Cognitive Psychology 86 (2016) 62–86

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Causal competition based on generic priors

Derek Powell [a,*], M. Alice Merrick [a], Hongjing Lu [a,b], Keith J. Holyoak [a]

[a] Department of Psychology, University of California, Los Angeles, United States
[b] Department of Statistics, University of California, Los Angeles, United States

ARTICLE INFO

Article history:
Accepted 1 February 2016

Keywords:
Causal learning
Simplicity
Cue competition
Generic priors
Bayesian models

ABSTRACT

Although we live in a complex and multi-causal world, learners often lack sufficient data and/or cognitive resources to acquire a fully veridical causal model. The general goal of making precise predictions with energy-efficient representations suggests a generic prior favoring causal models that include a relatively small number of strong causes. Such "sparse and strong" priors make it possible to quickly identify the most potent individual causes, relegating weaker causes to secondary status or eliminating them from consideration altogether. Sparse-and-strong priors predict that competition will be observed between candidate causes of the same polarity (i.e., generative or else preventive) even if they occur independently. For instance, the strength of a moderately strong cause should be underestimated when an uncorrelated strong cause also occurs in the general learning environment, relative to when a weaker cause also occurs. We report three experiments investigating whether independently-occurring causes (either generative or preventive) compete when people make judgments of causal strength. Cue competition was indeed observed for both generative and preventive causes. The data were used to assess alternative computational models of human learning in complex multi-causal situations.

* Corresponding author at: Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, United States.
E-mail address: derekpowell@ucla.edu (D. Powell).

http://dx.doi.org/10.1016/j.cogpsych.2016.02.001
0010-0285/

In real life there are no phenomena that have only one cause and have not been affected by secondary causes. Otherwise we should be living in a world of pure necessity, ruled by destiny alone.

[Spirkin, 1983, p. 90]

1. Coping with causal complexity

We live in a complex and multi-causal world. Consider how every modern commercial airplane is equipped with an in-flight recorder, or "black box." Should the airplane go down, the information provided by the recorder can help investigators diagnose the cause of the crash. Without such an aid, confident diagnosis might prove impossible, as many types of events, individually or in combination, can cause an aircraft crash—flight crew error, mechanical faults, the weather, terrorism, bird strikes, and so on (for a general review of causal thinking, see Lagnado, 2011).

Many more pedestrian causal relationships are no less complex, yet humans must continually learn and make inferences under severe constraints imposed by their limited attentional and memory resources (generally without the aid of external devices). Relative to the complexity of the actual physical and social world, the causal models formed by the human mind are inevitably simplified—more like rough sketches than high-resolution photographs. Causal simplification may play an especially important role in the early stage of learning, when the paucity of data makes it impossible to reliably estimate the strengths of all possible causes of an effect. Especially when data relevant to causal inference must be aggregated over time, memory limitations make it difficult to maintain representations of many alternative possible causal models. For example, when faced with five candidate causes for an effect, there are 32 possible causal models to consider (each including a particular combination of one to five effective causes). Faced with such a bewildering array of alternatives, the learner will likely be forced to make simplifying assumptions.

A general preference favoring simplicity has been proposed as a unifying principle for cognitive science (Chater & Vitányi, 2003). In the case of human causal learning and inference, several distinct simplicity constraints have been proposed. These are of three general types: (1) constraints on causal attribution within a known causal structure, (2) constraints on selection of causal structures, and (3) constraints on causal strengths associated with individual causal links. We will briefly consider each of these in turn, and then focus on the third.

1.1. Causal attribution

There is considerable evidence that people prefer to search their existing causal knowledge for a single explanation of a newly observed effect. As a consequence, multiple alternative explanations compete with one another. This type of competition gives rise to *causal discounting*, whereby the presence of one cause reduces the estimated probability that some other cause was active (Kelley, 1973). For example, if you observe wet grass in the morning, you might suspect it rained overnight. But if you find that there was a sprinkler on, you might attribute the wet grass to the sprinkler and discount the probability that the wet grass was caused by rain. Pearl (1988) showed that causal discounting is a normative consequence of reasoning with causal models. Moreover, the competitive nature of causal explanation follows from Thagard's (1989) theory of explanatory coherence, and is supported by a variety of experimental evidence (Holyoak, Lee, & Lu, 2010; Read & Marcus-Newhall, 1993). In addition to a general preference for single explanations, Read and Marcus-Newhall (1993) showed that people prefer explanations that cover a broader range of data (for additional evidence see Lombrozo, 2007, 2012).

1.2. Causal structure

Causal learners can also make judgments of causal *structure*, assessing which candidate causes in fact generate (or prevent) an effect (Griffiths & Tenenbaum, 2005). Interpreted within a graphical representation, a structure judgment is an evaluation of whether or not a link exists between a node representing a candidate cause to a node representing an effect. Here, simplicity constraints operate during the course of causal learning. The basic machinery of Bayesian inference yields a preference for graphs

that are less complex ("Bayesian Occam's razor"; see MacKay, 2003); hence graphs with fewer causal links will tend to be preferred. Similarly, Novick and Cheng (2004) argued that people make the default assumption that individual causes act independently to produce the effect (i.e., nodes representing conjunctions of multiple causes are excluded from causal graphs unless the data suggests a need to postulate conjunctive causes). Empirical evidence suggests that this type of simplicity constraint is honored both by adults (Liljeholm & Cheng, 2007) and children (Cheng, Liljeholm, & Sandhofer, 2013).

1.3. Causal strength

A causal structure judgment is an all-or-nothing decision as to whether a causal link exists, regardless of its strength. In contrast, causal strength (sometimes called causal *power*; Cheng, 1997) can be construed as the probability that the cause, acting alone, would yield the effect. As a probability, the value of causal strength is a quantity ranging from 0 to 1, where 0 implies the cause is ineffective and 1 implies it generates the effect in a deterministic manner. (Note that omitting a causal link is equivalent to assigning it 0 strength.) In Bayesian formulations, causal strength is coded by a probability distribution, which captures the learner's degree of uncertainty about the strength value. If simplicity is interpreted as a reduction in uncertainty about the expected strength distribution, then some *generic* prior may be preferred to an uninformative (i.e., uniform) prior on causal strength. In general, generic priors are based on systematic assumptions about the abstract quantitative properties of a system. In the case of motion perception, for example, human judgments of velocity are guided by the prior that motion tends to be *slow and smooth*, and this generic prior explains a wide range of visual illusions and motion perception phenomena (Lu & Yuille, 2006; Weiss, Simoncelli, & Adelson, 2002; Yuille & Grzywacz, 1988).

One natural generic prior that applies to causal strength can be derived from a preference for strength values that yield stable and precise estimates of the effect. Consider a simple situation in which one candidate cause yields one effect with a certain probability p (corresponding to the causal strength). Assuming the effect is a binary variable (it either does or does not occur), the effect's occurrence will follow a binomial distribution with parameter p. The variance of the effect variable is then given by $p(1-p)$, which is minimized (equivalently, precision of effect predictions is maximized) when p is at the extreme values of 0 or 1. Thus although causal strength is inherently probabilistic, people may prefer that causes be deterministic—either always yielding the effect (strength of 1) or being entirely ineffective (strength of 0). Evidence suggests that even preschoolers exhibit a preference for deterministic causes (Schultz & Sommerville, 2006).

A preference for causal determinism implies that the prior distribution over causal strength has peaks at 0 and 1 for each potential cause (including a general "background" cause representing an amalgam of potential causes other than the specific candidate causes under consideration). By itself, a deterministic prior does not imply any kind of competition between candidate causes—any number of alternative causes might each have the strength value of 1 (i.e., all causes might be strong). However, Lu, Yuille, Liljeholm, Cheng, and Holyoak (2008) suggested that causal strength has a generic prior that combines determinism with *sparseness*. Whereas a pure determinism constraint prefers all strengths to have the value 0 or 1 (and is otherwise indifferent), the *sparse-and-strong* (SS) prior prefers all strengths (for causes of the same polarity, generative or preventive) to have value 0 except for a single cause, which ideally has the value of 1. In the General Discussion we will elaborate on the possible rationale for a sparseness component of a generic prior on causal strength.

The general impact of SS priors will be to assist learners in quickly identifying the most potent individual cause, relegating weaker causes to secondary status or eliminating them from consideration altogether. Lu et al. (2008) reported a series of tests of alternative models of elemental causal learning in the simplest causal set-ups, involving a generative background cause coupled with either a single generative or preventive candidate cause. In general, a Bayesian model that incorporated SS priors into Cheng's (1997) power PC theory (i.e., assuming a noisy-OR combination rule for generative causes and a noisy-AND-NOT rule for preventive causes) provided the best overall account of human judgments for both causal structure and causal strength. Unlike an alternative model assuming uniform priors on strength values, a model incorporating SS priors was able to explain the dominance of strength and base rate of the effect over sample size in human judgments of causal structure, and also to account for subtle asymmetries in causal judgments across generative versus preventive causes.

2. Competition among multiple possible causes

The basic goal of the present paper is to test models incorporating SS priors that are extended to account for human judgments in more complex causal set-ups. The key qualitative prediction that follows from assuming SS priors is that multiple potential causes (of the same polarity) will compete with one another to acquire causal strength within a learner's emerging causal model, with apparently strong causes being preferred to weaker ones.

Importantly, SS priors predict a range of competition effects beyond those that have been previously reported in studies of causal learning. Various competitive dynamics are commonly observed in causal learning paradigms, including blocking (e.g., Shanks, 1985), overshadowing (e.g., Waldmann, 2001) and release from overshadowing (De Houwer & Becker, 2002). However, in all these paradigms competition arises 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. Note that in this design, cue B never occurs in the absence of cue A (implying a non-zero correlation between the occurrence of the two cues).

Under these circumstances, a cue competition effect would be observed when human learners assign a lower causal strength judgment for the blocked cue, B, due to greater uncertainty about the strength of cue B than of cue A. Thus, in some situations parsimony can be maintained as a byproduct of causal learning mechanisms that are otherwise indifferent to the complexity of causal models (Carroll, Cheng, & Lu, 2013), in accord with the Bayesian Occam's razor (MacKay, 2003). But without an informative prior, Bayesian models generally do not predict competition between causes that occur independently of one another (for a formal proof in the case of linear integration rules, see Busemeyer, Myung, & McDaniel, 1993a). However, if SS priors actively guide causal learning, then they should apply whenever multiple potential causes of a common effect are being assessed. Accordingly, causes should compete *even when they are uncorrelated in their occurrences*.

3. Goals of the present study

To our knowledge, the prediction of competition between independently-occurring causes has never been clearly tested.[1] More generally, relatively few studies of causal learning have used complex causal set-ups involving more than one or two candidate causes (but see, for example, Holyoak et al., 2010; Rehder, 2009). The present experiments were designed to determine whether multiple candidate causes would compete for causal strength, and to assess alternative computational models of human learning in multi-causal set-ups.

In the present paper we report three experiments that provide evidence for competition between causes during causal strength judgments within multi-causal set-ups. The first two experiments focus on generative causes only, and the final experiment deals with unknown causal polarities (i.e., situations in which candidate causes might generate or else prevent the effect). Then, in a modeling section, we present alternative Bayesian models with or without SS priors and compare their predictions to human data from the three experiments.

4. General method for experiments

4.1. Overview

In all of the present experiments, participants were asked to imagine themselves helping a doctor on an island resort. Some of the guests on the resort are falling ill, and the doctor thinks that some exotic vegetables that the guests have been eating may be involved. In Experiments 1 and 2, participants were told that the vegetables were generating the illness. In Experiment 3, participants were not

[1] Busemeyer, Myung, and McDaniel (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.

informed of the polarity of the causes. Instead, it was suggested that the vegetables might be making people sick, or might have medicinal properties that prevent the illness. Participants in all experiments then viewed a series of "case files" showing which combination of vegetables a guest had eaten, and whether or not they had fallen ill.

In each experiment, we created two conditions by manipulating the underlying causal strengths of each vegetable. Vegetables A and C were kept constant across conditions within an experiment, but the strength of vegetable B was manipulated to be either *weak* or *strong*. In all experiments cause C served as an observable generative "background" cause, as it was shown to be present on every trial. We were primarily concerned with participants' strength judgments for vegetable A, assessing whether judgments about the strength of A was influenced by variation in the strength of B. Critically, the occurrences of vegetables A and B were uncorrelated (see below).

4.2. Generating contingency data

In each experiment there were four possible combinations of vegetables: 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% of the time, and the correlation between the occurrence of A and B was 0. The numbers of guests who became sick after eating each combination were determined by the causal powers assigned to each vegetable.

Generating this contingency data requires the use of specific causal generative functions to integrate the expected causal influences of the multiple causes that happen to occur together on a given occasion. For binary variables of the sort used in the present experiments, the normative integration rules are based on noisy-logical functions derived from the power PC theory developed by Cheng (1997). Descriptively, numerous studies have confirmed that human judgments of causal strength follow the patterns predicted by the power PC theory (see Holyoak & Cheng, 2011, for a review). Moreover, the noisy-logical functions can be incorporated into Bayesian models of causal learning that captures uncertainty in strength judgments by updating entire strength distributions for each cue (Griffiths & Tenenbaum, 2005; Lu et al., 2008; Yuille & Lu, 2008). Therefore, in creating contingency data for the present experiments we simply adopted the integration rules posited by the power PC theory (Cheng, 1997), under the default assumption that each cause acts independently to generate or prevent an effect. Specifically, for generative causes, the noisy-OR integration rule was used, and for preventive causes, the noisy-AND-NOT rule was used. Tables 1 and 2 summarize the detailed contingency data used in the experiments reported in the present paper.

4.3. Experiential presentation of contingency data

An important methodological decision in all the present experiments was to sequentially present individual cases to participants, in a controlled laboratory setting. Many experiments on causal learning have employed summary presentations of contingency data, typically showing graphical representations of all cases to participants on a single page or computer screen, rather than allowing participants to encounter a sequence of individual cases (e.g., Liljeholm & Cheng, 2007; Lu et al., 2008; Yeung & Griffiths, 2015). The summary procedure is a simpler and less time-consuming methodology than providing a more experiential (i.e., sequential) presentation, but arguably much less natural as a model of everyday causal learning. People seldom are provided with such summary data, but rather are likely to encounter a series of relevant cases distributed over time (e.g., Danks, Griffiths, & Tenenbaum, 2003; Lu, Rojas, Beckers, & Yuille, 2015). For non-human animals, all causal learning is inevitably sequential. Also, although summary procedures have generally yielded orderly data for simple causal set-ups, we found in pilot work that for the more complex causal designs used in the present experiments, people found summary presentations of data difficult to grasp. When faced with summary presentations, participants would often resort to explicit arithmetic calculations, whereas sequential presentations appeared to more reliably elicit "natural" causal inference processes.

Table 1
Contingency data used in learning phase of Experiment 1, and in the first block of Experiment 2.

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

Table 2
Contingency data used in learning for one experimental block (40 trials) by trial type in Experiment 3.

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

Although arguably more basic (see Perales & Shanks, 2007), sequential presentation of probabilistic contingency data involving multiple causal cues inevitably makes it likely that people's causal learning will be influenced by general performance factors (e.g., lapses of attention or memory confusions) that can yield distortions in the effective inputs to an inductive learning mechanism. It therefore becomes necessary to account for possible input distortions when assessing formal models of causal learning. After presenting our three experiments, we will report tests of formal models of causal learning that incorporate uncertainty about contingencies attributable to noise arising from attentional and memory factors.

5. Experiment 1: competition between independently-occurring generative causes

The purpose of Experiment 1 was to assess whether competition arises between independently-occurring causes arising during initial learning.

5.1. Method

5.1.1. Participants

Participants were 90 undergraduate students at the University of California, 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.

5.1.2. 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 Fig. 1, 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 vegetables: 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 the occurrence of vegetable A and B was independent. A total of 44 cases (11 of each combination) 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 .09. 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. Cue 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 cue B. Cue C (causal power of .09) 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 their rating using a slider, inputting a value between 0 and 100 (see Fig. 1, bottom). Hence, a rating of 50 on the 100-point scale was interpreted as indicating an estimated causal strength of .50. 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).

5.2. Results and discussion

Data from one participant was excluded due to technical issues. Data from another three participants were excluded because they both entered extreme ratings of either 0 or 100 for cause A and entered 0 for cause C, suggesting accidental errors or a lack of engagement with the task.

Fig. 2 shows the mean estimated causal strength for each of the three cues, for the strong-B and weak-B conditions. We first focus on results for the critical cue A. Although the veridical causal strength of cue A was .50 in both experimental conditions (corresponding to an expected rating of 50), participants' strength estimates for cue A in the strong-B condition ($n = 43$; mean = 31.67; SD = 25.07) were significantly lower than participants' estimates in the weak-B condition ($n = 43$; mean = 45.88; SD = 30.22), $t(84) = 2.37$, $p = .02$. Hence, the estimate of causal strength of cue A was affected by the causal strength of independently-occurring cue B, suggesting causal competition between these two potential causes.

An important question is whether the effect of the strong-B cause on participants' ratings for cause A reflected a genuine difference in participants' estimates for the causal strength of A, or was somehow mediated by a superficial change in their use of the response scale. One might suppose, for example, that the rating for cue A might vary depending simply on whether the A rating was made before or after a rating was elicited for B. To assess this possibility, we compared ratings for participants who gave causal strength estimates for cue A before cue B to those of participants who rated B before A, performing a 2×2 (condition $\times$ order) ANOVA. These two variables did not interact ($F < 1$), indicating

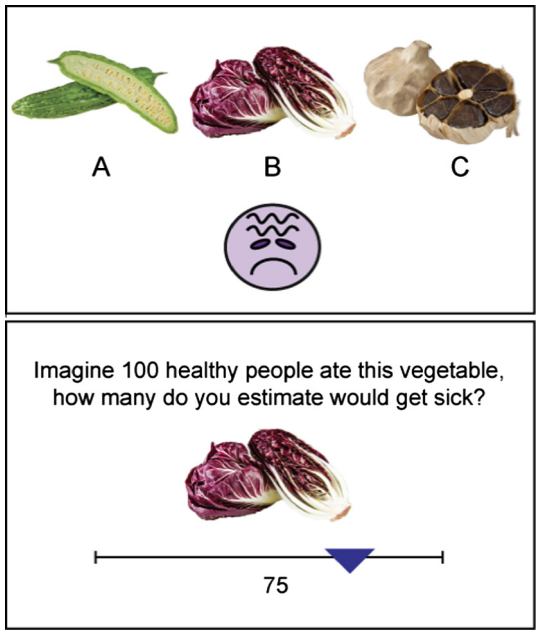

Fig. 1. Illustration of experimental stimuli. Top, example trial showing a guest who ate A, B, and C vegetables and became sick (top). Bottom, illustration of example response page.

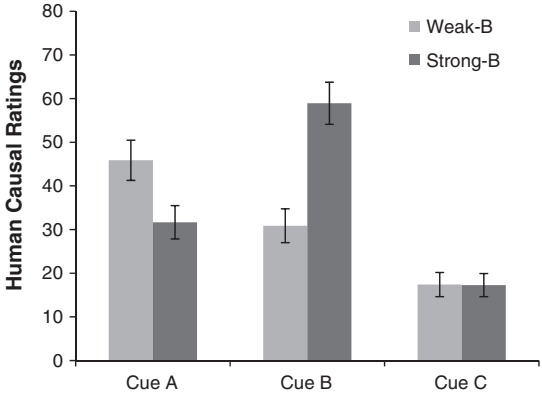

Fig. 2. Human causal strength ratings (0–100 scale) for cues A, B, and C in Experiment 1.

that the size of the competition effect was not influenced by the order in which the A and B cues were rated.

Further evidence of causal competition was provided by a correlational analysis of ratings for cues A and B *within* each condition. After accounting for individual participants' tendency to give high or low ratings overall (by partialing out average strength ratings for all three causes), we found a substantial negative partial correlation between strength ratings for cues A and B both within the strong-B ($r = -.62$, $p < .001$) and the weak-B conditions ($r = -.558$, $p < .001$). Thus, participants who considered A as a strong cause tended to rate B as a weak cause, and vice versa, even within each experimental condition.

We also analyzed the mean estimated strengths for cue B (which varied in veridical strength between the strong-B and weak-B conditions) and for cue C (the only cue that appeared by itself, rather than solely in combination with other cues). For cue B, as expected, participants' strength ratings were higher in the strong-B condition (mean = 58.95; SD = 31.62) than the weak-B condition (mean = 30.88; SD = 25.36; $t(84) = 4.54$, $p < .001$). It should be noted that participants' estimates for cue B were less extreme than the veridical causal power for each condition. That is, when B was weak (veridical causal power of .20), participants overestimated its strength, and when B was strong (power of .80), participants underestimated its strength. These deviations from veridical strength may reflect noise in processing the sequential contingency data due to attentional biases or memory errors. We will address these deviations in the modeling section.

Ratings for cause C were similar for both strong-B (mean = 17.9; SD = 17.86) and weak-B (mean = 18.0; SD = 18.76) conditions; however, participants overestimated the causal strength of cue C relative to its veridical causal power of .09, $t(85) = 4.374$, $p < .001$. As with cause B, this deviation from the veridical power may reflect input distortions in memory for contingency data.

6. Experiment 2: causal competition as a function of sample size

Experiment 2 examined whether the influence of competition between causes on strength judgments varies with sample size. In general, the influence of priors on Bayesian learning is expected to diminish as learners gather more data. Accordingly, a model assuming SS priors should predict that competition between causes will be strongest when participants have made few observations, and will decline as participants are exposed to more data. The design of Experiment 2 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 2×3 factorial design, with one between-subjects factor (causal strength of cue B) and one within-subjects factor (sample size).

The cover story was the same as 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."

6.1. Method

6.1.1. Participants

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

6.1.2. 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 (see Table 1). Order of presentation was randomized for each participant.

6.2. Results and discussion

Four participants were excluded using the same criteria as in Experiment 1. In addition, two participants were excluded because they entered the same response for all three cues in at least two blocks, suggesting a lack of understanding or motivation.

Fig. 3 shows mean causal strength ratings for the critical cue, vegetable A, after each learning block. To analyze the estimated strength of cue A, we conducted a repeated-measures ANOVA. We report results from multivariate tests because the data violated the sphericity assumption (Mauchly's test, $p = .001$). We found a significant interaction between condition (weak B vs. strong B) and learning block, $F(2, 105) = 4.09$, $p = .019$, reflecting an attenuated competition effect between independently-occurred cues with an increase in the number of observations. After the first block, participants

provided lower estimate for the strength of cue A when it was paired with a strong B cause ($n = 53$; mean = 45.95; SD = 24.70), relative to when it was paired with a weak B cause ($n = 55$; mean = 55.70; SD = 21.98; $F(1,106) = 4.69$, $p = .033$), replicating the causal competition effect reported in Experiment 1. The influence of the strength of B on ratings of A was no longer reliable after 88 or 132 trials (Fs < 1), supporting the hypothesis that the competition effect observed after one block of learning trials was primarily driven by people's priors on causal strength.

A correlational analysis of participants' A and B ratings in each block further supports this conclusion. Similar to Experiment 1, we computed the partial correlation between ratings for A and B when controlling for average ratings for all three causes (reflecting each participant's general preference for assigning low or high scores). In block 1, partial correlations between ratings for A and B were negative for the strong-B ($r = -.47$, $p < .001$) and weak-B conditions ($r = -.27$, $p = .053$), replicating the comparable partial correlations obtained in Experiment 1. These partial correlations were weaker and not significant in block 2 (strong-B: $r = -.22$, $p = .12$; weak-B: $r = -.16$, $p = .28$) or in block 3 (strong-B: $r = -.16$, $p = .26$; weak-B: $r = -.18$, $p = .20$), suggesting the causal competition effect diminished with increases in the number of observations.

Strength estimates for cues B and C are shown in Fig. 4. For cue B, as expected, we observed a significant main effect of condition, with higher ratings for the strong-B than for the weak-B condition ($F(1,106) = 86.79$, $p < .001$). After Block 1 the mean rating for B was 65.29 ± 23.08 in the strong-B condition, versus 39.13 ± 19.58 in the weak-B condition. This difference was maintained across the three learning blocks, as evidenced by the lack of interaction between condition and learning block, $F(2,105) = 1.83$, $p = .966$. However, as in Experiment 1 we found that participants' strength estimates for cue B showed a large deviation from the veridical causal power in each condition (i.e., ideal rating of 80 in the strong-B condition and 20 in the weak-B condition). There was no main effect of learning block on ratings for cue B, $F(2,105) < 1$. Also as in Experiment 1, the mean rating for cue C was somewhat higher (25.56) than the expected rating of 9 that would reflect the veridical causal power (.09) for cue C. Increasing sample size produced an upward shift in strength estimates for cue C, $F(2,105) = 4.71$, $p = .011$). The influence of condition (weak-B vs. strong-B) on judgments about cue C was not reliable, $F(1,106) = 1.23$, $p = .27$), nor was there any interaction between condition and learning block, $F(2,105) = 2.20$, $p = .12$).

Causal strength estimates for all three vegetables were somewhat higher (by a mean of 8.2 points on the 100-point rating scale) in the first block of Experiment 2 than in Experiment 1. This difference may reflect small changes in the rating-scale interface and instructions. In Experiment 1, the starting value for the slider was set at 0, whereas in Experiment 2 the slider did not appear until participants clicked on the rating scale. The initial starting point used in Experiment 1 may have anchored participants' responses at 0, yielding lower estimates. In addition, instructions in Experiment 2 emphasized the doctor's belief that the vegetables were indeed causing the illness.

7. Experiment 3: competition between independently-occurring preventive causes

In everyday life, causes can prevent as well as generate effects. Experiment 3 examined competition effects in a situation involving multiple *preventive* causes. In addition to previous findings that preventive causes exhibit several asymmetries relative to generative causes (Cheng, Novick, Liljeholm, & Ford, 2007; Lu et al., 2008), the general principle of SS priors implies that multiple preventive causes will also compete if they co-occur in preventing the presence of an effect. However, this prediction has so far gone untested, as the original formulation of SS priors (Lu et al., 2008) only considered the case of a single candidate cause coupled with the background cause. By default the background cause is assumed to be generative (since unless something generates the effect, it is impossible to assess whether other cues prevent it). Hence, in the case of a single candidate cause, SS priors predict competition effects for a generative candidate cause (which will compete with the generative background cause), but not for a single preventive candidate cause (which will not compete with the generative background, given the basic assumption that only causes of the same polarity compete). Extending the complex three-cause setups used in the first two experiments to preventive causes allows us test whether preventive causes compete in a fashion similar to that observed for generative causes.

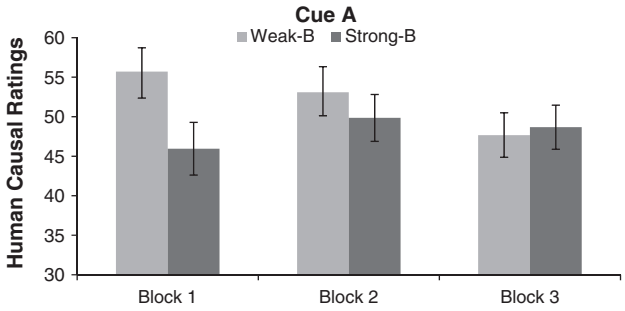

Fig. 3. Human casual strength ratings for cue A across learning blocks in Experiment 2.

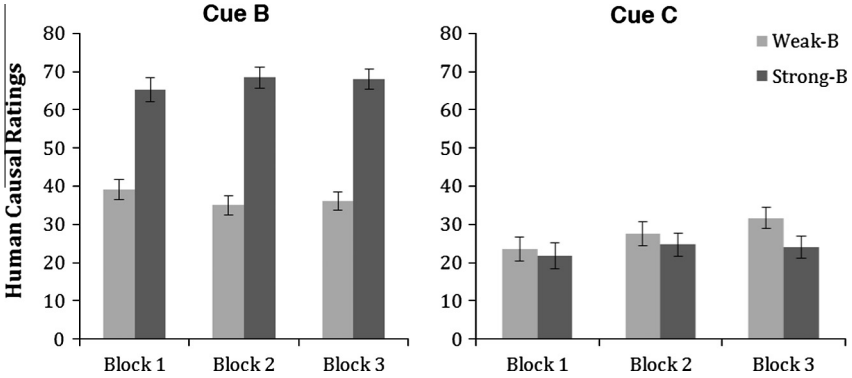

Fig. 4. Human causal strength ratings for cues B and C across learning blocks in Experiment 2.

In addition, in many real-life causal situations people may not know in advance the causal direction (preventive or generative) of potential causes. For example, a newly-developed experimental drug might either prevent or generate certain symptoms. Accordingly, Experiment 3 examined competition effects in a situation involving a mix of generative and preventive causes, where learners were not informed of causal directions beforehand.

7.1. Method

7.1.1. Participants

Participants were 57 UCLA undergraduate students who participated for class credit (82.5% female, mean age = 21.5 years old).

7.1.2. Materials and procedure

Experimental materials and procedure were very similar to those used in Experiment 2. Participants again viewed three blocks of trials that showed whether patients became sick or remained healthy after eating different combinations of three vegetables. New contingency data were created to reflect a causal structure where vegetable C produced or generated the illness and vegetables A and B prevented the illness. Vegetable C was a strong generative cause with strength .80, and Vegetable A was a preventive cause with strength −.50 for both conditions (using negative values to denote preventive strengths). In the strong-B condition, Vegetable B was assigned a causal strength of −.75, and in the weak-B condition it was assigned a strength of −.25. For each block, 40 trials of contingency data were generated to reflect these different underlying causal relations, as shown in Table 2.

Participants were not informed of the polarity of the causes they were to observe. Instead, participants were told, "One or more of the vegetables could be making the guests sick. However, not all of the guests who ate salad are getting sick. One or more of the vegetables may have medicinal properties that prevent the illness." After completing each learning block, participants were now asked two questions about each vegetable. For each vegetable, participants were first asked a causal polarity question. Participants were shown a picture of each vegetable along with a question asking whether it caused or prevented the illness. Participants pressed "C" to indicate that they thought the vegetable caused the illness, "P" to indicate that they thought it prevented the illness, or "N" if they thought it was not a cause at all.

If participants indicated that the vegetable caused or presented the illness, they advanced to the causal strength question. If they believed the vegetable was a generative cause, they were asked, "Suppose 100 people ate this vegetable, how many will get sick?" If they believed the vegetable was a preventer they were asked, "Suppose 100 people are about to get sick. If they all eat this vegetable, how many of the 100 will not get sick?" Participants made their rating using a slider as in Experiment 2. If a participant indicated that a vegetable was not a cause, the strength question was skipped and their strength estimate was scored as zero. The polarity question always preceded the strength question, but the order of the three vegetables was randomized for each participant and block. After responding to all questions in a block, participants were shown a summary of their responses and were asked to confirm that they had correctly entered their ratings.

7.2. Results and discussion

Four participants mistakenly indicated that cause C was a preventer during at least one of the three blocks. As it was unclear how their other responses are to be interpreted in light of this error, data from these participants were excluded from analyses. Participants' polarity judgments and strength ratings for each vegetable were combined into a single index. When participants indicated that a cause was generative, their strength rating was recorded as their score. When they indicated that a cause was preventive, their score was computed by multiplying this strength rating by -1.

Fig. 5 presents participants' strength estimates for the critical cue A, for which the veridical causal power was $-.50$ (corresponding to an ideal rating of $-.50$). Participants in the strong-B condition ($n = 27$) underestimated the preventive strength of vegetable A relative to participants in the weak-B condition ($n = 26$), $F(1,51) = 6.58$, $p = .013$. There was a marginally significant learning effect across blocks, with a trend toward a higher strength for preventive cue A with more observations, $F(2,50) = 2.87$, $p = .07$. However, there was no significant interaction between the experimental condition and learning block, $F(2,50) = .26$, $p = .77$. Rather, the competition effect remained statistically constant as participants were exposed to more data over the course of three blocks. Thus, a scenario that included both generative and preventive causes, but for which observers were not told the causal polarities of any individual cue, yielded a competition effect for multiple preventive causes that persisted across multiple blocks of learning trials.

A correlational analysis of participants' ratings for cues A and B within each block also suggests that the competition effect was maintained longer in Experiment 3 than in Experiment 2. We calculated the partial correlation between ratings for A and B (controlling for each participant's average absolute ratings for all three cues) for each condition separately. In block 1, partial correlation between ratings for cues A and B were negative for both the strong-B ($r = -.54$, $p = .004$) and weak-B conditions ($r = -.390$, $p = .054$). These partial correlations remained in block 2 (strong-B: $r = -.46$, $p = .017$; weak-B: $r = -.64$, $p = .001$); however, in block 3 a significant negative partial correlation was only obtained for the weak-B condition (strong-B: $r = .20$, $p = .32$; weak-B: $r = -.42$, $p = .038$).

Fig. 6 presents the mean strength ratings for the other two cues. Participants were sensitive to the different veridical causal strengths of cue B in the weak-B ($-.25$) and strong-B ($-.75$) conditions, as reflected in a main effect of condition, $F(1,51) = 40.96$, $p < .001$) for this cue. However, there was no main effect of learning block, $F(2,50) = .18$, $p = .833$), nor any interaction ($F(2,50) = .57$, $p = .57$). Causal strength estimates for the generative cue C were somewhat higher in the strong-B condition than the weak-B condition, $F(1,51) = 4.13$, $p = .047$. However, there was no main effect of block, $F(2,50) = .61$, $p = .55$, nor any interaction, $F(2,50) = 1.22$, $p = .30$.

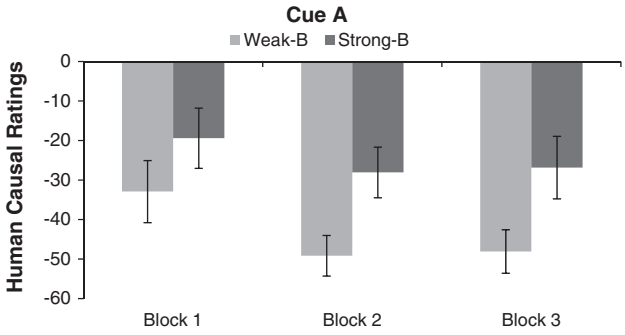

Fig. 5. Human causal strength ratings for preventive cue A across learning blocks in Experiment 3.

As in Experiment 2 (generative causes only), Experiment 3 (with both generative and preventive causes) yielded a causal competition effect. However, the impact of number of observations differed across the two experiments. In Experiment 2 the competition effect was only reliable for block 1, whereas in Experiment 3 the competition effect was basically stable across all three blocks. The sustained competition effect in Experiment 3 may reflect increased uncertainty due to the fact that the causal directions for the three cues were initially unknown, coupled with noise introduced by the experiential presentation of the contingency data. Particularly for cue B in the weak-B condition, with its low preventive power of −.25, participants may sometimes have mistaken the causal direction. The observed frequency of the effect when cues B and C co-occurred was just two less than the frequency for C-only cases. Given noise introduced by perception and memory processes, it is quite likely that participants would have sometimes mistakenly judged cue B to be non-causal or even a generative cause. If the causal direction of cue B was misperceived, causal competition would have a more extreme impact on the evaluation of cue A (increasing its apparent generative power). Thus, the sustained competition effect may reflect an exaggerated impact of SS causal priors due to ambiguity of causal direction for the weak-B cue. The simulations reported below will include a formal instantiation of this intuitive interpretation.

8. Modeling causal competition in strength judgments

Viewed from a Bayesian perspective, causal inferences are expected to be jointly determined by the likelihood function (evaluating how well observed covariation data can be explained by a given set of potential cause–effect relations, each associated with various possible strengths) and priors (expectations about causal relations that the learner brings to the task). An understanding of how causal relations operate and act together is encoded in the likelihood term, whereas generic causal priors constitute preferences for particular causal strengths or structures based on abstract properties (rather than domain-specific knowledge). Although some Bayesian models of causal learning have assumed uninformative priors (e.g., Griffiths & Tenenbaum, 2005), other models have incorporated substantive generic priors about the nature of causes. Here we focus on Lu et al.'s (2008) proposal that people have a preference for causes that are sparse and strong.

The qualitative signature of SS priors is a preference for one strong cause of a given polarity (generative or else preventive) with all other potential causes being weak. For example, a set of "ideal" causal strengths for three generative causes might be $w_A = 1$, $w_B = 0$ and $w_C = 0$ (where each w indicates a value of causal strength, ranging from 0 to 1, for a particular candidate cause). Note that the SS prior does not express any preference about *which* cause(s) are strong and which are weak. Instead, SS priors imply that multiple potential causes will compete with one another to acquire causal strength within a learner's emerging causal model, with apparently strong causes being preferred to weaker ones.

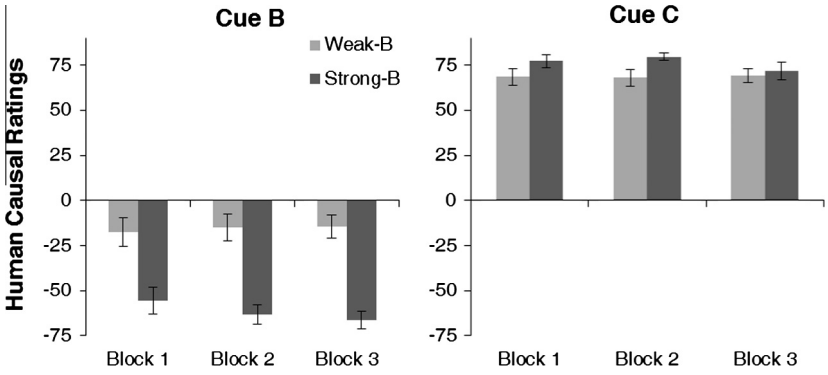

Fig. 6. Human causal strength judgments for preventive cue B and generative cue C across learning blocks in Experiment 3.

8.1. Generalizing SS priors to three generative causes

In formalizing SS priors, Lu et al. (2008) only considered the simple case of competition between one generative candidate cause and the (generative) background cause. The generalization of the sparse-and-strong prior for more than two causes is straightforward. For causal situations with only generative causes, the SS prior can be defined as in Eq. (1) on the basis of the causal strength of three candidate causes, *A*, *B*, and *C*, which are all generative:

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

Fig. 7 illustrates the SS prior for a generative three-cause situation. The prior favors one of three corners in the three-dimensional plot, each corresponding to one maximally-strong causal cue with strength of 1, and two maximally-weak cues with strength of 0.

8.2. Generalizing SS priors to three causes with polarity unknown

As we have emphasized, SS priors are defined under the general assumption that causes of the same polarity compete. When the causal polarity of certain causes are unknown, a generalized form of SS prior can be defined as shown in Eq. (2). The space of causal strength values is separated into four quadrants based on the possible causal polarities for cues A and B. Quadrants are labeled with a three-letter code specifying the polarity of each cause ("P" for preventive and "G" for generative) in the order A, B, C. For instance, the region representing A and B as preventers (where the two preventive cues compete) and C as a generator is labeled "PPG".

$$P(w_A, w_B, w_C) \propto \begin{cases} e^{-\alpha(1-w_C)}\left(e^{-\alpha w_A-\alpha(1-w_B)} + e^{-\alpha(1-w_A)-\alpha w_B}\right), & PPG \\ \left(e^{-\alpha(1-w_B)-\alpha w_C} + e^{-\alpha w_B-\alpha(1-w_C)}\right)\left(e^{-\alpha w_A} + e^{-\alpha(1-w_A)}\right), & PGG \\ \left(e^{-\alpha(1-w_A)-\alpha w_C} + e^{-\alpha w_A-\alpha(1-w_C)}\right)\left(e^{-\alpha w_B} + e^{-\alpha(1-w_B)}\right), & GPG \\ e^{-\alpha(1-w_A)-\alpha w_B-\alpha w_C} + e^{-\alpha w_A-\alpha(1-w_B)-\alpha w_C} + e^{-\alpha w_A-\alpha w_B-\alpha(1-w_C)}, & GGG \end{cases} \quad (2)$$

We adapted standard Bayesian accounts of causal reasoning to generate predictions for the two causal scenarios that we used in our experiments. As the design of Experiment 1 was identical to the first block of Experiment 2, we fit the model to the data from Experiment 2, which included three generative causes with known polarities. A more complex inference strategy was required to account for causal strength judgments in Experiment 3, which involved two preventive and one generative cause, but for which learners were not informed of the causal polarities. We will present the results of the model simulations of the human strength ratings separately for Experiment 2 and then Experiment 3.

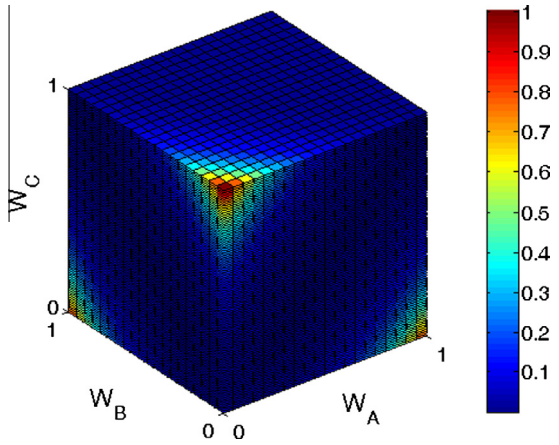

Fig. 7. Sparse-and-strong prior distribution over causal strengths of three potential generative causes. Colors indicate the values of prior probability (red corresponds to highest probability). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

8.3. Simulation results for Experiment 1 and 2 (generative causes only)

In Experiments 1 and 2, participants were explicitly informed of the causal polarity of candidate causes, with instructions stating that all vegetables potentially generated the illness. We primarily focus our simulations and model comparisons on Experiment 2, where we collected judgments for three blocks of learning trials, giving us a greater number of data points after averaging human judgments. We assumed that the underlying causal model follows the simple common-cause structure depicted in Fig. 8. To evaluate this assumption we computed the model evidence (i.e., the predictive posterior probability) for each of the four most probable causal structures (those in which cue C is present and cues A and B are each either present or absent). Based on the combinations of cues presented during the learning trials, model evidence after all learning blocks was greatest for the three-link structure shown in Fig. 8. This analysis thus confirmed that human data can be best fit by assuming people adopted a fixed model including all three cues as possible causes, rather than using model selection to drop some links. (However, we will later consider the possibility of using Bayesian model averaging across multiple structures.)

Using the identical fixed causal structure, we then compared two Bayesian models that use the same likelihood function but include different prior distribution on causal strength: either SS priors (favoring sparse-and-strong causes), or uniform priors, which imply no preference for any specific strength values. Generally, Bayesian models that adopt a fixed causal structure predict competition between independently-occurring causes only when SS priors are applied, and not when uniform priors are adopted.

The models applied Bayes' rule to compute the posterior probability distribution of causal strengths of the three cues, for a given set of contingency data D. For cue A this would be computed as:

$$P(w_A|D) = \iint \frac{P(D|w_A, w_B, w_C)P(w_A, w_B, w_C)}{P(D)} dw_B dw_C, \tag{3}$$

in which the prior term was defined using either the SS prior in Eq. (1), or a uniform distribution. The likelihood function $P(D|w_A, w_B, w_C)$ was defined as a noisy-OR function:

$$P(E = 1|C_A, C_B, C_C; w_A, w_B, w_C) = 1 - (1 - w_A C_A)(1 - w_B C_B)(1 - w_C C_C), \tag{4}$$

where the values of C variables are either 1 (indicating presence of the relevant causal cue) or 0 (indicating its absence), and values of w variables represent the causal strength of each causal cue.

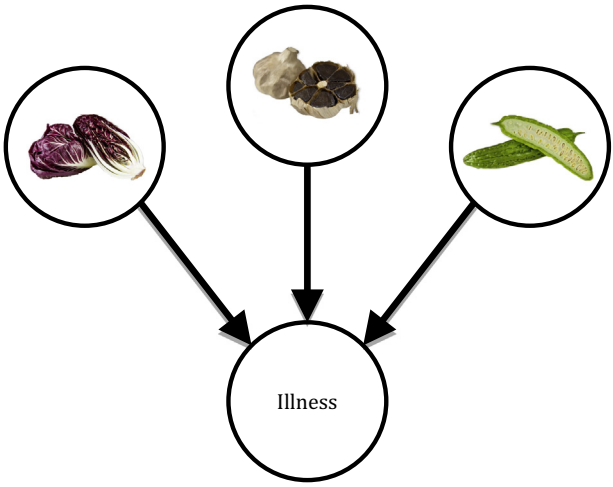

Fig. 8. Causal structure used in all Bayesian models.

Using the contingency data in Table 1 as the input to the two models, our simulation results confirmed that the model with SS priors predict the competition effect for cue A, with a higher estimated mean posterior strength in the weak-B condition than in the strong-B condition, even though the veridical power value for A was equated in the two conditions. For example, setting $\alpha = 5$ (the value estimated for the data sets reported by Lu et al., 2008), the model's estimates of the rating for cue A after the first block was 52.7 in the weak-B condition and 35.8 in the strong-B condition. This difference was attenuated as sample size increased across the three learning blocks in Experiment 2 (as the competition effect is largely driven by priors, the influence of which weakens as observations accumulate). After block 3, the predicted rating for cue A was 51.6 in the weak-B condition, and 44.6 in the strong-B condition). By contrast, the model assuming a uniform prior does not predict a competition effect for independently-occurring cues: predicted ratings for cue A after block 1 were the same (45.2) in both the weak-B and the strong-B conditions and this lack of competition was maintained for the subsequent blocks.

The model with SS priors predicted human ratings for cue A quite accurately, yielding a substantial correlation ($r_A = .71$) between averaged human ratings for cue A in Experiment 2 and the predictions of the model with SS priors. In contrast, the model with uniform priors did not predict any notable differences in the estimated strength of cue A across the two experimental conditions, in fact yielding a negative correlation with human ratings of cue A ($r_A = -.26$). However, neither model fared well in accounting for the human ratings of cues B and C. The Bayesian models predict ratings of B and C close to their veridical power values as determined by the actual contingency data provided to participants. However, human ratings for these two cues exhibited relatively large deviations from the veridical values. An obvious possibility, given the fact that the learning data involved several cue combinations and was presented sequentially, is that people's judgments were based on data distorted by noise, attributable to attentional or memory limitations.

In an effort to relate the learning models more closely to human performance limitations, we created versions in which the inputs to the induction process were noisy. To generate noisy inputs based on the actual contingency data, frequency values for the four cells in Table 1 in which the effect E is present were sampled from a truncated Gaussian distribution with the same mean as the veridical contingency value and a standard deviation of $2.5 * (1 + \log(\text{block number}))$. The scalar of $1 + \log(\text{block number})$ aims to capture the intuition that the larger cell frequencies generated with each added block of learning trials were associated with increased uncertainty about the frequency values. The noise scalar parameter was set to 2.5, selected using a grid search to maximize the fit of the uniform prior model for Experiment 2. The truncated Gaussian samples were bounded within the range of a

standard deviation lower than the mean and the maximum number of trials for each contingency condition (i.e., total number of trials divided by four, the number of contingencies included in the design). The frequencies of each cell in which the effect E was absent were set to the complement of the estimated frequency for the corresponding cell in which E was present, thereby ensuring that the mean proportion of trials of each type on which E occurred was equal to the veridical value. Each simulation was based on 1000 sampled inputs. In the model tests described below, we fix the noise parameter at 2.5 to simulate the noisy input to all the models used to predict the strength ratings obtained in Experiments 2 and 3.

Fig. 9 plots the simulation results based on noisy inputs for the estimated causal strength of each cue in Experiment 2, along with the observed human ratings. Table 3 summarizes results for three types of measures of model-fitting for both models. These measures are the Bayesian Information Criterion (BIC; Schwarz, 1978), which quantifies the mismatch between model predictions and observed ratings from individual participants, taking account of model complexity as defined by number of model parameters; Pearson's correlation coefficients (r) calculated between model predictions and human mean ratings, averaged across individual subjects; and root mean squared deviation (*RMSD*) calculated based on deviations between mean causal strength estimated by the models and averaged human causal ratings. In addition to reporting the fitting results for all three cues, we also report the r and *RMSD* values for cue A only, as these measures most precisely gauge whether the models can account for causal competition effects.

BIC is calculated according to Eq. (5), which combines the likelihood of observed individual data given the model predictions (L) with a penalty term based on the number of model parameters (k) and observations (n).

$$\mathrm{BIC} = -2 * \ln L + k * \ln(n). \tag{5}$$

We calculated the likelihood of individual participants' responses using the marginal distribution of the model estimates for each cue. The BIC's penalty term allowed us to penalize the 1-parameter SS priors model relative to the parameter-free uniform priors model. (Note that we did not estimate the value of the α parameter from our data, so this correction can be viewed as a conservative test of the SS priors model.) In both Experiments 1 and 2, participants' responses were better predicted by the SS priors model than the uniform priors model, as revealed by the negative difference scores between the BIC values for the two models (Experiment 1 ΔBIC = −63.57; Experiment 2 ΔBIC = −29.07). These results strongly favors the SS priors model over the uniform priors model, based on the comparison between predicted posterior distributions for each model and individual subject data (Kass & Raferty, 1995).

For Experiment 2, the overall fits of both models across all three cues were quite accurate and very similar. The overall r value was slightly lower for the model with the SS prior, perhaps due to our decision to select the noise parameter based on fits using a uniform prior (a decision that presumably favored the latter model.) However, only the model with SS priors ($\alpha = 5$) yielded a pattern of cue competition similar to that produced by human learners. Focusing on the critical A-cue in Experiment 2, we found a substantial correlation between averaged human ratings and estimates of the SS priors model (r_A = .45), but a negligible correlation between human ratings and the uniform priors model (r_A = −.003).

8.4. Simulation results for Experiment 3 (unknown causal polarity)

Experiment 3 involved two preventive causes (A and B) and one generative cause (C), but participants were not informed about the polarities of the causes (i.e., they had to learn which cues were preventive and which cues were generative). To simplify modeling of this situation, we assumed that learners would easily discover that vegetable C is generative, as it was presented alone with the effect occurring at a high frequency (yielding veridical generative power of .80). We assumed the other two cues, vegetables A and B, could be either preventive or generative, with causal strength in the range [−1, 1]. The orthogonal possibilities that each of cues A and B could be either preventive or generative yield four regions of causal strength. Density in positive regions represents probability density for generative strengths, whereas density in negative regions represents probability density for preventive

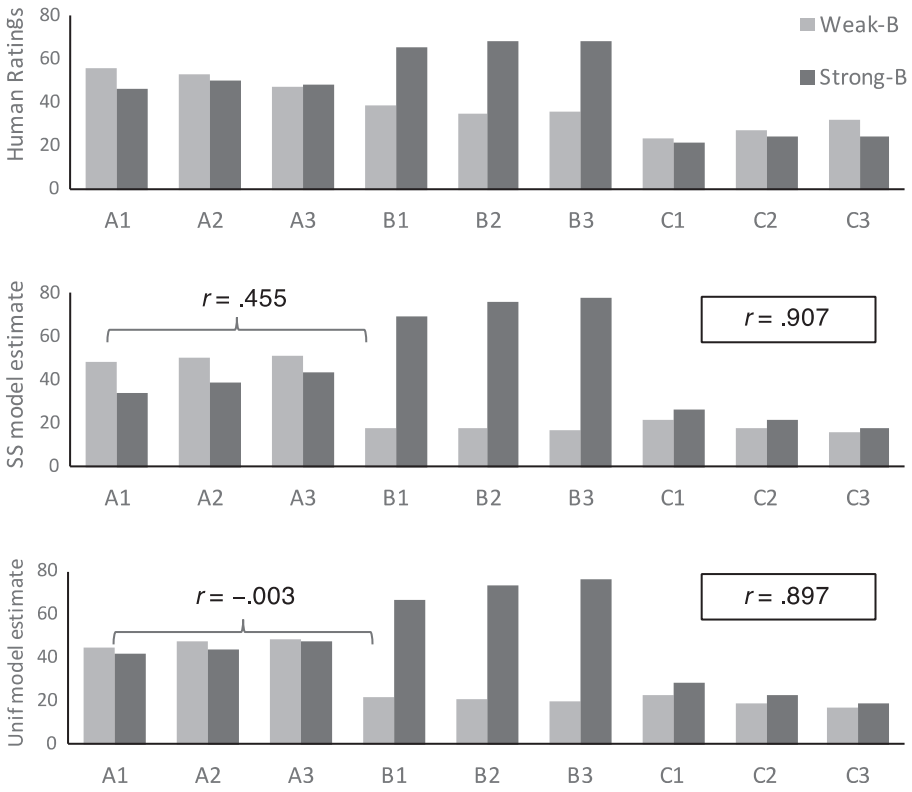

Fig. 9. Human ratings and model simulation results for Experiment 2 (all cues are generative). Modeling results are based on noisy inputs. Correlation values indicate correlations between model estimates and human data for cue A alone (indicated by brace) and for all cues (given in box on right).

Table 3
Comparison of model predictions based on noisy contingency data with human strength ratings in Experiments 2–3.

Exp.	Model	BIC	Overall *RMSD*	Overall *r*	A-cue *RMSD*	A-cue *r*
2	SS	**8248**	10.15	.897	8.25	**.455**
	Unif	8277	**6.69**	**.907**	**6.52**	–.003
3	SS	**5674**	**6.52**	**.995**	**6.43**	**.938**
	Unif	5861	7.30	.992	8.06	.843

Note: SS = model with SS priors, Unif = model with uniform priors. BIC is Bayesian Information Criterion which expresses degree of error in fit, therefore smaller numbers indicate better fit. Overall *RMSD* is root mean squared deviation calculated based on all three cues and A-cue *RMSD* is for cue A only. Smaller *RMSD* values indicate better fit. Similarly, overall *r* is correlation between model estimates and averaged human causal judgments for all three cues; A-cue *r* is for cue A only. Bold marks values that indicate superior fits (comparing SS vs Unif).

strengths. By generalizing the approach to unknown causal polarity developed by Lu et al. (2008), the likelihood function was specified over each of the four regions:

$$P(E=1|C_A,C_B,C_C;w_A,w_B,w_C)=\begin{cases} C_C w_C(1-C_A w_A)(1-C_B w_B), & PPG \\ (1-(1-C_B w_B)(1-C_C w_C))(1-C_A w_A), & PGG \\ (1-(1-C_A w_A)(1-C_C w_C))(1-C_B w_B), & GPG \\ 1-(1-C_A w_A)(1-C_B w_B)(1-C_C w_C), & GGG \end{cases} \quad (6)$$

where the values of C variables are either 1 (indicating presence of the relevant causal cue) or 0 (indicating its absence), and values of w variables represent the causal strength of each causal cue.

Using the veridical contingency data summarized in Table 2 as the input to the two models, and the three-link causal structure shown in Fig. 8, our simulations confirmed that the model with SS priors ($\alpha = 5$) predicts a competition effect for cue A, with a higher estimated mean posterior strength in the weak-B condition than in the strong-B condition, even though the veridical power value for cue A was equated in the two condition. For example, the model estimated the preventive strength for cue A after block 1 as −43.6 in the weak-B condition and −31.3 in the strong-B condition. In contrast, the model using a uniform prior did not predict a reliable competition effect for independently-occurring cues: estimated ratings for cue A after block 1 were similar in the weak-B (−43.8) and strong-B conditions (−41.5). This small competition effect disappeared with more observations after blocks 2 and 3.

As for Experiment 2, neither model successfully predicted the human ratings for cues B and C, which significantly deviated from their veridical powers. Accordingly, we introduced noisy inputs using truncated Gaussian samples, in the same manner as described for the simulations of data from Experiment 2. The parameter value for the standard deviation estimated for Experiment 2 (2.5) was used again to fit the data of Experiment 3.

The fits of the two resulting models to the data from Experiment 3 are depicted in Fig. 10, and measures of fit are summarized in Table 3 (bottom row). As was the case for Experiments 1 and 2, a comparison of BIC values for SS and uniform priors models yielded strong evidence in favor of the SS priors model (ΔBIC = −187.11). Using r as a measure of fit, across all three cues the fit of the model with SS priors to human ratings was excellent and slightly superior to that of the model with uniform priors (SS: r = .995, *RMSD* = 6.5; uniform: r = .992, *RMSD* = 7.3). After learning from noisy contingency inputs, the model with the SS prior ($\alpha = 5$) yielded a pattern very similar to that produced by human learners. For example, the estimated causal strength of cue A after block 1 was lower in the strong-B (−36.9) than the weak-B condition (−26.7). Importantly, this competition effect was maintained across the subsequent two blocks (−43.2 vs. −33.5 after block 2, and −45.5 vs. −37.2 after block 3). This pattern closely matches the averaged human ratings for cue A, yielding a good model fit based on the r measure ($r_A = .938$). These simulation results thus confirm that causal competition can be stable across multiple blocks when causal polarity is uncertain and the contingency data is noisy.

After learning with noisy input based on truncated Gaussian samples, the model with a uniform prior also yielded an apparent small competition effect for cue A (see Fig. 10). The reason again reflects the addition of noise to the contingency data. In particular, adding noise in the weak-B condition sometimes generated samples in which cue B (with veridical preventive power of −.25) was misperceived as a weak *generative* cause. In this situation, a Bayesian model will yield a higher estimate of the preventive power of cue A (which in essence is credited with preventive power that overrides the apparent generative power of A). Thus, the apparent small competition effect predicted by the model with uniform priors could be related to the phenomenon called "superconditioning" in the animal conditioning literature (Rescorla, 1971). However, the human data for cue A reveal a larger effect of the manipulation of the power of cue B than can be accounted for by the model with uniform priors.

8.5. Possible alternative models

In addition to the models described above, which assume noisy inputs to a single fixed causal structure, we also evaluated predictions from five alternative possible models. For simplicity we will report the simulation results for these alternative models only for Experiment 2 (as none showed promise of besting the models with SS priors described in the previous section).

8.5.1. Causal support

The first alternative we considered was to use a measure of causal support as a predictor of causal strength judgments, as advocated by Griffiths and Tenenbaum (2005) for the case of simple one-cue designs. For such designs, the support measure failed in a direct comparison with the SS priors model (Lu et al., 2008). As an account of the present findings with more complex multi-causal designs, causal support has a number of apparently fatal problems. First, the query used in the present study was

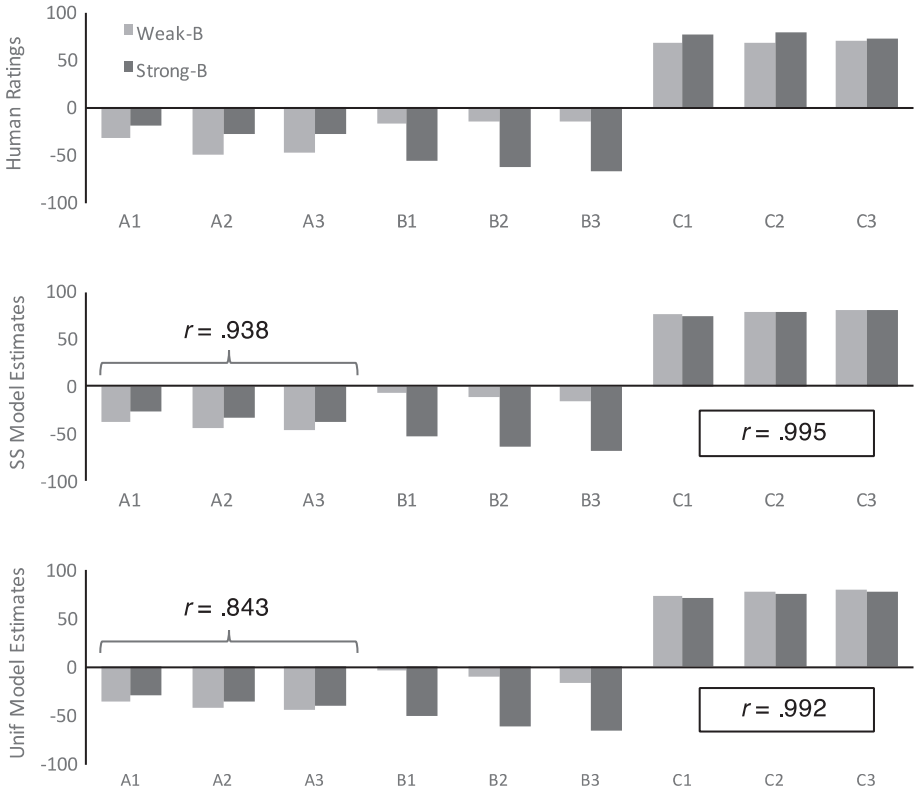

Fig. 10. Human ratings and model simulation results for Experiment 3 (causes A and B are preventive, C is generative). Modeling results are based on noisy inputs. Correlation values indicate correlations between model estimates and human data for cue A alone (indicated by brace) and for all cues (given in box on right).

aimed to elicit strength judgments, not structure judgments. Second, structure is defined by a ratio of the probabilities of graphs with and without a given causal factor. Thus a support value cannot be calculated for cue C (or takes a value of infinity), because although it is low in causal strength, cue C was constantly present in all our designs (generating a support ratio with 0 in its denominator). Third, a qualitative prediction from the support model is that the competition effect on cue A will *increase* across blocks. For Experiment 2, for example, the difference in support values between the weak-B and strong-B condition for cue A is 1.02 in block 1, 3.10 in block 2, and 4.39 in block 3. This predicted increase across blocks reflects the increase in certainty regarding a causal structure that accompanies increasing numbers of observations. This prediction is clearly inconsistent with human ratings.

8.5.2. Model averaging

A much more viable version of the same basic idea as causal support (i.e., assuming that participants may have considered multiple causal structures, rather than a single fixed structure) is model averaging. This alternative approach calculates expected causal strength weighted by the posterior probabilities of multiple structural models (i.e., models assuming that different combinations of cues A, B and C were effective causes). We implemented this model-averaging approach using both SS and uniform priors. For the data from Experiment 2, the version of model averaging based on SS priors outperformed that based on uniform priors (SS: $r = .994$; uniform: $r = .991$). However, neither variant yielded a clear advantage over the model using SS priors on a fixed causal structure (see Table 3), despite the greater computational complexity of models that assume multiple causal structures are

considered. For example, although the model-averaging approach predicted a competition effect for cue A, it had difficulty in accounting for the pattern observed for cue B. Participants overestimated strength in the low-B condition (−17.5) and underestimated it in the high-B condition (−55.5); by comparison model averaging assuming SS priors yielded predicted values of −6.4 and −53.2, respectively.

8.5.3. Jeffreys prior

A third alternative model we considered involved adopting a different non-informative prior, the Jeffreys prior, which is proportional to the square root of the determinant of the Fisher information regarding the model variables (corresponding in the present designs to causal strengths). This non-informative prior is considered "objective" because it does not depend on the specific set of parameter variables that is chosen to describe the parameter space (Jeffreys, 1946). In the present case, the Jeffreys prior for each cue follows a beta distribution of its causal strength, with two distribution parameters of .5. The Jeffreys prior thus corresponds to a preference for deterministic causes, favoring a causal strength of either 0 (certainly not a cause) or 1 (certainly generates the effect).

The model assuming the Jeffreys prior yielded a small competition effect in Experiment 2, for example, predicting a rating of 41.7 for cue A in block 1 for the weak-B condition and 39.2 in the strong B condition. However, rather than competition diminishing across blocks as shown in human ratings, the competition effect predicted by the model with Jeffreys prior actually increased modestly. For block 3 the Jeffreys-prior model predicted a rating of 49.7 for cue A in the weak-B condition and 45.6 in the strong-B condition. Based on this qualitative mismatch, the model with the Jeffreys prior can therefore be ruled out as unpromising. Quantitatively, simulation results yielded an entirely negligible correlation between the model prediction and human ratings for the critical cue A ($r_A = -.009$).

8.5.4. Strong-only prior

The fourth alternative model we considered uses a "strong-only" preference as the prior term. The strong-only prior favors causes with greater causal strength, and can be modeled using the equation $P(w_A, w_B, w_C) \propto e^{-\alpha(1-w_A)-\alpha(1-w_B)-\alpha(1-w_C)}$. However, the strong-only prior does not capture competition between causes, and therefore failed to predict the competition effect for the critical cue A ($r_A = -.08$ between model prediction and human ratings).

8.5.5. Linear regression

As discussed at the outset, the present experiments were designed using materials based on noisy-logical integration functions (Yuille & Lu, 2008). An alternative integration rule is the linear function that underlies linear regression, the delta-*p* model and the Rescorla–Wagner learning model (Danks, 2003; Stone, 1986). Extensive model comparisons performed in previous work (notably, using data from a meta-analysis reported by Perales & Shanks, 2007) have shown that models based on a linear function are generally less successful in accounting for human causal judgments based on binary variables (Lu et al., 2008). Nonetheless, we also modeled the data from Experiment 2 using coefficients from a linear regression performed on the contingency data. The resulting correlations with means for human judgments were substantial ($r = .91$, $r_A = .60$). As mentioned earlier, the present study employed a design that equated causal power, but not delta-*p* values, across conditions. Hence, the high correlations between human ratings and linear regression results are not surprising. However, a more refined quantitative examination revealed that the deviation scores for the fit of linear regression ($RMSD = 16.24$, $RMSD_A = 16.98$) were higher than those obtained for the best-fitting models shown in Table 3. More importantly, the linear regression model does not account for changes in performance across blocks.

In summary, none of the alternative models we considered proved competitive as an account of the data from the present study. Of course we cannot rule out the possibility that some other model we did not consider might explain our findings, but no obvious plausible alternative is apparent.

9. General discussion

9.1. Competition in multi-causal learning situations

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

The three experiments we report demonstrated competition between independently-occurring causes in causal strength judgments, for both generative and preventive causes. We also compared human performance to the predictions of alternative Bayesian models. Overall, the present findings support Lu et al.'s (2008) proposal that people's causal learning is guided by a generic prior that causes are sparse-and-strong. After a relatively small number of observations, participants underestimated a generative cause of moderate strength when a competing (but uncorrelated) cause was strong (Experiment 1). As learning continued, however, participants' causal strength judgments for a generative cause became less affected by competing causes (Experiment 2). This competition dynamic cannot be explained by naïve Bayesian models that assume uninformative priors (Busemeyer et al., 1993a). Experiment 3 examined potential competition among multiple preventive causes in a set-up in which generative and preventive causes were intermixed, and the polarity of each cause was initially unknown. Competition was found between preventive causes, which persisted across three blocks of learning trials.

We examined several alternative Bayesian models of causal learning, focusing on a comparison between models incorporating SS priors versus uninformative priors. Because the experimental designs involved complex multi-causal set-ups, with sequential presentation of learning trials, it proved necessary to model the possibility of input errors (i.e., distortions of the veridical contingencies due to attentional or memory factors). Relative to models assuming perfect inputs (i.e., the veridical contingency data), overall fits to the human pattern of causal judgments were considerably improved by modeling input errors based on a truncated Gaussian function. After taking account of input noise, the overall model fit was comparable based on SS versus uniform priors for the data of Experiments 2 and 3. But SS priors provided a better account to judgments about the critical cue A, which in all experiments was judged to be stronger in the weak-B than the strong-B condition, even though its veridical strength was identical in both conditions. Thus, formal models provided additional support for the competitive nature of causal strength judgments.

The present findings of competition between independently-occurring causes cannot be explained by accounts of causal learning based on either uniform priors or on alternative generic priors that do not predict causal competition. An example of the latter would be a preference for strong (approximately deterministic) causes unaccompanied by a preference for fewer causes. Yeung and Griffiths (2015) derived a prior of this sort by an empirical method of iterative learning applied to the simplest possible causal set-up (one candidate cause coupled with a generative background cause). The estimated prior appeared to favor a strong candidate cause, but was indifferent to the strength of the background cause. Yeung and Griffiths reported that their empirically-generated prior provided a good fit (superior to SS priors) for data from human judgments of causal strength across a wide range of causal contingencies, elicited in two experiments based on summary presentation of contingency data. The derived prior was not assessed against earlier data sets available from other labs (e.g., the meta-analysis of data from experiments using sequential presentation reported by Perales & Shanks, 2007).

Although iterative learning has provided an elegant empirical measure of generic priors for some learning situations (e.g., Kalish, Griffiths, & Lewandowsky, 2007), in Yeung and Griffiths' (2015) study the strength estimate did not converge for the background cause. The authors report (p. 10, footnote 6) that a mathematical analysis shows that obtaining sufficient human data to reach convergence would be impractical. Since iterative learning failed to yield a stable solution for one of the two causal cues in

the simplest possible design and their subsequent experiments did not elicit judgments about strength of the background cause, it is somewhat difficult to interpret these results. In addition, the potential for extending the iterative-learning method to more complex multi-causal set-ups of the sort investigated in the present paper (involving multiple candidate causes) appears to be inherently limited. Nonetheless, Yeung and Griffiths' findings are certainly consistent with the hypothesis that a preference for strong causes constitutes an important component of a generic prior on causal strength.

9.2. Simplicity as a constraint on priors for causal strength

As we discussed earlier, the general constraint of simplicity manifests itself in several distinct ways within the domain of causal judgments The sparse-and strong prior implies that effective causes will be individually strong but few in number. A causal model based on a small number of strong causes has several potential advantages. In the limiting case, assigning a strength of zero to a candidate cause is equivalent to determining that the cue is not an effective cause at all (i.e., making a structure judgment that excludes a link from the cue to the effect). The "sparse" component of sparse-and-strong priors can thus be viewed as a kind of soft approximation to an all-or-nothing judgment about whether a causal link exists. By encouraging a clearer separation between the strengths of the strongest and weakest causes, the competitive prior potentially prepares the reasoner to make heuristic decisions using on a reduced pool of potential causes. For example, under conditions that impose speed pressure or high cognitive load, a reasoner may elect to base causal inferences on just a small number (perhaps one) of strong causes, in lieu of invoking a more normative but resource-intensive mechanism that integrates information from a larger number of causes, most of them weaker.

An additional consideration is based on a common underlying principle involved in sensory information processing: information is represented by a relatively small number of simultaneously active neurons out of a large population, commonly referred to as *sparse coding*. In the area of vision, neural models that assume sparse coding have had considerable success (e.g., Graham & Field, 2006; Olshausen & Field, 1996, 1997). Among a variety of arguments favoring sparse coding (Olshausen & Field, 2004), a simplicity-related consideration is that sparse neural codes are energy efficient, as they minimize total neural activity (Levy & Baxter, 1996). In general, virtually all models of quantitative encoding in the nervous system imply that representations of higher values on a continuous dimension will involve a greater amount of neural firing than will lower values (e.g., Gati & Tversky, 1982; Restle, 1959; Stevens, 1967). For example, in Dehaene and Changeux's (1993) neuronal model of number representation, higher values will generate greater neural activity in the "summation cluster" that computes total numerosity.

Since causal strength is an instance of a quantitative value on a continuous dimension, the basic considerations that support the neural efficiency of sparse coding imply a preference for lower values of causal strength, with the "ideal" value being 0. Of course, if an effect occurs at all, then at least one cause must have non-zero strength in order to predict it (and as we noted earlier, a value of 1 maximizes precision in predicting the occurrence of the effect). Hence, sparse-and-strong priors represent a compromise between the efficiency of sparse coding and the predictive precision of strong causes.

9.3. Limitations and future directions

The present study represents an initial effort to specify detailed constraints that modulate the acquisition of causal knowledge in situations involving multiple candidate causes that may be either generative or preventive in nature. Our findings demonstrate that causal competition, previously observed in situations in which multiple causes are positively correlated with one another, can also be obtained when the causes are uncorrelated. The specific designs used in the present experiments are only examples of many possible contingency relationships that may arise among multiple candidate causes. It would be desirable, for example, to investigate whether competition emerges even among causal cues that are *negatively* correlated with one another (in the limit, for two candidate causes that never co-occur). In its current formulation, SS priors predicts potential competition even in such cases. However, the impact of priors will always tend to be more pronounced when the actual

data are ambiguous regarding the status of alternative causes (e.g., when the number of learning experiences is small, or in situations where candidate causes are seldom or never presented in isolation).

Another important direction for future research will be to explore possible differences in causal priors across judgment tasks, knowledge domains, individuals, and cultural groups. With respect to judgment tasks, Lu et al. (2008) found that the deterministic component of the sparse-and strong prior was given greater weight in judgments of causal structure than of causal strength. With respect to knowledge domains, Yeung and Griffiths (2015) found evidence that people tend to view causes as more deterministic in physical than in social domains. A limitation of the present study is that all the experiments used a single basic scenario and set of stimuli, with a medical concern (food poisoning) as the effect of interest. Holding the scenario constant enabled clearer comparisons among the design variations (e.g., generative versus preventive causes); however, no single scenario is sufficient to establish the generality of sparse-and-strong priors across different types of causal domains (e.g., those that involve causal factors that are primarily physical, biological, or social in nature). Further research will be required to assess the role of generic priors across different types of domains.

In the fields of social and organizational psychology, measures of individual differences in cognitive complexity have been shown to have consequences for performance in a variety of social judgment tasks. For example, individuals who score low on a measure of cognitive complexity tend to make more extreme ratings of target persons than do individuals who score higher in complexity (Frauenfelder, 1974). Conversely, those who score high in cognitive complexity tend to provide less extreme evaluations of both in-groups and out-groups (Ben-Ari, Kedem, & Levy-Weiner, 1992). Such findings raise the possibility that people may systematically vary in their generic priors, and that SS priors may be strongest for those least able or willing to grapple with the actual causal complexity of the world. It is also possible, as suggested by research on East–West differences in reasoning propensities (e.g., Choi, Nisbett, & Norenzayan, 1999), that different cultures foster alternative expectations about the simplicity of causal explanations (e.g., people from Asian cultures may be more willing to accept multiple weak causes of an effect). Rather than simply reflecting invariant constraints on causal induction, generic priors may be subtly shaped by prior experience, situational constraints, and more general cognitive propensities.

Author note

Preliminary versions of some of the work reported here were presented at the 35th Annual Conference of the Cognitive Science Society (Berlin, August 2013) and the 36th Annual Conference of the Cognitive Science Society (Quebec City, July 2014).

References

Ben-Ari, R., Kedem, P., & Levy-Weiner, N. (1992). Cognitive complexity and intergroup perception and evaluation. *Personality and Individual Differences, 13*, 1291–1298.

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. (2013). Inferential dependencies in causal inference: A comparison of belief-distribution and associative approaches. *Journal of Experimental Psychology: General, 142*, 845–863.

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.

Cheng, P. W., Liljeholm, M., & Sandhofer, C. M. (2013). Logical consistency and objectivity in causal learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 2034–2039). Austin, TX: Cognitive Science Society.

Cheng, P. W., Novick, L. R., Liljeholm, M., & Ford, C. (2007). Explaining four psychological asymmetries in causal reasoning: Implications of causal assumptions for coherence. In M. Rourke (Ed.). *Topics in contemporary philosophy: Explanation and causation* (Vol. 4, pp. 1–32). Cambridge, MA: MIT Press.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*(1), 47–63.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology, 47*, 109–121.

Danks, D., Griffiths, T. L., & Tenenbaum, J. B. (2003). Dynamical causal learning. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 67–74). Cambridge, MA: MIT Press.

Dehaene, S., & Chengeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*, 390–407.
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.
Frauenfelder, K. J. (1974). A cognitive determinant of favorability of impression. *Journal of Social Psychology, 94*, 71–81.
Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 325–340.
Graham, D. J., & Field, D. J. (2006). Sparse coding in the neocortex. In J. H. Kaas & L. A. Krubitzer (Eds.). *Evolution of the nervous system: Mammals* (Vol. 3). San Diego, CA: Academic Press.
Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334–384.
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.
Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General, 139*, 702–727.
Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 186*(1007), 453–461.
Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review, 14*(2), 288–294.
Kass, R. E., & Raferty, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–794. http://dx.doi.org/10.1002/0471667196.ess0985.
Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28*, 107–128.
Lagnado, D. A. (2011). Causal thinking. In P. McKay-Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 129–149). Oxford, UK: Oxford University Press.
Levy, W. B., & Baxter, R. A. (1996). Energy efficient neural codes. *Neural Computation, 8*, 531–543.
Liljeholm, M., & Cheng, P. W. (2007). When is a cause the "same"? Coherent generalization across contexts. *Psychological Science, 18*, 1014–1021.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*, 232–257.
Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. M. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 260–276). New York: Oxford University Press.
Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2015). A Bayesian theory of sequential causal learning and abstract transfer. *Cognitive Science*. http://dx.doi.org/10.1111/cogs.12236.
Lu, H., & Yuille, A. L. (2006). Ideal observers for detecting motion: Correspondence noise. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.). *Advances in neural information processing systems* (Vol. 18, pp. 827–834). Cambridge, MA: MIT Press.
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.
MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.
Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111*, 455–485.
Olshausen, B. A., & Field, D. J. (1996). Emergence of simple cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.
Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research, 37*, 3311–3325.
Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology, 14*, 481–487.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin & Review, 14*, 577–596.
Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65*, 429–447.
Rehder, B. (2009). Causal-based property generalization. *Cognitive Science, 33*, 301–343.
Rescorla, R. A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation, 2*, 113–123.
Restle, F. A. (1959). A metric and an ordering on sets. *Psychometrika, 24*, 207–220.
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.
Schultz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development, 77*, 427–442.
Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.
Spirkin, A. (1983). *Dialectical materialism*. Moscow: Progress Publishers.
Stevens, S. S. (1967). On the psychophysical law. *Psychological Review, 64*, 153–181.
Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435–502.
Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8*, 600–608.
Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*, 598–604.
Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology, 76*, 1–29.
Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, 333*, 71–74.
Yuille, A. L., & Lu, H. (2008). *The noisy-logical distribution and its application to causal inference. Advances in neural information processing systems* (Vol. 20). Cambridge, MA: MIT Press.