Cognitive Psychology 65 (2012) 381–413

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

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

Causal imprinting in causal structure learning

Eric G. Taylor *, Woo-kyoung Ahn

Yale University, Psychology Department, 2 Hillhouse Avenue, New Haven, CT 06520, United States

ARTICLE INFO

Article history:
Accepted 1 July 2012

Keywords:
Causal structure learning
Causal reasoning
Belief revision
Cognitive biases
Bayesian modeling

ABSTRACT

Suppose one observes a correlation between two events, B and C, and infers that B causes C. Later one discovers that event A explains away the correlation between B and C. Normatively, one should now dismiss or weaken the belief that B causes C. Nonetheless, participants in the current study who observed a positive contingency between B and C followed by evidence that B and C were independent given A, persisted in believing that B causes C. The authors term this difficulty in revising initially learned causal structures "causal imprinting." Throughout four experiments, causal imprinting was obtained using multiple dependent measures and control conditions. A Bayesian analysis showed that causal imprinting may be normative under some conditions, but causal imprinting also occurred in the current study when it was clearly non-normative. It is suggested that causal imprinting occurs due to the influence of prior knowledge on how reasoners interpret later evidence. Consistent with this view, when participants first viewed the evidence showing that B and C are independent given A, later evidence with only B and C did not lead to the belief that B causes C.

1. Introduction

Imagine you are a scientist investigating the causes of myopia (near-sightedness). You discover a correlation between using a nightlight as a child and developing myopia as an adult, and based on this correlation, you infer that nightlights are a cause of myopia. To support this inference, you also develop a theory that explains how light exposure during the evening could alter the eye's physiology (Quinn, Shin, Maguire, & Stone, 1999). A year later, however, new evidence emerges that a third factor—whether the child's parents have myopia—explains away the original correlation (Gwiazda, Ong,

* Corresponding author. Fax: +1 203 432 7172.
E-mail address: eric.taylor@yale.edu (E.G. Taylor).

0010-0285/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cogpsych.2012.07.001

Held, & Thorn, 2000). Parents who have myopia both tend to use a nightlight in their child's room and tend to pass along myopia to their children. For children whose parents do not have myopia, the original correlation disappears, undermining the belief that nightlights are a cause of myopia.

Examples like the one above (taken from a series of articles published in the journal *Nature*) are commonplace in science and in everyday causal reasoning. Presumably due to their abundance, students are repeatedly told in research methods and statistics courses that correlation does not imply causation and that they should look out for a hidden common cause that accounts for the observed correlation. Yet, as noted in the above examples, people frequently neglect such warnings. The main question we address in the current work is: Given that people frequently do infer causality from correlations without considering potential common causes, what happens to those beliefs when a hidden common cause is revealed?

One possibility is that people reconsider the initial evidence in light of the common cause, which leads them to discard their initial belief. For example, as the scientists did in the myopia scenario, reasoners might use their knowledge of parent myopia to discard their belief that nightlights cause child myopia. We refer to this behavior as *belief revision*. Another possibility is that reasoners do not reconsider the initial evidence, but instead maintain their initial belief and merely add the common cause relations. We refer to this behavior as *causal imprinting*, based on the idea that the initial evidence imprints a belief into the reasoner's mind, making it difficult to dispel despite the later evidence.

In what follows, we first review previous studies on causal structure learning and discuss the reasons to expect either belief revision or causal imprinting. Then, we present our empirical framework for distinguishing between these two outcomes, including a Bayesian analysis of the data we present to participants. Our Bayesian analysis shows under what conditions causal imprinting is normative, and the current study empirically tests these conditions.

1.1. Causal structure learning

Many previous studies have examined how people use covariation evidence to judge whether two events are causally related (Buehner, Cheng, & Clifford, 2003; Cheng, 1997). For example, one can use the data from a number of medical patients to judge whether taking a pill causes or prevents a headache. When learning only a single causal relation, people tend to give approximately normative judgments, consistent with statistical models of causal inference (Buehner et al., 2003; Griffiths & Tenenbaum, 2005, 2009; Rottman, Ahn, & Luhmann, 2011).

However, many causal judgments involve situations where there are more than just two events, and in these tasks people behave less normatively (Lagnado & Sloman, 2004, 2006; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). For example, Steyvers et al. (2003) found that fewer than half of their participants could distinguish a common cause structure (A causes B and C) from a common effect structure (B and C cause A) based on observations alone. Inferring the causal relations between multiple events is difficult because it requires verifying which events are merely correlated, and which are genuinely causally related (Pearl, 2000). As in the myopia example, a set of three events may all be correlated with each other because one is the common cause of the other two. In this case, one may perform numerous computations to verify the true causal structure: first, the overall contingency between each pair of events, and then, the conditional contingency of each pair of events, given the value of the third event (Scheines, Spirtes, Glymour, & Meek, 1994).

Despite these apparent difficulties, people do have knowledge of complex, real world causal systems, such as economic trends, weather patterns, and social hierarchies. Such learning may be possible because people assemble their causal models piece by piece, learning one relation at a time, rather than attempting to learn them all at once (Ahn & Dennis, 2000; Fernbach & Sloman, 2009).

If the learning of causal structures often occurs incrementally, then it is crucial to know how people update their initial causal beliefs when faced with new evidence. This is especially true when the new evidence reveals previously hidden causal factors that lead to different interpretations of the initial evidence. Though some work has addressed how people reason about hidden causes (Rottman et al., 2011), the current studies are the first to examine how learners update their beliefs when these hidden causes are revealed.

1.2. Belief revision or causal imprinting

Now we return to the myopia example, which illustrates the paradigm we will use to discriminate belief revision and causal imprinting. To rephrase this example using the abstract notations to be used in this paper, a learner is first presented with a positive contingency between events B and C, which leads the learner to believe that B causes C.[1] Later the learner is presented with contingencies between events B, C, and a new event A, which was not observed in the first phase. These data suggest that B is independent of C, conditional on A. Our primary question regards how people respond to the later evidence where the status of A was known. Do they engage in belief revision, reinterpreting the initial evidence based on the contingency patterns from the later evidence, or do they avoid such reinterpreting and instead show causal imprinting? The prior literature suggests that both outcomes are plausible.

First we consider reasons why people may show belief revision in the above paradigm. Note that in order for belief revision to take place based on the later evidence, a learner should first be able to notice that B and C are independent, conditional on A. Indeed, previous work has shown that people are adept at noticing when two events are spuriously correlated due to a confounding cause (Spellman, 1996; Waldmann & Hagmayer, 1995). For instance, Spellman (1996) had participants rate whether two liquids, one red and one blue, caused or prevented flower growth. Both liquids were correlated with flower growth, but the liquids were also confounded with one another. When controlling for the blue liquid, the red liquid was actually independent of flower growth. Learners seemed aware of this confound, as they rated the causal strength of the red liquid as much lower than the blue liquid (and generally null). Thus, these results suggest that people are sensitive to conditional contingencies. (For more direct demonstrations of common-cause learning, see the results from the control conditions reported in the current experiments.) Given that people are able to adjust their causal strengths ratings between two events based on conditional contingencies, the question is now whether they use this type of data when it appears more recently (i.e., during the second half of learning) to revise their initial beliefs about the two correlated effects.

Studies on the learning of causal relations between just two events have shown that causal strength estimates are often affected more by recent evidence than earlier evidence (Fernbach & Sloman, 2009; Glautier, 2008; Lopez, Shanks, Almaraz, & Fernandez, 1998). For instance, participants in Fernbach and Sloman (2009, Experiment 2) viewed five trials of contingency data between three events, and the fourth trial showed evidence inconsistent with a causal relation implied by the earlier trials. Participants tended to exclude the relation implied by the earlier trials in their causal judgments, showing a recency effect.

Though Fernbach and Sloman (2009) used learning sequences with only five trials, others have shown recency effects with much larger data sets, more similar to our myopia example. For example, Lopez et al. (1998) manipulated the order of two consecutive blocks, one in which an event (say, X) was a good predictor of another event (say, Y) and another in which X was a poor predictor of Y. Participants rated the relationship between X and Y to be stronger when they viewed that X was a good predictor recently than when they viewed that X was a poor predictor recently, suggesting that they had weighted the recent evidence more heavily. These studies suggest that in the context of the myopia example, people may give more weight to recent evidence, leading to weaker belief that B causes C after observing that A causes both B and C.

Not only do people give more weight to recent evidence, they also sometimes spontaneously revise their prior beliefs in light of more recent evidence. For example, in backward blocking paradigms (Shanks, 1985) learners first observe that two cues, X and Y, are positively associated with an outcome, Z. Then in a second phase, they observe that X is also positively associated with Z in the absence of Y. Measures of the learned association between Y and Z tend to decrease from the first phase to the second, suggesting revision. Adults, children, and even rats show this tendency (Miller & Matute, 1996; Shanks, 1985; Sobel, Tenenbaum, & Gopnik, 2004). Along with the previous studies, these findings suggest that people may be willing to revise their prior beliefs based on later evidence.

[1] Positive covariation between B and C can also imply that C causes B, but for simplicity, we only consider the case in which people can readily infer only one causal direction, as in the myopia example where B temporarily preceded C.

Nonetheless, we predicted that people would show causal imprinting rather than belief revision, based on a simple principle that people learn things in order to use that knowledge in the future. Thus, once a person acquires a certain causal belief, it will be used to interpret later evidence (e.g., Luhmann & Ahn, 2011), including evidence that should be used to revise that belief. For example, in our paradigm the initial evidence will lead to a belief that B causes C, which will then lead to biased interpretations of the later evidence. In effect, this would delay the realization that B and C are independent, given A, and reduce the likelihood that learners would reinterpret the initial evidence based on the later evidence. In general, we claim that this use of prior knowledge creates an asymmetry in how initial and later pieces of evidence are used in learning. The initial evidence will have a greater impact because it leads to the formation of prior beliefs, which then affect how later evidence is interpreted.

There are a number of previous demonstrations of such an asymmetry in learning and reasoning tasks. For instance, in the well-known studies of "confirmation bias", participants often ignore or rationalize later contradictory evidence in an attempt to confirm their original hypotheses (for a review, see Nickerson, 1998). Prior beliefs also affect the learning of categories and concepts. For example, the first category we assign to an item may affect what features and causal relations we encode about the item, and these may bias us from seeing other possible categorizations (e.g., Moreau, Lehmann, & Markman, 2001).

Also, in associative learning, studies of forward blocking (Kamin, 1968) show that prior associations between a cue and an outcome impede the learning of additional associations between other cues and the same outcome. Applied to our paradigm, if people first infer that B causes C, this may block their learning that A causes C, as A and B represent two alternative causes for the same effect (Waldmann & Holyoak, 1992). Failing to learn that A causes C would impede learning of the common cause structure that A causes B and C (the true structure in our paradigm), and if the common cause structure is not inferred, then there is no basis for discounting the correlation between B and C.[2]

Studies on the learning of causal relations in particular have also found primacy effects, where initial evidence influences causal judgments more heavily than later evidence. For instance, Dennis and Ahn (2001) showed participants a set of data where the overall contingencies between two events changed from the first half to the second half (either from a positive contingency to a negative contingency, or the reverse). Collapsing across the two halves, the data revealed no contingency between the candidate cause and the effect. However, causal strength judgments taken at the end of learning were positive when a positive contingency was presented in the first half, and negative when a negative contingency was presented in the first half. That is, people showed a primacy effect (see also Hogarth & Einhorn, 1992).

More relevant to our claims regarding interpretation is a recent study by Luhmann and Ahn (2011) showing that primacy effects are due in part to learners forming a causal belief based on the first half of the data, which then creates a bias when interpreting the second half. For instance, they found that when a learner believed that one event X caused another event Y, a later trial in which X was followed by the absence of Y was interpreted as the operation of a separate inhibitory cause or lack of enabling condition, but not as evidence that X prevents Y. In addition, Marsh and Ahn (2006) argued that the aforementioned recency effects in Lopez et al. (1998) were obtained because the task was too complex, inhibiting learners' ability to form and maintain the initial causal belief.

While primacy effects (Dennis & Ahn, 2001) and dynamic interpretations during causal and category learning (Luhmann & Ahn, 2011; Moreau et al., 2001) are consistent with the current claim, none of these studies necessarily imply that primacy effects will occur in the myopia scenario, due to qualitative differences in this scenario in the way that the later evidence contrasts with earlier evidence. In the previous studies demonstrating primacy effects in causal learning for instance (e.g., Dennis & Ahn,

[2] The relations between forward blocking and causal imprinting are more conceptual than literal. Blocking has been shown in a variety of ways (e.g., Chapman & Robbins, 1990; Dickinson, Shanks, & Evenden, 1984; Kruschke & Blair, 2000), but in typical studies the two competing cues are confounded, such that the later cue cannot increase predictive accuracy (Rescorla & Wagner, 1972). In contrast, events A and B in our experiments were not confounded, and adopting A as a cause of C would have increased predictive accuracy. Nonetheless, blocking may occur in a more general sense in our scenario. For example, the initial account for the B–C contingency (i.e., B is the cause of C) may block the learning of the latter common cause account. To our knowledge, such forms of blocking have not been empirically demonstrated.

2001), participants viewed two consecutive blocks of data showing opposite contingency patterns between the same two events. In contrast, in the current paradigm, like the myopia scenario, the contingency between the two initial events, B and C, does not change over time, but instead, a hidden variable is revealed that leads to an alternative interpretation of that contingency. In cases like this, people may be more open to revising their initial beliefs as they should later realize that it was based on incomplete information.

To summarize, one set of studies (e.g., learning of conditional contingencies, backward blocking, and recency effects in causal learning) suggests that people may show belief revision rather than causal imprinting. On the other hand, another set of studies (e.g., confirmation bias effects, forward blocking, primacy and interpretation effects in causal and category learning) suggests that people may have a bias to maintain their initial beliefs when learning causal structures. That is, people may show causal imprinting rather than belief revision. Indeed, some investigators have even found evidence for both primacy and recency effects in the same experiment (Danks & Schwartz, 2005, 2006). Most importantly, none of the previous studies have utilized a paradigm like the one depicted in the myopia scenario. As discussed earlier, the later discovery of a common cause is recurrent in real life and scientific reasoning, and this sequence of events seems more ecologically valid than the method of joining two blocks of data with opposing contingencies used in studies of primacy and recency effects. Given that belief revision and causal imprinting are each consistent with some prior research, and that none of the previous paradigms directly apply to the current one, new empirical tests are needed to distinguish between these hypotheses.

2. The current studies

In this section, we describe the key aspects of the current studies in order to derive the specific predictions of causal imprinting and belief revision. Participants observed sets of contingency data and then judged three possible causal relations among events A (Ablique virus), B (Burlosis condition), and C (Caprix condition), as shown in Fig. 1. Specifically, they viewed 40 trials (see the left panel of Fig. 2), each depicting an individual person's status on the two or three variables, depending on the condition.

The two critical conditions utilized two blocks of contingency data. The joint frequency distributions across the 20 trials in each block are summarized in Fig. 2. One block of data depicted only B and C (henceforth, the BC block), and participants were told nothing about event A. The BC block corresponds to a case in which a researcher observes a positive contingency between the use of nightlights and development of myopia in children. A common measure of contingency, ΔP (Jenkins & Ward, 1965), reveals a positive relation between events B and C.

The other block depicted A, B, and C (henceforth, the ABC block) and corresponds to a case in which a researcher also recorded whether the children's parents had myopia. In this block, the overall ΔP between B and C was identical to the BC block. However, for trials where A was present, and separately for trials where A was absent, the ΔP between B and C was near zero. Thus, the data from the ABC block suggested that B was not a cause of C. Additionally, the ΔP between A and B, and between A and C, was highly positive, suggesting that A was the common cause of B and C.

To distinguish between belief revision and causal imprinting, we manipulated which block(s) participants viewed: either the BC block followed by the ABC block (henceforth, the BC–ABC condition), or

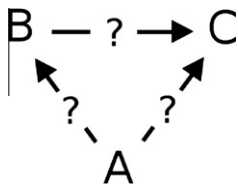

Fig. 1. Potential causal relations among events A, B, and C that participants in Experiments 1–4 had to judge by viewing contingency data.

Trial	A	B	C
1	1	1	1
2	1	1	1
3	0	0	0
4	1	1	1
5	0	1	0
6	0	0	0
7	0	0	0
8	1	1	1
9	0	0	0
10	0	0	1
11	1	1	1
12	1	1	0
13	0	0	0
14	0	0	0
15	0	0	0
16	1	0	1
17	1	1	1
18	1	1	1
19	0	0	0
20	1	1	1

BC block contingency

	C	~C
B	8	2
~B	2	8

$$\Delta P_{BC} = P(C|B) - P(C|{\sim}B) = \frac{8+0}{8+0+1+1} - \frac{1+1}{1+1+0+8} = 0.6$$

ABC block contingencies

	A, C	~A, C	A, ~C	~A,~C
B	8	0	1	1
~B	1	1	0	8

$$\Delta P_{BC|A} = P(C|A\&B) - P(C|A\&{\sim}B) = \frac{8}{8+1} - \frac{1}{1+0} = -0.11$$

$$\Delta P_{BC|{\sim}A} = P(C|{\sim}A\&B) - P(C|{\sim}A\&{\sim}B) = \frac{0}{0+1} - \frac{1}{1+8} = -0.11$$

$$\Delta P_{AB} = P(B|A) - P(B|{\sim}A) = \frac{9}{9+1} - \frac{1}{1+9} = 0.8$$

$$\Delta P_{AC} = P(C|A) - P(C|{\sim}A) = \frac{9}{9+1} - \frac{1}{1+9} = 0.8$$

Fig. 2. Contingencies used in Experiments 1–4 for the BC and ABC blocks. The table on the left shows presence (value 1) or absence (value 0) of the three events on each of the 20 trials. The BC and ABC blocks were identical except that in the BC block, values of A were missing (indicated by gray shading in the table). Calculations of ΔP are provided for all causal relations depicted in Fig. 1. The subscripts used with ΔP (e.g., ΔP_{BC}) indicate that it is a measure of the second factor (C) being contingent on the first factor (B). For the ABC block, ΔP for the B-causes-C relation is calculated separately for trials with A present and with A absent, indicating the lack of evidence for this relation when conditionalizing on A.

two consecutive ABC blocks (henceforth, the ABC–ABC condition). The second ABC block was included to control for the total sample size. Then, after viewing the contingency data, participants judged which of the causal relations they believed were true (Experiment 1) or how strongly each of the causes led to their effects (Experiment 2), yielding both holistic, structural judgments and more fine-tuned, strength judgments (see Griffiths & Tenenbaum, 2005).

We now present our predictions for the two conditions. First, we predicted that participants in the ABC–ABC condition would infer that A causes both B and C, and that B does not cause C (e.g., Spellman, 1996; Waldmann & Hagmayer, 1995). We also acknowledge that conditionalizing on a third variable is difficult and that some learners may fail to conditionalize on A (e.g., Steyvers et al., 2003), or perhaps do so for only some of the trials, resulting in a weak belief in B-causes-C. Both of these structures are illustrated in the top left panel of Fig. 3, where we summarize our predictions.

Second, if learners in the BC–ABC condition engage in belief revision, then we predicted that they would give similar causal judgments to those in the ABC–ABC condition. For example, they might assume that the contingencies from the BC block would have been similar to the ABC block if the values of A were visible. In this case, they would likely reinterpret the BC block based on the contingency patterns from ABC block and revise their initial belief that B causes C.

Third, if learners in the BC–ABC condition show causal imprinting, then they should be more likely to believe that B causes C than those in the ABC–ABC condition. This would occur if the initial evidence truly imprints learners with the belief that B causes C, making them hesitant to revise this belief.[3]

[3] Another reason that the causal imprinting pattern may occur is if learners are simply uncertain about the missing values of A in the BC block even after they observe the ABC block. If so, they should avoid reinterpreting the BC block, meaning that the data from this block would still support the belief that B causes C. We address this possibility in the modeling section below and in Experiment 2.

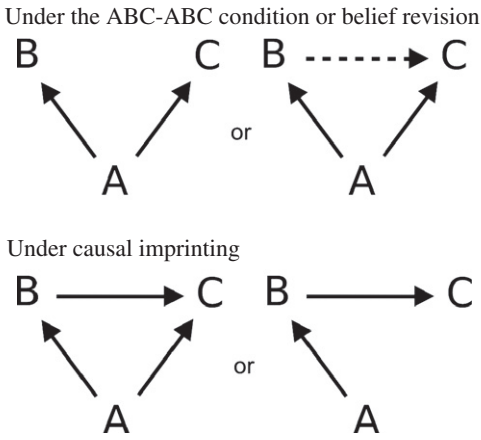

Fig. 3. A summary of the descriptive predictions according to belief revision or for the ABC–ABC condition (above) and causal imprinting (below). The dotted line represents a specifically weak causal relation.

Finally, let us consider the causal inferences regarding event A if causal imprinting occurs. Because A positively correlates with B and C in the ABC block, participants will have to account for these contingencies. There are two ways for learners in the BC–ABC condition to account for the positive A–B and A–C contingencies, while also maintaining their belief that B causes C. One is to infer that A causes both B and C, in addition to B strongly causing C (i.e., the bottom left panel of Fig. 3). Another is to infer that A causes only B, and that B strongly causes C, resulting in a causal chain structure (i.e., the bottom right panel of Fig. 3). This structure may occur due to blocking of the belief in A causes C, given the prior learning of B causes C (Waldmann & Holyoak, 1992). Note that even though the chain excludes A-cause-C, it still accounts for all observed pairwise contingencies, because A would also share a positive contingency with C due to the causal chain relationship from A to B to C. Although this structure and the previous one do account for the additional positive contingencies between A and B, and A and C, both ignore the fact that the contingency between B and C disappears when conditionalizing on A.

To summarize, learners in the ABC–ABC condition and those in the BC–ABC condition engaging in belief revision should infer the common cause structure, or all three causal relations endorsing B-causes-C as only a weak causal relation. Under causal imprinting, however, learners should infer all three causal relations or the causal chain structure, both with a strong B-causes-C relation. Thus, comparing the top and bottom panels of Fig. 3, the two structures that best distinguish between belief revision and causal imprinting are the common cause structure and the causal chain structure. And most critically, the endorsement of B-causes-C should be generally much stronger under causal imprinting than under belief revision (or for the ABC–ABC condition).

3. Bayesian analyses

3.1. Preview of Bayesian analyses

In this section, we show how a Bayesian learner would assign likelihoods to the various possible causal structures, given the data in the ABC–ABC and BC-ABC conditions. As in previous literature (Anderson, 1990; Griffiths & Tenenbaum, 2009), we treat Bayesian analyses as providing rational, or normative standards for human judgments. For readers who may wish to skip the details of the analyses, we first provide a brief overview of the results and the normative criteria they suggest.

The Bayesian analysis of the ABC–ABC condition reveals that the common cause structure is most likely, followed by the structure with all three causal relations. These normative predictions are the same as the descriptive predictions shown in Fig. 3.

In contrast, the analysis of the BC–ABC condition differs depending on how we assume that the learner utilizes the data from the BC block. We considered three general methods for utilizing the BC block, two normative and one boundedly normative.

The first normative method corresponds to belief revision, where the learner reinterprets the BC block based on the contingencies from the ABC block, assuming that both the BC and ABC block are representative samples of the general contingency patterns between events A, B, and C. This assumption is sensible, given that the contingency between B and C is identical across the two blocks. Thus, the learner can assume that the BC block would have been similar (or identical) to the ABC block had the values of A been visible. The results from such an analysis would thus be similar (or identical) to the analysis of the ABC–ABC condition, and the normative predictions for the BC–ABC condition would differ from the causal imprinting predictions summarized in Fig. 3. Experiment 1 tests whether people deviate from these normative predictions.

The second normative method corresponds to a case where the learner remains uncertain about the missing values of A in the BC block, and thus, avoids inferring any specific values of A based on the contingencies from the ABC block, even after observing the ABC block. Instead, the learner considers both possible values of A and collapses across these to determine the overall likelihood of the BC trials. This is normative if the learner cannot be sure that the BC block would have been similar to the ABC block. When using this method, the common cause structure is slightly less likely than in the ABC–ABC analysis, and both the causal chain and the structure with all three relations are slightly more likely than the ABC–ABC analysis. In other words, the normative inferences from the BC–ABC condition assuming uncertainty for A values in the BC block result in a minor tendency towards the causal imprinting pattern shown in Fig. 3. Critically, however, even this small shift should occur only if a learner is truly uncertain about the missing A values after viewing the ABC block. Experiment 2 tests this possibility.

The third method, which we termed a "bounded" Bayesian method, is similar to the fully Bayesian method in that the learner remains uncertain about the missing values of A in the BC block even after observing the ABC block. However, unlike the fully Bayesian method, the bounded Bayesian learner is cognitively limited, and thus during the BC block the learner only infers whether or not B causes C, without considering both possible values of A. This means that the learner does not consider the possibility that the positive B–C contingency during the BC block may have been due to a common cause. Consequently, the bounded Bayesian analysis leads to a stronger belief about B-causes-C during the BC block than does the fully Bayesian analysis, which leads to the final inferences being much more similar to causal imprinting than in the fully Bayesian analysis. However, as with the fully Bayesian analysis, the ability of this model to justify causal imprinting rests on a crucial assumption that the values of A are uncertain during the BC block. Thus, if the missing values of A during the BC block were revealed to be same as the values of the ABC block, then learners should not show causal imprinting as predicted by our bounded Bayesian model. Experiment 2 tests this possibility.

3.2. Modeling methods

Our approach follows previous work on causal learning using causal graphical models, or "Bayes nets" (Pearl, 2000), a powerful way to represent causal knowledge and perform related computations. A Bayes net consists of nodes, which stand for events, and directed edges, which stand for causal relations between events. When a node becomes active (e.g., event A occurs), its effects also become active (e.g., event B occurs) with some probability defined for their causal relation. These basic properties allow one to use Bayes nets to make statistical inferences, e.g., to compute the probability that a particular event (or set of events) will occur, the probability that a causal relation exists between two events, or the likely strength of a causal relation (Cheng, 1997; Griffiths & Tenenbaum, 2005, 2009).

In this section, we use Bayes nets to infer the likelihoods of the causal relations depicted in Fig. 1, given the data provided in our experiments. Inferences from a Bayes net are made according to probability theory. Bayes' rule states that a learner's prior belief in a particular causal hypothesis, or a set of causal relations, h_i, should be updated based on the data, D, in the following way:

$$P(h_i|D) = \frac{P(D|h_i)P(h_i)}{\sum_i P(D|h_i)P(h_i)}, \quad (1)$$

Table 1
The posterior probabilities of each causal structure, separately for the ABC–ABC condition and the three analyses for the BC–ABC condition. Bolded values are maxima.

								
ABC–ABC (No missing values)	0.00	0.00	0.00	**0.72**	0.00	0.00	0.00	0.28
BC–ABC (Reinterpretation)	0.00	0.00	0.00	**0.72**	0.00	0.00	0.00	0.28
BC–ABC (Fully Bayesian)	0.00	0.00	0.00	**0.56**	0.00	0.00	0.03	0.41
BC–ABC (Bounded Bayesian)	0.00	0.00	0.00	0.01	0.00	0.02	0.18	**0.79**

where $P(D|h_i)$ is the likelihood of the data according to hypothesis h_i, and $P(h_i)$ is the prior probability of hypothesis h_i, or the degree of belief in h_i before viewing the data. The numerator computes the weighted likelihood of each hypothesis, and the sum in the denominator normalizes these weighted likelihoods so that they sum to one. The result, $P(h_i|D)$, is the *posterior* probability of hypothesis h_i after observing data D.

Given the three causal relations shown in Fig. 1, there were eight possible hypotheses (see Table 1), consisting of each possible combination of the causes. In our analyses, we set all $P(h)$ to 1/8, representing the assumption that the hypotheses were equally likely before viewing the data.[4]

The data, D, in our experiments consist of multiple trials. In standard Bayesian inference, one computes the posteriors by applying Bayes' rule iteratively across trials. Specifically, the posterior for trial t, $P(h_i|dt,\ldots,d_1)$, is used as the prior probability, $P(h_i)$, on trial $t+1$. Conveniently, the result of this iterative process is equivalent to setting $P(D|h_i)$ to the joint likelihood of all trials, given hypothesis h_i. This joint likelihood is simply the product of the likelihoods of all individual trials. Note that in computing the product, the order in which the data are presented becomes irrelevant for Bayesian updating (see Kruschke, 2006; Slovic & Lichtenstein, 1971).

We explain how we computed the likelihood $P(d_t|h_i)$ using an example trial in which A and B are present but C is absent. We compute the likelihood for the specific hypothesis 'A-causes-C and B-causes-C.'

First, we take the product of the likelihoods of each event being present or absent, given whether or not the direct causes of those events were present or absent:

$$P(d_t|h_i) = \prod_{e\in\{A,B,C\}} P(e|\text{causes}_e), \tag{2}$$

where e represents the status of a given event (i.e., present or absent), and causes_e represents the status of the direct causes of e indicated by hypothesis h_i. In our example, this equals $P(A^+)P(B^+)P(C^-|A^+B^+)$, where A^+ and B^+ indicate that A and B are present, and C^- indicates that C is absent. In our example, hypothesis h_i does not suggest any candidate causes of A and B, meaning that $P(A^+)$ and $P(B^+)$ represent the probability that A and B are caused by some other event in the causal background (i.e., an event other than A, B, and C).

Second, to compute the likelihood of an event being present, conditional on the status of its causes, we used a noisy-OR function[5]:

$$P(e^+|\text{causes}_e) = 1-(1-b_e)\prod_{e\in\text{causes}_e}(1-m_{ce})^{c_{present}}, \tag{3}$$

where c is a direct cause of e, m_{ce} is the probability that the causal mechanism from cause c to effect e succeeds (*the causal power*; Cheng, 1997), c_{present} is an indicator variable set to 1 when cause c is

[4] We also considered prior distributions reflecting a preference for hypotheses with either more or fewer causes (see Fernbach & Sloman, 2009). These different priors did not fundamentally alter the relative differences between the model predictions (see Appendix A).

[5] The noisy-OR function is a method for combining the influence of multiple causes when they lead to the same effect. An important alternative is the linear method (for discussion, see Griffiths & Tenenbaum, 2005), where $P(e^+|\text{causes}_e) = b_e + \Sigma_c m_{ce} c_{\text{present}}$. Explorations of the model using the linear method produced similar results to those with noisy-OR (see Appendix A).

present and 0 when c is absent, and b_e is the probability that event e is caused by an event in the causal background.

Thus, according to the hypothesis 'A-causes-C and B-causes-C,' the likelihoods of the events in our example are: $P(A^+) = 1 - (1 - b_A)$, $P(B^+) = 1 - (1 - b_B)$, and $P(C^-) = 1 - P(C^+) = (1 - b_C)(1 - m_{AC})(1 - m_{BC})$. The joint likelihood of these events is: $[1 - (1 - b_A)][1 - (1 - b_B)][(1 - b_C)(1 - m_{AC})(1 - m_{BC})]$.

To ensure that our results did not depend on any particular values of the b and m parameters, we marginalized across these parameters using Monte Carlo (for a similar application, see Griffiths & Tenenbaum, 2005). Specifically, we drew 100,000 samples of the parameter set $\theta = \{b_A, b_B, b_C, m_{AB}, m_{CA}, m_{CB}\}$, each parameter drawn from a uniform distribution with values ranging from 0 to 1.[6] With each set of sampled parameters, θ_j (where j refers to the sample number) we computed the likelihood, $P(D|\theta_j, h_i)$, for each hypothesis. To approximate the likelihoods required by Eq. (1), $P(D|h_i)$, we took the average of the sampled likelihoods, as follows:

$$P(D|h_i) \approx \frac{1}{m}\sum_{j=1}^{m} P(D|\theta_j, h_i), \tag{4}$$

where m = 100,000. The likelihoods for all eight hypotheses were used to compute the posteriors of each hypothesis according to Eq. (1).

3.2.1. Different methods of utilizing the BC block

In typical applications of causal Bayes nets, and in the ABC–ABC condition, the presence or absence of all variables in question (i.e., A, B, and C) is specified on all trials. However, in the BC–ABC condition the status of A was unknown during the BC block. We consider three different methods of utilizing the BC block in light of the missing values of A. We also discuss under what circumstances each of these models would be considered normative.

Our first normative method captures the belief revision approach. Under this method, the learner assumes that the contingencies from the BC block would have been similar to the ABC block had the values of A been visible. Accordingly, the learner reinterprets the BC block based on the evidence from the ABC block. A simple way to apply this method in our case would be for the learner to assume that the BC block was in fact identical to the ABC block, since the B–C contingencies were identical across the two blocks.[7] Note that this approach does not require an exact memory for the BC block, but rather, only an assumption that if the learner were to re-view the BC block with the values of A visible, it would appear identical to the ABC block. Thus, for this method the analysis was conducted using two ABC blocks, yielding results identical to the ABC–ABC analysis.

Our second normative method captures the assumption that learners remain uncertain about the values of A in the BC block. The fully Bayesian method of honoring this uncertainty is to compute the likelihood of a given BC trial under both possibilities (i.e., A^+ and A^-) and take the sum[8]:

$$P(B, C|h_i) = P(B, C, A^+|h_i) + P(B, C, A^-|h_i) \tag{5}$$

Intuitively, the learner considers both possible settings of A and then collapses across them to assess the overall likelihood of a given structure leading to the observed values of B and C.

[6] We also drew samples of θ restricting the range of the b parameters from to 0 to 0.1 (see Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008). These samples led to a slightly stronger preference for the structure with all three relations (see Appendix A), but the relative differences between analyses remained the same.

[7] More formally, one may use the contingencies from the ABC block to derive inferences for the missing status of A during the BC block (e.g., Anderson, 1991). For example, one may reason that the probability of A being present (though hidden) on a BC block trial with B present and C absent equals $N(A^+, B^+, C^-)/N(B^+, C^-)$, where $N(A^+, B^+, C^-)$ is the number of trials from the ABC block with A present, B present, and C absent, and $N(B^+, C^-)$ is the total number of trials from the ABC block with B present and C absent, ignoring the values of A. Note, these are only point estimates and are somewhat uncertain based on the low number of trials. Yet, they are also the maximum likelihood estimates based on the ABC block. Using this method to compute the expected number of trials with A present, per trial type (i.e., setting of B and C), leads to inferring a copy of the ABC block, and is thus, equivalent to the conceptually simpler method of assuming that the blocks would have been identical.

[8] We thank an anonymous reviewer for suggesting this approach.

Both methods can be considered normative, but each under different circumstances. Specifically, if the learner is told that the BC and ABC blocks are both representative samples of the contingency patterns between events A, B, and C, then strict uncertainty regarding A is not necessary, and indeed, may underutilize the information about the A–B–C contingencies from the ABC block. For example, after reading the later myopia study showing that nightlights and myopia are conditionally independent, it seems to us that the normative inference is that nightlights do not cause myopia, given that both research teams evaluated large random samples of the population. Hence, the former method may be considered normative. However, if the learner is told nothing about where the two blocks of data came from, then some uncertainty regarding A is prudent and the latter method would be considered normative. We acknowledge that not all circumstances will fall neatly into one classification or another, but some cases can clearly identify one method as uniquely normative. Experiment 2 utilizes such a case where causal imprinting is unambiguously non-normative in order to test whether participants are sensitive to these distinctions.

Finally, our bounded Bayesian analysis captures a learner that neither thinks back to the BC block to infer values of A, nor incorporates uncertainty regarding A by considering both possible values of A, due to cognitive limits. We argue that these are reasonable assumptions to make about human cognition, and that within these limits a normative Bayesian approach can still be applied. Instead of considering the possible values of A when it was missing, the learner utilizes the BC block by computing posteriors for only two hypotheses: B-causes-C, and B does not cause C. Then, when starting the ABC block the learner uses these posteriors as the prior probabilities for the eight new structures that include all three causes (shown as column labels in Table 1). Specifically, the prior probabilities for the four structures with B-causes-C (the last four structures of Table 1) are set to the posterior of the structure with B-causes-C obtained from the BC block. Likewise, the priors for the four structures excluding B-causes-C (the first four structures of Table 1) are set to the posterior of the structure where B does not cause C obtained from the BC block. This method will have direct consequences for the common cause structure and the structure with all three relations. For example, suppose the posterior for the structure with B-causes-C was 0.75, and the structure where B does not cause C was 0.25. Then at the start of the ABC block, 0.75 would become the prior probability of the structure with all three relations. Likewise, 0.25 would become the prior for the common cause structure. This will cause the learner to favor structures with B-causes-C, a bias that as we will see may not be fully overcome by viewing the data from ABC block.

3.3. Modeling results

The posterior probabilities of the eight hypotheses are presented in Table 1. Row 2 represents the predictions for the ABC–ABC data, and rows 3–5 represent the three ways of utilizing the BC block in light of the missing values of A.

For the ABC–ABC condition and the BC–ABC condition assuming that the BC block is reinterpreted, the most likely structure was the common cause structure (0.72), followed by the structure with all three relations (0.28). We note that these two normative predictions match the descriptive predictions of the ABC–ABC condition and belief revision as summarized in Fig. 3.

One may wonder why, from a normative perspective, the structure with all three relations is likely at all for the ABC–ABC condition, given that B–C contingency is near zero, conditional on A. A closer look at our 100,000 samples revealed that this structure received a high likelihood mostly for samples in which the causal strength parameter for B-causes-C was low. A structure that assigns a low causal strength for B-causes-C (e.g., less than 0.1) yields very similar predictions in terms of what trials are likely to appear compared to a model with no B-causes-C relation at all, such as the common cause structure. Crucially, when the causal strength parameters for all three relations were high, the structure with all three relations received a very low likelihood.

For the BC–ABC condition assuming uncertainty about the missing values of A (shown as "Fully Bayesian" in Table 1), the most likely structure was the common cause (0.56), followed by the structure with all three relations (0.41), followed by the causal chain, 'A-causes-B and B-causes-C' (0.03). The fully Bayesian analysis predicts a pattern similar to the ABC–ABC analysis but with a slightly greater preference for structures with B-causes-C (i.e., all three relations and causal chain), and also

a slightly reduced preference for the common-cause structure compared to the ABC–ABC analysis. This occurs because during the BC block, there is no direct evidence that B and C are independent, given A, and thus the conditional independence implied from the ABC block is obscured in the final judgments. Yet, even assuming uncertainty about missing values of A, this method leads to a clear preference for the common cause, which is unlikely if causal imprinting occurs.

Finally, for the bounded Bayesian analysis of the BC–ABC condition, the most likely structure was the structure with all three relations (0.79), followed by the causal chain (0.18). The posterior for the common cause was very low (0.01). Thus, this method leads to a more pronounced form of causal imprinting, as all of the emphasis is now away from the common cause, and directed toward the two models including B-causes-C (similar to the descriptive predictions for causal imprinting in Fig. 3).

This last method shows that causal imprinting may be considered boundedly normative, given that a similar pattern results from a normative Bayesian analysis when adding the assumption that learners neither reinterpret the foregone BC block, nor consider both possible values of A, due to cognitive limits. Note, however, that if these assumptions about cognitive limitations no longer provided a valid justification for causal imprinting, it would be difficult to argue for the normative basis of causal imprinting. For instance, if the missing values of A in the BC block were later revealed, and the learner was informed that the BC block was in fact the same as the ABC block that they were about to view, then causal imprinting should not occur. In this case, the normative predictions should be given by the ABC–ABC analysis. We explore this possibility directly in Experiment 2.

4. Experiment 1: Causal structure judgments

In Experiment 1, participants viewed contingency data regarding events A, B, and C, either in the BC–ABC order or the ABC–ABC order. Then, they selected which of the eight possible causal structures from Table 1 they believed were most likely to be true.

4.1. Participants

Eighty workers from Amazon's Mechanical Turk website (http://www.mturk.com/) participated for $1.33. The benefits and reliability of experimental data collected from Mechanical Turk have been previously documented (Paolacci, Chandler, & Ipeirotis, 2010). Only workers residing in the United States and with above 90% approval ratings[9] on Mechanical Turk were allowed to participate. In addition, workers were not allowed to participate in more than one of the experiments reported in this paper. They were screened based on their unique worker ID assigned by Mechanical Turk. Participants in Experiment 1 were randomly assigned to one of two conditions: BC–ABC ($N = 42$) or ABC–ABC ($N = 38$).

4.2. Stimuli and design

Participants learned about the relations between two fictitious medical conditions and a fictitious virus in a number of individuals. The possible event settings were: (A) has the "Ablique" virus or does not, (B) has "Burlosis condition" or does not, and (C) has the "Caprix" condition or does not. In addition, we purposively chose the content of the events to be one virus and two conditions such that B causing A, or C causing A would be extremely unlikely, making the task more manageable for participants given the general difficulty of causal structure learning tasks (e.g., Steyvers et al., 2003).

Sample pictures used to illustrate the events are shown in Fig. 4. Absence of an event was always displayed in blue, and the presence of an event was always displayed in bolded red, so that participants would be less likely to mistake them.

There were two blocks of trials, one showing contingencies between events B and C (the BC block), and the other showing contingencies between all three events (the ABC block). The left panel of Fig. 4 illustrates a trial from the BC block, and the right panel of Fig. 4 illustrates a trial from the ABC block.

[9] Approval ratings correspond to the percentage of times a worker's submissions have been approved by a requestor (the person posting the assignment).

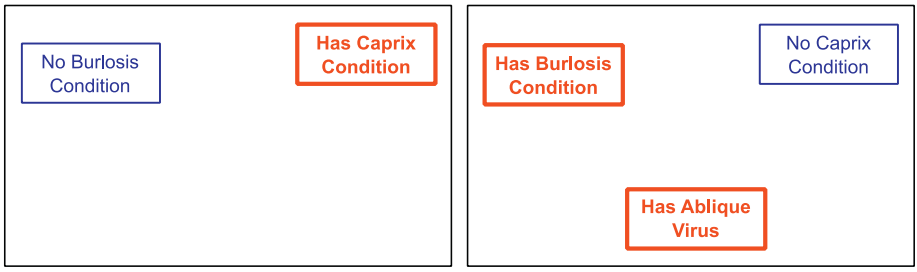

Fig. 4. Pictures used to illustrate the presence and absence of events A, B, and C in Experiments 1–4. (*Note:* Red font used in the experiments is shown in bold.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As shown in the left panel of Fig. 4, the Ablique virus was not visible at all during the trials of the BC block, nor was this factor mentioned prior to the BC block in the instructions, just as in the myopia case where people were not considering parent myopia as a possible common cause. The contingencies of the blocks were summarized above in Fig. 2.

There were two between-subject conditions, the BC–ABC and ABC–ABC conditions. Participants in the BC–ABC condition viewed the BC block first and the ABC block second. Participants in the ABC–ABC condition viewed the ABC block twice.

4.3. Procedure

The experiment was conducted using Qualtrics survey software at www.qualtrics.com. Participants were forwarded to the Qualtrics page after they agreed to participate on the Amazon Mechanical Turk website. Before starting the experiment, participants indicated their informed consent.

Before the main task, participants completed a training session as in Dennis and Ahn (2001), where their task was to find out whether an exotic plant causes a physical reaction or not, by viewing several cases where a plant was either ingested or not and where the person either had a physical reaction or not. After the training, participants began the main task. In both the BC–ABC and ABC–ABC conditions, participants were first told about the types of events they would see during the first block (i.e., events B and C in the BC–ABC condition, and events A, B, and C in the ABC–ABC condition). Then, all participants were told, "As in all medical cases, there may be other relevant factors for these patients that we are simply unaware of. Just to be clear, we are not implying that these factors are absent." These instructions, although presented to all participants, were provided as a point of clarification to make sure that participants in the BC–ABC condition did not later infer that A was absent during the first block just because the status of A was not presented to them.

Then, we specified for the participants which causal relations to keep track of so that the task would be manageable. Specifically, participants in the BC–ABC condition were told to consider only whether the Burlosis condition caused the Caprix condition (B-causes-C), as the causal relation could not possibly work in the other direction. Similarly, participants in the ABC–ABC condition were told to consider only three causal relations (A-causes-B, A-causes-C, and B-causes-C).

Then, all participants were told that they would observe 20 individuals whose descriptions would be presented on the same screen. Participants were told to view these individuals in the order presented by proceeding from the top of the screen to the bottom. They were further told that they could review the trials they had already seen by scrolling back up and then returning to their current position. We chose this presentation format because previous studies have shown that learning causal relations among more than two variables is generally difficult (Lagnado & Sloman, 2004; Steyvers et al., 2003; White, 2006). By modifying the traditional trial-by-trial presentation format in this way, participants may go back to review or reconsider any trials they might have missed, which should help to reduce working memory load. In addition, in order to encourage participants to take

Table 2
The average scores and percentages of 1st choices for each structure, separately for each condition. Bolded values are maxima.

								
Average scores								
BC–ABC cond.	0.64	0.24	0.40	0.88	0.64	0.48	1.02	**1.69**
ABC–ABC cond.	1.03	0.24	0.45	**1.55**	0.26	0.55	0.45	1.47
% of 1st choices								
BC–ABC cond	0.17	0.05	0.05	0.12	0.10	0.02	0.12	**0.38**
ABC–ABC cond.	0.32	0.00	0.03	**0.32**	0.00	0.03	0.05	0.26

the task seriously, they were told that they would receive a $1 bonus if their ratings were within 10 points of the correct values.[10]

After these instructions for the first block, participants in the BC–ABC condition viewed 20 trials of the BC block, and those in the ABC–ABC condition viewed 20 trials of the ABC block. These were in fact the same 20 trials, but for the BC block, the status of A was missing (see Fig. 2). The order of the trials for each block was randomized, and the same order, as shown in Fig. 2, was used across all participants to minimize any additional variance due to within-block order effects. Each trial had a unique patient number to indicate that these were all different individuals.

The 20 trials were presented in a single column with two adjacent trials separated by a horizontal line. Thus, a trial that looks like right panel of Fig. 4 was presented with a patient number on the top, a horizontal line underneath that trial, followed by another trial with a different patient number, and so on.

When the first block was finished, participants received the instructions about the second block on a new screen. Participants in the BC–ABC condition were told that scientists had discovered a new virus, and part of their job now was to determine how the virus relates to the two conditions. They were now told to evaluate relations A-causes-B and A-causes-C, in addition to B-causes-C. Then they viewed the ABC block. In the ABC–ABC condition, participants were told that they would view some additional patients to make sure they had enough evidence to judge the causal relations. Then they viewed the ABC block for a second time. The order of trials was identical to the first block.

When participants completed the second block, they proceeded to a different page and were told to choose which of the eight possible causal structures they thought best described the causal structure among the three events. A figure was shown to depict each causal structure (similar to the ones used in Table 1). The figures were labeled 1–8, starting with zero causal relations (1), to the one-relation structures (2–4), to the two-relation structures (5–7), and the three-relation structure (8). All of these eight figures were presented on a single page so that participants could browse through them before making their responses. Participants made three consecutive choices, indicating which structure corresponded to the first most likely, second most likely, and third most likely structure. They were required to choose three different structures. The questions appeared in the order described, but participants could change their responses multiple times before all three ratings were submitted.

4.4. Results

The data are summarized in Table 2. To make use of all three choices, we gave each participant a "score" for each structure corresponding to their ranking of the structure. If a structure was chosen 1st, 2nd, and 3rd, the scores were 3, 2, and 1, respectively. If a structure was not chosen, the score was 0. In addition, we also report the percentages of participants who selected each structure as their first choice.

[10] The normatively correct values depend on a number of variables (e.g., the prior distributions assumed, scaling of posteriors), so we chose a range of 0–10 for B-causes-C and 80 to 100 for A-causes-B and A-causes-C. These ranges were not disclosed to participants after the experiment to ensure that they were not shared with other Mechanical Turk workers.

Experiment 1 was designed to test whether participants who discover an initially unobserved cause later in learning would revise their prior causal beliefs in light of this new evidence, or show causal imprinting by failing to revise their prior beliefs. As summarized in Fig. 3 and Table 1, if participants engage in belief revision in the BC–ABC condition, they should be most likely to select the common-cause structure or the structure with all three causal relations. However, if participants show causal imprinting in the BC–ABC condition, they should be most likely to select the causal chain structure or the structure with all three relations. As shown in Table 2, the results were consistent with causal imprinting. The structure with all three relations was the most popular 1st choice and received the highest scores in the BC–ABC condition. In comparison, the common cause structure was among the two most popular 1st choices and received the highest scores in the ABC–ABC condition.[11]

First we present our analyses with the scores dependent measure. To examine whether the conditions differed overall, we conducted an 8 (structures) $\times$ 2 (conditions) mixed ANOVA, with structure as a within-subject factor and condition as a between-subject factor. The interaction between structure and condition was significant, $F(7,546) = 2.82$, $p < 0.01$, $\eta = 0.03$.[12] Independent samples t-tests showed that the BC–ABC condition had significantly lower scores for the common cause structure ($M = 0.88$, SD = 1.09) than the ABC–ABC condition ($M = 1.55$, SD = 1.25), $t(78) = -2.58$, $p = 0.01$, $d = 0.57$, but significantly higher scores for the causal chain ($M = 1.02$, SD = 1.05) than the ABC–ABC condition ($M = 0.45$, SD = 0.86), $t(78) = 2.67$, $p < 0.01$, $d = 0.60$. Also, while both normative (i.e., reinterpretation and fully Bayesian) analyses of the BC–ABC condition predicted a preference for the common cause over the causal chain and over the structure with all three causal relations (see Table 1), scores for the common cause structure in the BC–ABC condition ($M = 0.88$, SD = 1.09) did not differ significantly from scores for the causal chain ($M = 1.02$, SD = 1.05), $t(41) = -0.56$, $p = 0.58$, $d = 0.09$, and were significantly lower than scores for the structure with all three causal relations ($M = 1.69$, SD = 1.28), $t(41) = -3.20$, $p < 0.01$, $d = 0.49$. Thus, the judgments in the BC–ABC condition failed to support normative predictions.

Another important test for causal imprinting is whether participants in the BC–ABC condition were more likely to choose structures including the B-causes-C relation. Note that we cannot evaluate this by considering the structure with only B-causes-C, as this structure also excludes A-causes-B and A-causes-C, which is highly unlikely to be endorsed by either condition given the strong positive A–B and A–C contingencies. Instead, to evaluate this difference, for each participant we took the average scores for structures with B-causes-C (i.e., the last four structures shown in Table 2) and compared these across conditions. The BC–ABC condition had significantly higher average scores ($M = 0.96$, SD = 0.38) than the ABC–ABC condition ($M = 0.68$, SD = 0.41), $t(78) = 3.11$, $p < 0.01$, $d = 0.71$, consistent with causal imprinting.

Next, we examine overall differences in the 1st choices across the conditions using logistic regression, including structure, condition, and their interaction as predictor variables. To assess the interaction between condition and structure, we used a model comparison technique by comparing the fit of a regression model with all interaction terms (there were multiple interaction terms due to the categorical nature of the structure variable) to the fit of a model with no interaction terms. The model excluding the interaction terms provided a significantly worse fit than the model with the interaction terms, $\chi^2(7) = 17.81$, $p = 0.01$. Thus, participants differed between the conditions in their pattern of 1st choices. In particular, the common cause was a less frequent 1st choice in the BC–ABC condition (0.12) than in the ABC–ABC condition (0.32), $\chi^2(1, N = 80) = 4.61$, $p = 0.03$. In contrast, the causal chain was a more frequent 1st choice in the BC–ABC (0.12) than in the ABC–ABC condition (0.05), though this

[11] The other most popular first choice in the ABC–ABC condition was the structure with no causal relations. We speculate that the reason for this preference was that the ABC–ABC condition was more difficult overall, given that all three events were present throughout. Hence, participants gave this response due to confusion or being overwhelmed with choices. Experiments 2 and 3, which used the same study procedure but causal strength estimation of individual links rather than structure judgments, do not show lower ratings overall in the ABC–ABC condition. These findings suggest that the preference for no causal relations is not a general property of the ABC–ABC condition.

[12] We also performed a more conservative analysis to avoid the assumption made by ANOVA that the scores are normally distributed. Rather than weighting the 1st, 2nd, and 3rd choices differently, we assigned each structure a score of 1 if it was included in the top three and a score of 0 otherwise. A logistic regression assessing the interaction between condition and structure revealed a significant interaction for this coding scheme. In addition, Chi-square tests used for the main effects analyses yielded similar significance values to the t-tests presented below.

difference did not reach significance, $p = 0.44$, using Fisher's exact test. These results are analogous to those using the scores dependent measure, and are both consistent with causal imprinting.

Finally, a significantly greater percentage of participants in the BC–ABC condition indicated a structure with B-causes-C as their 1st choice (0.62) than in the ABC–ABC condition (0.34), $\chi^2(1, N = 80) = 6.12$, $p = 0.01$, consistent with causal imprinting. Indeed, participants in the BC–ABC condition were nearly twice as likely to choose structures with B-causes-C, which is impressive, given that the two conditions viewed the same B–C contingencies, with the only exception that A was absent during the first block.

Overall, Experiment 1 showed ample evidence for causal imprinting. Participants in the BC–ABC condition were less likely to choose the common cause structure than participants in the ABC–ABC condition, more likely to choose the causal chain structure, and more likely to choose structures including B-causes-C. Participants in the BC–ABC condition also had a slight preference for the causal chain over the common cause and a large preference for the structure with all three relations over the common cause, inconsistent with belief revision and with the reinterpretation and fully Bayesian analyses.

Though our findings are consistent with causal imprinting, our analyses were limited in several ways due to our use of causal structure judgments. First, we note that the structure with all three relations was one of the most popular choices in both conditions, representing about one third of the top choices. Unfortunately, these choices were not useful in testing our hypotheses, because the structure with all three relations is consistent with both belief revision and causal imprinting (see Fig. 3). Our primary analyses focused on other structures, and as a result, our demonstrations of causal imprinting may have been less impressive. For instance, although all of our comparisons between conditions were in the right direction, the comparison of the 1st choices for the causal chain did not reach significance. In addition, our comparisons of the common cause structure were significant, but somewhat weak (e.g., using Bayesian *t*-tests, the evidence is only suggestive; Rouder, Speckman, Dongchu, Morey, & Iverson, 2009).

Second, although we predicted that participants in both conditions would choose the structure with all three relations (see Fig. 3), we also predicted that those in the ABC–ABC condition or those who engaged in belief revision should have believed in a weaker B-causes-C relation than those who showed causal imprinting. Causal structure judgments do not allow us to detect such important differences between the inferred strength of the causal relations. A more direct way of testing this quantitative difference would be to ask participants to estimate the causal strength of each link. This is one of the modifications made in the next experiments.

5. Experiment 2: Causal strength judgments with same vs. different tokens

The main goal in Experiment 2, aside from the change to causal strength ratings just mentioned, was to understand why participants in the BC–ABC condition in Experiment 1 failed to revise their initial causal belief. As we discussed in the introduction, we propose that causal imprinting occurs because of an asymmetry in how initial and later pieces of evidence are used during learning. Namely, because people have a tendency to apply previously acquired knowledge when interpreting the later evidence, they will be slow to reinterpret or revise that knowledge based on the later evidence. Thus, people will act as if they were imprinted with their initial causal belief. Yet, there may be other reasons that we obtained what appears to be causal imprinting. Experiment 2 attempts to rule out two theoretically important alternative accounts, which we derived from our Bayesian analyses.

Recall that both the fully and bounded Bayesian analyses of the BC–ABC condition showed greater preferences for structures including B-causes-C, compared to the ABC–ABC analysis. In the fully Bayesian analysis, this occurred due to the reasonable assumption that learners may have been uncertain about the missing values of A during the BC block. To make this idea of uncertainty concrete, learners may have been unsure about whether the individuals from the BC block were taken from the same *context* as those from the ABC block (Liljeholm & Cheng, 2007). Causal relations among events may change depending on temporal or spatial context, because of differences in the presence or absence of other interacting causes. For instance, in a cold climate B may cause C regardless of presence of virus

A, but in a hot climate B may share a positive contingency with C only because of virus A. Participants may have also assumed that the data presented in the first block occurred before the data in the second block, and that the context changed as time passed. Thus, if learners were uncertain about the underlying contingencies in the BC block, they would have been justified in not revising their initial belief, as the two sets of data could have been obtained from different contexts.

A pattern similar to causal imprinting also occurred in the bounded Bayesian analysis, where the learner neither reinterprets the BC block after viewing the ABC block, nor considers possible values of A during the BC block, due to cognitive limits. Instead, the BC block is used only to update the belief in B-causes-C, which leads to a greater belief in structures including this relation and a much lower belief in the common cause structure. That is, causal imprinting might have happened in Experiment 1 simply due to cognitive limitations.

To examine both possibilities, Experiment 2 included a new BC–ABC "same tokens" condition where participants were told that the ABC block represented the exact same individuals as the BC block. More specifically, they were told after the BC block that scientists recently discovered information about presence or absence of virus A in the individuals that they had just observed, and that now they would be observing these individuals again along with this information. This new condition provides a direct test of the fully and bounded Bayesian accounts of causal imprinting. First, the new instructions in the same tokens condition remove all uncertainty about the missing values of A, meaning that learners should not show causal imprinting as predicted by the fully Bayesian analyses. Second, because the BC block will now be repeated, and this time with the actual values of A revealed, the rational behavior is to use this information *in lieu of* the contingencies from the BC block. Thus, cognitive limitations are also no longer a valid excuse for not revising the initial beliefs. If this new BC–ABC condition continues to show causal imprinting, then it could not be motivated by uncertainty or cognitive limitations. Instead, it would imply a non-normative bias to maintain the belief that was first imprinted in the reasoner's mind.

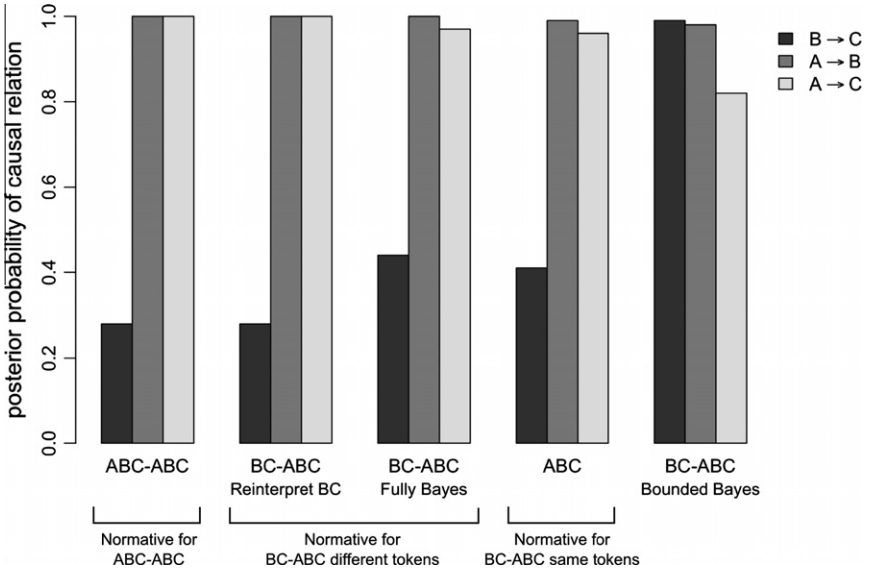

Fig. 5. Causal strength ratings derived from the normative Bayesian analyses of the ABC–ABC and BC–ABC conditions presented in Table 1, separately for B-causes-C (B → C), A-causes-B (A → B), and A-causes-C (A → C). Sub-headings beneath the graph panels indicate for what conditions each set of predictions should be considered normative. The bounded Bayesian analysis of the BC–ABC condition may also be considered normative for the BC–ABC different tokens condition (though not the same tokens condition), but only when taking into account cognitive limitations. The bounded Bayesian analysis also shows a pattern very similar to what we would expect if causal imprinting occurs.

5.1. Causal strength predictions

The descriptive predictions summarized in Fig. 3 shown in Section 2 were given in terms of causal structure judgments, not causal strength ratings. However, we can restate these predictions by assuming that relations appearing more often in the predicted structures from Fig. 3 will be given higher causal strength ratings. Using this method, if causal imprinting occurs, then the central prediction is that ratings for B-causes-C should be higher in the BC–ABC condition than in the ABC–ABC condition.

In addition, if participants in the BC–ABC condition infer all three relations or the chain due to causal imprinting (see Fig. 3), then their ratings for B-causes-C should not differ much from their ratings for A-causes-B. In contrast, participants in the ABC–ABC condition should give lower ratings for B-causes-C than for A-causes-B. Also, if causal imprinting leads some participants in the BC–ABC condition to infer all three relations and others to infer the causal chain, then ratings for A-causes-C could be lower in the BC–ABC condition than in the ABC–ABC condition. Note, however, that the lack of a difference in ratings for A-causes-C would not necessarily provide evidence against causal imprinting, as people in the BC–ABC condition might primarily infer all three relations.

We can also obtain normative predictions for causal strength ratings based on the posterior probabilities of the causal structures from the Bayesian analyses (see Table 1 in Section 3.3). Following previous work (Friedman & Koller, 2003; Griffiths & Tenenbaum, 2005), we computed the causal strength for a specific causal relation by summing the posterior probabilities of the causal structures including that relation. For example, to obtain the posterior probability of the B-causes-C relation, we summed the posteriors for the four hypotheses in the rightmost columns of Table 1, which all include B-causes-C.

Fig. 5 shows the strength for each causal relation, separately for each of the analyses considered in Table 1. Fig. 5 also includes the normative predictions for the BC–ABC same tokens condition, where posterior probabilities are computed based only on one ABC block. As can be readily seen in the figure, all of the normative analyses for the BC–ABC conditions (including the analysis for the BC–ABC same tokens condition) predict much lower ratings for B-causes-C than for A-causes-C, as do the analyses for the ABC–ABC condition. Finally, we note that the bounded Bayesian analysis again represents a form of causal imprinting, but in Experiment 2 this is non-normative for the BC–ABC same tokens condition.

5.2. Participants

One hundred eighty-one workers from Amazon's Mechanical Turk website participated as in Experiment 1. Workers that participated from Experiment 1 were excluded from Experiment 2 based on their unique worker ID assigned by Mechanical Turk. Participants were randomly assigned to one of three conditions: BC–ABC different tokens ($N = 63$), BC–ABC same tokens ($N = 59$), or ABC–ABC ($N = 59$).

5.3. Stimuli, design, and procedure

The stimuli and procedure were similar to Experiment 1, except for the addition of the new BC–ABC same tokens condition, the use of causal strength ratings, and changes to the training procedure to reflect the use of causal strength ratings.

The training procedure was the same as in Experiment 1, except that after participants viewed the sample cases, they were introduced to the scale for the causal strength ratings. Participants were told that a rating of 0 indicated that Event 1 has no effect on Event 2, and 100 indicated that Event 1 very strongly causes Event 2. They were further provided with examples to give an intuition for how to use the scale, as in Dennis and Ahn (2001).

The new BC–ABC same tokens condition was identical to the BC–ABC different tokens condition (i.e., the BC–ABC condition from Experiment 1), except for the instructions prior to the ABC phase. Participants in this condition were told that the same individuals from the BC block were shown in the ABC block, though now the status of event A (the Ablique virus) would be shown. Participants were told, "The scientists tested the EXACT SAME 20 individuals for the Ablique virus, using the EXACT SAME

blood samples that led to the Burlosis and Caprix diagnoses. You will now re-view these SAME 20 individuals' descriptions, but this time you will see whether or not each individual had the Ablique virus."

After viewing the two blocks of descriptions about patients that correspond to their condition, all participants gave three causal strength ratings (namely, A-causes-B, A-causes-C, and B-causes-C) using the same scale from the training session. A picture of the scenario was shown, similar to the right panel of Fig. 4, but with the boxes colored in black and only the event names without their presence/absence being stated (e.g., event B read simply, "Burlosis Condition"). Arrows between the events were added to clarify the direction of the causal relations.

All three causal strength rating questions were given on the same screen, so participants could make their ratings in any order and change them multiple times before all three were submitted. The order in which the three causal strength ratings was displayed from top to bottom was counterbalanced by a factorial combination of two factors: whether the B-causes-C rating was displayed before the two common cause ratings (A-causes-B and A-causes-C), and which of the two common cause ratings was displayed first. The differences between the conditions we report below did not depend on the order of the three causal strength ratings.

5.4. Manipulation check

If the causal imprinting effects obtained in Experiment 1 were due to uncertainty about the missing A values or to cognitive limitations, then there should be no causal imprinting in the BC–ABC same tokens condition. Thus, these accounts are ruled out if causal imprinting still occurs in this condition. However, one might argue that causal imprinting in this condition could happen simply because participants fail to attend to or properly encode the instructions.[13] To exclude this possibility we conducted a separate experiment with only the two BC–ABC conditions to verify that participants read and understood the instructions.

For this manipulation check, a separate group of 31 participants from Amazon's Mechanical Turk website were randomly assigned to the BC–ABC same ($N = 17$) or different ($N = 14$) tokens conditions, and viewed the exact same instructions and blocks of data as participants in the main Experiment 2. However, after viewing the ABC block, participants did not give causal ratings, but rather, answered a question about the instructions they read prior to the ABC block. The question stated, "You have now seen two sets of 20 individuals. Did the instructions state that the second set represented the same individuals as the first set or different individuals?" They responded either "same" or "different." In the same tokens condition, 16 of 17 (94%) answered "same," which would not be expected if they misunderstood the instructions and were responding at chance ($p < 0.01$, binomial test). In contrast, in the different tokens condition, only 7 of 14 (50%) answered "same." The difference in proportion of "same" responses between conditions was significant ($p = 0.01$, Fisher's exact test). Hence, the instructions in the BC–ABC same tokens condition was effective in establishing that the individuals from the BC and ABC blocks were indeed the same.

We also note that the chance responding in the different tokens condition (50% accuracy) is sensible given that the instructions (i.e., "You will now observe 20 more individuals.") did not strongly emphasize different individuals. Incidentally, this lower accuracy also strengthens our account of the causal imprinting effects in our other experiments. That is, if the answers to our manipulation check for the different tokens condition correspond to reasoners' actual beliefs about the individuals presented during learning, then roughly half of the participants in all of our experiments were implicitly in the same tokens condition, where they should not have shown causal imprinting according to normative accounts.

5.5. Results

Overall, the results showed causal imprinting in both the same and different tokens BC–ABC conditions, suggesting that causal imprinting is in fact non-normative. Fig. 6 shows the average causal strength ratings for the three conditions.

[13] We thank an anonymous reviewer for this suggestion.

For statistical analyses, we first examine whether the conditions differed overall in their patterns of ratings. A 3 (causal relation) × 3 (condition) mixed ANOVA with causal strength rating as the dependent variable, causal relation as a within-subjects factor, and condition as a between-subjects factor, revealed a significant interaction between causal relation and condition, $F(4,356) = 7.32$, $p < 0.01$, $\eta^2 = 0.07$. We then examined the differences between each pair of conditions by conducting separate 3 (causal relation) × 2 (condition) mixed ANOVAs. The ANOVA comparing the BC–ABC same and different tokens conditions revealed no interaction between condition and cause, $F(2,240) = 0.07$, $p = 0.93$, $\eta^2 < 0.01$. However, there were significant interactions from the ANOVAs comparing the BC–ABC different tokens condition and the ABC–ABC condition, $F(2,240) = 10.92$, $p < 0.01$, $\eta^2 = 0.08$, and comparing the BC–ABC same tokens condition and the ABC–ABC condition, $F(2,232) = 11.87$, $p < 0.01$, $\eta^2 = 0.09$.

Next, we examined which of the specific causal relations led to the different patterns of ratings in the BC–ABC conditions and the ABC–ABC condition. In support of causal imprinting, ratings for B-causes-C were significantly higher in both the BC–ABC different tokens condition ($M = 58.05$, SD = 31.22) and the BC–ABC same tokens condition ($M = 57.15$, SD = 30.92) than in the ABC–ABC condition ($M = 39.98$, SD = 30.72), $t(120) = 3.22$, $p < 0.01$, $d = 0.58$, and $t(120) = 3.03$, $p < 0.01$, $d = 0.56$, respectively.

Further analyses addressed the two other causal relations. First, if causal imprinting lead some participants in the BC–ABC condition to infer the causal chain, then ratings for A-causes-C should be lower in the BC–ABC condition than in the ABC–ABC condition. Consistent with this prediction, ratings for A-causes-C were significantly lower in both the BC–ABC different tokens condition ($M = 48.24$, SD = 33.79) and the BC–ABC same tokens condition ($M = 45.42$, SD = 29.31) than in the ABC–ABC condition ($M = 58.92$, SD = 25.40), $t(120) = -1.97$, $p = 0.05$, $d = 0.36$, and $t(116) = 2.67$, $p < 0.01$, $d = 0.49$, respectively.

Second, if participants in the BC–ABC conditions had inferred the chain, then ratings for A-causes-B should not differ between the BC–ABC and ABC–ABC conditions. In fact, ratings for A-causes-B were somewhat lower in both the BC–ABC different tokens condition ($M = 55.33$, SD = 32.95) and the BC–ABC same tokens condition ($M = 55.03$, SD = 27.36) than in the ABC–ABC condition ($M = 64.19$, SD = 24.61), $t(120) = -1.67$, $p = 0.10$, $d = 0.30$, and $t(116) = -1.91$, $p = 0.06$, $d = 0.35$, respectively. We note, however, that these slight differences may have occurred only because participants in the ABC–ABC condition viewed two blocks of trials with both A and B visible, whereas participants in

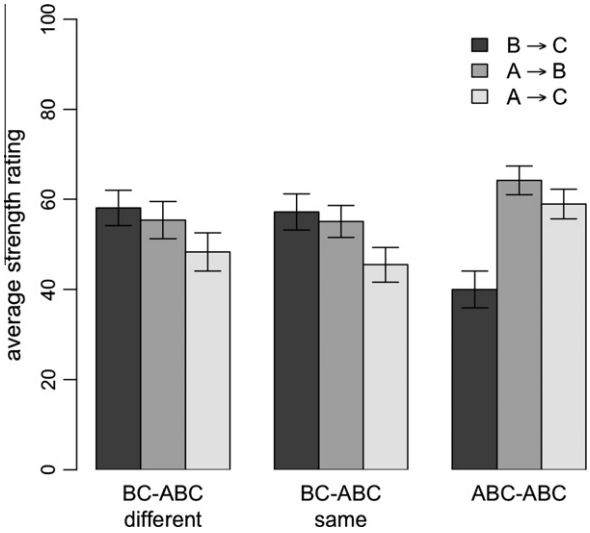

Fig. 6. Average causal strength ratings for the three conditions in Experiment 2, separately for B-causes-C (B → C), A-causes-B (A → B), and A-causes-C (A → C).

the BC–ABC conditions viewed only one block. In Experiment 3 we add a new ABC condition that presents only one ABC block to see whether the difference in ratings for A-causes-B would disappear, while the difference in ratings for A-causes-C would still occur (suggesting causal chain inferences).

Finally, we compared ratings of B-causes-C and A-causes-B within each condition. If causal imprinting occurred, then participants in the BC–ABC conditions should have given similar ratings for these relations, whereas participants in the ABC–ABC condition should have given much lower ratings for B-causes-C (see Fig. 3). Consistent with causal imprinting, in the BC–ABC different tokens condition ratings for B-causes-C ($M = 58.05$, SD = 31.22) were not significantly different from A-causes-B ($M = 55.33$, SD = 32.95), $t(62) = 0.51$, $p = 0.61$, $d = 0.06$. Similarly, in the BC–ABC same tokens condition ratings for B-causes-C ($M = 57.15$, SD = 30.92) were not significantly different from A-causes-B ($M = 55.03$, SD = 27.36), $t(58) = 0.39$, $p = 0.70$, $d = 0.05$. In contrast, in the ABC–ABC condition ratings for B-causes-C ($M = 39.98$, SD = 30.71) were significantly lower than for A-causes-B ($M = 64.19$, SD = 24.61), $t(58) = -5.10$, $p < 0.01$, $d = 0.66$. Furthermore, the difference in ratings in the ABC–ABC condition was significantly greater than that of the BC–ABC different tokens condition, $t(120) = 3.76$, $p < 0.01$, $d = 0.68$, and the same tokens condition, $t(116) = 3.65$, $p < 0.01$, $d = 0.67$.

Overall, the results from Experiment 2 provide further support for causal imprinting and suggest that it is highly unlikely to be due to uncertainty or cognitive limitations. Furthermore, the lack of a significant difference in the BC–ABC conditions between ratings for B-causes-C and A-causes-B fails to support the predictions of the normative models presented in Fig. 5. Thus, the causal imprinting observed in the BC–ABC conditions appears to be non-normative.

6. Experiment 3: Causal strength judgments with additional controls

In Experiments 1 and 2, participants in the BC–ABC conditions gave strong endorsement and high ratings for B-causes-C. In Experiment 2 in particular, B-causes-C ratings in the BC–ABC conditions were as high as their ratings for A-causes-B, and higher than the ratings for B-causes-C in the ABC–ABC condition. Though consistent with causal imprinting, these specific results do not necessarily imply that participants in the BC–ABC condition held the same level of belief in B-causes-C across the BC block and the ABC block. Instead, they might have begun to reduce their belief in B-causes-C during the ABC block, even if not fully to the levels of participants in the ABC–ABC condition. Consistent with this observation, a Bayesian analysis on just the data from the BC block predicts a much higher posterior probability of B-causes-C (0.76) than does the fully Bayesian analysis of the BC–ABC block (0.28). If ratings of B-causes-C in the BC–ABC condition did decline after observing the ABC block, then our demonstrations of causal imprinting would be weaker than we initially thought. To test this possibility, Experiment 3 compares ratings for B-causes-C in the BC–ABC condition after the ABC block to a new BC condition, where participants would only view the BC block and then immediately give causal strength ratings.

Another goal of Experiment 3 was to further examine ratings for A-causes-B and A-causes-C. If participants in the BC–ABC conditions had inferred the causal chain, then ratings for A-causes-C would be lower in the BC–ABC condition than in the ABC–ABC condition, but ratings for A-causes-B would be similar in these two conditions. However, in Experiment 2 we found that ratings for both relations were lower in the BC–ABC condition. We attributed the difference for A-causes-B to the greater number of trials in the ABC–ABC condition where both events A and B were visible. Yet, this greater number of trials would also invalidate the expected finding that A-causes-C was lower in the BC–ABC, because the BC–ABC condition also viewed fewer trials where both events A and C were visible. Thus, while the ABC–ABC condition was useful in controlling for the number of trials with both B and C, the results from the ABC–ABC condition are difficult to interpret in relation to the ratings for A-causes-B and A-causes-C. To control for the number of trials with A and B, and A and C, Experiment 3 utilized a new ABC condition, where participants would only view the ABC block once and then immediately give causal strength ratings. By comparing the BC–ABC condition to the ABC condition, we can determine whether participants in these conditions show equal learning of A-causes-B, but give lower ratings for A-causes-C.

6.1. Participants

One hundred forty-three workers from Amazon's Mechanical Turk website participated as in Experiments 1 and 2. Workers that participated from Experiments 1 and 2 were excluded from Experiment 3 based on their unique worker ID assigned by Mechanical Turk. Participants were randomly assigned to one of three conditions: BC ($N = 47$), BC–ABC ($N = 52$), or ABC ($N = 51$).

6.2. Materials, design, and procedure

The materials, design, and procedure were identical to Experiment 2 except for the following. There were three conditions: BC–ABC, BC, and ABC conditions. The BC–ABC condition was identical to the BC–ABC different tokens condition in Experiment 2. The BC condition viewed only the BC block. The ABC condition viewed only the ABC block. The instructions for the BC condition were identical to the instructions for the BC–ABC condition prior to the BC block. The instructions for the ABC condition were identical to the instructions in Experiment 2 for the ABC–ABC condition prior to the ABC block.

6.3. Results

Fig. 7 shows the average causal ratings for the three conditions. To preview, participants in the BC–ABC condition did not appear to have lowered their belief in B-causes-C at all, even after viewing the ABC block. Also, the results from the BC–ABC condition suggest that some participants inferred the causal chain.

We first examined whether the conditions differed overall. A 3 (causal relation) × 3 (condition) mixed ANOVA with causal strength rating as the dependent variable, causal relation as a within-subjects factor, and condition as a between-subjects factor, revealed a significant interaction between causal relation and condition, $F(4,294) = 11.64$, $p < 0.01$, $\eta^2 = 0.14$. The 3 (causal relation) × 2 (condition) mixed ANOVA comparing the BC and BC–ABC conditions revealed a significant interaction between condition and cause, $F(2,194) = 5.12$, $p < 0.01$, $\eta^2 = 0.05$, as did the ANOVA comparing the BC–ABC and ABC conditions, $F(2,202) = 10.65$, $p < 0.01$, $\eta^2 = 0.09$.

Next we compared ratings for specific relations across the BC and BC–ABC conditions. If participants in the BC–ABC condition revised their beliefs in B-causes-C based on the ABC block, their ratings

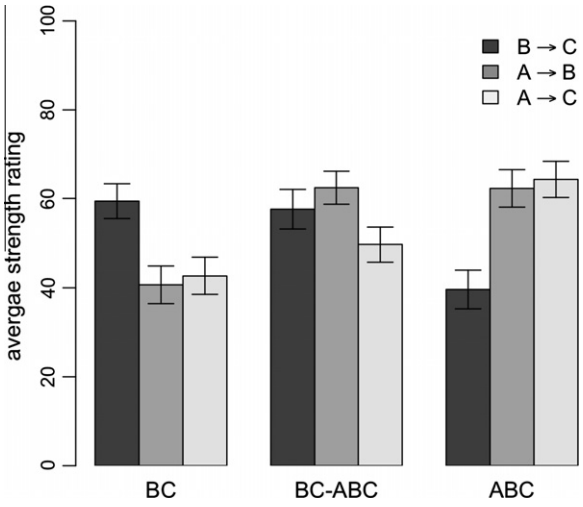

Fig. 7. Average causal strength ratings for the three conditions in Experiment 3, separately for A-causes-B (A → B), A-causes-B (A → B), and A-causes-C (A → C).

should be lower than the BC condition. Yet, ratings for B-causes-C in the BC–ABC condition ($M = 57.58 = 32.08$) were not significantly different from ratings in the BC condition ($M = 59.38$, SD = 26.85), $t(97) = -0.30$, $p = 0.76$, $d = 0.06$, failing to provide any evidence for belief revision. In contrast, ratings for A-causes-B were significantly greater in the BC–ABC condition ($M = 62.40$, SD = 26.73) than the BC condition ($M = 40.68$, SD = 29.04), $t(97) = 3.88$, $p < 0.01$, $d = 0.78$, suggesting that participants did update their beliefs in some way based on the ABC block. Ratings for A-causes-C did not differ significantly between the BC–ABC condition ($M = 49.65$, SD = 28.12) and the BC condition ($M = 42.66 = 28.29$), $t(97) = 1.23$, $p = 0.22$, $d = 0.25$. This finding is consistent with the hypothesis that some participants in the BC–ABC condition may have inferred the causal chain, which excludes this relation.

We also compared the BC–ABC and ABC conditions. If participants in the BC–ABC condition inferred the causal chain, then ratings for A-causes-C should be lower in this condition than in the ABC condition, but ratings for A-causes-B should be similar across these conditions. As predicted, ratings for A-causes-C were lower in the BC–ABC condition ($M = 49.65$, SD = 28.12) than in the ABC condition ($M = 64.27$, SD = 29.03), $t(101) = -2.60$, $p < 0.01$, $d = 0.51$, but ratings for A-causes-B were not significantly different between the BC–ABC condition ($M = 62.40$, SD = 26.73) and the ABC–ABC condition ($M = 62.25$, SD = 30.08), $t(101) = 0.026$, $p = 0.98$, $d = 0.01$. Hence, the difference in Experiment 2 for A-causes-C was not likely due to the lower number of trials in the BC–ABC condition with both A and C visible. The difference in ratings for A-causes-B, however, disappears when controlling for the number of such trials.

Finally, if causal imprinting occurred, then ratings for B-causes-C should be higher in the BC–ABC condition than in the ABC condition, and the difference in ratings between B-causes-C and A-causes-B should be smaller in the BC–ABC condition than in the ABC condition, as we found in Experiment 2. Indeed, ratings for B-causes-C were higher in the BC–ABC condition ($M = 57.58$, SD = 32.08) than in the ABC condition ($M = 39.63$, SD = 31.05), $t(101) = 2.88$, $p < 0.01$, $d = 0.57$. We note, however, that in Experiment 3 participants in the BC–ABC condition also viewed more trials with both B and C than the ABC condition (due to our controlling for the number of trials with both A and B/C visible), which may have contributed to this difference (though see Experiments 2 and 4). In addition, ratings for B-causes-C in the BC–ABC condition ($M = 57.58$, SD = 32.08) were not significantly different from A-causes-B ($M = 62.40$, SD = 26.73), $t(51) = -1.02$, $p = 0.31$, $d = 0.14$, whereas ratings for B-causes-C in the ABC condition ($M = 39.63$, SD = 31.05) were significantly lower than for A-causes-B ($M = 62.25$, SD = 30.08), $t(50) = -4.38$, $p < 0.01$, $d = 0.61$. The differences in these two ratings for the BC–ABC condition was also significantly smaller than that of the ABC condition, $t(101) = 2.54$, $p = 0.01$, $d = 0.50$, providing further support for causal imprinting.

7. Experiment 4: Causal strength judgments with order manipulation

In Experiment 4 we test a final prediction of our account of causal imprinting. We have argued that causal imprinting occurs because the evidence from the BC block leads to imprinting of the belief that B causes C, which then leads to biased interpretations of the ABC block. An essential aspect of this account is that causal imprinting stems from viewing evidence in a specific order, with the BC block before the ABC block. This ordering is critical, because if participants were to view the ABC block before the BC block, they should infer structures with weak B-causes-C (as demonstrated in the previous experiments), which in turn would lead to biased interpretations of BC block by the same principle. Hence, a crucial test of our account is that causal imprinting should occur only when learners view the BC block before the ABC block, and a similar pattern should not occur for the reverse ordering.

In Experiment 4 we considered only two conditions, the BC–ABC condition and a new ABC–BC condition, which viewed the BC block after viewing the ABC block. We predicted that although participants in the two conditions would have observed identical trials by the time they make causal strength judgments, those in the ABC–BC condition would not show causal imprinting, while those in the BC–ABC would. Thus, participants in the ABC–BC condition would give lower ratings for B-causes-C than those in the BC–ABC condition, and would give lower ratings for B-causes C than for A-causes-B, whereas participants in the BC–ABC condition would give equally high ratings for B-causes-C and A-causes-B.

7.1. Participants

Sixty-one workers from Amazon's Mechanical Turk website participated as in Experiments 1–3. Workers that participated from Experiments 1–3 were excluded from Experiment 4 based on their unique worker ID assigned by Mechanical Turk. Participants were randomly assigned to one of two conditions: BC–ABC ($N = 30$) or ABC–BC ($N = 31$).

7.2. Materials, design, and procedure

The materials, design, and procedure were identical to Experiment 2 except for the following. There were only two conditions: BC–ABC and ABC–BC. The BC–ABC condition was identical to the BC–ABC different tokens condition in Experiment 2. Participants in the ABC–BC condition read the same instructions prior to the ABC block as the ABC–ABC condition from Experiment 2, but a new set of instructions prior to the BC block: "Unfortunately, due to computer errors, [scientists] lost all the records about whether or not these individuals had been infected with the Ablique virus." These instructions were designed to be plausible, given our cover story of a research setting, and to ensure that participants did not misinterpret the missing information about values of A as the absence of A, given that they had just observed A during the first block.

7.3. Results

Fig. 8 shows the average causal ratings for the two conditions. To preview, the pattern of results from the BC–ABC condition suggests causal imprinting much more than that from the ABC–BC condition. A 3 (causal relation) $\times$ 2 (condition) mixed ANOVA with causal strength rating as the dependent variable, causal relation as a within-subjects factor, and condition as a between-subjects factor, revealed a significant interaction between causal relation and condition, $F(2,118) = 3.25$, $p = 0.04$, $\eta^2 = 0.05$.

Ratings for B-causes-C were higher in the BC–ABC condition ($M = 60.07$, SD = 25.19) than in the ABC–BC condition ($M = 41.19$, SD = 31.77), $t(59) = 2.57$, $p = 0.01$, $d = 0.66$, but there was no difference in ratings for A-causes-B (BC–ABC condition: $M = 62.00$, SD = 27.54; ABC–BC condition: $M = 63.35$, SD = 31.77), $t(59) = -0.19$, $p = 0.85$, $d = 0.02$, or for A-causes-C (BC–ABC condition: $M = 52.53$, SD = 27.89; ABC–BC: $M = 56.26$, SD = 30.38), $t(59) = -0.50$, $p = 0.62$, $d = 0.13$. As we mentioned earlier,

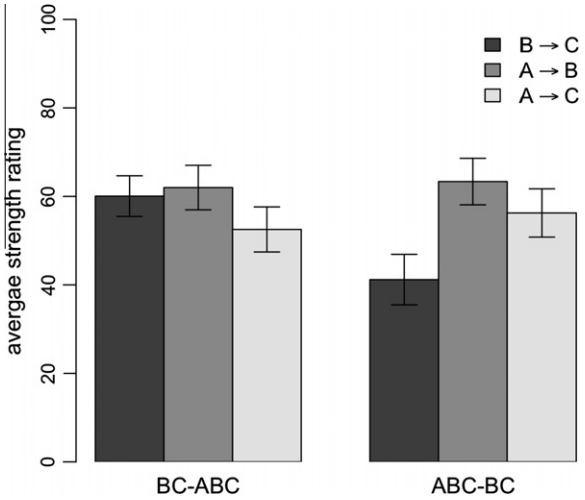

Fig. 8. Average causal strength ratings for the two conditions in Experiment 4, separately for A-causes-B (A → B), A-causes-B (A → B), and A-causes-C (A → C).

the lack of a difference in ratings for A-causes-C does not undermine our results for causal imprinting, though it does suggest that participants in the BC–ABC condition of the current experiment were more likely to infer the structure with all three relations, as opposed to the causal chain.

In addition, for participants in the BC–ABC condition ratings for B-causes-C ($M = 60.07$, SD = 25.19) were not significantly different from their ratings for A-causes-B ($M = 62.00$, SD = 27.54), $t(29) = -0.31$, $p = 0.76$, $d = 0.06$, whereas in the ABC–ABC condition, ratings for B-causes-C ($M = 41.19$, SD = 31.77) were significantly lower than their ratings for A-causes-B ($M = 63.35$, SD = 29.27), $t(30) = -3.04$, $p < 0.01$, $d = 0.55$. The difference between these two ratings in the BC–ABC condition was significantly smaller than the difference in the ABC–BC condition, $t(59) = 2.11$, $p = 0.04$, $d = 0.54$, consistent with causal imprinting.

8. General discussion

We often learn causal structures incrementally over time. In some cases the later viewed evidence provides information about an initially hidden variable, and the conditional contingency involving this hidden variable calls into question our original causal beliefs. The goal of this paper was to examine whether people use the later evidence in these circumstances to reinterpret the initial evidence that led to their original causal beliefs, or whether they show "causal imprinting" instead, which we define as the tendency to avoid such reinterpreting and to maintain the original causal belief. We examined this issue in the context of learning about a common cause after observing a positive contingency between the two effects of the common cause. Across four experiments using different dependent measures and control comparisons, we found consistent evidence in support of causal imprinting.

In Experiment 1, participants in the BC–ABC condition observed a positive contingency between two effects (events B and C), followed by a positive contingency between a common cause (event A) and the two effects. In contrast, participants in the ABC–ABC condition observed a positive contingency between the common cause and its two effects from the outset. We found that participants in the BC–ABC condition inferred the correct common cause structure less often than the ABC–ABC condition and inferred the causal chain and other structures including B-causes-C more often. Both results suggest that causal imprinting had occurred.

In Experiment 2, we found further evidence for causal imprinting using causal strength ratings rather than causal structure judgments. Ratings for the B-causes-C relation were higher in the BC–ABC condition than in the ABC–ABC condition. Furthermore, we tested whether the failure to revise the belief in B-causes-C in Experiment 1 was due to learners being uncertain about the missing values of A during the BC block or to limited cognitive abilities. This test was critical given that a fully Bayesian and bounded Bayesian analysis of the BC–ABC condition yielded strength ratings tending towards causal imprinting when incorporating uncertainty and cognitive limits. In Experiment 2 we removed uncertainty and cognitive load by telling participants in a new BC–ABC condition that the BC and ABC blocks represented the same exact individuals. In this case, the normative behavior is not causal imprinting, but participants in this condition continued to show imprinting, giving higher ratings for B-causes-C than the ABC–ABC condition.

Experiment 3 demonstrated the robustness of causal imprinting using the original BC–ABC condition, plus a new BC condition (presenting only the BC block) and ABC condition (presenting only the ABC block). Ratings for B-causes-C were nearly identical in the BC and BC–ABC conditions, suggesting that participants in the BC–ABC condition did not revise their original belief at all, even after observing data showing that the positive contingency between B and C disappears when conditionalizing on A. In addition, ratings for A-causes-B (part of the common cause structure) did not differ between the BC–ABC and ABC conditions, while ratings for A-causes-C were weaker in the BC–ABC than in the ABC condition. This pattern suggests that the participants in the BC–ABC condition were more likely to have induced the causal chain than those in the ABC condition, another signature of causal imprinting.

Finally, in Experiment 4 we compared the BC–ABC condition to a new ABC–BC condition, where participants viewed the same two blocks of data but in the reverse order. Ratings for B-causes-C were higher in the BC–ABC condition than in the ABC–BC condition despite the fact that they observed identical B–C contingencies under identical contexts (one block without A and the other with A). This

result demonstrates that the BC block appearing before the ABC block is critical in establishing a strong belief in B-causes-C. Thus, causal imprinting appears to result specifically from the influence of prior knowledge on causal inferences.

8.1. A flexible interpretations account of causal imprinting

We have argued that causal imprinting results from the tendency of learners to flexibly interpret the data they observe based on their prior causal beliefs. This claim builds on previous work supporting the idea of flexible interpretations. Specifically, Luhmann and Ahn (2011) provided direct evidence for such interpretations to account for primacy effects in the learning of single causal relations (Dennis & Ahn, 2001; Einhorn & Hogarth, 1986; Marsh & Ahn, 2006; Yates & Curley, 1986). That is, contradictory evidence was explained away while maintaining the original belief, rather than being treated as a reason to undermine the original belief.

Applying this flexible interpretation account to the current studies, participants in the BC–ABC condition who came to believe that B causes C might continue to interpret the ABC block as evidence that B causes C, and this may slow their realization that B and C are independent, given A. This interpretation effect seems especially likely in our task, given that B and C were still correlated during the ABC block if not considering A, which is at some level consistent with the belief that B causes C. Indeed, this evidence might have actually prevented participants from seeking alternative interpretations of the B–C contingency, such as the common cause structure.

This account is most strongly supported by the results from Experiment 4, where participants in the ABC–BC condition viewed the exact same data as those in the BC–ABC condition, but gave lower ratings for B-causes-C. If prior causal beliefs are used to interpret later contingencies, this is precisely the pattern one would expect. First, for the BC–ABC condition, the belief in B-causes-C would impede the realization that these factors are independent, as indicated in the ABC block. Second, for the ABC–BC condition, the belief that A is the common cause of B and C would seem to inoculate learners from making the faulty inference that B causes C based on their correlation during the BC block.[14]

8.2. Normative claims, Bayesian models, and causal imprinting

One of our central goals was to address whether causal imprinting is a normative, or rational response to new evidence that conflicts with prior causal beliefs. To this end, we presented normative Bayesian analyses of the data our participants viewed in the BC–ABC and ABC–ABC conditions. Our analyses revealed that maintaining a somewhat higher belief in B-causes-C in the BC–ABC condition than in the ABC–ABC condition was normative when learners are justified in remaining uncertain about the status of A during the BC block after viewing the ABC block. Furthermore, a more pronounced form of causal imprinting may be considered boundedly normative if learners were unable to consider both possible values of A during the BC block, due to cognitive limits. Experiment 2 provided a direct test of both of these assumptions and showed that causal imprinting persisted even when neither assumption could be validated. Participants in one of our BC–ABC conditions were told that the BC block was identical to the ABC block. Thus, they should no longer have been uncertain about the values of A during the BC block, nor should they have allowed their biases formed during the BC block to influence their causal judgments. Yet, causal imprinting still occurred, suggesting that it is a non-normative bias.

In addition to supporting our claim that causal imprinting is non-normative, there is another sense in which our modeling analyses may be useful in interpreting our results. Specifically, one might argue based on our results that human learning is consistent with the general Bayesian framework for causal learning and reasoning (Griffiths & Tenenbaum, 2009), with the caveat that sometimes Bayesian inference operates in non-normative ways as well. For example, Bayesian inference may be non-normative when learners compute posteriors using inappropriate theories or prior knowledge, as in our BC–ABC same tokens condition from Experiment 2 (see Griffiths & Tenenbaum, 2009, p. 662). Indeed, when

[14] We thank Eric–Jan Wagenmakers for pointing this out.

allowing for such non-normative forms of Bayesian inference, our causal imprinting results are entirely consistent with Bayesian inference, especially with the predictions of the bounded Bayesian analysis (see Fig. 5).

In fact, when treated as an algorithmic-level (Marr, 1982), or descriptive processing account, the bounded Bayesian analysis actually provides an alternative to our flexible interpretations account of causal imprinting that we presented earlier. Specifically, the bounded Bayesian analysis suggests that causal imprinting occurred in the BC–ABC same tokens condition because participants failed to disregard their prior belief in B-causes-C, even after reading the instructions prior the to the ABC block stating the equivalence of the BC and ABC individuals. This alternative account differs from the flexible interpretation account in that the former does not assume that leaners flexibly interpret the trials from the ABC block. Though technically possible, we view this alternative as somewhat implausible given that the manipulation check in Experiment 2 showed that learners did comprehend the instructions, which would have encouraged disregarding of the BC block. Furthermore, there is already evidence that learners with prior beliefs about a causal relation do flexibly interpret contingency data (Luhmann & Ahn, 2011). Nevertheless, with our current data we are unable to determine which of the processes suggested by these accounts is more accurate given that our causal judgments were all taken at the end of learning, whereas interpretation effects are most discernable in trial-by-trial dynamics (e.g., Luhmann & Ahn, 2007, 2011). Future studies would benefit from examining trial-by-trial dynamics in order to better understand the processes underlying causal imprinting effects.

8.3. Phenomenon closely related to causal imprinting

In broad strokes, the current experiments showed that people were hesitant in revising their prior causal beliefs based on later evidence. Based on this characterization, we note that causal imprinting relates to a number of other interesting phenomena where people exhibit difficulty in attempting to unlearn something. This includes not only the learning of causal relations, but also of general associations and facts.

For instance, studies of conditioning have shown that animals are often slower to learn that a previously reinforced stimulus is no longer reinforced (e.g., to stop pushing a lever when it no longer produces food) than they are to learn that a previously non-reinforced stimulus is now reinforced (e.g., to start pushing a lever when it begins to produce food; Rescorla, 2002). These findings support the more general claim that learning rates are slower during extinction than during initial acquisition (Bush & Mosteller, 1951; Lovejoy, 1968; Wagner, Logan, & Haberlandt, 1968). That is, unlearning seems to be generally more difficult than learning.

As another example, studies from social psychology with humans have shown that participants who are debriefed after an experiment persist in believing the false ideas conveyed in the experiment, even when they were fully discredited in the debriefing (Anderson, Lepper, & Ross, 1980; Ross, Lepper, & Hubbard, 1975). For example, participants in Ross et al. (1975) attempted to distinguish authentic from fake suicide letters and received feedback on their performance. In fact, the experimenter determined the feedback randomly, and participants were told so in the debriefing. Nevertheless, after the debriefing those who had received mostly positive feedback during the task rated their abilities in this and similar tasks as higher than those who had received more negative feedback. The current studies are similar in that they also found belief persistence despite the later evidence that invalidated the earlier made inferences.

The resistance to unlearning is also consistent with philosophical and artificial intelligence views, which argue for conservative belief revision (Doyle, 1992; Gärdenfors, 1992a; Harman, 1988; Quine, 1986). Though most of this work has addressed revisions of logical/deterministic predicates, similar arguments can be applied to revisions in probabilistic causal knowledge. One argument for conservative belief revision is that keeping track of where our beliefs come from (i.e., the evidence they are based on) is computationally expensive (Gärdenfors, 1992b), and thus, the benefits of retaining beliefs that are not obviously contradicted by new evidence outweigh the costs of storing the sources of the beliefs in case they come under dispute. In our experiments, participants in the BC–ABC conditions may have stored their beliefs in B-causes-C independently from the contingency data, essentially purging memory for individual trials in the BC block after it was complete. If so, the later evidence

in the ABC block showing that B and C were conditionally independent should not necessarily negate the prior belief entirely, given that one cannot refute evidence that is not remembered.

Note, however, that the previous argument does not hold for the BC–ABC same tokens condition in Experiment 2, where the source of the original belief was re-presented (i.e., the values of A from the BC block were revealed and presented alongside the original BC block), and yet, we also found causal imprinting. Given this finding, perhaps causal imprinting stems from an over-generalized tendency to avoid belief revision in the interests of conservatism.

8.4. Moderating factors for causal imprinting

We want to clarify that while causal imprinting is an apparently robust phenomenon, there are likely to be boundary conditions on the effects we have shown in this paper. For one, we do not argue that causal imprinting is a permanent consequence of initial learning or that learners will never dismiss their prior beliefs. Rather, in the current studies, we have shown that when viewing a single block of contradicting evidence of the same length as the block of initial evidence, the initial evidence seemed to carry more weight or have more influence, presumably due to affecting how the later evidence was interpreted. However, this result does not guarantee that if the later evidence were repeated numerous times or given more emphasis that learners imprinted with their original belief would not eventually dismiss that belief.

Indeed, one study found that participants in a scenario similar to the BC–ABC condition, but with extended learning and a form of feedback, did eventually dismiss their prior belief (Taylor, 2010). Recency effects are also observed in studies of category learning, where feedback about the correct categorization of an item is provided on each trial (Jones, Love, & Maddox, 2006; Jones & Sieck, 2003). However, we do not know whether causal imprinting will spontaneously disappear in the absence of feedback with only repeated presentations of the ABC block. Furthermore, Taylor (2010) showed that the prior belief continued to exist at a more implicit level when measuring learners' expectations without referring explicitly to causal relations. Further examining the robustness of causal imprinting and whether residual beliefs persist even after extended learning would be important in understanding how we can dispel causal imprinting.

Another boundary condition on causal imprinting concerns the types of initial evidence we expect to result in causal imprinting. We do not claim that causal imprinting always occurs when a learner views a correlation between two events in the absence of their common cause. Instead, it should happen only when the learner firmly believes that a causal relation exists between the correlated events. If the initial evidence is very weak or consists of very few trials, as in the 5-trial studies of Fernbach and Sloman (2009), learners may be more willing to revise their initial beliefs based on later evidence. In addition, the plausibility of the causal relation implied by the initial evidence may play a mediating role. For example, if one observes that ice cream consumption and drowning rates are correlated, they may not immediately infer that ice cream consumption causes drowning (or vice versa), given the lack of a plausible mechanisms linking these two events. To the contrary, in these circumstances learners may even infer the common cause despite having no direct evidence, thereby preventing the causal imprinting effect from occurring. In general, learners with an awareness or skepticism that a common cause may be lurking in the background may be less likely to show causal imprinting effects.

Similarly, the extent to which later evidence contradicts the initial belief may also moderate the causal imprinting effect. For instance, in previous studies showing primacy effects in causal learning (e.g., Dennis & Ahn, 2001; Marsh & Ahn, 2006), participants maintained their initial causal beliefs even after viewing blatantly opposing contingencies. Such primacy effects may be less robust and enduring than the causal imprinting, as participants in these studies were likely aware that they had to justify these opposing contingencies. Indeed, in Fernbach and Sloman (2009), the contradictory evidence was even more obvious than in the current study—a single trial where the cause appeared in the absence of the effect, casting doubt on a causal relation between these factors. This direct conflict actually led to a recency effect. In contrast to these previous studies, in the current paradigm the conflict in the later evidence was much less obvious (i.e., the conditional independence of B and C), and some participants may have failed to notice it entirely. Hence, our order effects may be fundamentally different and perhaps even much more persistent, given that the form of conflicting evidence is much more subtle.

Finally, much previous work on causal learning has shown that whether primacy or recency effects are observed depends greatly on the specific methods used, such as presentation format, working memory load, frequency of elicited judgments, and perhaps even the total number of trials (Fernbach & Sloman, 2009; Glautier, 2008; Hogarth & Einhorn, 1992; Marsh & Ahn, 2006). Other factors such as the temporal delay between causes and effects also have a notable impact on causal structure and strength judgments (Greville & Buehner, 2010; Lagnado & Sloman, 2004, 2006; Shanks, Pearson, & Dickinson, 1989). In light of these dependencies, we decided to prioritize ecological validity by modeling our experiments after a real world scenario, the myopia scenario, while aligning with recent work on causal structure learning where trial by trial presentation format was used (Lagnado & Sloman, 2004, 2006; Steyvers et al., 2003). At the same time, because previous studies have shown suboptimal learning of causal structures among more than two variables, we attempted to reduce working memory load by specifying possible causal directions, and allowing participants to re-view all trials within each block. With this presentation format, we obtained robust causal imprinting effect across Experiment 1–4, with both structure and strength judgments, and using numerous different control conditions. Yet, the amount or strength of causal imprinting may vary when the details of these methods change. In addition, by manipulating presentation orders at the block level rather than trial level, our presentation format does not allow us to infer whether causal imprinting operates at a trial-by-trial level or at a more global level, such as block-by-block level (see Greville & Buehner, 2010 for evidence of global processing).

Future research can examine other potential boundary conditions for causal imprinting. Such research can also shed light on implementing intervention techniques for overcoming or preventing causal imprinting. For example, a domain where it is critical to avoid causal reasoning biases is legal reasoning. Order effects may occur in courtrooms if some initial evidence is presented in favor of a ruling, which then biases the jurors' and judge's interpretations of later evidence (Lagnado, 2011). Such examples underscore the importance of research on human reasoning biases and ways to correct them.

9. Conclusions

In four experiments, we provided evidence for causal imprinting during causal structure learning, whereby participants failed to revise their prior belief in a causal relation based on conflicting evidence. Causal imprinting occurred using a variety of dependent measures and control conditions, and also appeared in cases where it was non-normative. We argue that causal imprinting occurs due to the influence of prior knowledge on how people interpret later contingency evidence, which slows the process of belief revision. Future studies are needed to explore the scope of causal imprinting and to explore the conditions that might lead people to revise, rather than persist in their initial causal beliefs.

Acknowledgment

This work was supported by a grant from NIH.

Appendix A

Here we present the results from several additional normative analyses, including those using a restricted range for the b parameters, complexity and simplicity priors, and an alternative functional form. First, we present our analyses when restricting the range of the b parameters to $[0, 0.1]$, in light of recent evidence showing that people tend to assume that background causes are weak (Lu et al., 2008).

As shown in Table A1, the main difference in these analyses and those in the main text is that the posteriors for the common cause decrease slightly overall, while the posteriors for the structure with all three relations increase slightly. Crucially, the relative patterns remain the same, and thus, the pos-

Table A1
Bayesian analyses with restricted range for the *b* parameter. Bolded values are maxima.

								
Posterior probabilities with b parameters sampled from [0,0.1]								
ABC–ABC (No missing values)	0.00	0.00	0.00	**0.71**	0.00	0.00	0.00	0.29
BC–ABC (Reinterpretation)	0.00	0.00	0.00	**0.71**	0.00	0.00	0.00	0.29
BC–ABC (Fully Bayesian)	0.00	0.00	0.00	0.27	0.00	0.00	0.02	**0.71**
BC–ABC (Bounded Bayesian)	0.00	0.00	0.00	0.00	0.00	0.00	0.08	**0.92**

Table A2
Bayesian analyses with alternative priors. Bolded values are maxima.

								
Posterior probabilities with $\theta = 0.5$ (preference for complex)								
ABC–ABC (No missing values)	0.00	0.00	0.00	**0.57**	0.00	0.00	0.00	0.43
BC–ABC (Reinterpretation)	0.00	0.00	0.00	**0.57**	0.00	0.00	0.00	0.43
BC–ABC (Fully Bayesian)	0.00	0.00	0.00	0.39	0.00	0.00	0.02	**0.58**
BC–ABC (Bounded Bayesian)	0.00	0.00	0.00	0.01	0.00	0.01	0.10	**0.88**
Posterior probabilities with $\theta = 20$ (preference for simple)								
ABC–ABC (No missing values)	0.00	0.02	0.01	**0.98**	0.00	0.00	0.00	0.02
BC–ABC (Reinterpretation)	0.00	0.02	0.01	**0.98**	0.00	0.00	0.00	0.02
BC–ABC (Fully Bayesian)	0.00	0.02	0.01	**0.91**	0.00	0.00	0.05	0.03
BC–ABC (Bounded Bayesian)	0.01	0.02	0.02	0.04	0.22	0.05	**0.53**	0.12

Table A3
Bayesian analyses with a linear method for combining multiple causes. Bolded values are maxima.

								
Posterior probabilities using the linear method of combining multiple causes								
ABC–ABC (No missing values)	0.00	0.00	0.00	**0.86**	0.00	0.00	0.02	0.12
BC–ABC (Reinterpretation)	0.00	0.00	0.00	**0.86**	0.00	0.00	0.02	0.12
BC–ABC (Fully Bayesian)	0.00	0.00	0.00	**0.77**	0.00	0.00	0.15	0.07
BC–ABC (Bounded Bayesian)	0.00	0.00	0.00	0.03	0.00	0.00	**0.85**	0.12

sible accounts of causal imprinting based on the fully Bayesian and bounded Bayesian analyses remain valid.

Second, we present our analyses when incorporating prior distributions over hypotheses to reflect a preference for simpler or more complex causal structures. We used the method from Fernbach and Sloman (2009), where the priors were determined using the function $\theta^{-l(h)}/\Sigma\theta^{-l(h)}$, where $l(h)$ is the number of causal links in hypothesis h, and the value of θ varies to determine the preference for simpler hypotheses with few links (values of θ greater than 1) or more complex hypotheses with many links (values of θ between 0 and 1).

We considered 0.5 and 20 as values of θ to demonstrate that even at these relative extremes the normative analyses leads to analogous predictions for B-causes-C relative to the analyses in the main text. As shown in Table A2, for both simplicity and complexity preferences, the fully Bayesian analysis and bonded Bayesian analysis lead to greater belief in structures with B-causes-C than the other two analyses. That is, these analyses continue to show a trend toward causal imprinting relative to the ABC–ABC analysis and the reinterpretation analysis of the BC–ABC condition.

Third, we present analyses using a linear function for combining the influence of multiple causes on a single effect (e.g., as in A-causes-C and B-causes-C both leading to effect C). The following linear function can be substituted for Eq. (3):

$$P(e^{+}|\text{causes}_e) = b_e + \sum_{c \in \text{causes}_e} m_{ce} c_{\text{present}} \quad (A.1)$$

The linear method is not as often used in Bayesian modeling of causal inference, because restrictions are required when choosing the b and m parameters. Specifically, the b and m parameters for a given effect must not sum to a value greater than 1 (e.g., when C is the effect, then $b_C + m_{BC} + m_{AC}$ must not exceed 1). Otherwise, the "probability" of the effect would exceed 1, which is invalid. In our explorations of the linear method, when the sum of the randomly sampled b and m parameters for a given effect exceeded 1, we normalized these parameters so that summed to 1.

The results from these analyses are presented in Table A3. As can been seen, the common cause structure receives the largest posterior for all analyses but the bounded Bayesian analysis, where the posterior for the common cause drops to near zero. Overall, these results are very similar to those from the noisy-OR method. One unique result using the linear method is that the structure with all three relations receives a lower posterior than using the noisy-OR function. Indeed, the causal chain receives the highest posterior according to the bounded Bayesian analysis, rather than the full structure. Crucially, however, this does not change whether and under what conditions these analyses mimic causal imprinting. Just as for the noisy-OR method, a small trend toward causal imprinting is present for the fully Bayesian analysis, and a much larger trend is present for the bounded Bayesian analysis.

References

Ahn, W., & Dennis, M. (2000). Induction of causal chains. In *Proceedings of the 22nd annual meeting of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1990). *The adaptive character of thought*. Lawrence Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037–1049.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1119.

Bush, R. R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review, 58*(6), 413.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition, 18*(5), 537–545.

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

Danks, D., & Schwartz, S. (2005). Causal learning from biased sequences. In *Proceedings of the 27th annual meeting of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Danks, D., & Schwartz, S. (2006). Effects of causal strength on learning from biased sequences. In *Proceedings of the 28th annual meeting of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dennis, M. J., & Ahn, W. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition, 29*(1), 152–164.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology, 36*(1), 29–50.

Doyle, J. (1992). Reason maintenance and belief revision: Foundations vs. coherence theories. In P. Gärdenfors (Ed.), *Belief revision* (pp. 26–51). Cambridge University Press.

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

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 678.

Friedman, N., & Koller, D. (2003). Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning, 50*(1), 95–125.

Gärdenfors, P. (1992a). *Belief revision. Cambridge tracts in theoretical computer science*. Cambridge University Press.

Gärdenfors, P. (1992b). The dynamics belief systems. Foundations versus coherence theories. In C. Bicchieri & M. L. Dalla Chiara (Eds.), *Knowledge, belief, and strategic interaction*. Cambridge, England: Cambridge University Press.

Glautier, S. (2008). Recency and primacy in causal judgments: Effects of probe question and context switch on latent inhibition and extinction. *Memory & Cognition, 36*(6), 1087–1093.

Greville, W. J., & Buehner, M. J. (2010). Temporal predictability facilitates causal learning. *Journal of Experimental Psychology: General, 139*(4), 756.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*(4), 334–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*(4), 661.

Gwiazda, J., Ong, E., Held, R., & Thorn, F. (2000). Vision: Myopia and ambient night-time lighting. *Nature, 404*, 144.

Harman, G. (1988). *Change in view: Principles of reasoning*. MIT Press.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*(1), 1–55.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General & Applied, 79*.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 316.

Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 626.
Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–33). Coral Gables, FL: University of Miami Press.
Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113*(4), 677.
Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review, 7*(4), 636–645.
Lagnado, D. A. (2011). Thinking about evidence. In P. Dawid, W. Twining, & & M. Vasaliki (Eds.). *Evidence, inference and enquiry* (Vol. 17, pp. 183–223). Oxford University Press/British Academy.
Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 856.
Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 451.
Liljeholm, M., & Cheng, P. W. (2007). When is a cause the "same"?: Coherent generalization across contexts. *Psychological Science, 18*(11).
Lopez, F. J., Shanks, D. R., Almaraz, J., & Fernandez, P. (1998). Effects of trial order on contingency judgments: A comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 672.
Lovejoy, E. (1968). *Attention in discrimination learning: A point of view and a theory*. San Francisco, CA: Holden-Day.
Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*(4), 955.
Luhmann, C. C., & Ahn, W. (2007). BUCKLE: A model of unobserved cause learning. *Psychological Review, 114*(3), 657.
Luhmann, C. C., & Ahn, W. (2011). Expectations and interpretations during causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 568.
Marr, D. (1982). *Vision: A computational approach*. San Francisco, CA: W.H. Freeman.
Marsh, J. K., & Ahn, W. (2006). Order effects in contingency learning: The role of task complexity. *Memory & Cognition, 34*(3), 568–576.
Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General, 125*(4), 370.
Moreau, C. P., Lehmann, D. R., & Markman, A. B. (2001). Entrenched knowledge structures and consumer response to new products. *Journal of marketing research, 38*(1), 14–29.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175.
Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgment and Decision Making, 5*(5), 411–419.
Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.
Quine, W. V. O. (1986). *Philosophy of logic*. Harvard University Press.
Quinn, G. E., Shin, C. H., Maguire, M. G., & Stone, R. A. (1999). Myopia and ambient lighting at night. *Nature, 399*, 113.
Rescorla, R. (2002). Comparison of the rates of associative change during acquisition and extinction. *Journal of Experimental Psychology: Animal Behavior Processes, 28*(4), 406.
Rescorla, R., & Wagner, A. (1972). Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*(5), 880.
Rouder, J. N., Speckman, P. L., Dongchu, S., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.
Rottman, B., Ahn, W., & Luhmann, C. C. (2011). When and how do people reason about unobserved causes? *Causality in the Sciences, 150*.
Scheines, R., Spirtes, P., Glymour, C., & Meek, C. (1994). *TETRAD II: Tools for causal modeling*. Pittsburgh PA: Carnegie Mellon University.
Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology, 37*(1), 1–21.
Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology, 41*(2), 139–159.
Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6*(6), 649–744.
Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3), 303–333.
Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7*(6), 337.
Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*(3), 453–489.
Taylor, E. G. (2010). Learning and restructuring causal concepts. In *Proceedings of the 32st annual meeting of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.
Wagner, A. R., Logan, F. A., & Haberlandt, K. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*, 171–180.
Waldmann, M. R., & Hagmayer, Y. (1995). Causal paradox: When a cause simultaneously produces and prevents an effect. In *Proceedings of the 22nd annual meeting of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.
Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*(2), 222.

White, P. A. (2006). How well is causal structure inferred from cooccurrence information? *European Journal of Cognitive Psychology, 18*(03), 454–480.
Yates, J. F., & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta Psychologica, 62*(3), 293–302.