

Results and discussion

One participant failed to make any responses during two of the experimental conditions and was dropped from the analysis. There were no outliers, in terms of causal ratings. The same exclusion criteria as for the previous experiments were applied for the analyses of response frequency and timing.

Causal ratings As in the previous experiments, no significant Block Order $\times$ Trial Structure interaction was evident, $F(1, 31) = 0.627$, $MSE = 675.843$, $p = .435$, nor a significant Block Order $\times$ Trial Structure $\times$ Trial Length interaction, $F(1, 31) = 0.422$, $MSE = 374.880$, $p = .521$. Subsequent analyses were therefore collapsed across block orders.

Figure 5 clearly shows that causal ratings were lower for trials 5 s in length, as compared to those 2 s in length, when trial structure was not apparent. However, a corresponding decline was not seen for the apparent conditions; essentially, no effect of trial length occurred when the trial structure was

Fig. 5 Mean causal ratings in Experiment 3 as a function of trial length with either apparent trial structure or no apparent structure. Error bars show standard errors

apparent. Causal judgments were also generally higher when trial structure was apparent than when it was not. The analysis corroborated these observations, revealing significant main effects of trial structure, $F(1, 32) = 7.660$, $MSE = 1065.595$, $\eta_p^2 = .193$, and trial length, $F(1, 32) = 4.719$, $MSE = 420.799$, $\eta_p^2 = .129$, suggesting that making structure apparent improved judgments of causality, and that judgments declined as trial length increased. Most tellingly, the significant interaction between trial structure and trial length, $F(1, 32) = 4.702$, $MSE = 368.111$, $\eta_p^2 = .128$, indicated that delay only exerted a detrimental effect on judgments when structure was not apparent. Paired-samples t tests confirmed this dichotomy, with mean judgments being significantly higher for trials of 2 s ($M = 40.24$) rather than 5 s ($M = 25.24$) when no structure was provided, $t(32) = 3.02$, $p < .01$, but not differing significantly when structure was apparent ($M = 48.73$ and $M = 48.21$ for 2 and 5 s, respectively), $t(32) = 0.107$, $p = .92$. Taken together, these results suggest that the provision of trial structure information attenuated the deleterious impact of delay, consistent with the structural-awareness hypothesis. Once again, however, it was necessary to determine whether such effects corresponded to changes in response-timing patterns or were due to a true understanding of trial structure.

Response frequency and timing As is shown at the bottom of Table 2, we obtained the expected main effects of trial length on both total responses, $F(1, 27) = 41.364$, $MSE = 879.036$, $\eta_p^2 = .605$, and response–outcome interval, $F(1, 29) = 432.653$, $MSE = 0.215$, $\eta_p^2 = .937$. No significant main effect of trial structure was evident on either response–outcome interval, $F(1, 29) = 0.434$, $MSE = 0.112$, $p = .515$, or total

responses, $F(1, 27) = 0.855$, $MSE = 927.411$, $p = .363$; neither did we find significant interactions between trial length and trial structure for total responses, $F(1, 27) = 0.695$, $MSE = 493.667$, $p = .412$, or response–outcome interval, $F(1, 29) = 1.273$, $MSE = 0.122$, $p = .268$. Thus, the observed differences in the ratings between apparent and not-apparent trials are not explained by differences in response frequency or timing.

These findings suggest that causal learning in real time can, under certain conditions, be approached as a contingency-based learning task. When the trial structure is apparent, contingency information can easily be discerned. Under such circumstances, delays do not interfere with learning, and the problem of causal induction reduces to a simple contingency estimation task. Indeed, in this case, the judgments closely matched actual ΔP : Using values of .7 and .2 for $P(e | c)$ and $P(e | \neg c)$, respectively, $\Delta P = .5$, and the mean ratings were 48.73 and 48.21 out of 100 for the 2- and 5-s conditions, respectively, when trial structure was apparent. Of course, other, more sophisticated probabilistic models (e.g., Griffiths & Tenenbaum, 2005, 2009; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008) might provide an equally good, if not better, fit to the present set of results than ΔP ; our goal here, however, was not to provide evidence favoring one specific model over another, but rather to demonstrate the effects of structural awareness. We infer that one way in which reinforcement delays can impair causal learning is by introducing ambiguity concerning response–outcome pairings, and that this can be alleviated by awareness of trial structure.

General discussion

Our first experiment indicated the potential of providing cues to indicate trial structure in order to enhance judgments of causality in a real-time causal-learning task. A control experiment ruled out enhanced outcome salience as a competing explanation. A further experiment confirmed that by making the provision of structural information explicit, the detrimental effect of temporal separation between action and outcome could be completely abolished. Analysis of response patterns confirmed that this effect was not due to changes in objective contingencies or experienced response–outcome delays. We thus concluded that by conveying to learners that continuous time was carved into discrete learning trials, the learning process remained unaffected by the experienced response–outcome delays, consistent with the structural-awareness hypothesis.

To completely eliminate the effects of delay, it was necessary not just to provide low-level cues to participants, but also to explicitly inform them what the presence of these cues denoted. This need for explicit reminders to make use of available structural information

parallels results found by Gick and Holyoak (1980) in analogical problem solving. Participants, attempting to solve Duncker's (1945) radiation problem, were provided with a structurally analogous problem (attacking a fort) and its solution. In the absence of explicit encouragement to make use of the structural analogy, performance did not increase markedly. However, when the analogy was pointed out to participants, they readily used it to map the structures between problems and produce a correct solution. It seems that participants in our experiments likewise needed to be explicitly told what the trial cues meant in order to fully utilize trial structure. Even so, the presence of the cues alone notably improved judgments of causality in Experiment 1, and though we found no significant interaction between trial markers and trial length, Fig. 2 suggests at least a limited moderation of delay effects. It would be interesting to consider ways in which this moderating effect might be augmented without having to resort to verbal instructions, which could in turn provide insight as to how higher-level structural knowledge might be constructed from bottom-up stimuli.

A series of experiments by Wasserman, Chatlosh, and Neunaber (1983) may indicate a possible means by which this could be achieved. Wasserman et al. examined the effect of increasing the sampling interval during a contingency judgment task, contrasting trials 1 and 4 s in duration in their second experiment. Interestingly, no significant effects of adjusting the sampling interval were found on either response rates or judgments. However, Wasserman et al.'s focus in this study was purely on event contingencies, and they did not report the timing of participant responses. It is therefore impossible to conclude from their experiments whether adjusting the sampling interval actually increased response–outcome intervals, and thus degraded contiguity, or whether participants modified their behavior in response to the change in sampling interval, thus neutralizing the impact of trial length (and, hence, the necessity of the present study). Wasserman et al., however, speculated that participants might indeed have detected the sampling interval due to the high incidence of the outcome. Since five out of nine conditions in the experiment had either $P(e | c)$ or $P(e | \neg c)$ set at .875, “considerable temporal regularity in the onset of the outcome light would be common, thus effectively signaling the sampling interval” (p. 423). Expanding on this idea, it is possible that if signals denoting trial structure were provided in tandem with a high incidence of the outcome, the regular occurrence of the outcome at the same point in time as the trial marker would reveal both that a trial structure existed and that the outcome occurred at the end of each trial, without need for explicit verbal instructions. In future research, we intend to pursue this question as to whether instruction is always necessary for low-level cues to be fully utilized, and furthermore to see how changes in $P(e | c)$ and $P(e | \neg c)$ might interact with the trial structure effect.

Theoretical implications

The findings of this article build on the work of Buehner and May (2002, 2003, 2004), demonstrating that detrimental effects of delays in causal learning may be overcome. This conflicts with views of causal learning where it is argued that contiguity is necessary for the formation of an association (e.g., Arcediano & Miller, 2002). However, other theories of learning have considered a more complex role for temporal information. Gibbon's (1977) scalar expectancy theory (SET), for instance, postulates that temporal intervals are in fact the sole determinant of conditioning (Gallistel & Gibbon, 2000). This model was developed to account for the timing of the conditioned response (CR) in animals when there is some temporal separation between the CS and US. At the heart of this theory is the idea of a temporal accumulator that monitors the time until reinforcement. When a reinforcement is received, the latency is written to memory. At the onset of the CS, the currently elapsing interval (t_e) is compared to the remembered latency (t^*). When this ratio exceeds a threshold (β), the animal responds; hence, this ratio $t_e : t^*$ is known as the *decision variable*. Since the CR is an anticipatory response, the when-to-respond threshold β is somewhat less than 1. To summarize in the simplest of terms, the timing of the CR depends on when the animal expects the US to be delivered.

This model could feasibly be extended to account for the effects of contingency structure reported here. Through repeated presentation of the tone, regardless of whether or not the effect occurred, it becomes apparent that successive trials are separated by the same temporal interval, and this interval can be recorded in memory, analogously to the t^* signal specified by SET. Coupled with the knowledge that the effect can only occur at the end of a trial, there thus develops a clear expectancy of points in time at which an outcome can occur. Meanwhile, the outcome is not expected at other times. Attention can then be more closely directed to the point at which the outcome is anticipated—or, in terms of SET, when the currently elapsing interval t_e approaches the remembered interval t^* . The decision is thus simplified to comparing the rate or frequency of outcome occurrences during intervals with and without responses. In other words, the temporal predictability (Greville & Buehner, 2010) of the outcome appears to facilitate the attribution process.

Other work from our lab might, at first glance, appear to conflict with the results obtained here. Greville and Buehner (2007) presented participants with unambiguous data in tabular format, summarizing the effects of a particular treatment on the death of bacterial cultures over a 5-day period. In each case, it was indicated whether the culture was killed off, and if so, on which day. This described rather than experienced presentation format was free from the burdens

that delays place on cognitive resources in real-time causal induction. Yet, participants still took note of the temporal distribution of effects in making their causal judgments; merely advancing (or postponing) the time of the effect occurrence was sufficient to generate causal (or preventive) impressions. This seemingly conflicts with our finding here that delays do not matter when contingency structure is apparent, because structure was maximally apparent in Greville and Buehner's (2007) stimuli. How, then, could temporal information still influence judgments? A key consideration is the type of response solicited from participants. Greville and Buehner (2007) asked participants how *effective* the particular treatment was at killing bacteria, rather than *whether* it killed off the bacteria. Participants were thus directed to focus on causal efficacy rather than pure contingency. Here, on the other hand, we specifically asked participants only to consider the effect of their action on *whether* or not the triangle lit up—by implication, ignoring *when*, and thus focusing on contingency. While such a distinction may seem trivial, it can have a profound influence on reported scores (Buehner, Cheng, & Clifford, 2003; Tenenbaum & Griffiths, 2001). To illustrate, if a hypothetical patient who suffers from migraines tries two different medications, both of which provide relief from headaches with equal reliability, whichever produces the effect fastest is likely to be preferred, and to be considered more causally effective, despite the fact that the patient is quite able to attribute the effect of pain relief to both medications. Because our experiment did not involve considerations of the utility associated with faster delivery of the effect, our participants based their responses purely on the contingency that they perceived.

In summary, our research indicates the potential of stimulus cues in the environment to reveal some hidden structure that governs the time frame linking causes to their effects. Such cues appear to assist causal attribution by creating an awareness of trial structure, without necessitating an intermediary step of altering response frequency or timing. When structure is maximally apparent, the detrimental effect of delay is eliminated completely, and delays no longer adversely affect causal judgments. In other words, the results support a structural-awareness hypothesis. Models of conditioning such as SET, which acknowledge that animals may acquire representational knowledge of temporal intervals, could also be applied in order to account for the effects of trial structure.

Appendix: General instructions for all experiments

Your task in this experiment is to judge the extent to which you can cause something to happen on the computer screen. There will be a triangle on the screen. Below it you will see

a button which you may press at any time and as often as you like. Your task is to find out whether pressing the button has any effect on whether or not the triangle lights up. You will work on four separate problems, each lasting a few minutes. In each, the triangle will light up of its own accord some of the time. In addition to this, pressing the button may also cause the triangle to light up, may prevent the triangle from lighting up or it may have no effect at all. You can press the button as often or as rarely as you like. However, because of the nature of the task, it is to your advantage to press it some of the time and not to press it some of the time. The relation between pressing the button and whether or not the triangle lights up will be constant within each problem, but may well differ between problems.

References

- Ahn, W.-K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, 54, 299–352. doi:10.1016/0010-0277(94)00640-7
- Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, 114, 435–448.
- Arcediano, F., & Miller, R. R. (2002). Some constraints for models of timing: A temporal coding hypothesis perspective. *Learning and Motivation*, 33, 105–123.
- 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.
- 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, 1119–1140.
- Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking and Reasoning*, 8, 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, 865–890. doi:10.1080/02724980244000675
- Buehner, M. J., & May, J. (2004). Abolishing the effect of reinforcement delay on human causal learning. *Quarterly Journal of Experimental Psychology*, 57B, 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, 42–80.
- Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking and Reasoning*, 12, 353–378.
- Chatlosh, D. L., Neunaber, D. J., & Wasserman, E. A. (1985). Response–outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation*, 16, 1–34. doi:10.1016/0023-9690(85)90002-5
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405.
- Dickinson, A. (2001a). Causal learning: An associative analysis. *Quarterly Journal of Experimental Psychology*, 54B, 3–25.
- Dickinson, A. (2001b). Causal learning: Association versus computation. *Current Directions in Psychological Science*, 10, 127–132.

- Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(5, Whole No. 270).
- Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3–19. doi:10.1037/0033-2909.99.1.3
- Gallistel, C. R., & Gibbon, J. (2000). *The symbolic foundations of conditioned behavior*. Mahwah, NJ: Erlbaum.
- Garcia, J., Ervin, F. R., & Koelling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science*, 5, 121–122.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325. doi:10.1037/0033-295X.84.3.279
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306–355.
- Greville, W. J., & Buehner, M. J. (2007). The influence of temporal distributions on causal induction from tabular data. *Memory & Cognition*, 35, 444–453.
- Greville, W. J., & Buehner, M. J. (2010). Temporal predictability facilitates causal learning. *Journal of Experimental Psychology: General*, 139, 756–771.
- Grice, G. R. (1948). The relation of secondary reinforcement to delayed reward in visual discrimination learning. *Journal of Experimental Psychology*, 38, 1–16.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 334–384. doi:10.1016/j.cogpsych.2005.05.004
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116, 661–716. doi:10.1037/a0017201
- Hammond, L. J. (1980). The effect of contingency upon the appetitive conditioning of free-operant behavior. *Journal of the Experimental Analysis of Behavior*, 34, 297–304. doi:10.1901/jeab.1980.34-297
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62, 135–163. doi:10.1146/annurev.psych.121208.131634
- Hume, D. (1888). A treatise of human nature. In L. A. Selby-Bigge (Ed.), *Hume's treatise of human nature*. Oxford, UK: Clarendon Press. Original work published 1739.
- Jenkins, H., & Ward, W. (1965). Judgment of contingencies between responses and outcomes. *Psychological Monographs*, 7, 1–17.
- Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, 115, 955–984. doi:10.1037/a0013256
- Reed, P. (1992). Effect of a signaled delay between an action and outcome on human judgment of causality. *Quarterly Journal of Experimental Psychology*, 44B, 81–100.
- 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, 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 research and theory* (pp. 64–99). New York, NY: 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, 139–159.
- Solomon, P. R., & Groccia-Ellison, M. E. (1996). Classic conditioning in aged rabbits: Delay, trace, and long-delay conditioning. *Behavioral Neuroscience*, 110, 427–435.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13, pp. 59–65). Cambridge, MA: MIT Press.
- Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response–outcome contingencies under free-operant procedures. *Learning and Motivation*, 14, 406–432.
- Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior*, 46, 15–35.
- Wolfe, J. B. (1921). The effect of delayed reward upon learning in the white rat. *Journal of Comparative Psychology*, 17, 1–21.