Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 6, 1432–1448

0278-7393/11/$12.00 DOI: 10.1037/a0024829

Effect of Grouping of Evidence Types on Learning About Interactions Between Observed and Unobserved Causes

Benjamin Margolin Rottman and Woo-kyoung Ahn
Yale University

When a cause interacts with unobserved factors to produce an effect, the contingency between the observed cause and effect cannot be taken at face value to infer causality. Yet it would be computationally intractable to consider all possible unobserved, interacting factors. Nonetheless, 6 experiments found that people can learn about an unobserved cause participating in an interaction with an observed cause when the unobserved cause is stable over time. Participants observed periods in which a cause and effect were associated followed by periods of the opposite association ("grouped condition"). Rather than concluding a complete lack of causality, participants inferred that the observed cause does influence the effect (Experiment 1), and they gave higher causal strength estimates when there were longer periods during which the observed cause appeared to influence the effect (Experiment 2). Consistent with these results, when the trials were grouped, participants inferred that the observed cause *interacted* with an unobserved cause (Experiments 3 and 4). Indeed, participants could even make precise predictions about the pattern of interaction (Experiments 5 and 6). Implications for theories of causal reasoning are discussed.

Keywords: causal reasoning, causal learning, time, unobserved causes

Many events are produced through interactions involving multiple factors. For instance, John's driving to work this morning depends not only on his turning the ignition key but also on the presence of oxygen in the air, gas in the fuel tank, and the battery being charged. Furthermore, it depends upon all factors that contributed to the existence and functioning of John and his car up to that moment, such as John's salary being sufficient for him to own a car.

Understanding specific ways in which causes interact can be crucial. For instance, Waldmann (2007) examined cases in which multiple causes average or add up to produce an effect. Suppose that one observes that Drug A alone increases a person's heart rate by 20 beats per minute compared with normal and observes that the combination of Drug A and Drug B increases a person's heart rate by 20 beats per minute compared with normal. If one believes that the effect of drugs combine additively, one may infer that Drug B does not influence heart rate at all, whereas if one believes the effect of drugs combine by averaging, one may infer that the second drug still has causal efficacy in influencing heart rate. In a classic article, Kelley (1972) described multiple necessary and multiple sufficient cause schemata, which aid in various causal inferences involving multiple causes.

While understanding how multiple causal factors interact is crucial for causal inference, people cannot reason or learn about all possible causal interactions, not just because it is beyond their computational limitations but also because only a fraction of the interacting causes are observable. There have been normative accounts for computing the causal strength of interacting causes (e.g., Novick & Cheng, 2004), but such accounts are based on *observing* how interacting causes behave and cannot capture many real-life situations in which interacting causes are *unobservable* or initially *unattended*.

Few previous studies have examined whether people make inferences about how unobserved causes interact with observed causes. Instead, previous models of causal learning made simplified assumptions about how unobserved and observed causes combine. The goal of the current study is to empirically demonstrate that under certain conditions people deviate from these traditional assumptions and learn that observed and unobserved causes may interact in complicated ways.

The outline of this introduction is as follows. We first briefly review the two most prominent assumptions that previous models have made about the relationship between observed and unobserved causes. Then we describe a case in which observed and unobserved causes interact in a different way. Finally, we describe conditions in which people would infer unusual (or nontraditional in the causal learning literature) causal interactions, despite the interacting cause being unobserved.

This article was published Online First August 8, 2011.

Benjamin M. Rottman and Woo-kyoung Ahn, Department of Psychology, Yale University.

This research was supported by a National Science Foundation Graduate Research Fellowship to Benjamin M. Rottman and National Institutes of Mental Health Grant R01 MH57737 to Woo-kyoung Ahn. The authors thank Rachel Litwin for assistance with data collection and Alan Wagner, Frank Keil, Brian Scholl, and Laurie Santos for helpful comments on the studies presented in this article.

A discussion of parts of Experiment 3 can be found in a section of the book chapter "When and How Do People Reason About Unobserved Causes?" (Rottman, Ahn, & Luhmann, 2011). Experiments 1 and 3 were presented at the 31st Annual Conference of the Cognitive Science Society (Rottman & Ahn, 2009a).

Correspondence concerning this article should be sent to Benjamin M. Rottman, who is now at University of Chicago, 5841 S. Maryland Ave., MC5000, Chicago, IL 60615. E-mail: benjaminrottman@uchicago.edu

Assumptions About How Observed and Unobserved Causes Combine

There are two prominent assumptions made by existing theories of causal learning in terms of how unobserved causes interact with observed causes.

Additivity, or Linear Integration Function

Some models of causal learning have assumed that causes combine additively (e.g., Jenkins & Ward, 1965; Rescorla & Wagner, 1972). For example, if two cues each have associative strengths of .5, when presented simultaneously, the Rescorla-Wagner algorithm, not assuming a configural cue, suggests that the subject would predict the probability of the outcome to be the sum of the associative strengths of the individual elements, 1. This algorithm treats unobserved causes the same way: A background cue, which is often interpreted as the sum of all unobserved causes (Hagmayer & Waldman, 2007; Shanks, 1989), contributes to the total associative strength in an additive manner. Griffiths and Tenenbaum (2005) have demonstrated that another model of causal strength, ΔP (Jenkins & Ward, 1965), is the maximum likelihood estimate of causal strength if one observed cause and one unobserved cause combine linearly.

Noisy-OR and Noisy-AND-NOT Integration Function

Other models have assumed a "noisy-OR" integration (e.g., Cheng, 1997; Pearl, 1988, 2000), which essentially describes situations involving multiple sufficient causes. In noisy-OR, the presence of an effect is determined by the union of independent causes. Thus, the likelihood that an effect occurs is the sum of the likelihood that an observed cause, C, produces the effect and the likelihood that all other alternative (observed or unobserved) causes, A, produce the effect, minus the likelihood that C and A together produce the effect. Likewise, a noisy-AND-NOT integration has been used for inhibitory causes. (See Novick & Cheng, 2004, for probabilistic conjunctive causal interactions among multiple observed causes.)

Reasoning About Other Combinations?

So far, we briefly reviewed many models of causal learning that assume that unobserved causes combine with observed causes either through additive or noisy-OR functions. There are two limitations to this approach. First, one of the most intuitive ways that observed and unobserved causes may interact is neither additive nor noisy-OR; instead, an unobserved cause may be a *necessary enabling condition* for the observed cause. For instance, if a car battery is uncharged (unobserved enabling condition), the car will not start, even if the ignition key is turned (observed cause). It is likely that people understand that all observed causes have enabling conditions, suggesting that people know other ways that causes combine.[1]

Second, it is also unlikely that people assume that all observed and unobserved causes combine in the same way. The three types of interactions we discussed so far (i.e., additive, noisy-OR, and multiple necessary) are all plausible forms of interactions among causes, and it seems quite implausible that people only use one of these three. Indeed, when it comes to *observed* causes, recent studies have uncovered that people can flexibly learn that different causes combine through different functions (Beckers, De Houwer, Pineño, & Miller, 2005; Lu, Rojas, Beckers, & Yuille, 2008; Lucas & Griffiths, 2010). Yuille and Lu (2008) have developed a "grammar" to express the form of any probabilistic logical combination of causes.

Given that there are several plausible ways in which observed and unobserved causes can interact and that people flexibly learn these different combinations involving *observed* causes, it seems possible that people would also flexibly learn different ways that *observed and unobserved* causes combine. Nonetheless, it is also possible that people may not be able to flexibly learn different ways that *observed and unobserved* causes combine. Learning interactions between one observed and one unobserved causes is more difficult than with two observed causes, because the state of the unobserved cause needs to be inferred. Thus, whether people can infer how observed and unobserved causes interact warrants empirical study.

To demonstrate that people can flexibly infer specific ways in which observed and unobserved causes combine, we investigated the "biconditional" interaction between an observed and an unobserved cause as a paradigm case. Given the extensive research on linear and noisy-OR interactions and the fact that biconditional interactions are fairly complicated, it is unlikely that people would use the biconditional interaction as a default assumption for how causes combine. Thus, if people do learn about biconditional interactions, it would suggest that people make novel inferences about how observed and unobserved causes interact. In the next section, we describe the biconditional interaction in more detail.

The Biconditional Interaction

Consider a case of two causes (C_1 and C_2) and an effect (E), which can take the values 0 or 1. In a biconditional interaction (analogous to an XOR relation and negative patterning),[2] $E = 1$ if both causes have the same value (i.e., $C_1 = C_2 = 1$, or $C_1 = C_2 = 0$), and $E = 0$ if the two causes have different values (i.e., $C_1 = 0$ and $C_2 = 1$, or $C_1 = 1$ and $C_2 = 0$).

The biconditional interaction has a unique property compared with the linear and noisy-OR integration functions. According to the other two functions, the probability of an effect is different given alternative states of the observed cause (e.g., $C = 1$ or $C = 0$).[3] But if two causes combine through a biconditional relationship, this is not necessarily true. As demonstrated in the extended example below, this property of the biconditional interaction poses challenges for causal inference.

One famous interaction example in the psychology literature is the relationship between parenting styles and child development (see Darling & Steinberg, 1993, for a review). "Authoritative"

[1] Research from the animal-conditioning literature has investigated similar cases. For example, an "occasion-setting" cue may signal periods of time during which another cue is paired with an outcome. And "positive-patterning" describes situations when an outcome only occurs if two cues are simultaneously present.

[2] Because 0 and 1 reflect alternative states of the causes and effects in this article, the biconditional interaction can also be described as an XOR interaction and negative patterning with the 0 and 1 values exchanged.

[3] This is true for noisy-OR with the caveat of ceiling and floor effects.

parents give emotional support, set high standards, and allow autonomy, whereas "authoritarian" parents require strict adherence to rules. Within middle-class, European American families, authoritative parenting leads to the most academic success. Yet within minority populations, authoritarian parenting often leads to the most academic success. The pattern of results can be summarized in Figure 1, which reports the outcomes of 50 European American (white cells) and 50 minority (gray cells) families, each with 25 authoritative and 25 authoritarian families. If we know that race is a relevant variable, we can look within the different colored cells and notice the opposite relationships between parenting style and academic success for the two populations, easily identifying the interaction.

However, when one of the interacting causes is unknown, interactions of this sort pose a problem for causal inference. For example, consider Figure 1 again pretending that race information is unknown (i.e., by ignoring the shades of the cells). Since race is not observed, it is impossible to test for an interaction between race and parenting style. Furthermore, the probability of academic success is the same given the two forms of parenting, suggesting that parenting style is not a cause of academic success.

In sum, if people do not consider that an observed cause interacts with an unobserved cause, they may make incorrect inferences.[4] Yet it is difficult to always consider interactions with unobserved causes. For example, researchers initially found a correlation between authoritative parenting and academic success using a sample primarily composed of European American families (Baumrind, 1967). However, because there was no a priori reason to test whether race interacted with parenting style, this interaction was only discovered later (Baumrind, 1972).

Learning a Biconditional Interaction When an Unobserved Factor Is Stable

Although discovering a biconditional interaction involving an unobserved cause is difficult, we propose that people may overcome this challenge if the unobserved factor is fairly stable over time. To understand the influence of stability, we illustrate an example of a biconditional interaction between two causes of an effect, where the unobserved cause is either stable or unstable over time (see Figure 2). Suppose there are two light switches connected to one light, and the light is on if both switches are up or down, but the light is off if one switch is up, and the other is down. In this scenario, an observer can only see one of the switches and the light, and the contingency between the switch and the light is presented to the observer in the order in which the events take place. Even if the observer is unaware of one of the switches, he or she may be able to use temporal information to infer an interaction with an unobserved factor if it is fairly stable.

	Academic Success	
	Yes	No
Authoritative	25	25
Authoritarian	25	25

Figure 1. Hypothetical parenting on academic achievement data. Note: White cells reflect European American families, and gray cells reflect minority families.

Grouped

Steps	1	2	3	4	5	6	7	8
Switch	0	1	0	1	1	0	1	0
Light	0	1	0	1	0	1	0	1
U	1	1	1	1	0	0	0	0

Ungrouped

1	2	3	4	5	6	7	8
0	1	1	0	0	1	1	0
0	0	1	1	0	0	1	1
1	0	1	0	1	0	1	0

Trial Summary

Switch \ Light	1	0
1	2	2
0	2	2

Figure 2. Double light switch scenario. Note: For the switch, 0 = "down" and 1 = "up." For the light, 0 = "off" and 1 = "on." "U" is an unobserved biconditional interacting factor. Cells are white for the groups composed of ($C = 1$, $E = 1$) and ($C = 0$, $E = 0$) trials. Cells are gray for the groups composed of ($C = 1$, $E = 0$) and ($C = 0$, $E = 1$) trials.

For example, suppose you enter a room for the first time and observe that when you flip a switch up, a light goes on, and when you flip it down, the light goes off, as illustrated in Steps 1–4 of the "Grouped" table in Figure 2. If you assume that other potential causes of the light are fairly stable and did not happen to change at the same instant you flipped your switch, you would infer that the observed switch influences the light. At Step 5, however, the light turns off without anyone touching the observed switch, because an unobserved cause (*U*) in Figure 2 changed. (For instance, another person may have pulled down the second switch unbeknownst to you.) Afterward, when the observed switch is down, the light is on, and the light is off when the switch is up (Steps 5–8). From this scenario, you might be very confident that your switch influences the light; there were two long periods when the status of the switch correlated with the status of the light. In addition, because the light mysteriously turned off, and because the two periods had opposite associations between the switch and light, you might infer an unobserved factor that interacts with your switch, explaining the overall zero contingency between the switch and light. Stated in a different way, if people understand that the observed cause interacts with an unobserved cause to produce an effect, they may still think that the observed cause influences the effect despite a lack of overall contingency.

However, if the unobserved interacting factor is unstable and changes frequently, inferring the interaction may be much harder. Suppose the same data we just discussed are rearranged as in the Figure 2 "Ungrouped" table, such that *U* changed frequently. Initially, the switch is down and the light is off (Step 1). In Step 2 the switch is flipped up, but the light still stays off. In order to believe that the switch is causally efficacious, one must infer that at the moment the switch was flipped, an unobserved factor coincidentally changed and counteracted the effect of the observed switch, as specified under column "*U*" (unobserved interacting factor). Then, in Step 3, the light turns on without flipping the switch, and so on. Thus, for the situation shown in Figure 2 Ungrouped, it would be extremely difficult to infer the switch to be causally efficacious; the switch cannot be the sole cause of the light because there is zero contingency with the light. Furthermore, it would be difficult to infer it as part of an interaction because doing so would require inferring an unobserved factor operating as specified under row "*U*," which is counterintuitive; the unobserved interacting factor is highly unstable and (intuitively) exceedingly

[4] Simpson's paradox presents a related but distinct problem. In Simpson's paradox, if the learner takes into consideration the second cause, the learner will make a very different inference than if the second cause is ignored (Spellman, Price, & Logan, 2001).

complicated to track. Finally, it is highly coincidental that *U* repeatedly changes at the same instant the observed switch was flipped. Instead, it seems likely that people would infer an unobserved factor that is entirely responsible for the light.

These two examples were meant to demonstrate that if observations are organized into groups of trials with the same contingency, reflecting relatively stable background causes (e.g., Figure 2 Grouped), rather than intermixed, reflecting highly unstable and coincidental background causes (e.g., Figure 2 Ungrouped), people may be more likely to infer that an interaction is taking place and that the observed cause is still efficacious. Alternatively, if one simply uses the overall contingency between the observed switch and light, one would conclude that there is no causal relationship in both the grouped and ungrouped scenarios. As shown in Figure 2 Trial Summary, there is no contingency between the observed switch and light, and thus, an unobserved factor could be entirely responsible for the light and the observed switch could be irrelevant.

We suggest that in the real world, unobserved causes often are stable, and assuming stability would be a rational assumption that would facilitate learning. For example, if you press a button on a television remote control and the channel changes, you would likely infer that your button press caused the channel to change and that that channel change was not caused by some other unobserved factor that coincidentally changed at the moment you pressed your button. (If, by chance, your sibling did have a second remote and pressed other buttons whenever you press a button, they would likely succeed in confusing you.) Or when you are trying to learn which foods you are allergic to by trying different foods on different days, it is safe to assume that your general health condition remains fairly constant as you try different foods. Thus, if you do get an allergic reaction, you would likely attribute it to a new food. Indeed, the temporal assumption that unobserved causes are fairly stable and do not happen to change at the exact moment as observed causes is similar to the nontemporal assumption underlying the efficacy of interventions in causal learning, that interventions are independent of other causes of an effect (e.g., Pearl, 2000; Woodward, 2003).

In six experiments, we investigated the role of stability on inferring an interaction with an unobserved cause. We manipulated the grouping of observations reflecting stable versus unstable unobserved causes. In grouped conditions, the trials supporting an association between one state of the cause and effect (white cells in Figure 2) were grouped together, and those supporting the opposite association (gray cells in Figure 2) were grouped together in a trial-by-trial presentation. In ungrouped conditions, these two types of observations were intermixed. Although the data are identical between the two conditions except for different orders, participants may be more likely to infer an unobserved interacting factor in the grouped than ungrouped conditions.

In Experiments 1 and 2, we tested whether people would infer that an observed cause is still causally efficacious in a grouped condition despite a low correlation between the observed cause and effect. In the remaining experiments, we examined whether people appeal to interactions with unobserved causes when asked to explain the grouped conditions. In Experiment 3, we tested whether participants would infer an interaction with an unobserved factor in the grouped condition. In Experiment 4, we tested whether people indeed consider this unobserved factor to be a cause rather than a noncausal cue. In Experiment 5, we tested the specificity of peoples' inferences in an interaction—whether people's beliefs in an interaction corresponded to the proposed biconditional integration function. In Experiment 6, we used a different paradigm to ensure that the results of Experiment 5 were not due to memory demands.

Experiment 1

Experiment 1 compared the grouped and ungrouped conditions across three levels of contingency.

Method

Participants. Thirty-six undergraduates from Yale University participated, either for payment at $10 per hour or for partial fulfillment of an introductory psychology course requirement.

Procedure and design. Participants first read a cover story explaining that they would observe machines with a lever producing different shaped blocks over time. Participants were instructed to determine whether the position of the lever affects the shape of the blocks.

Next, participants saw six scenarios. During each scenario, participants viewed a video of a lever changing position between left and right and blocks changing between two shapes (e.g., square or triangle; see Figure 3). Each of these binary values are denoted as 0 and 1 henceforth. For the cause (*C*), 0 and 1 each represents that the lever was set to the left and right, respectively. For the effect (*E*), 0 and 1 each represents that the machine produced one of two shapes, respectively. Across the six conditions, different shapes were used, but here we just refer to squares and triangles for simplicity. Hereon, a (1, 0) trial denotes that $C = 1$ and $E = 0$. Each scenario had 16 trials, each of which appeared for 2 s, followed immediately by the next trial. The six scenarios were ordered in a Latin square design, such that each of the six scenarios appeared first for some participants.

The six scenarios were created by crossing two levels of Grouping (grouped vs. ungrouped) × three levels of Contingency, $\Delta P = P(E = 1|C = 1) - P(E = 1|C = 0) = .25, .5,$ or $.75$. The frequencies of trial types used to create the three levels of contingency are summarized under "Trial Summary" in Figure 4. In each contingency level, the grouped and ungrouped conditions had the same set of 16 trials, but they were presented in different orders. In the grouped conditions, (1, 1) and (0, 0) trials appeared in one cluster (e.g., Trials 1–10 for $\Delta p = .25$ in Figure 4). The (1, 0) and (0, 1) trials appeared in another cluster (e.g., Trials 11–16 for $\Delta p = .25$ in Figure 4). These two different groups of trials suggest different associations between the cause and effect during different

Figure 3. An image of the lever when set to the left and producing a square block.

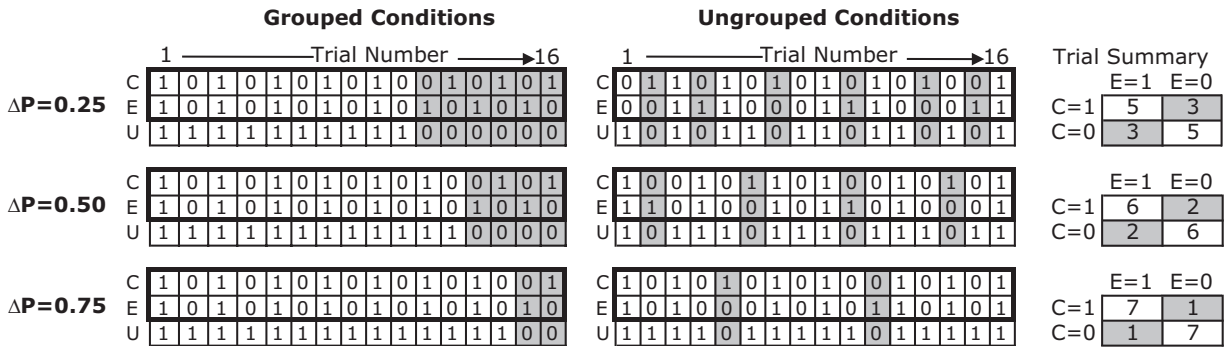

Figure 4. Summary of stimuli in Experiment 1. Note: "C" represents the cause (lever). "E" represents the effect (shape of block). "U" represents an unobserved, interacting, biconditional factor not shown to participants. Cells are white for the groups composed of (1, 1) and (0, 0) trials. Cells are gray for the groups composed of (0, 1) and (1, 0) trials.

periods. In the ungrouped conditions, the four types of trials were intermixed. In Figure 4, the "*U*" rows show what the value of an unobserved factor would need to be in order to postulate that the observed cause and unobserved factor participate in a biconditional relationship to produce the effect. As illustrated in Figure 4, in the ungrouped conditions, if one inferred an unobserved factor for a biconditional interaction, it would have to be highly unstable, which would be very difficult for participants to track and highly coincidental. However, in the grouped conditions, inferring *U* would be more likely, because it would only change once.

Because people often base causal efficacy ratings more on initial than final trials (e.g., Dennis & Ahn, 2001), the trials were presented in the reverse order for half the participants. Psychologically, the main difference in this manipulation was simply when the change between the two groups of data (white vs. gray cells) occurred in the grouped conditions. In the order presented in Figure 4, the change between the two periods occurred in the second half of the trials (e.g., between Trials 14 and 15 in the $\Delta p = .75$, stable condition). In the reverse orders, the change between the two periods occurred in the first half of the trials.

After each scenario, participants answered one causal efficacy question, "To what extent does the lever affect the shape of the blocks?" on a sliding scale from "The lever did not affect the shape at all" to "The lever strongly affected the shape of the blocks," later recoded to 0–100 for analysis. This question is intended as a general measure of causal influence including interaction effects, rather than as a specific measure of simple generative or inhibitory influence. (See Experiments 3–6 for other measures.) If participants believed that a cause interacts with an unobserved cause to produce an effect, they would likely agree that the cause affects the effect.

Results and Discussion

Participants' average causal efficacy ratings for the six scenarios are presented in Figure 5. The pattern of results is consistent regardless of the order of the six scenarios and regardless of whether the order of trials within a scenario was reversed; all analyses collapse across these factors. For all experiments, α was set at .05, and all p values are two-tailed.

The most dramatic finding is that participants gave much higher causal efficacy ratings for the grouped than ungrouped conditions. In a 2 (Grouping) $\times$ 3 (Contingency) repeated-measures analysis of variance (ANOVA), the main effect of grouping was significant, $F(1, 35) = 75.90$, $p < .01$, $\eta_p^2 = .69$. Furthermore, the main effect of contingency was significant, $F(2, 70) = 7.23$, $p < .01$, $\eta_p^2 = .17$, and was significant when tested as a linear trend, $F(1, 35) = 13.59$, $p < .01$, $\eta_p^2 = .28$. There was no interaction between grouping and contingency, $F(2, 70) < 1$.

Follow-up tests reveal that for each of the three contingencies, participants gave higher causal efficacy ratings for the grouped than ungrouped conditions, all ts$(35) > 5.48$, all ps $< .01$. In addition, even though there was no significant interaction between grouping and contingency, we performed separate one-way ANOVAs and linear trend analyses for the grouped and ungrouped conditions to see which one was driving the effect. There was a significant main effect of contingency for the ungrouped condition, $F(2, 70) = 5.84$, $p < .01$, $\eta_p^2 = .143$, and this effect was significant as a linear trend, $F(1, 35) = 9.92$, $p < .01$, $\eta_p^2 = .221$. However, there was no significant main effect of contingency in the grouped condition, $F(2, 70) = 1.93$, $p = .15$, $\eta_p^2 = .05$, and this effect was not significant as a linear trend, $F(1, 35) = 2.41$, $p = .13$, $\eta_p^2 = .06$. (But note that these two effects are not significantly different.)

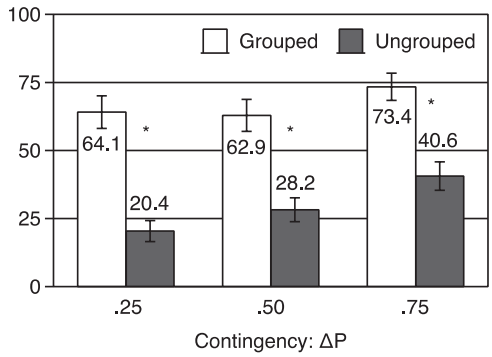

Figure 5. Mean causal strength ratings and standard errors in Experiment 1. * $p < .01$.

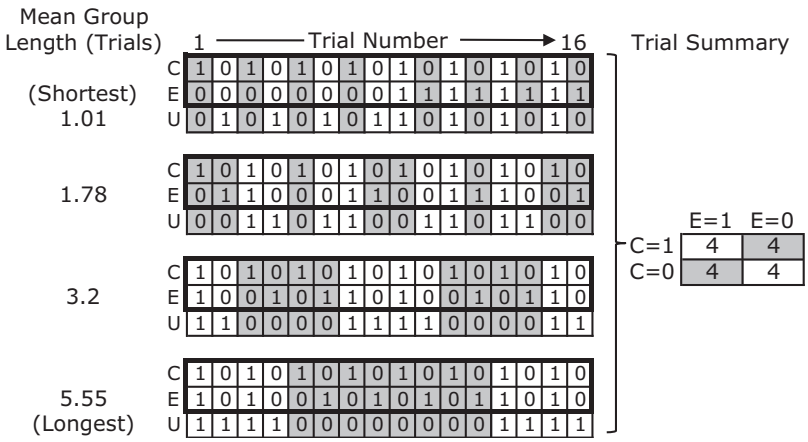

Figure 6. Summary of stimuli in Experiment 2. Note: "C" represents the cause (lever). "E" represents the effect (shape of block). "U" represents an unobserved, interacting, biconditional factor not shown to participants. Cells are white for the groups composed of (1, 1) and (0, 0) trials. Cells are gray for the groups composed of (0, 1) and (1, 0) trials.

In summary, Experiment 1 found that participants inferred higher causal strengths when the data were grouped than ungrouped, even though participants observed identical trials in the two conditions. Experiments 3–6 investigate whether people appeal to interactions with unobserved causes when asked to explain the grouped conditions.

Experiment 2

Experiment 2 extended the findings of Experiment 1 in two ways. First, Experiment 2 tested the most extreme version of a biconditional interaction—when there is zero correlation between the observed cause and effect. By testing whether participants are willing to infer causality even in the absence of any simple contingency, Experiment 2 is a more robust test of the phenomenon demonstrated in Experiment 1.

Second, while Experiment 1 only used two levels of grouping, Experiment 2 employed four levels of grouping to examine whether participants' causal attributions are a continuous function of the amount of grouping. We predicted that people would give higher causal strength estimates for the observed cause with higher degrees of stability of the unobserved cause. Thus, the longer the groups are, the more time there is for participants to accumulate evidence and become confident that the observed cause does in fact influence the effect.

Method

Participants. Twenty-four participants from the same population as in Experiment 1 completed the study.

Procedure and design. The cover story was the same as in Experiment 1. There were four conditions, each of which had 16 trials with zero correlation between the observed cause and effect.

The only difference between the four conditions was the order of the trials. The four conditions comprised a range of average group length, where a group indicates an uninterrupted stream of (1, 1) or (0, 0) trials, or an uninterrupted stream of (1, 0) or (0, 1) trials. The four conditions had an average group length of 1.01, 1.78, 3.20, or 5.33 trials (see Figure 6). Longer groups reflect a more stable unobserved factor.

All four conditions were presented to each participant, and the order of the four conditions was counterbalanced between subjects in a Latin-square design. After each scenario, participants answered the same causal strength question from Experiment 1.

Results

The pattern of results described below held up when only looking at the first condition seen by each participant; thus, the following analyses collapse across different orders of the four conditions.

As can be seen in Figure 7, there is a monotonic increase in causal strength ratings with increasing mean group length, despite that the overall contingency was constant. A repeated-measure one-way ANOVA confirmed the main effect of mean group length, $F(3, 69) = 11.29$, $p < .01$, $\eta_p^2 = .33$. Furthermore, testing a linear contrast revealed that the pattern across conditions can be described well by a sloped line, $F(1, 23) = 32.02$, $p < .01$, $\eta_p^2 =$

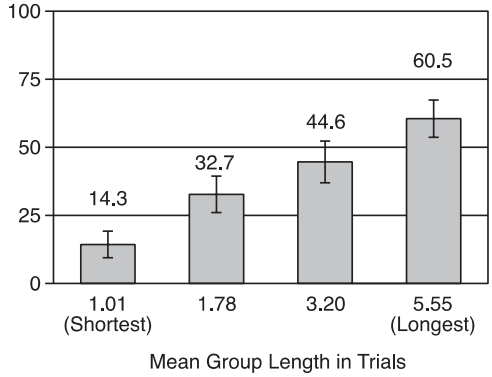

Figure 7. Mean causal strength ratings and standard errors in Experiment 2.

.58. In sum, the longer the groups of trials reflecting a more stable unobserved cause, the higher the causal strength ratings.

There are three important theoretical questions left unanswered from Experiments 1 and 2: (a) whether people actually infer an *interaction* with an unobserved factor; (b) if so, whether this unobserved factor has to be an unobserved *cause*; and (c) exactly *how* people believe the observed and unobserved causes interact. These questions are addressed in the Experiments 3–6.

Experiment 3

Do people actually infer an interaction with an unobserved factor? Experiments 1 and 2 demonstrated that people give higher causal strength ratings when the data are more grouped, which is consistent with what people would do, had they inferred an interaction with an unobserved stable biconditional cause. In Experiments 1 and 2, however, we asked questions only about the observed cause. On one hand, asking only about the observed cause avoids the demand characteristic of alerting participants to the possibility of an interaction with an unobserved factor. Asking about the observed cause has also been the most common paradigm in causal learning experiments (e.g., Cheng, 1997; Jenkins & Ward, 1965; but see Hagmayer & Waldman, 2007; Luhmann & Ahn, 2007) highlighting the uniquely high causal strength ratings in the grouped conditions in Experiments 1 and 2.

On the other hand, since we asked questions only about the observed causes in Experiments 1 and 2, we cannot determine whether participants inferred an interaction with an unobserved cause or whether they believed that the observed cause alone influenced the effect (a main effect of the observed cause). Participants might have inferred higher causal strengths in the grouped than ungrouped conditions in Experiments 1 and 2 without inferring an interaction with an unobserved factor if they used a strategy similar to the classic win-stay lose-shift (WSLS) strategy (e.g., Harlow, 1949; Nowak & Sigmund, 1993).[5] In the grouped scenario, a learner using WSLS would have long periods of time believing one hypothesis (when the switch is left the shape is a square and when the switch is right the shape is a triangle) and then switch to the opposite hypothesis. Within each of these relatively long periods of time, the cause appears to influence the effect. However, in the ungrouped scenario, a learner using WSLS would have to switch more frequently between the two hypotheses, and there are never any long periods during which the cause seems to influence the effect.

Our claim, however, is different from WSLS because we propose that people switch the hypothesis not just because their hypothesis is disconfirmed (as in WSLS) but also because they infer an unobserved interacting cause. In Experiment 3, as an initial attempt to examine whether people appeal to an interaction with an unobserved factor to explain the grouped conditions, we asked at the end of learning the degree to which participants agreed with three statements: that the observed cause (a) alone influenced the effect, (b) interacted with other factors, and (c) had no influence on the effect.[6]

The predictions follow the same reasoning as those in Experiments 1 and 2. That is, in the grouped condition, participants would infer an unobserved cause interacting with an observed cause. Thus, participants would give low ratings on (a) "the observed cause is the only influence on the effect," high ratings on (b) "the observed cause interacts with other factors," and low ratings on (c) "the observed cause has no influence on the effect" (since it influences through an interaction effect). However, in the ungrouped condition, participants would not infer an unobserved cause interacting with an observed cause and would more likely believe that the observed cause had no influence on the effect, because there is zero correlation between the two. Thus, participants would give low ratings on (a) "the observed cause alone influenced the effect." Furthermore, participants would give lower ratings on (b) "the observed cause interacts with other factors" and higher ratings on (c) "the observed cause has no influence on the effect" compared with the grouped condition.

Method

Participants. There were 29 participants from the same population as in Experiment 1.

Procedure and design. There were two conditions, grouped and ungrouped, which had the same 20 trials with zero contingency. Both conditions were presented to all participants in a counterbalanced order. As before, the only difference between the conditions was the order of the trials (see Figure 8).

In Experiment 3, we also made some minor modifications in the cover story to make the stability manipulation more valid. Specifically, the grouped versus ungrouped orders were intended to manipulate the stability or instability of an unobserved cause *over time*, but previously, it was left somewhat ambiguous whether the trial order corresponded to the actual temporal order of the events. Also, to convey the maximum amount of stability in the grouped condition, the machine had to be the same machine over time. That is, if the machines varied across trials, unobserved causes would also vary, even in the grouped condition. To clarify the cover story for participants, that they were observing the behavior of *one machine over time*, we included a picture of one machine for the 20 trials within a scenario (see Figure 9). Different pictures were used for the grouped and ungrouped conditions.

After each scenario, to understand the relationship between their beliefs about the interaction and the main effects of the observed cause, participants rated their agreement with three statements from 1 (*absolutely disagree*) to 9 (*absolutely agree*). The statements were as follows:

1. "The lever alone influenced the shape of the blocks."
2. "A combination of the lever and some other factor influenced the shape of the blocks."[7]

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

[6] We are not necessarily claiming from this study that people make such inference to an unobserved cause during learning. See the General Discussion for more discussion on this issue.

[7] It is likely that our participants would not know the precise meaning of the word "interaction," so we used "combination" for simplicity. Experiments 5 and 6 clarify exactly how participants thought that the two causes combined.

Figure 8. Summary of stimuli in Experiment 3. Note: "C" represents the cause (lever). "E" represents the effect (shape of block). "U" represents an unobserved, interacting, biconditional factor not shown to participants. Cells are white for the groups composed of (1, 1) and (0, 0) trials. Cells are gray for the groups composed of (0, 1) and (1, 0) trials.

3. "The lever had no influence on the shape of the blocks."[8]

These three statements can be interpreted in terms of the three possible causal structures as shown in Figure 10. Statement 1 reflects a structure in which only C causes E. Statement 2 reflects a causal structure in which both C and U cause E. Statement 3 reflects a causal structure in which only U causes E. Note that no two of these questions are exact opposites of one another. In general, the more a person agrees with one, the less he or she should agree with the other two, so they are partially dependent. These questions are designed to allow participants to show which of the three options they agree more with, and participants may potentially be agnostic across the three.

Results

The pattern of results was consistent regardless of the order of the conditions; thus, the following analyses collapse across order. Because the three dependent variables were somewhat dependent upon one another, a doubly repeated measures general linear model was performed (see Kerr, Hall, & Kozub, 2002, for an introduction). There was a significant multivariate difference in a linear combination of the three dependent variables across the grouped versus ungrouped conditions, $F(3, 26) = 5.91$, $p < .01$, $\eta_p^2 = .41$. Since there was a significant difference, the three statements are analyzed in turn below. Given that we report the results of nine t tests, we also performed a Bonferroni adjustment of the alpha level. All the significant results reported below are still significant after the Bonferroni adjustment.

Statement 1: Lever alone influenced shape. In the grouped conditions, almost every time the cause changed, the effect also changed. However, participants predominantly disagreed that the lever alone influenced the shape in both conditions (see Figure 10a). One-sample t tests revealed that the means of both conditions were significantly less than the middle of the scale (5): for grouped, $t(28) = 7.42$, $p < .01$; for ungrouped, $t(28) = 6.97$, $p < .01$. In addition, a paired t test revealed that there was no significant difference between the grouped and ungrouped conditions, $t(28) < 1$. These results demonstrate that people do not simply infer that the observed cause is entirely responsible for the effect in the groped conditions. Statements 2 and 3 clarify their inferences.

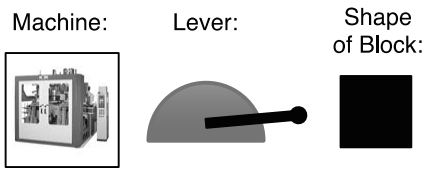

Figure 9. Image of one trial in Experiment 3.

Statement 2: Combination of lever and other factor influenced shape. Participants inferred an interaction with an unobserved factor much more in the grouped than ungrouped condition, $t(28) = 4.20$, $p < .01$ (see Figure 10b). In the grouped condition, participants agreed with this statement; the mean was significantly above the middle of the scale, $t(28) = 6.50$, $p < .01$. However, when an unobserved biconditional cause is very unstable, it is much more difficult to infer an interaction. Participants in the ungrouped condition were ambivalent about this statement; the mean was not significantly different from the middle of the scale, $t(28) < 1$.

Statement 3: Lever had no influence on shape. Participants were much more likely to infer that the lever had no influence on shape in the ungrouped than grouped condition, $t(28) = 3.64$, $p < .01$ (see Figure 10c). In the grouped condition, participants significantly disagreed that the lever had no influence shape; the mean was below the middle of the scale, $t(28) = 3.95$, $p < .01$. This is presumably because they believed that the lever influenced the shape through an interaction as demonstrated with their judgments on Statement 2. But in the ungrouped condition, there was less reason to believe that the lever influenced shape; after all, there was zero correlation between the observed cause and effect.

Participants in the ungrouped condition were ambivalent about this statement; the mean was not significantly different from the middle of the scale, $t(28) = 1.00$, $p = .33$. This finding replicates the results of Experiments 1 and 2 with opposite wording of the question.

In sum, Experiment 3 demonstrated three points. First, people do not simply infer a stronger *main effect* of the observed cause in grouped than ungrouped conditions. Second, people have some understanding that an unobserved factor can flip the contingency between an observed cause and effect and use such an interaction to explain the grouped data. Third, people did not explain the

[8] Afterward, participants predicted the shape of the block given that the lever was set to the left/right. These questions tested hypotheses that are better addressed by Experiments 5 and 6. We do not think they had any effect on the current results as the findings are consistent with a between-subjects analysis of questions answered prior to the unreported questions. Thus, they are not further discussed.

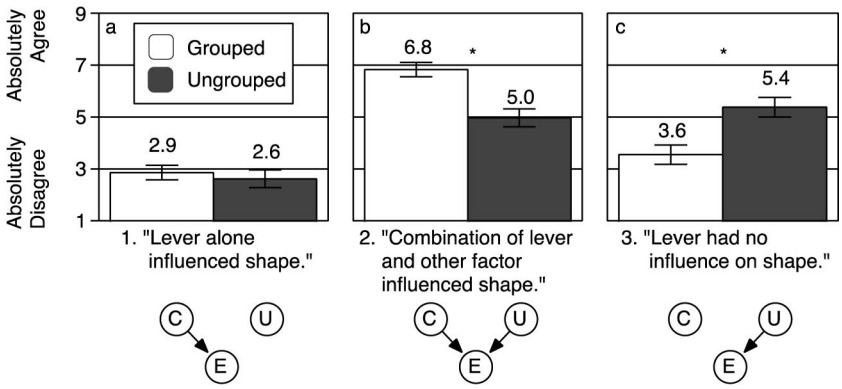

Figure 10. Mean agreement with causal structure statements and standard errors in Experiment 3. * $p < .01$.

ungrouped data with an interaction, which is a possible explanation if they admitted the possibility of the unobserved factor being unstable. This suggests that they believe that unobserved factors are fairly stable and don't happen to coincidentally change at the same instant as an observed cause. These findings do not necessarily imply that people inferred an unobserved interacting cause in Experiments 1 and 2, although it is consistent with such an interpretation. This is discussed further in the General Discussion.

Experiment 4

The results from Experiment 3 suggest that in the grouped condition people inferred that the observed cause interacted with another unobserved factor to produce the effect. One limitation of Experiment 3, however, is that the only possible option for an interaction effect was to presume that the unobserved factor was a cause. A majority of participants might have chosen this option simply because this option refers to an interaction effect, even though they might not have necessarily agreed that the unobserved interacting factor was a cause. Perhaps when participants agreed that the lever and some unobserved factor influenced the shape, they merely used the unobserved factor as a nominal label for the alternating periods of contingency but did not believe that the unobserved factor was the *cause* of the two periods.

In Experiment 4, we tested whether people specifically understand the unobserved factor as a *cause*. In two conditions, participants were told about a machine and were asked whether they thought the machine was responsible for producing the toy blocks, or whether another machine was responsible. In both conditions, the observed machine had one lever and potentially produced one block. Most critically, in both conditions, the contingency between the lever and block was grouped into periods. However, in the "two potential causes condition" (see Figure 11a), the machine had a second unobserved lever (U), whereas in the "one potential cause" condition (see Figure 11b), the machine was described as having only one control mechanism. In order to make the unobserved factor as salient in the one potential cause condition as in the two potential causes condition while keeping it noncausal in the one potential cause condition, we described U as a second unobserved block. If people simply treat U as an associative cue and do not distinguish between causes and effects, then they would not distinguish between these two conditions (see, e.g., Waldmann, 1996; Waldmann, Holyoak, & Fratianne, 1995, for similar manipulations and predictions derived for the associative account). However, if people think that the grouped pattern of data can only be explained by a second unobserved *cause* that interacts with the observed cause, then they would more likely believe that the

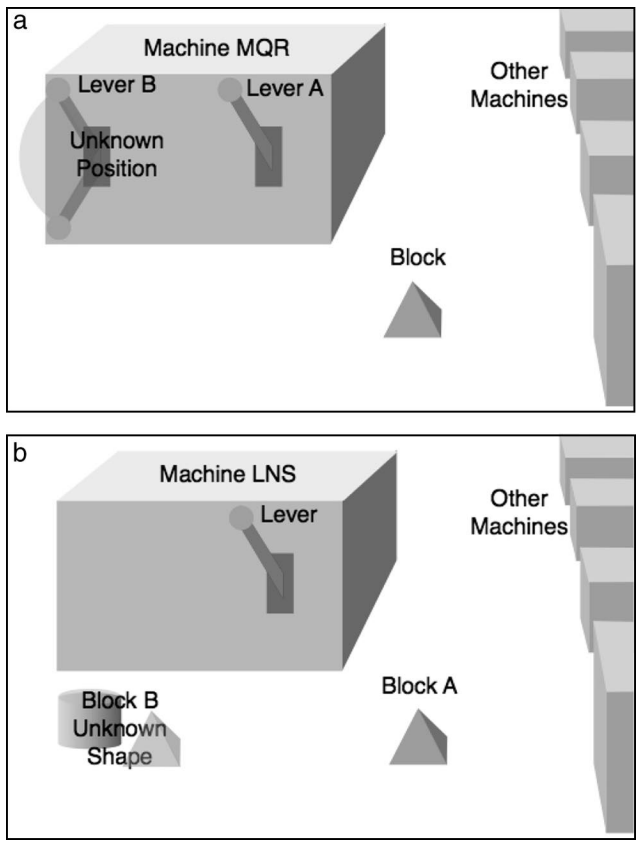

Figure 11. Stimuli from Experiment 4. Note: A is from the "two potential causes" condition, in which the machine had a second lever in an unknown position. B is from the "one potential cause" condition, in which the machine only had one lever and two blocks were produced, but the shape of Block B was unknown. In both conditions, participants chose to attribute the blocks to the target machine or one of the other machines.

machine in the two potential causes condition produced the blocks than the machine in the one potential cause condition. Restated, the grouped pattern can be explained by another unobserved lever but not by another unobserved block (or any other noncausal factor).

Method

Participants. Thirty-nine people were recruited through Amazon Mechanical Turk to participate in an online experiment administered through Qualtrics online survey software. Participants were paid one dollar for approximately 6 min of time. Participants were, on average, 36 years old (SD = 13), 21 were female, all but one had a high-school degree, and 24 had a college degree or higher. One participant skipped one of the two dependent variable questions, and his data were omitted from these demographics and the results below.

Procedure and design. Participants worked with both the two potential causes and the one potential cause conditions in a counterbalanced order. Within each condition, participants were first asked to pretend that they work in a factory with several machines that produce toy blocks. Then they were introduced to the particular machine. In the two potential causes condition (see Figure 11a), participants were told that Machine MQR had two levers, Levers A and B, but they could not see the position of Lever B. The unknown position of Lever B was visually represented by the lever being semitransparent and simultaneously in both the up and down positions. Each day the factory produced one block, either a triangle or cylinder. In the one potential cause condition (see Figure 11b), participants were told that Machine LNS had only one lever. In order to facilitate reasoning about an unknown factor that is not a cause in the one potential cause condition, participants were further told that each day the factory produces two blocks, A and B, although B was unobserved. The unknown shape of Block B was visually represented by a semitransparent cylinder and triangle appearing next to each other. In sum, in both conditions, U was unobserved, but in one case, it was framed as another cause (Lever B) and in another it was framed as another effect (Block B).

Participants then observed a sequence of 20 trials, described as 20 consecutive days. Each day was represented by a picture like those in Figures 11a and 11b; the observed lever could be up or down and the observed block could be a triangle or cylinder. Participants never knew the state of the unobserved cue. All 20 pictures were presented on one webpage in a column, numbered by the day. Participants were told that they could scroll up and down through the days and were asked to look through the 20 days in order. In both conditions, the data for the 20 days was the same as that in the grouped condition in Experiment 3 (see Figure 8).

At the bottom of the webpage below the 20 pictures, participants were asked, "Do you think that Machine X produced the blocks, or do you think that another machine produced the blocks? Please remember that all the blocks over the 20 days were either produced by Machine X, or all were produced by another machine," where X was MQR or LNS in the two conditions, respectively. Participants selected their choice on a scale from 0 (*Machine X produced the blocks*) to 100 (*Another machine produced the blocks*). We used this new question of which machine was responsible for producing the blocks because it still captures causal efficacy but also more clearly captures the interaction effect. If people merely use U to label the alternating periods of contingency, then in both conditions, they could agree that the target machine was responsible for producing the observed block. Alternatively, if people infer that an unobserved cause U must interact with the observed cause, then they would be more likely to answer that the machine in the two potential causes condition produced the block, because the two levers can interact. People should be less likely to answer that the machine in the one potential cause condition produced the block, because there is no other possible cause to interact in the machine.

Results

The pattern of results holds regardless of the order that participants worked with the two conditions, so we collapsed across order. Participants in the two potential causes condition were more likely to conclude that the target machine produced the shape (M = 45, SD = 30) than were participants in the one potential cause condition (M = 63, SD = 33), $t(38) = 2.35$, $p = .02$. That is, despite that both conditions had the identical grouped sequence, participants thought that the sequence of data was more likely produced by a machine with two potentially interacting causes than to a machine in which such interaction was not feasible. In sum, combining the results of Experiments 3 and 4, people believe that the grouped pattern of data is explained by an interaction with a second unobserved *cause*, not merely an associative cue.

Experiment 5

The purpose of Experiment 5 was to determine more precisely people's inferences of how observed and unobserved causes interact to produce an effect. Do people just have vague notions that

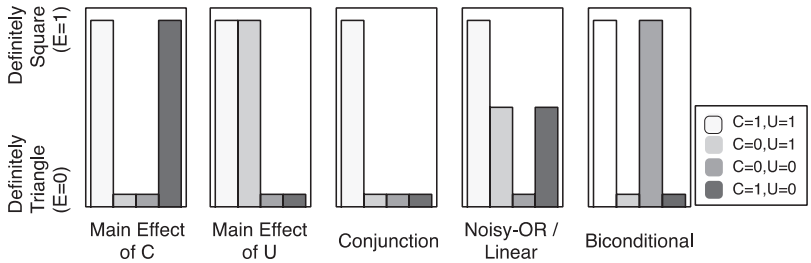

Figure 12. The signature patterns of five ways that C and U could combine.

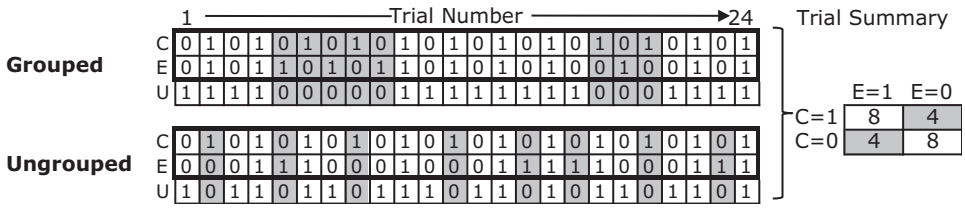

Figure 13. Summary of stimuli in Experiment 5. Note: "C" represents the cause (lever). "E" represents the effect (shape of block). "U" represents an unobserved, interacting, biconditional factor not shown to participants. Cells are white for the groups composed of (1, 1) and (0, 0) trials. Cells are gray for the groups composed of (0, 1) and (1, 0) trials.

observed causes can interact with unobserved factors, or do they actually believe that observed and unobserved causes combine through a biconditional integration function?

To study this question, we had participants predict the effect given the four combinations of the two states of the observed and unobserved causes, as shown in the legend of Figure 12. Similar to Experiments 1–3, participants observed a series of trials with a lever producing different shapes of blocks. On the last trial, a second lever, the previously unobserved factor, was revealed. On this last trial, participants saw the states of both levers and the effect; $C = 1$, $U = 1$, and $E = 1$. From this last trial, participants could work backward to predict what shape would have been produced under the other combinations of the two levers *based on their inference of how the two causes combine*.

Figure 12 displays five likely ways that participants might think that C and U combine. First, C may be the only cause that has a main effect such that $E = 1$ if $C = 1$, and $E = 0$ if $C = 0$, regardless of U. Second, U may be the only cause that has a main effect such that $E = 1$ if $U = 1$, and $E = 0$ if $U = 0$, regardless of C. Third, C and U could combine through a conjunction (multiple necessary causes), in which case $E = 1$ only if $C = U = 1$. Fourth, C and U might also combine in a "noisy-OR" or "linear" fashion. Both of these parameterizations are probabilistic, and they suggest that $E = 1$ is most likely if both $C = 1$ and $U = 1$, less likely if either $C = 1$ or $U = 1$, and $E = 1$ is least likely if $C = 0$ and $U = 0$. Many models of causal learning have assumed a linear or noisy-OR interaction (see the introduction). Fifth, C and U might combine through a biconditional relationship such that $E = 1$ if $C = U = 1$ or if $C = U = 0$, but otherwise $E = 0$.

If participants in the *grouped* condition in Experiments 1–3 actually thought that C and U combined through a biconditional relationship, then the pattern of their predictions would resemble the Biconditional in Figure 12 rather than the other plausible combinations. In contrast, participants in the *ungrouped* condition in Experiment 3 agreed more that the observed factor had no influence on the effect, suggesting that an unobserved factor is primarily responsible. If so, participants' responses would resemble those outlined in Figure 12, Main Effect of U. Such findings would demonstrate that people can learn that observed and unobserved causes may combine through unusual interactions such as a biconditional interaction and that people can make specific predictions about how they interact.

However, there are other plausible outcomes. People in both the grouped and ungrouped conditions may infer that the observed Lever C has a weak main effect (ΔP was .33 for Experiment 5). Or they could infer in both conditions that U is primarily responsible, or some combination like linear or noisy-OR. Or people in *both* conditions could infer that C and U interact in a biconditional interaction. After all, in both the grouped and ungrouped conditions sometimes when the lever was left, it produced a square and right produced a triangle, and sometimes the opposite occurred. Experiment 5 provides the most explicit test of whether people infer a biconditional interaction and do so primarily in the grouped condition.

Method

Participants. There were 16 participants from the same population as in Experiments 1–3.

Procedure and design. There were two conditions: grouped and ungrouped. Both conditions had 24 trials and a ΔP of .33. The only difference between the two conditions was the order of the trials (see Figure 13). All participants received both conditions in a counterbalanced order.

Each scenario initially proceeded like Experiments 1–3: Participants observed a lever being flipped between the left and right positions, and they observed the shape of blocks produced by the machine. While the 24th trial was still visible, a second lever was revealed. Participants were instructed that the revealed lever "may have influenced the shape of the block for the previous 24 trials."

In order to assess how participants thought that the two levers interacted to produce the shape of block, participants answered four counterfactual questions about the 24th trial.[9] Specifically, participants read, "Suppose on the 24th trial, the levers were set like this," and were shown pictures of the four combinations of the states of the two levers. For each of the four combinations, they were asked, "Do you think that the machine would produce a Square or Triangle?"

Results

The primary comparison of interest was whether participants' predictions resembled the biconditional integration function more in the grouped than ungrouped conditions. Figure 14 presents the means for the four prediction questions separated by the grouped and ungrouped conditions. As can be easily seen from Figure 14, the pattern of results from the grouped condition resembles the biconditional predictions of Figure 12, whereas the pattern from the ungrouped condition resembles predictions of a main effect of the unobserved factor.

[9] We asked about the 24th trial in order to hold the context from the data observations constant as much as possible.

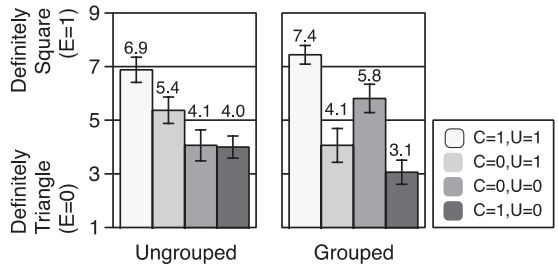

Figure 14. Mean ratings on prediction questions and standard errors in Experiment 5. Note: "U" represents the unobserved cause that is revealed on the 24th trial.

A 2 (Grouping; grouped vs. ungrouped) × 4 (Prediction Questions) repeated-measures ANOVA tested for differences between the conditions. There was a significant main effect of the four prediction questions, $F(3, 45) = 11.32$, $p < .01$, $\eta_p^2 = .43$, but no main effect of grouping, $F(1, 15) < .01$. Most important, there was a significant interaction between grouping and the prediction questions, $F(1.93, 28.88) = 5.12$, $p = .01$, $\eta_p^2 = .26$,[10] suggesting that participants believed the observed and unobserved causes combined in different ways for the grouped and ungrouped conditions.

Follow-up tests were preformed to look for differences between participants' predictions on particular questions in the grouped vs. ungrouped conditions. Compared with the other three combinations in Figure 12, the biconditional interaction function is the only one that predicts $E = 1$ when $C = U = 0$; all the other strategies predict that $E = 0$ when $C = U = 0$. Participants' predictions for this question were higher (i.e., closer to the biconditional prediction) in the grouped than ungrouped condition, $t(15) = 2.45$, $p = .03$. The prediction for when $C = 0$ and $U = 1$ also provides an informative contrast. For this question, a main effect of U is the only combination out of the four in Figure 12 that predicts $E = 1$. For this question, participants gave higher ratings (i.e., closer to the main effect of U prediction) for the ungrouped than grouped condition, $t(15) = 2.17$, $p = .05$.

Another analysis was conducted to examine whether the overall pattern of participants' predictions resembled the biconditional integration function more in the grouped than ungrouped conditions. This analysis can be formalized by computing sum square error (SSE), the sum of the squared differences between a participant's four predictions compared with the four predictions made by the biconditional combination in Figure 12. The formula below represents SSE for a particular participant,where X is a given score and β is the biconditional prediction on question i.

$$\mathrm{SSE} = \sum_{i=1}^{4} (X_i - \beta_i)^2.$$

The SSE was calculated for each participant separately for the grouped and ungrouped conditions. On average, SSE was lower in the grouped ($M = 41.38$, $SD = 36.33$) than ungrouped ($M = 71.56$, $SD = 26.57$) conditions, $t(15) = 3.59$, $p < .01$.

In sum, the overall pattern of participants' four predictions was closer to the biconditional in the grouped than ungrouped condition. This experiment provides further evidence that people can learn that observed and unobserved causes can combine in nontraditional ways; our participants simultaneously understood that sometimes the lever produced one outcome, and other times it produced the opposite, and they attributed this difference to the unobserved lever.

Experiment 6

The purpose of Experiment 6 was to rule out the possibility that participants had different memories for the events in the grouped and ungrouped conditions and that these different memories were responsible for the different inferences. It is possible that participants in the grouped condition could more easily chunk the trials of one contingency together, potentially resulting in a different memory of the trials. In Experiment 6, the trials were presented on index cards, and participants had access to all the trials while answering the questions predicting whether E would be present or absent, thus eliminating any memory demands. While Experiment 4, which presented all trials on a single webpage that could be scrolled up and down, already partially ruled out the memory difference as a potential alternative account, Experiment 6 attempts to bolster this finding by using the dependent measures used in Experiment 5.

Method

Participants. There were 29 participants from the same population as in Experiments 1, 2, 3, and 5.

Procedure and design. There were two conditions: grouped and ungrouped. Both conditions had the same 36 trials and a ΔP of zero. The only difference between the grouped and ungrouped conditions was the order of the trials (see Figure 15). A different set of trials was used from Experiment 5 to test the more extreme case of zero contingency. We also presented more trials to allow participants to gather more evidence, given that memory limitation would not be an issue in this experiment. All participants received both conditions in a counterbalanced order.

The instructions were slightly modified from Experiment 5 in two ways. First, instead of referring to trials, we called each trial a "day" of observing the machine, which was intended to make the packet of 36 index cards easier to understand. (The same modification was made in Experiment 4, which also presented all the trials simultaneously.) Participants were told that "from day to day, each lever may stay at the same position or may flip to the other position, and the shape of the block may stay the same or change." In addition, participants were told from the beginning of the scenario about the second lever. This instruction was added because of concern that perhaps participants in Experiment 6 only attributed the change in contingency between the observed cause and effect to the revealed lever after the fact. In the current experiment, participants were told, "Each machine has two levers, Lever A and Lever B, and produces two shapes of blocks. . . However, for the 36 days, you are only able to observe Lever A; Lever B is hidden." Thus, participants could view the unobserved lever as a potential cause throughout the experiment.

Then participants worked through a packet of 36 index cards comprising the 36 days. In order to reduce any memory demands, participants were told that they could read through the

[10] Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2 = 12.71$, $p = .03$; therefore, the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity, $\varepsilon = .62$.

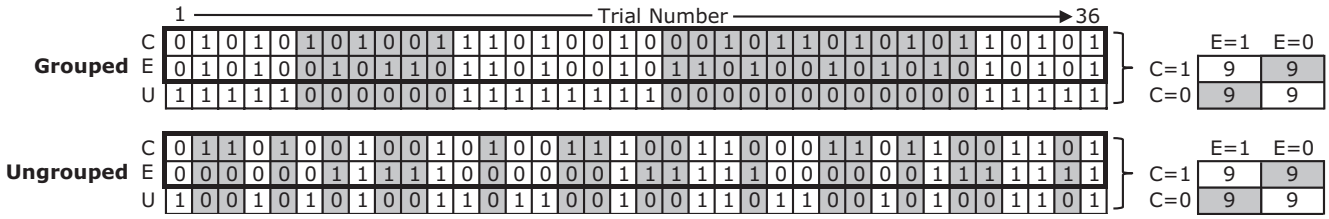

Figure 15. Summary of stimuli in Experiment 6. Note: "C" represents the cause (lever). "E" represents the effect (shape of block). "U" represents an unobserved, interacting, biconditional factor not shown to participants. Cells are white for the groups composed of (1, 1) and (0, 0) trials. Cells are gray for the groups composed of (0, 1) and (1, 0) trials.

packet of cards as many times as they want, can flip back and forth between the days and can look at the packet while answering the questions. The cards were 13.97 cm tall by 21.59 cm wide. Each card had the number of the day, pictures of both levers, and the shape produced. The picture of the observed lever was either left or right. The position of the unobserved lever was denoted with a question mark.

After working through the 36th index card, participants saw one more card revealing that on Day 36, Lever B was set to the right. Then participants proceeded on to the same set of questions as in Experiment 5 in which they predicted the shape produced from the four combinations of the two levers.

Results

A 2 (Grouping; grouped vs. ungrouped) × 4 (Prediction Questions) repeated-measures ANOVA tested for differences between the conditions (see Figure 16).[11] There was a significant main effect of the four prediction questions, $F(2.17, 60.98) = 56.15$, $p < .01$, $\eta_p^2 = .67$, but no main effect of grouping, $F(1, 28) = .11$. There was a significant interaction between grouping and the prediction questions, $F(1.77, 49.70) = 3.76$, $p = .04$, $\eta_p^2 = .26$, suggesting that participants believed the observed and unobserved causes combined in different ways for the grouped and ungrouped conditions.

The most critical test was whether the grouped and ungrouped conditions differed on the $C = U = 0$ question. Compared with the other four combinations in Figure 12, the biconditional interaction function is the only one that predicts $E = 1$ when $C = U = 0$; all the other strategies predict that $E = 0$ when $C = U = 0$. Participants' predictions for this question were higher (i.e., closer to the biconditional prediction) in the grouped than ungrouped condition, $t(28) = 2.52$, $p = .02$.

In addition, when $C = U = 1$, participants gave higher scores for the grouped than ungrouped condition, $t(28) = 2.15$, $p = .04$. This could reflect general uncertainty if participants believed that the effect was harder to predict in the ungrouped condition. (Note that on the last trial, which was still available for participants to view as they answered the questions, participants saw that $C = U = E = 1$. Thus, if they believe that C and U explain E, then they should give high responses for E, as they did in the grouped condition. However, if they felt that C and U do not explain E, and E is determined by other unknown causes, then they would be more ambivalent as they were in the ungrouped condition.) There were no significant differences for the other two questions (ts < 1.49, ps $> .15$).

We performed the same SSE analysis as described in Experiment 5. On average, SSE was significantly lower in the grouped ($M = 45.93$, $SD = 38.53$) than ungrouped ($M = 65.34$, $SD = 40.35$) conditions, $t(28) = 2.03$, $p = .05$. This suggests that the overall pattern was closer to the biconditional interaction in the grouped than ungrouped condition.

In sum, the results suggest that memory demands were unlikely to be a factor driving the different inferences in Experiment 5. Even when memory demands were minimized by allowing participants access to all the trials while answering the questions, participants still inferred a biconditional interaction more in the grouped than ungrouped condition. This experiment also replicated the basic effect of inferring a biconditional interaction using a different set of trials and a different presentation format.

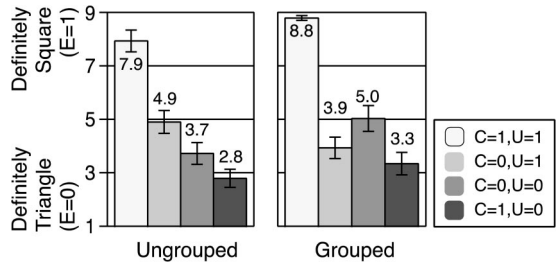

Figure 16. Mean ratings on prediction questions and standard errors in Experiment 6.

General Discussion

We began this article by asking whether people can learn nontraditional causal interactions between observed and unobserved factors. In particular, we investigated whether people are able to learn a "biconditional" interaction between an observed and unobserved cause. Because it is unlikely that people would use the biconditional interaction as a default assumption for how causes

[11] Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2 = 19.65$, $p < .01$ for the four predictions, and $\chi^2 = 32.89$, $p < .01$ for the interaction. Therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity, $\varepsilon = .72$ for the four predictions, and $\varepsilon = .59$ for the interaction.

combine, if people do learn about such interactions, it would suggest that they engage in sophisticated learning about how observed and unobserved causes interact. In contrast, many existing theories of causal learning suggest that people make simplifying assumptions that unobserved factors do not interact with observed causes and instead combine through a noisy-OR function (e.g., Cheng, 1997) or combine linearly (e.g., Rescorla & Wagner, 1972).

Nonetheless, these experiments demonstrated that people can learn about biconditional interactions and make specific inferences about unobserved interacting causes if the unobserved causes are stable. In grouped conditions, the trials supporting an association between one state of the cause and effect were grouped together, and those supporting the opposite association were grouped together in a trial-by-trial presentation. Such a pattern could arise in the real world from a relatively stable unobserved interacting cause that occasionally changes state. In ungrouped conditions, these two types of observations were intermixed. Such a pattern would arise from an unobserved interacting cause that *frequently and coincidentally* changes at the same instant as the observed cause, which would be less plausible. We predicted that participants would be more likely to infer a biconditional interaction in grouped than ungrouped conditions.

In Experiment 1, participants gave higher causal strength judgments for the observed cause in grouped than ungrouped conditions. In Experiment 2, participants judged the cause to be more efficacious to the extent that the data were more grouped. In fact, even when there was zero correlation between a cause and effect, participants judged the causal strength to be as high as 60 out of 100. Such responses are sensible, because when an observed cause interacts with an unobserved cause through a biconditional interaction, there may be a weak or zero correlation between the observed cause and effect, even if it truly is a cause (but see the alternative explanations below).

Experiment 3 more directly demonstrated that participants were more likely to believe that the observed cause *interacted with an unobserved factor* to produce an effect in a grouped than ungrouped condition. Experiment 4 further demonstrated that participants thought of the unobserved factor as a *cause*, not just an associative cue; they were more likely to believe that a machine with two causes could produce the grouped pattern of data than a machine with only one cause and another noncausal cue.

Participants in Experiment 5 inferred precisely how the observed and unobserved causes combine to produce the effect, and these inferences more frequently corresponded to the proposed biconditional integration function in the grouped than ungrouped condition. Experiment 6 ruled out the possibility that participants inferred a biconditional integration function merely because of different memories in the two conditions.

Challenges for Existing Models

Existing models fail to explain the current results. "Rule-based" models, based on measures such as ΔP, compute causal strengths by aggregating over all observations regardless of order (see Hattori & Oaksford, 2007, for a comprehensive review of 41 such models). That is, such models are simply not intended to explain any order effects, including the difference between the ungrouped and grouped conditions found in the current experiments.

There are "trial-by-trial" models that continually update their causal efficacy estimate after each trial. Yet they also fail to explain the current results in their current form (e.g. Luhmann & Ahn, 2007; Rescorla & Wagner, 1972). For example, if we run the Rescorla-Wagner learning algorithm (Rescorla & Wagner, 1972) without making any assumptions about unobserved causes, using only the ever-present background cue, then the model would never settle down with a strong associative strength between the observed cause and the outcome in the grouped condition. During some groups ([1, 1] and [0, 0] trials), the model would increase the causal efficacy estimate, and during the opposite groups ([1, 0] and [0, 1] trials), it would decrease the causal efficacy estimate. Consequently, if the data were grouped, the association between C and E would cycle between positive and negative values indefinitely.

Of course, many trial-by-trial, associative models with compound cues can learn biconditional interactions between two *observed* cues and an outcome (e.g., positive and negative patterning; Pearce, 2002; Pearce & Bouton, 2001; Wagner & Rescorla, 1972). Thus, one natural way to model the current experiments is to turn the unobserved cause into an observed cause, by inferring that it is present or absent exactly how U was specified in the stimuli figures. For instance, whenever ($C = 1$ and $E = 1$) or ($C = 0$ and $E = 0$), then $U = 1$, and whenever ($C = 0$ and $E = 0$) or ($C = 1$ and $E = 0$), then $U = 0$. In this case, such associative models would learn that there is an association between the observed cause and effect. That is, we are not suggesting that the Rescorla-Wagner model or other associative models can never learn biconditional interactions per se. Most important, however, what is needed is an account of how people infer the unobserved cause and under what conditions, which is what the current results are about. Such inferences can be added to these associative models but are not predicted a priori.[12]

To summarize, the current results provide new challenges to the existing models of causal induction. Stronger causal attributions in

[12] There are at least two other aspects of the current study that pose problems in directly applying many existing models. First, some models like Rescorla-Wagner (Rescorla & Wagner, 1972) treat cues differently if they are present versus absent. Specifically, Rescorla-Wagner only updates the association strength of a cue if it is present on a given trial. However, in the current experiments, the cause (lever) was either left or right, not present versus absent, so it is unclear how exactly to model these scenarios. Second, without positive and negative states, the relationship between the cause and effect cannot be described as "generative" or "inhibitory" or "positive" or "negative." Yet almost all models of causal learning infer causal strength on a scale from a negative relationship to a positive relationship. In the current experiments, however, the dependent measure was on a scale from the cause did not affect the effect at all to the cause strongly affected the effect. We asked this question because it is possible to agree that a cause affects an effect even if it is through an interaction with an unobserved cause. Taking the absolute value of the output from existing models also does not solve this problem. For "rule-based models" the final causal strength would be zero in the grouped conditions of Experiment 2, but people inferred high causal strengths. For trial-by trial models, the absolute value solution could produce a high final causal strength for some of the grouped stimuli used in the current study given some parameter values. However, for other sets of stimuli used here and other parameter values, the absolute value of the final associative strengths are still close to zero. In sum, there does not appear to be an obvious and principled way to circumvent these issues.

the grouped compared with the ungrouped condition add to a growing number of order effects that the rule-based models are incapable of explaining. The associative models may be able to model the current results by making just the right kinds of representational assumptions about unobserved factors, but they cannot predict or explain these assumptions about unobserved causes.

Unresolved Issues Regarding Inferences About the Unobserved Cause

We have proposed that learners make inferences about an unobserved cause, which interacts with the observed cause. This proposal appears to best explain why participants gave so much higher causal strength ratings in the grouped than ungrouped conditions in Experiments 1 and 2. The participants in the grouped conditions would have believed that an unobserved factor interacted with the observed factor to produce an effect. In Experiment 3, participants indeed were more likely to agree that a combination of the observed cause and some other factor influenced the effect in the grouped than ungrouped condition. In Experiments 5 and 6, when an unobserved cause was revealed, participants were more likely to infer a biconditional interaction in the grouped than ungrouped conditions. Yet there are a number of unresolved issues in terms of the details of when such inferences are made. We discuss two possibilities: (a) people retroactively infer an unobserved cause (e.g., when they were prompted about the possibility of an unobserved cause), and (b) people spontaneously infer unobserved cause online while they are making observations.

The first possibility is that learners reason about an unobserved cause retroactively, after observing all the learning trials and/or only when they were alerted about the possibility of an unobserved cause (as in Experiments 3, 5, and 6). Without such a prompt, people may not have spontaneously inferred an unobserved cause. Thus, according to this position, the results from Experiments 3, 5, and 6, which demonstrate that people infer a biconditional interaction with an unobserved cause, are experimental artifacts of forcing them to think about an unobserved cause. Furthermore, the higher causal strength ratings found in the grouped condition in Experiments 1 and 2 were not obtained because people inferred an unobserved cause but, rather, because they might have used an alternative reasoning process such as the win-stay lose-shift strategy explained in the introduction to Experiment 3. Namely, in the grouped conditions of Experiments 1 and 2, there were long periods of time in which participants were often able to predict the outcome (shape of the block) based on the cause (position of the lever), and it is this steady period, rather than spontaneous inferences about an unobserved cause, that led people to give stronger causal strength judgments.

Alternatively, people might have spontaneously reasoned about an unobserved biconditional cause while observing the learning trials. We believe that there are three points that are suggestive that participants in these studies may have reasoned about the unobserved cause online. First, previous research has found that people spontaneously reason about unobserved causes and their dynamic inferences about unobserved causes that changed during learning influence their inferences about observed causes (Hagmayer & Waldmann, 2007; Luhmann & Ahn, 2007, 2011; Rottman, Ahn, & Luhmann, 2011). Although that research involved scenarios designed to favor a noisy-OR interpretation of how the observed and unobserved causes interact, it seems at least possible that people would reason dynamically about biconditional unobserved causes as well. Second, intuitively, we believe that it is easier to reason about unobserved causes while learning. For example, if the lever is set to the left and the machine produced a square, and then the machine starts to produce a triangle, it seems easier to explain away this anomalous change in the effect by inferring a change in an unobserved cause. It would be highly cognitively demanding for a reasoner, upon being asked about an unobserved cause, to retroactively retrieve all the previous trials and make inferences about an unobserved cause from memory. Third, if people only reason about the unobserved cause after observing all the trials, it is unclear why they would infer a biconditional interacting cause. Instead, when there was a weak or zero contingency between the observed cause and effect, it would be logical to just infer an unobserved cause that entirely explains the effect and conclude that the observed cause has no influence on the effect.

Yet the current experiments fall short of providing a definitive answer to these two possibilities, because in the current experiments, we asked questions about the unobserved cause only after the end of the learning trials. Future research investigating whether and how people reason about unobserved biconditional causes dynamically may provide additional insights into the reasoning processes examined here.

Rationality of Inferring an Unobserved Interacting Cause

Is inferring an interaction with unobserved causes in grouped conditions rational? On one hand, inferring an unobserved cause may appear to be an irrational form of motivated reasoning. For example, in a grouped condition, a cause may initially appear to generate an effect and later appear to inhibit the effect. Concocting an unobserved factor that flips the relationship between the observed cause and effect could simply be a way to perpetuate the initial hypothesis (e.g., the cause generated the effect) in the face of contrary evidence.

On the other hand, inferring an unobserved, interacting factor in grouped conditions may be rational. In grouped conditions with zero contingency, if the observed cause is truly unrelated to the effect, why would it display such long periods during which it appeared to influence the effect? Grouped conditions likely appear too coincidental for people to conclude that the observed cause is unrelated to the effect. Furthermore, when the contingency reverses after some time, it would be rational to have some kind of explanation for the reversal. Inferring a stable unobserved interacting cause appears to be a parsimonious reason why the pattern of data had long groups of trials with the same contingency and why the pattern reverses for other long groups of trials.

Similar reasoning can explain why people do not infer an unobserved, interacting cause in the ungrouped conditions. In the ungrouped conditions, inferring an unobserved interacting biconditional factor would require the unobserved cause to frequently change *at the same time* that the observed cause changes, negating the effect of the observed cause. For example, in the double light switch example (see Figure 2, Ungrouped), from Step 1 to Step 2, both the observed and unobserved switches get flipped simultaneously, resulting in the light staying off. Inferring an unobserved interacting factor in an ungrouped condition requires many more

of these coincidences than in a grouped condition. For example, in the condition with shortest groups in Experiment 2 (see Figure 6), the observed lever was flipped 15 times, and inferring an unobserved interacting factor would require it to change simultaneously 14 out of these 15 times. In contrast, in the condition with the longest groups in Experiment 2 (see Figure 6), the observed lever was flipped 15 times, but an unobserved biconditional factor would only change simultaneously twice. Thus, inferring an unobserved interacting factor in an ungrouped condition may seem too coincidental and unlikely. Similar theories involving coincidences have been used to explain inference in other domains such as explanation (Hacking, 1983), vision (Barlow, 1985; Binford, 1981; Feldman, 1997; Knill & Richards, 1996; Witkin & Tenenbaum, 1983), and word learning (Xu & Tenenbuam, 2007; also see Griffiths & Tenenbaum, 2007).

Some researchers have recently proposed a theoretical account of renewal in animal learning similar to our account of inferring unobserved interacting causes (Gershman, Blei, & Niv, 2010; Redish, Jensen, Johnson, & Kurth-Nelson, 2007).[13] In renewal experiments, an animal initially experiences a contingency between a cue and outcome, then experiences a second phase during which there is no longer a contingency between the cue and outcome, and finally experiences a third phase with the initial contingency. Animals acquire an association in the first phase, then "extinguish" the association in the second phase and quickly reacquire the association in the third phase. Most theories of learning suggest that in the second phase, the animal unlearns the initial association, but they cannot explain why the initial association is often more quickly relearned in the third phase than the initial acquisition. The new theories propose that when an animal has high prediction error, such as at the beginning of the second phase, the animal may infer a new "state." Instead of unlearning the initial association, extinction involves learning about this new state and how it is different from the initial state. When the animal then experiences the third phase with the contingency between the cue and outcome, the animal infers that it is back in the initial state, and thus, the animal quickly renews the association. This theoretical account is similar to ours in that an unobserved state or cause must be inferred, and this unobserved factor moderates the contingency between the observed cue and outcome. However, it is unclear how these theories would handle Experiment 4, in which framing *U* as a cause versus effect influences the inferred interaction.

Reasoning With Causal Scenarios That Unfold Over Time

One important general point made by the current experiments is that people fluently reason about causal phenomena that occur over time. Many standard theories of causal learning were meant to handle scenarios with "independent" or "between-subjects" events (e.g., a person taking or not taking medicine and developing or not developing heart disease). Even in studies that present trials sequentially, each trial typically presents a separate case, and the trials are often randomized. In the current study, one machine with one lever is repeatedly tested across a period of time (a "within-machine" or "repeated-measures" scenario). Tracking transitions between trials allows for rich inferences. For example, if a lever is flipped, and there is simultaneously a change in the shape of block, this suggests that the lever produced the change in the shape. If a lever is flipped, but the machine continues to produce the same shape of block, this suggests that the lever does not influence the shape of the block. In formal statistics, we use different procedures for independent and dependent data. One intriguing possibility is that people also engage in different processes when reasoning about independent versus dependent data (e.g., Rottman, 2011; Rottman & Ahn, 2009b; Rottman & Keil, 2011).

An important direction for future research is to study how people reason with other causal phenomena that unfold over time. For example, it has been proposed that when predicting sequences of binary events over time (e.g., whether a basketball player will make or miss his next shot based on whether he made the prior shot), people use complex theories involving the nature of the underlying mechanisms involved (Oskarsson, Van Boven, McClelland, & Hastie, 2009). Exploring how people reason with temporal causal phenomena will help inform theories of causal reasoning more generally.

Conclusions

In everyday causal reasoning, unobserved and observed causes frequently interact in ways complicating causal inference. The current article demonstrated how people use the temporal sequence of events to learn about complicated interactions between observed and unobserved causes. However, existing models of causal learning cannot make such inferences because they are not appropriately sensitive to the sequence of trials. Future research is needed to explore how people reason about other kinds of interactions and how to incorporate this reasoning into a general model of human causal learning.

[13] We thank Dr. Allan Wagner for directing us to this literature.

References

Barlow, H. (1985). Cerebral cortex as a model builder. In D. Rose & V. G. Dobson (Eds.), *Models of the visual cortex* (pp. 37–46). Chichester, England: Wiley.

Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs, 75,* 43–88.

Baumrind, D. (1972). An exploratory study of socialization effects on Black children: Some Black–White comparisons. *Child Development, 43,* 261–267. doi:10.2307/1127891

Beckers, T., De Houwer, J., Pineño, O., & Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 238–249. doi:10.1037/0278-7393.31.2.238

Binford, T. O. (1981). Inferring surfaces from images. *Artificial Intelligence, 17,* 205–244. doi:10.1016/0004-3702(81)90025-4

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405. doi:10.1037/0033-295X.104.2.367

Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological Bulletin, 113,* 487–496. doi:10.1037/0033-2909.113.3.487

Dennis, M. J., & Ahn, W. (2001). Primacy in causal strength judgments. *Memory & Cognition, 29,* 152–164.

Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology, 41,* 145–170. doi:10.1006/jmps.1997.1154

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review, 117,* 197–209. doi:10.1037/a0017808

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 334–384. doi:10.1016/j.cogpsych.2005.05.004

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition, 103,* 180–226. doi:10.1016/j.cognition.2006.03.004

Hacking, I. (1983). *Representing and intervening.* Cambridge, England: Cambridge University Press.

Hagmayer, Y., & Waldmann, M. R. (2007). Inferences about unobserved causes in human contingency learning. *Quarterly Journal of Experimental Psychology, 60,* 330–355. doi:10.1080/17470210601002470

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56,* 51–65. doi:10.1037/h0062474

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science, 31,* 765–814. doi:10.1080/03640210701530755

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General & Applied, 79,* 1–17.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanounse, H. H. Kelley, R. E. Nissbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.

Kerr, A. W., Hall, H. K., & Kozub, S. A. (2002). *Doing statistics with SPSS.* London, England: Sage.

Knill, D. C., & Richards, W. A. (1996). *Perception as Bayesian inference.* Cambridge, England: Cambridge University Press.

Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. (2008). Sequential causal learning in humans and rats. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 185–190). Austin, TX: Cognitive Science Society.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science, 34,* 113–147.

Luhmann, C. C., & Ahn, W. K. (2007). BUCKLE: A model of unobserved cause learning. *Psychological Review, 114,* 657–677. doi:10.1037/0033-295X.114.3.657

Luhmann, C. C., & Ahn, W. K. (2011). Expectations and Interpretations during Causal Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 568–587.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111,* 455–485. doi:10.1037/0033-295X.111.2.455

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature, 364,* 56–58. doi:10.1038/364056a0

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin, 135,* 262–285. doi:10.1037/a0014821

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior, 30,* 73–95. doi:10.3758/BF03192911

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111–139. doi:10.1146/annurev.psych.52.1.111

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, England: Cambridge University Press.

Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review, 114,* 784–805. doi:10.1037/0033-295X.114.3.784

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

Rottman, B. M. (2011). *Causal learning over time* (Unpublished doctoral dissertation). Yale University, New Haven, CT.

Rottman, B. M., & Ahn, W. (2009a). Causal inference when observed and unobserved causes interact. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 1477–1482). Austin, TX: Cognitive Science Society.

Rottman, B. M., & Ahn, W. K. (2009b). Causal learning about tolerance and sensitization. *Psychonomic Bulletin & Review, 16,* 1043–1049. doi:10.3758/PBR.16.6.1043

Rottman, B. M., Ahn, W., & Luhmann, C. (2011). "When and how do people reason about unobserved causes? In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 150–183). Oxford, England: Oxford University Press.

Rottman, B. M., & Keil, F. C. (2011). Learning causal direction from repeated observations over time. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33th annual conference of the Cognitive Science Society* (pp. 1847–1852). Austin, TX: Cognitive Science Society.

Shanks, D. R. (1989). Selectional processes in causality judgments. *Memory & Cognition, 17,* 27–34. doi:10.3758/BF03199554

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory & Cognition, 29,* 193–208. doi:10.3758/BF03194913

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. *Inhibition and Learning,* 301–336.

Waldmann, M. R. (1996). Knowledge-based causal induction. *Psychology of Learning and Motivation, 34,* 47–88. doi:10.1016/S0079-7421(08)60558-7

Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science, 31,* 233–256.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124,* 181–206. doi:10.1037/0096-3445.124.2.181

Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481–543). New York, NY: Academic Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* New York, NY: Oxford University Press.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114,* 245–272. doi:10.1037/0033-295X.114.2.245

Yuille, A., & Lu, H. (2008). The noisy-logical distribution and its application to causal inference. *Advances in Neural Information Processing Systems, 20,* 1673–1680.

Received February 27, 2010
Revision received April 19, 2011
Accepted April 25, 2011 ■