In M. Knauf, M. Pauven, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1157-1162). Austin, TX: Cognitive Science Society. (2013)

Generic Priors Yield Competition Between Independently-Occurring Causes

Derek Powell[1] (derekpowell@ucla.edu)

M. Alice Merrick[1] (m.a.merrick@gmail.com)

Hongjing Lu[1,2] (hongjing@ucla.edu)

Keith J. Holyoak[1] (holyoak@lifesci.ucla.edu)
Departments of Psychology[1] and Statistics[2], University of California, Los Angeles
Los Angeles, CA, USA

Abstract

Recent work on causal learning has investigated the possible role of generic priors in guiding human judgments of causal strength. One proposal has been that people have a preference for causes that are *sparse and strong*—i.e., few in number and individually strong (Lu et al., 2008). Evidence for the use of sparse-and-strong priors has been obtained using a maximally simple causal set-up (a single candidate cause plus unobserved background causes). Here we examine the possible impact of generic priors in more complex, multi-causal set-ups. Sparse-and-strong priors predict that competition can be observed between candidate causes even if they occur independently (i.e., the estimated strength of cause A will be lower if the strength of uncorrelated cause B is high rather than low). Experiment 1 revealed such a cue competition effect in judgments of causal strength. Experiment 2 showed that, as predicted by a Bayesian learning model with sparse-and-strong priors, the impact of the prior diminishes as sample size increases. These findings support the importance of a preference for parsimony as a constraint on causal learning.

Keywords: causal learning; generic priors; causal strength; parsimony; Bayesian modeling

Introduction

Prior Beliefs in Causal Learning

Humans (and other intelligent organisms) are able to extract causal knowledge from patterns of covariation among cues and outcomes. Viewed from a Bayesian perspective, causal inferences are expected to be a joint function of likelihoods (the probability of observing the data given potential causal links of various possible strengths) and priors (expectations about causal links that the learner brings to the task). For relatively simple causal set-ups involving binary variables, human causal judgments can be described quite accurately by the power PC theory (Cheng, 1997), which uses a noisy-OR likelihood function to integrate the influences of multiple generative causes (Griffiths & Tenenbaum, 2005; Lu et al., 2008; see Holyoak & Cheng, 2011, for a review).

Prior beliefs about causal relationships can also be formulated within a Bayesian framework for causal learning. Generic causal priors can be thought of as preferences for certain types of causal explanations, without relying on domain-specific knowledge. Some Bayesian models have assumed uninformative priors (e.g., Griffiths & Tenenbaum, 2005); however, other models have incorporated substantive generic priors about the nature of causes. In particular, Lu et al. (2008) proposed that people have a preference for causes that are *sparse and strong*: i.e. a preference for causal models that include a relatively small number of strong causes (rather than a larger number of weak causes). Such priors can be viewed as a special case of a more general pressure to encourage parsimony (Chater & Vitanyi, 2003), which implies a combination of simplicity and explanatory power. The preference for parsimony has a number of expressions elsewhere in causal learning phenomena and theory. For instance, causal learners appear to make the default assumption that causes act independently in producing an effect, rather than interacting (Cheng, 1997; Novick & Cheng, 2004). Moreover, people generally prefer simpler explanations to equally accurate but more complex explanations (Lombrozo, 2007).

Generic Prior: Sparse-and-Strong (SS) Causes

Lu et al. (2008) formalized the "SS power" model with sparse-and-strong (SS) priors for simple causal models with a single candidate cue and a constantly-present background cause. When the candidate cause generates (rather than prevents) the effect, there is an expectation that the candidate cause is strong (strength close to 1) and the background is weak (strength close to 0), or vice versa. A single free parameter, α, controls the strength of the prior (when $\alpha = 0$, the distribution is uniform).

The possible role of generic priors in causal strength

Table 1. *Contingency learning data for one experimental block (44 trials) by trial type*

Conditions		C	AC	BC	ABC
Weak-B	E absent	1	6	3	7
	E present	10	5	8	4
Strong-B	E absent	1	6	9	10
	E present	10	5	2	1

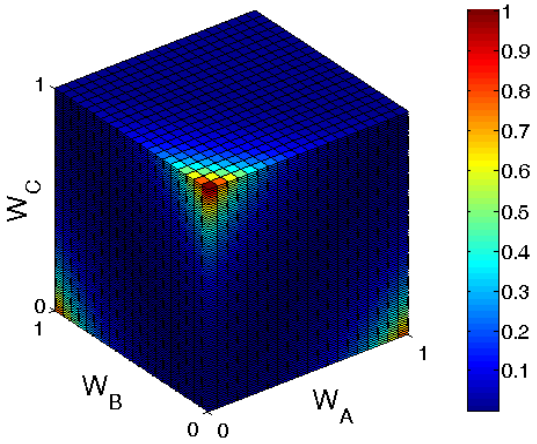

Figure 1: Sparse-and-strong prior distribution over causal strengths of three causes. Colors indicate the values of prior probability (red corresponds to highest probability).

judgments has so far only been examined for very simple causal graphs (e.g., one generative candidate cause and a constantly-present background cause). Lu et al. (2008) fit several causal learning models to parametric data for human strength judgments. They found the best fit was provided by a Bayesian implementation of the power PC theory that incorporated SS priors with an α value of 5 (not 0), implying a human preference for sparse-and-strong causes. When α value is set to zero, the prior distribution would be equivalent to a uniform distribution.

The generalization of the sparse-and-strong prior for more than two causes is straightforward. For the experiments reported here, the SS prior is constructed on the basis of three candidate causes, *A*, *B*, and *C*, which are all generative. The SS prior can be defined as,

$$P(w_A, w_B, w_C) \propto e^{-\alpha(1-w_A)-\alpha w_B-\alpha w_C} + e^{-\alpha w_A-\alpha(1-w_B)-\alpha w_C} + e^{-\alpha w_A-\alpha w_B-\alpha(1-w_C)}. \quad (1)$$

in which *w* denotes causal strength for different causes.

Figure 1 illustrates the sparse-and-strong prior in the three-cause situation. A signature of SS priors is the preference for one strong cause coupled with other weak causes, i.e., a set of "ideal" causal strengths for the three causes might be w_A=1, w_B=0 and w_C=0. This preference instantiated in SS priors implies a key empirical prediction: competition effects in judgments of causal strength when multiple causes co-occur. Strength competition implies that if a candidate cause appears along with another cause of greater strength (as defined by likelihoods), then the strength of the weaker candidate cause will be underestimated. This prediction goes beyond competition effects predicted by the likelihood function alone (i.e., a model assuming uninformative priors). The goal of the present paper is to test this key empirical prediction in a situation requiring inference based on multiple causes.

Competition Between Causes

Various competitive dynamics are commonly observed in causal learning paradigms, including blocking (e.g., Shanks, 1985), overshadowing (e.g., Waldmann, 2001) and un-overshadowing (De Houwer & Becker, 2002). However, in all these paradigms the competition is between cues that co-occur in a systematic way. For example, blocking is typically obtained when cue A is first shown to produce the effect by itself, and then the compound cue AB is introduced and also followed by the effect. From a Bayesian perspective, a lower causal strength judgment for the blocked cue, B, is entirely rational, as the learner has no opportunity to observe what happens when B is presented without A (i.e., there will be greater uncertainty about the strength of B than of A). More generally, Bayesian models with uniform priors can readily account for a wide range of competition effects that involve cues occurring in a correlated fashion (Carroll, Cheng & Lu, in press).

However, sparse-and-strong priors are unique in predicting competition between independently-occurring causes (e.g., the occurrence of cue A is uncorrelated with the occurrence of cue B). We will show simulation results confirming that when alternative causes A and B occur independently, a Bayesian model with uniform priors predicts that judgments of the strength of A will not be influenced by the strength of B, or vice versa (also see Busemeyer, Myung & McDaniel, 1993a). In contrast, an otherwise-identical model incorporating sparse-and-strong priors predicts that early in learning (when the impact of priors is maximal), independently-occurring causes will compete for strength (e.g., the strength of A will be judged to be lower if B is strong rather than weak).

The present experiments include two conditions based on a set of contingency data, *D,* shown in Table 1. The occurrences of causes A and B are independent in both conditions. The causal power of A is held constant across the two conditions (0.5), but the causal power of B varies from one condition (0.2, weak-B condition) to the other (0.8, strong-B condition).

For this set of contingency data the model computes the mean of estimated causal strength derived from the posterior distribution:

$$\overline{w}_A = \int_0^1 w_A P(w_A \mid D)\cdot \quad (2)$$

The posterior distribution $P(w_A \mid D)$ is obtained by applying Bayes rule to combine likelihood function and priors as

$$P(w_A \mid D) = \iint \frac{P(D \mid w_A, w_B, w_C) P(w_A, w_B, w_C)}{P(D)} dw_B\, dw_C. \quad (3)$$

In our simulations, we employed the noisy-OR likelihood function (Cheng, 1997), since binary causes and effects were used in the experiments:

$$P(E = 1|C_A, C_{\boldsymbol{B}}, C_C; w_A, w_{\boldsymbol{B}}, w_C) = 1 - (1 - w_A C_A)(1 - w_{\boldsymbol{B}} C_{\boldsymbol{B}})(1 - w_C C_C) \quad (4)$$

where E and C indicate the presence or absence of effect and causes, respectively.

Figure 2 shows the model predictions for causal strength of A in the two conditions. The Bayesian model with SS prior (center bards in Figure 3) predicts different estimates of w_{A} across conditions due to competition between causes A and B, even though the two cues occur independently. In contrast, a model with uniform prior (right bars in Figure 3) predicts that w_{A} will not vary across the two conditions. The latter simulation result confirms that a Bayesian model with uniform priors does not predict competition between independently-occurring causes when the likelihood function is a noisy-OR, extending the similar negative conclusion for the case in which the likelihood function is linear (Busemeyer et al., 1993a).

Testing these opposing predictions provides a means to discriminate between alternative possible priors for causal inference with multiple cues. The prediction of competition between independently-occurring causes has never been clearly tested. Busemeyer et al. (1993b) reported an experiment that obtained competition between uncorrelated cues, in a paradigm that may have drawn on causal learning mechanisms. However, this competition effect was observed only when participants were informed that the two cues would be of different strengths, one strong and one weak (see their Footnote 5, p. 194). It is possible that this instruction suggested to subjects that the cues were expected to be competitive. In general, relatively few studies of causal learning have used complex causal set-ups involving more than one or two candidate causes. The present experiments were designed to determine whether multiple candidate causes would compete for causal strength, and whether such effects can be modeled by assuming people have priors that causes will be sparse and strong.

Experiment 1

Method

Participants Participants were 90 undergraduate students at the University of California at Los Angeles (UCLA) who participated for class credit (80% female, mean age = 20 years). Half were assigned to the strong-B condition and half to the weak-B condition.

Procedure Participants read a cover story, as follows: "Imagine that you are assisting a doctor at a new island resort. Many of the guests at this new resort have become ill, and you are charged with helping to determine the cause of the illnesses. Every day, at dinner, the resort provides a complimentary salad for its guests. The salads can be made with different exotic vegetables. The salads always have at least one exotic vegetable, and can be ordered with up to three different exotic vegetables. The resort's doctor thinks one or perhaps several of these exotic vegetables may be causing the illness <pictures of three vegetables are shown>. You will be reviewing a number of case files that describe what a guest ate and whether they became sick. After viewing these files you will be asked to give your assessment of which vegetable or vegetables are the culprits. Please pay attention to each case…. When you are done reviewing the cases you will be asked to estimate how many people each vegetable is likely to affect negatively."

These vegetables were labeled A, B, and C, and were shown as photographs of exotic vegetables (see Figure 2, top). These photographs depicted the actual vegetables radicchio, bitter melon, and black garlic. The assignment of vegetables to the labels A, B and C was randomized across participants. During the learning phase, participants viewed "case files" for individual guests, showing which combination of vegetables they had eaten, and whether or not they had fallen ill.

There were four possible combinations of fruits: each guest had either eaten vegetable C alone, vegetables A and C, vegetables B and C, or all three vegetables A, B, and C. These four combinations were presented in equal number, such that A and B both occurred 50 percent of the time, and the correlation between the occurrence of A and B was 0. A total of 44 cases (11 of each type) was the minimum number required to reflect the underlying causal powers in the presented distribution of cause combinations and their associated outcomes.

The numbers of guests who became sick after eating each combination were determined by the causal powers assigned to each vegetable, calculated according to the noisy-OR likelihood function under the default assumption that each cause acts independently to produce the effect (Cheng, 1997). In both conditions, vegetable A was assigned a causal strength of .50, and vegetable C was assigned a causal strength of .10. In the strong-B condition, vegetable B was assigned a causal strength of .80, whereas in the weak-B condition, vegetable B was assigned a causal strength of .20. Cause A was the focus of the study, as we were interested in whether its judged strength would be influenced by the variation in the strength of cause B. Cause C served as an observable "background" cause, as it was shown to be present on every trial. The resulting contingency data is summarized in Table 1.

The 44 cases were presented sequentially in a different random order for each participant. After viewing all 44 learning trials, participants were asked to give a causal strength rating for all three vegetables. Participants were shown a picture of each vegetable along with text that read, "Imagine 100 healthy people ate this vegetable; how many do you estimate would get sick?" Participants then made

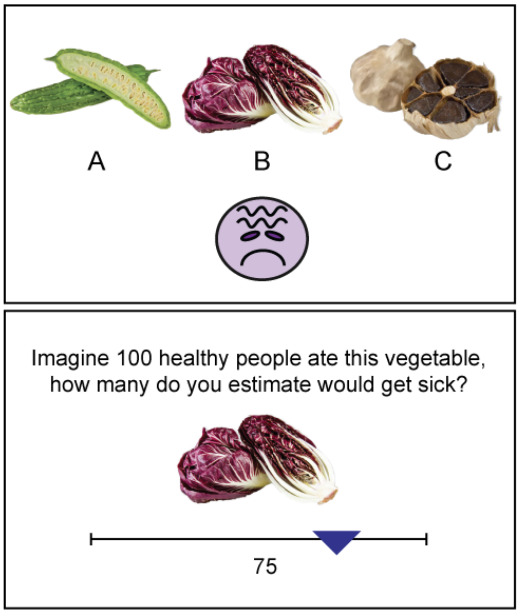
A
B
C
Imagine 100 healthy people ate this vegetable, how many do you estimate would get sick?
75

Figure 2: Example trial showing a guest who ate A, B, and C vegetables and became sick (top). Example response trial (bottom).

their rating using a slider, inputting a value between 0 and 100 (see Figure 3, bottom). The order of the three questions was randomized for each participant. After making all three ratings, participants were shown a summary of their responses and were asked to confirm that they had correctly entered their ratings. Participants were randomly assigned to one of two experimental conditions (weak-B or strong-B).

Results and Discussion

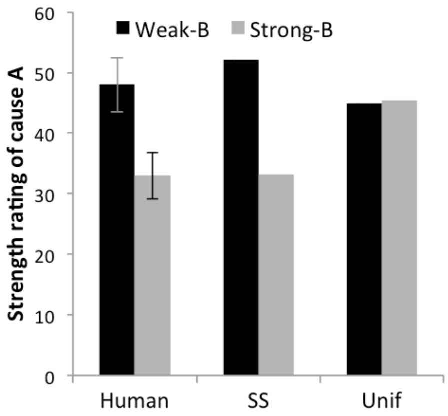

Figure 3: Observed human strength ratings (0-100 scale) and predictions based on sparse-and-strong (SS) versus uniform priors for cause A (0-100 scale) in Experiment 1.

Data from two participants were excluded due to technical issues. Data from another eight were excluded because they entered responses of zero to both cause A and cause B, suggesting errors or a lack of engagement with the task. Figure 4 shows the data for the critical A cue, along with the predictions derived from the SS power model and an otherwise-identical model with uniform priors. Participants in the strong-B condition underestimated the strength of cause A relative to participants in the Weak-B condition (mean of 34.05 versus 46.95), $t(79) = 2.17$, $p < 05$. The data were fit using Lu et al.'s (2008) "SS power" model, which provides a Bayesian formalization of sparse-and-strong priors. For modeling purposes we simply set $\alpha = 5$ (the value estimated for the data sets reported by Lu et al., 2008), thus avoiding any need to fit a free parameter to the present data. The SS power model predicts the observed difference in the judged strength of A in the weak-B versus strong-B conditions, whereas the model with uniform priors does not.

Table 2. Observed human strength ratings (0-100 scale) and predictions based on sparse-and-strong (SS) priors for three different cues in Experiment 1.

	A (.50)		B (.20, .80)		C (.10)	
	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.
Weak-B	52	47	16	34	13	18
Strong-B	35	34	78	63	16	18

Table 2 presents the mean ratings of causal strength obtained for three different cues, and Figure 3 plots the human data with predictions from the two models assuming different priors. Across all cues and conditions, the SS power model provides a good overall fit to the human data ($R = .95$, root mean square deviation, RMSD = 9.1).

Although the overall fit of the SS power model is quite good, it bears noting that the predictions of the SS power model for cue B were more extreme than the estimates given by participants. That is, when B was weak participants overestimated its strength relative to the model with SS priors; when B was strong participants underestimated it relative to SS priors. The estimates of the model using uniform priors deviate from the observed data in a similar (though marginally smaller) fashion. We speculate that these discrepancies may be due to memory limitations. Whereas the models assume perfect memory for contingency data, participants are likely to forget presented instances on some proportion of the trials, and therefore to have greater uncertainty in their strength estimates than predicted by the models. The models' estimates are computed from the mean of the posterior distribution, so increased uncertainty would lead to less extreme strength estimates for cue B (i.e., estimates closer to 50). Uncertainty would be expected to have less impact on estimates for cue A, for which the veridical strength in fact corresponds to a rating of 50.

Experiment 2

It is a natural feature of Bayesian models that the influence of priors diminishes as learners gather more data. Thus, the SS power model (Lu et al., 2008) predicts that competition between causes should be strongest when participants have made few observations, and will diminish as participants are exposed to more data.

Experiment 2 examined competition between causes after varying amounts of experience. The design was identical to that of Experiment 1, but added a second independent variable: sample size. Participants in both strong- and weak-B conditions were asked to make judgments of causal strength three times, after viewing 44, 88 and finally 132 total cases. This resulted in a 2 x 3 factorial design, with one between-subjects factor (causal strength of cue B) and one within-subjects factor (number of cases observed).

The cover story was the same as it in Experiment 1, except for one sentence: "The resort's doctor thinks one or perhaps several of these exotic vegetables may be causing the illness" (Experiment 1) was revised to read, "The resort's doctor thinks these exotic vegetables may be causing the illness."

Method

Participants Participants were 114 UCLA undergraduate students who participated for class credit (76% female, mean age = 20 years).

Procedure Experimental materials were identical to those used in Experiment 1. Participants in Experiment 2 went through three blocks of learning trials, making causal strength estimates after 44, 88 and 132 learning trials. The distribution of types of cases (combinations of causes and outcome) were identical within each block. Order of presentation was randomized for each participant.

Results and Discussion

One participant gave the same response on every trial, and six responded with extreme ratings of 0 or 100 for cause A, or ratings of 100 for Cause C. Data from these seven participants were excluded from analyses.

Figure 4 shows mean causal strength ratings for each vegetable at the end of each of the three learning blocks. A factorial repeated-measures ANOVA found no overall effect of increasing sample size, $F(2, 210) = 2.31$, $p = .10$, or of condition, $F(1, 105) = 1.16$, $p = .28$). However, the analysis revealed a significant interaction between condition and learning block, $F(2, 210) = 5.61$, $p < .01$. As shown in Figure 4, Experiment 2 replicated the competition effect observed in Experiment 1 after 44 trials. After the first block, participants underestimated the strength of A when it was paired with a strong B cause, relative to when it was paired with a weak B cause (means of 56 versus 46; $t(105) = 2.26$, $p < .05$). As predicted, this difference disappeared with an increase in sample size, supporting the hypothesis that the observed competition effect is due to people's priors. The effect of the strength of B on ratings of A was not significant after 88 or 132 trials, $t(105) = 0.75$, $p = .46$, and $t(105) = -0.26$, $p = .79$, respectively. Assuming $\alpha = 5$ as before, the SS power model (Lu et al., 2008) provides a good fit to the human data across all cues and conditions ($R = .96$, RMSD = 12.87). For cause A, the human data for the weak-B and strong-B conditions converge on the veridical value (50) more quickly than does the model's predictions (see Figure 4), perhaps reflecting the additional uncertainty participants experienced due to their fallible memory for the observations.

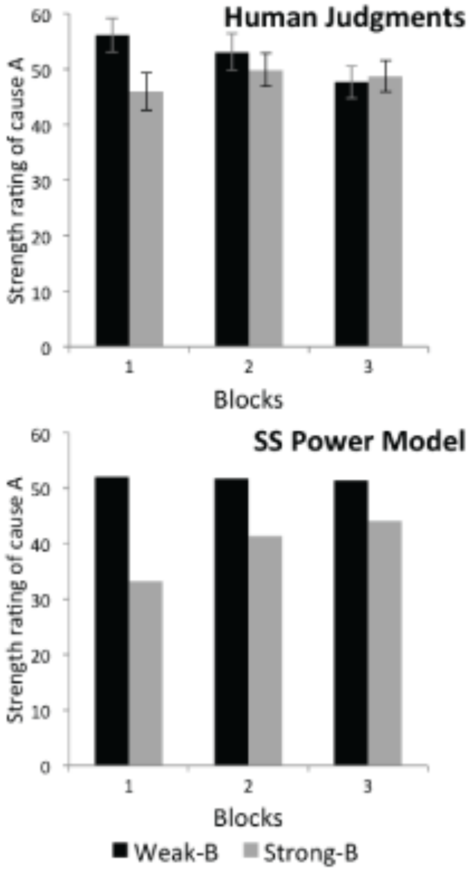

Figure 4. Observed human strength ratings (0-100 scale, top) and predictions of SS power model (bottom) for cause A across blocks in Experiment 2.

Causal strength estimates for all three vegetables were somewhat higher in Experiment 2 than Experiment 1. This difference may be due to a small change in instructions, which in Experiment 2 emphasized the doctor's belief that the vegetables were indeed causing the illness.

General Discussion

The experiments reported here provide evidence for competition between independently-occurring causes in causal strength judgments, as predicted by a Bayesian model of causal learning that assumes sparse-and-strong priors. After participants had made a relatively small number of observations, a cause of moderate strength was judged to be weaker when a competing (but uncorrelated) cause was strong than when the competing cause was weak. After additional cases were presented, the two conditions converged. This competition dynamic cannot be explained by naïve Bayesian models that assume uninformative priors (Busemeyer et al., 1993a), nor can such models explain why the competition effect diminishes as data is accumulated. The present results support the hypothesis that causal learners have generic prior expectations about causal relationships, and that a sparse-and-strong prior accurately characterizes these expectations.

The experiments presented here go beyond most previous investigations on causal learning by examining a more complex causal situation, one that included three observed generative causes. Examining a causal situation with multiple causes enabled a novel test of predictions that discriminated between alternative possible priors. Moreover, the relatively complex situation examined here may be more representative of the actual situations that causal learners encounter in the real world.

Using an iterative-learning method, Yeung and Griffiths (2011) empirically derived a different (but non-uniform) prior that was suggestive of a preference for strong causes, but that lacked the competitive pattern associated with the sparse prior. However, since the iterative method did not fully converge for the background cause, their results are open to multiple interpretations. Our task with multiple cues may provide a good way to further evaluate the generalization of empirical priors derived from the iterative-learning paradigm.

Lu et al. (2008) formalized sparse-and-strong priors for both generative and preventive causes. However, the preference for "sparseness" only applies across causes of the same polarity. In the generative case, sparseness is an influential factor even for simple causal set-ups, in which a single observed cause competes with an unobserved background cause (assumed by default to be generative). However, in the preventive case of the sparse-and-strong prior, competition dynamics are not evident when there is only a single preventive cause, as the observed cause is preventive whereas the background cause remains generative. The influence of sparseness, and hence the possibility of competition, is also predicted to arise in complex causal situations involving multiple preventers. As previous investigations have only examined the simplest cases, further research with more complex causal set-ups is needed to examine the possible impact of sparse-and-strong priors for preventive causes.

Acknowledgments

Preparation of this paper was supported by grant N000140810186 from the Office of Naval Research.

References

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (1993a). Cue competition effects: Theoretical implications for adaptive network learning models. *Psychological Science*, *4*, 196–202.

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (1993b). Cue competition effects: Empirical tests of adaptive network learning models. *Psychological Science*, *4*, 190-195.

Carroll, C., Cheng, P. W., & Lu, H. (in press). Inferential dependencies in causal inference: A comparison of belief-distribution and associative approaches. *Journal of Experimental Psychology: General*.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in cognitive sciences*, *7*, 19-22.

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

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology*, *55B*, 289–310.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, *62*, 135–63.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 334–384.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive psychology*, *55*, 232–57.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological review*, *115*, 955-84.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, *111*, 455–85.

Shanks, D. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *37*, 1–21.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*, *8*, 600–8.

Yeung, S., & Griffiths, T. L. (2011). Estimating human priors on causal strength. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1709-1714). Austin, TX: Cognitive Science Society.