Title
A Causal Power Approach to Learning with Rates

Permalink
https://escholarship.org/uc/item/0168n34d

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN
1069-7977

Author
Buehner, Marc J.

Publication Date
2006

Peer reviewed

A Causal Power Approach to Learning with Rates

Marc J Buehner (BuehnerM@Cardiff.ac.uk)
School of Psychology, Cardiff University,
Tower Building, Park Place, Cardiff, CF10 3 AT, Wales, UK

Abstract

Current models of causal induction are seriously compromised because they cannot represent variations in cause-effect timing. Theoretical considerations and empirical evidence converge, suggesting that cause-effect timing influences induction beyond mere interference, in line with predictions of psychophysical models of rate comparison. Rather than accepting two distinct cognitive processes for causal induction from rate vs. probability data, this paper shows that a current normative theory of probabilistic causality (Cheng, 1997) can be extended to encompass rate data. Causal induction in "experienced vs. described" situations (Shanks, 1991) may be rooted in a unified process after all.

Keywords: Causal Learning; Rates; Induction; Time;

Introduction

Cognitive Science has inherited David Hume's (1739/1888) approach to causal inference: causal knowledge (and with it the capacity to predict and control our environment) has to be derived from non-causal input to our sensory system. Hume has identified three core principles which need to be present in order to license causal conjecture: i) temporal priority of the cause c before the effect e, ii) temporal and spatial contiguity between c and e, and iii) constant conjunction between c and e. Empirical and modeling efforts within cognitive science have taken the first two principles largely for granted or self-evident, and have mainly focused on the last principle, also referred to as *contingency*. More specifically, how contingency gives rise to causal impressions has been the subject of a hot debate in the field. Early suggestions (e.g. Allan & Jenkins, 1980) to use contingency (ΔP) - calculated by the difference between the two conditional probabilities: $P(e|c)-P(e|\neg c)$ – as a direct measure of causal strength were followed by more sophisticated judgment rules (e.g. Anderson & Sheu, 1995; White, 2003). An alternative suggestion (Shanks & Dickinson, 1987) was that causal learning recruits the same principles thought to underlie associative learning and Pavlovian conditioning (e.g. Rescorla & Wagner, 1972). Debates between these various accounts were mostly carried out by means of model-fitting, with models having the highest number of free parameters often ending up as the "better" models, because they could account for a wider range of empirical findings.

More recently, however, the focus has shifted towards a normative *understanding* of the inductive problem: what is the *goal* of causal inference? Cheng (1997) has drawn a parallel between perception (Marr, 1982) and causal inference. Just as the goal in perception is to appreciate features of the distal world through an analysis of the proximal stimulus on the retina, the goal of causal inference is to infer distal, unobservable causal powers by means of analyzing the proximal stimulus (observable covariation patterns). Cheng has shown that all competing approaches (which do not entail the notion of unobservable causal power) fail to represent causal power as a distal variable, unbound (Holyoak & Hummel, 2000) from its observable manifestation via contingency patterns.

A complementary approach to Cheng's power PC theory is a Bayesian structural model of causal induction (Griffiths & Tenenbaum, 2005). This latter model views causal induction as a problem of structural inference: the reasoner's primary goal is to decide *whether* a causal relation between c and e exists. According to this approach, the *strength* of the relation (which is calculated according to the principles outlined in Cheng's theory) is of secondary importance ("Structure before Strength"). Cheng and her colleagues have shown, however, that structural inferences only yield normative solutions if they are grounded in a normative understanding of strength. In other words, even though structural understanding has a stronger rational appeal, it is preceded computationally by a normative understanding of strength (Cheng, Novick, Liljeholm, & Ford, in press).

Time is of the Essence: How covariational approaches are severely limited in their explanatory scope

The debate about the computational basis and goals of causal inference has clouded another fundamental problem of causal inference: how to represent its interaction with time. Early, non-computational psychological theories of causal induction (Einhorn & Hogarth, 1986; Young, 1995) have recognized the role of cause-effect timing: everything else being equal, contiguous event sequences have a stronger causal appeal than delayed ones, in line with Hume's second principle (but see Buehner & May, 2002 for top-down malleability of this principle).

A simple explanation of the deleterious influence of cause-effect delays on causal induction could be that delayed regularities are harder to detect than immediate ones (e.g. Buehner & May, 2003). Events need to be kept in memory for longer, intervening events open the possibility for multiple regularities that need to be compared against each other, etc. Given that our

computational resources are limited, it is easy to see how delayed relations could appear less causal than immediate ones. Note that this explanation offers a rationale of why reasoners might deviate from the *normative* perspective of causal power, based on a contingency view of causation. In other words, the normative response to a delayed relation would be to view it as just as causal as an equivalent immediate relation, but computational complexities introduced through the delay might interfere with this assessment, producing a non-normative, weaker judgment.

An alternative approach to the influence of cause-effect timing would be to accept that regularity is not the sole determinant of causal strength, but that time, even from a normative perspective, carries causal information. That the role of time in causal inference might not be limited to interference, but instead could also be shaping the normative quality of the inference can best be illustrated with an example: Imagine that George suffers from recurring bouts of headaches. The headaches usually start in the morning, and tend to recede around noon. If George takes Aspirin, however, he already feels better by mid-morning. Can Aspirin be said to relieve George's headaches? A contingency-based assessment of the situation reveals two different solutions, depending on the size of the temporal window that is used to carve the continuous stream of events into discrete pieces of evidence that can be fed into a contingency-based model: If the window spans the entire day, then Aspirin would be seen as ineffective in relieving George's headaches: His headaches are just as likely to disappear on days when he takes Aspirin than when he does not. However, if the temporal window is narrower, spanning only a few hours, then Aspirin clearly offers relief: George's headaches go away after a few hours if he takes Aspirin, but if he does not, he continues to suffer for a few more hours.

Hagmayer and Waldmann (2002) have shown that reasoners indeed interpret the very same statistical information differently, depending on the size of the temporal window they think is appropriate for the causal relation in question. Hagmayer and Waldmann's (2002) results are evidence for an interaction between top-down influences of prior knowledge (see also Buehner & May, 2002; Buehner & McGregor, in press) and event timing. More specifically, Hagmayer and Waldmann deliberately created their materials to contain ambiguous statistical information: the quantitiy and sign of the cause-effect contingency was dependant on the timeframe over which it was calculated.

However, time could be a carrier of information on its own, independent of specific top-down assumptions about the timeframe of the causal relation in question. More specifically, consider situations where contingency information is unambiguously available to the reasoner, as is the case in accumulated longitudinal epidemiological data. Such data contains information not only about *whether* the effect occurred in an individual, but also

when it occurred. Using unambiguous tabular data, Greville and Buehner (in press) have recently demonstrated that contingency and contiguity interact in shaping causal inference. The remainder of this paper presents a novel analysis of this data, not contained in the forthcoming original report[1].

Greville and Buehner's (in press) Data
Participants in Greville and Buehner's (in press) study were presented with tables containing information about the occurrence of an effect in an experimental group (where the cause had been administered once on day 0) and in a control group. Each row represented an individual, and each column represented a day in the five-day period of a hypothetical study. Occurrence of the effect was marked with an X in the appropriate cell. The difference between the total numbers of Xs in the experimental and the control table thus allowed an easy calculation of the cause-effect contingency. In addition, the location of the Xs within a row (days 1-5) contained temporal information: whether the effect occurred close to the administration of the cause on day 0, or further away from it. Variation in the frequency *distribution* (while leaving the frequency itself constant) thus allowed a manipulation of cause-effect contiguity, while contingency remained at a constant, unambiguous value. Figure 1 shows an excerpt of a sample stimulus used in these studies. It is evident that a) contingency is clearly conveyed by the number of Xs in each table, and that b) peaks in frequency distribution near or far from day 1 in the experimental group convey strong or weak cause-effect contiguity, respectively.

Results showed that participants took both contingency and contiguity into account when making causal inferences, such that identical contingencies were attached with higher causal effectiveness when frequency distributions peaked near the beginning of the study than when they peaked near its end (see Table 1). Moreover, the mere advancing or postponing of the effect in time was attached with causal significance, even when the cause did not increase the overall probability of the effect. In other words, zero-contingencies were interpreted as indicative of a generative or preventive influence, depending on the frequency distribution. They were judged as non-causal only when the distribution of effects was random in both the experimental and the control group.

The influence of temporal distributions in non-contingent conditions is particularly interesting when compared to normative accounts of causation. Any normative model (e.g. see Cheng, 1997) postulates that the absence of contingency signals the absence of causation (bar a few exceptions concerning ceiling effects). On a probabilistic level, lack of contingency implies that the cause makes no difference to the occurrence of the effect: the effect is just as likely when

[1] An online version of the empirical report can be found at http://www.cardiff.ac.uk/psych/home/buehnerm/pubs/index.html

the cause is present as when it is absent. In non-contingent conditions involving random distributions of events over time participants followed this normative principle, and correctly inferred that the cause made no difference to the occurrence of the effect. When the temporal distribution of effects in the experimental groups of non-contingent conditions had a discernible peak either near or far from the administration of the cause, however, participants did not think that the cause made no difference to the occurrence of the effect, even though, overall, the effect occurred just as often in the experimental as in the control group. The mere temporal regularity, i.e. that effects tended to cluster soon or late after administration of the cause, was sufficient to create impressions of generative or preventive causality, even though there was no statistical regularity, at least not when considered over the entire data range of five days.

Figure 1. Sample stimulus materials used by Greville and Buehner (in press).

PG13 Bacteria exposed to **Gamma** treatment on day 0

	Day 1	Day 2	Day 3	Day 4	Day 5
1			x		
2	x				
3		x			
4		x			
5			x		
6		x			
7	x				
8			x		

PG13 bacteria with **no exposure** to any radiation treatment

	Day 1	Day 2	Day 3	Day 4	Day 5
1				x	
2					
3		x			
4					x
5					
6	x				
7					
8					

Table 1. Design and Results from Greville and Buehner (in press). Strong and Weak contiguity was implemented by event distributions of data pertinent to P(e|c) peaking, respectively, near day 1 or 5 of the experimental period. Data for P(e|¬c) was always randomly distributed over the 5 day period. Participants provided causal ratings on a scale ranging from -100 to +100, where -100 meant *c* strongly promotes *e*, 0 meant that *c* has no effect on *e*, and +100 meant *c* strongly prevents *e*

Scenario	P(e|c)	P(e|¬c)	Delta-P	Contiguity	Exp 1 (N=38)		Exp 2 (N= 35)		R(e|c)	R(e|¬c)	power
					M	S.D.	M	S.D.			
A	0.25	0.25	0.00	Strong	-9.18	31.42	-1.26	35.19	40	30	0.06
B	0.50	0.25	0.25	Strong	-27.13	33.50	-46.60	30.15	80	28	0.30
C	0.75	0.25	0.50	Strong	-47.55	35.02	-79.43	13.22	120	32	0.52
D	1.00	0.25	0.75	Strong	-67.50	30.49	-89.63	14.03	160	26	0.77
E	0.25	0.25	0.00	Weak	18.26	33.48	21.77	31.40	20	32	-0.38
F	0.50	0.25	0.25	Weak	-8.84	39.72	-10.86	38.26	40	28	0.07
G	0.75	0.25	0.50	Weak	-7.20	53.38	-33.26	35.79	60	28	0.19
H	1.00	0.25	0.75	Weak	-26.84	55.76	-53.46	40.66	80	31	0.29
I	0.50	0.50	0.00	Strong	-8.82	25.36	-25.14	24.06	80	55	0.17
J	0.75	0.75	0.00	Strong	-26.82	36.89	-36.49	24.78	120	90	0.27
K	1.00	1.00	0.00	Strong	-27.34	34.65	-45.46	32.60	160	120	0.50
L	0.50	0.50	0.00	Weak	15.74	23.93	13.29	16.04	40	60	-0.33
M	0.75	0.75	0.00	Weak	22.18	36.46	23.49	22.38	60	90	-0.33
N	1.00	1.00	0.00	Weak	22.95	41.95	26.26	22.53	80	120	-0.33
O	0.25	0.25	0.00	Random	2.89	27.20	5.91	21.44	30	30	0.00
P	0.50	0.50	0.00	Random	1.42	14.75	1.57	21.69	60	60	0.00
Q	0.75	0.75	0.00	Random	-1.97	16.71	-9.14	10.47	90	90	0.00
R	1.00	1.00	0.00	Random	-1.18	16.25	-7.86	12.14	120	120	0.00

Rate-based Accounts of Causal Influence

A qualitatively similar finding regarding temporally advanced or postponed effects that would occur anyway has been reported by Anderson and Sheu (1995) and Wasserman and Neunaber (1986). Both studies employed free operant procedures, and in both pressing a button did not actually change the probability of the effect, but merely advanced or postponed its delivery. As in Greville and Buehner's (in press) studies, this was sufficient to create impressions of generative or preventive causal power. One important difference between these studies and ours, however, is that the nature of the free-operant task by definition makes the calculation of contingency very difficult, and in fact dependent on the (subjective) length of the learning trial.

Anderson and Sheu's Experiment 4, for instance, employed learning trials of 1s, 2s, and 4s for trials when participants did not press the button; trials on which participants did press the button could be 250ms, 500ms, 1s, 2s, 4s, and 8s. Regardless of trial length, and regardless of whether or not the participant pressed the button, an effect was delivered at the end of every trial. The effect thus was just as likely when participants pressed the button than when they did not press the button, i.e. P(e|c) = P(e|¬c) = 1.0, ΔP = 0. In other words, the effect happened anyway (at a steady rate, e.g. once every 2 seconds), but if participants pressed the button

they could advance or postpone its delivery (depending on the specific combination of trial lengths). Moreover, since the total time spent in each condition was constant, pressing the button also increased or decreased the total overall number of effects accordingly. Higher ratings for more contiguous conditions (i.e. conditions where trials on which the participant responded were shorter than trials on which no response was made) in Anderson & Sheu's studies thus were entirely normative, even from a purely statistical perspective.

On a probabilistic (or contingency) level, of course, pressing the button made no difference, since every trial saw the delivery of an effect irrespective of whether participants pressed the button. Contingency accounts thus fail to accommodate Anderson and Sheu's or similar results. Such models can only represent the relative difference in outcomes, depending on the presence and absence of the cause, which – when ignoring trial length considerations – was zero in all conditions of Anderson & Sheu's Experiment 4.

Anderson and Sheu suggested a rate-contrast model

$$G = \frac{R(O\,|\,R) - R(O\,|\,\neg R)}{R(O\,|\,R) + R(O\,|\,\neg R)}$$

where R(O|R) and R(O|¬R) are the rates of the outcome occurrence given the presence and absence of a response, respectively, to account for their results. Because rates are calculated per unit of time, they are of course sensitive to variations of trial length, which are outside the scope of any contingency model to date. The grating contrast model provided an excellent fit to Anderson & Sheu's data, and accounted for over 90% of the variance.

Extending Cheng's power PC theory to Rate Data

According to Cheng's (1997) power PC theory, the causal power q of a cause c to produce an effect e and the power p to prevent e are given by

$$q = \frac{P(e\,|\,c) - P(e\,|\,\neg c)}{1 - P(e\,|\,\neg c)} \qquad p = -\frac{P(e\,|\,c) - P(e\,|\,\neg c)}{P(e\,|\,\neg c)}$$

This means that ΔP is normalized by 1 (the maximum probability of the effect) minus the base-rate (the probability that the effect would occur anyway) for generative causal power, and simply by the base rate for preventive power. Which equation applies is readily determined by the sign of the contingency.

Most studies involving causal learning in continuous time have no upper limit of outcome rate. In the absence of an upper limit, an extension of Cheng's (1997) theory towards such data is not straightforward, as rates cannot be compared and normalized in the same way as probabilities. More specifically, the causal power approach postulates that generative and preventive causes respectively increase and decrease the occurrence of the effect *by some proportion of the distance to the upper and lower boundary*. Consider a preventive power of .5. This value means that, everything else being equal, the introduction of the preventive cause c reduces the occurrence of the effect by one half. For example, if e occurs with probability 1.0 in the control group, adding c will reduce this probability to .5; if e occurs with probability of .5 in the control group, adding c will reduce this probability to .25. Note that the lower bound used to calculate causal effectiveness here is 0: e cannot happen less than not at all. Thus, a preventive power of 1.0 means that c prevents e everytime (regardless of the value of P(e|¬c)), which is captured in the power equations: whenever P(e|c) = 0, preventive power p will be 1.0.

Analogously, a generative cause will increase the occurrence of the effect relative to the upper limit. If generative power is .5 and e never occurs in the control group, adding c will increase P(e|c) to .5; if e already occurs with P(e|¬c) = .5, then the remaining portion of entities (1.0 - .5 = .5) in the sample will be affected by c, i.e. P(e|c) will increase by another .25 to a total of .75. The upper limit of 1.0 means that the effect cannot occur more than once in every entity. Thus, a generative power of 1.0 means that c produces e everytime it has a chance to do so, which is likewise captured in the power equations: whenever P(e|c) = 1, q will be 1 also[2].

While the lower bound of 0 is preserved for causal relations expressed as rates (the effect cannot occur less than 0 times per unit of time), the upper bound is not. An effect could easily occur more than once per unit of time. Thus, a generative causal interpretation of rates, which entails reasoning about a *proportional* rather than an *additive* increase, is only possible when there is an upper bound.

Note that Greville and Buehner's (in press) design employed a clear upper limit: The highest rate *Rmax* that was possible was that all 40 samples showed the effect on day 1. A cause could thus show its effectiveness (to increase the rate of effect occurrence) in two ways: a) by increasing the overall frequency of the effect in the 5 day period, and b) by advancing in time the occurrence of effects that would otherwise have occurred later in the period. How could one quantify rates in Greville & Buehner's design? One way is to consider the number of days an individual effect could be present; if an effect happens on day 1, it will be present and noticeable for all 5 days of the period, while an effect occurring on day 5 will only be present for 1 day. Thus, the maximum number of "effect-days" in this design is 40 effects x 5 days = 200 effect-days. It is straightforward to see how a number of effect-days can easily be calculated for each condition of Greville & Buehner, simply by considering the actual event frequency distributions contained in the original report (see Footnote 1). Multiplying quantity with time of course differs from the standard concept of rate (*quantity / time)*, but a ratio could easily be obtained (effect-days / observation period) without changing the outcome of the analysis to come. Because of the clear definition of an upper limit (*Rmax*=200), causal effectiveness in Greville & Buehner's design can be

---

[2] The exceptions to both cases are ceiling effects and their preventive analogs, where causal power is undefined, see Wu & Cheng (1999).

expressed proportionally, thus licensing the application of Cheng's power PC theory:

$$q_r = \frac{R(e \mid c) - R(e \mid \neg c)}{R_{max} - R(e \mid \neg c)} \qquad p_r = -\frac{R(e \mid c) - R(e \mid \neg c)}{R(e \mid \neg c)}$$

Which equation applies can be determined by the sign of the rate contrast. Table 1 lists the *power* values obtained via the above calculations. Note that in conditions E, L, M, and N, the rate contrast is negative, although the contingency is zero, and consequently p_r applies. In line with conventions, preventive estimates are expressed as negative numbers in Table 1. Causal power calculated this way fits the data from both experiments extremely well; the extension of the power PC theory to rates can account for 95% of the variance in Experiment 1 and 92% of the variance in Experiment 2.

Representing time of effect occurrence via multiplication with units of time of course corresponds to simple linear weighting by a negative function, with the impact of effects decaying, the further away in time they are from the cause. Note, however, that this weighting was achieved without any free parameters. Instead, the weighting was obtained simply by considering the maximum impact an effect could have, or, to express it differently, by taking into account the maximum possible distance between *c* and *e* (in this case 5 days). If we limit ourselves to situations where a given cause can only produce one effect (as opposed to multiple instantiations of the same effect), we thus can easily calculate *Rmax* for any paradigm, as soon as we know the maximum temporal distance between *c* and *e*. In many cases, this information will be available via prior knowledge, but it can of course also be observed empirically.

In Anderson & Sheu's Experiment 4, for example, the maximum temporal distance between *c* and *e* was 8s. Continuing the logic outlined above, we can apply a weight of 1 to an effect occurring at 8s, a weight of 8 to an effect occurring at 1s, with the maximum possible weight of 9 applied to an effect occurring immediately (the shortest interval participants experienced in Anderson & Sheu's design was only 250ms, corresponding to a weight of 8.75). Thus, the observed maximum possible distance between *c* and *e* allows an observer to calculate increases of effect occurrence proportional to a maximum effectiveness (i.e. instantaneous effect delivery), licensing a causal power interpretation, again without recourse to any free parameters. Causal power calculated according to these principles can account for 94 % of the variance in Anderson & Sheu's Experiment 3, and 91% of the variance in Experiment 4, a fit comparable to *G* (91% and 94%, respectively).

Described versus Experienced Events: One or Two Cognitive Architectures?

Studies involving summary data (such as Hagmayer and Waldmann, 2002; or Greville and Buehner, in press) are sometimes criticized for lacking ecological validity.

More specifically, Shanks (1991) proposed that causal induction from described events is based on fundamentally different processes than the ones involved in causal learning in real time. It is certainly true that some approaches to causal learning in continuous time (for example associative learning theory) cannot be applied to summary data. Other theories, however, have a considerably wider explanatory scope. Contingency and probability based theories, or statistical approaches in general, are largely agnostic to the stimulus format they require as input. It does not matter in principle whether estimates of P(e|c) and P(e|¬c) have to be gleaned from discrete learning trials presented one-by-one, in a list, or whether these values are directly provided (the more difficult the assessment of probabilities, though, the larger the scope for error, and thus noise and bias in the data, see Buehner et al., 2003).

It would seem unparsimonious to propose two distinct cognitive architectures for doing exactly the same task, particularly because we often and routinely switch between both modes of learning (experienced vs. described) for the same problem. An epidemiologist, for instance, might build up a causal model of a disease, based on statistical records from past years (described), and update this model in the light of new data on a case-by-case basis. Prior research converges to show that a normative, probabilistic approach to causality provides a better model of causal induction from discrete events and summary data than competing approaches such as associative learning or decision rules (Buehner et al., 2003; Cheng, 1997; Wu & Cheng, 1999). Greville and Buehner's (in press) data show, however, that probability-based approaches to causal induction are seriously compromised because of their inability to represent variations in contiguity. Rate-based approaches on the other hand, by definition take both contingency and contiguity into account, and so far have provided a good fit to data obtained from continuous paradigms where both factors varied. Rather than proposing two separate cognitive architectures for doing essentially the same task with slightly different input, the analysis offered here suggests that, under certain boundary conditions, rate-data, just like probabilistic or frequency data can be interpreted in terms of computational causal power. These boundary conditions include knowledge of the maximum outcome rate (i.e. maximum causal efficiency) against which a candidate cause should be compared.

Conclusion

Temporal spacing influences causal learning and inference in principled ways, going beyond mere interference. Thus, a comprehensive, ecologically sound approach to causal learning has to take into account temporal spacing in addition to regularity. Current normative theories of causal induction cannot represent temporal information, however. Rather than proposing separate cognitive architectures for causal induction involving discrete, probabilistic versus continuous, rate-

based data, the dilemma could be overcome by extending the normative account (Cheng, 1997) to also include rate-based data. This extension is only possible when one additional constraint is met: that there is a (known) upper limit of the rate with which e can occur. When this constraint is met, Cheng's equations can be applied to rate data.

Previous research on cause-effect timing has shown that the otherwise deleterious impact of temporal delays can be overcome by prior knowledge of potential delays (Buehner & May, 2002; 2003); given appropriate expectations about timing, delays can even facilitate causal inference, while immediacy can impair it (Buehner & McGregor, in press). These top-down influences on causal event parsing could easily be implemented in a causal power framework for rates. The linear decay function would simply be reversed (representing a deleterious influence of immediacy), or replaced with a uniform function (representing no impact of delay). How such functions can be acquired in a bottom-up manner would need to be determined by future research.

Anderson and Sheu (1995) reported that participants in their studies commented on the "perceptual quality" of the c-e pairings they experienced, and concluded that causal inference from real time data is best represented as a perceptual process, governed by the laws of psychophysics. Anderson and Sheu might have, unknowingly, prefigured Cheng's analogy between perception and causal inference. Causal induction, like perception, is not infallible. If certain constraints are not met in the environment, we cannot go beyond tracking features of the proximal stimulus, and thus fail to achieve a proper representation of the distal stimulus. Crucially, though, the apparatus is equipped to handle input in various formats or textures, as would be expected from an adaptive system.

Acknowledgments

References

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

Anderson, J. R., & Sheu, C. F. (1995). Causal inferences as perceptual judgments. *Memory and Cognition, 23*(4), 510-524.

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

Buehner, M. J., & May, J. (2003). Rethinking Temporal contiguity and the judgment of causality:Effects of Prior Knowledge, Experience, and Reinforcement Procedure. *Quarterly Journal of Experimental Psychology, 56A*(5), 865-890.

Buehner, M. J., & McGregor, S. (in press). Temporal Delays can Facilitate Causal Attribution: Towards a General Timeframe Bias in Causal Induction. *Thinking and Reasoning*.

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

Cheng, P. W., Novick, L. R., Liljeholm, M., & Ford, C. (in press). Explaining four psychological asymmetries in causal reasoning: Implications of causal assumptions for coherence. In M. O'Rourke (Ed.), *Topics in contemporary philosophy (Volume 4): Explanation and Causation*. Cambridge, MA: MIT Press.

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

Greville, W. J., & Buehner, M. J. (in press). The Influence of Temporal Distributions on Causal Induction from Tabular Data. *Memory & Cognition*.

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

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

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. Markman (Eds.), *Cognitive Dynamics: Conceptual change in humans and machines.* (pp. 229-263). Mahwah, NJ: Erlbaum.

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

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

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

Shanks, D. R. (1991). On Similarities Between Causal Judgments in Experienced and Described Situations. *Psychological Science, 2*(5), 341-350.

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

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 710-727.

Young, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin and Review, 2*(1), 83-104.