Sensitivity to Confounding in Causal Inference: From Childhood to Adulthood

E. Christina Ford (christis@ucla.edu)
Department of Psychology, Box 951563
Los Angeles, CA 90095-1563 USA

Patricia W. Cheng (cheng@psych.ucla.edu)
Department of Psychology, Box 951563
Los Angeles, CA 90095-1563 USA

Abstract

A necessary condition for correctly assessing causality is the absence of confounding causes. This paper reports a pair of experiments that investigate whether people are sensitive to confounding when they infer causation. Two stories were constructed, one in which two candidate causes perfectly covaried with each other (*confounded*), and another in which the two candidate causes occurred independently of each other (*unconfounded*). In the confounded story, both causes covaried perfectly with an outcome; in the unconfounded story, only one of the two candidates covaried with the outcome. If people control for alternative causes while they evaluate a candidate cause, then subjects in the confounded condition should indicate that it is impossible to determine causality for either candidate alone, whereas those in the unconfounded condition should be able to judge that one of the candidates is causal and the other not. If people are not sensitive to confounding, however, subjects in the confounded condition should attribute causality to both candidates, and their judgments for these candidates should be the same as those for the target causal candidate in the unconfounded condition. Two experiments were conducted respectively with children and adults: Children received one or the other story, while adults received both. Both children and adults distinguished between confounded and unconfounded candidate causes when making attributions of causality. Our results show that children are able to state the indeterminacy of confounded candidate causes at an age much earlier than previously documented.

Introduction

One view of how children learn is that they approach the world as scientists and form theories about the world using information about variation and covariation to establish causal connections (e.g. Gelman, 1996; Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004; Gopnik, Sobel, & Schulz, 2001). Further, they intervene upon the world in order to discover these relationships (Schulz, 2003). Although children may have misconceptions in their explanations, as when a child states that he thinks God made the sun out of gold and lit it with fire (Siegler, 1998), the presence of such misconceptions does not mean that children are unable to use the data present in the environment to form correct causal attributions. Given that adults have had many more experiences than children, we should not expect children's theories to be the same as adult's theories, especially for complex phenomena. What is important is whether the same process is utilized when determining causality. In particular, this paper seeks to examine whether both children and adults are sensitive to confounding when there are two candidate causes for a novel outcome.

In addition to the potential implications for improving science instruction, assessing children's sensitivity to confounding is also important for differentiating between two types of models of causal learning. The first type is instantiated in the *unconditional* ΔP model (Jenkins & Ward, 1965); the second type consists of models that tease apart the influence of a candidate cause from the influences of alternative causes (e.g., Cheng, 1997; Cheng & Novick, 1992; Glymour, 2001; Gopnik et al., 2004; Novick & Cheng, 2004; Pearl, 2000; Spirtes, Glymour, & Scheines, 2000; Tenenbaum & Griffiths, 2001).

Under the unconditional ΔP model, people contrast the frequency of e, an effect of interest, when c, a potential cause, is present, with the frequency of e when c is absent:

$$\Delta P = P(e|c) - P(e|{\sim}c)$$

If ΔP is equal or close to 0, then c is considered noncausal; if it is noticeably greater than 0, then c is thought to cause e, and if it is noticeably less than 0, c is thought to prevent e. The unconditional ΔP model implies that people ignore confounding and pool over all the information known about the candidate cause. Thus, if two candidate causes perfectly covary with each other and the effect, then both candidates will be judged as causal.

Under the alternative approach, a definite causal judgment can result from the above contrast only when alternative causes are controlled (i.e., they occur independently of the candidate cause). One simple variant of this approach is the *conditional* ΔP model: the same ΔP formula is applied to a focal set of events in which alternative causes occur independently of the candidate (Cheng & Novick, 1992). For example, if there is a situation in which there are two possible causes of an event, one way in which a person could determine the causality of the individual candidates would be to compare the frequency of e in the situation in which only one candidate is present to a situation in which no candidate is present, holding the other candidate constantly absent. If people utilize conditional ΔP, they would be unable to draw a definite causal conclusion when there is confounding because no focal set of events could be formed.

Although unconditional ΔP has fallen into disfavor in the adult causal learning literature, previous studies of children suggest that they do not withhold judgments of causality in the presence of confounded variables (e.g., Kuhn, Amsel, & O'Loughlin, 1988). In other words, they seem to behave as predicted by the unconditional ΔP model. Our study focusses on young children. If the ability to reason causally is an unlearned fundamental human process, then it should be present at an age much earlier than indicated by prior research.

One study looking at third, sixth and ninth graders, as well as non-college young adults and undergraduate college students found that before the ninth grade, students were unlikely to state that there was insufficient evidence to determine causality when there is confounding (Kuhn, Amsel, & O'Loughlin, 1988). In fact, not a single subject suggested indeterminacy as the correct answer until the 9th grade. In one condition, 20 subjects in each age group was asked to determine the whether a feature of a ball (namely, texture) caused a ball to be bouncier in the presence of a perfectly confounded covariate (namely, color). None of the 3rd or 6th graders, one 9th grader, 2 non-college adults, and 5 college subjects proposed indeterminacy as the correct answer. But, these experiments involved causes for which the students were likely to have prior theories, and people interpret ambiguous data in ways that are consistent with their prior beliefs (Darley & Gross, 1983). Kuhn et al. (1988) do not indicate whether students who did not notice the indeterminacy were answering in a manner consistent with their prior theory. Also, because their studies focus on the coordination of theory and evidence, one of criteria used for assessing students' answers was their ability to justify their responses. But, if causal learning is an unconscious process, students might be sensitive to confounding, yet unable to justify their responses. In the present study, the task is made simpler, by presenting subjects with a novel effect, thereby reducing the relevance of prior causal beliefs, and by measuring subjects' causal attribution without asking for a justification.

Data from two pilot experiments are presented. Both experiments test whether people differentiate between confounded and unconfounded candidate causes. In one experiment, the subjects were undergraduates, while in the other the subjects were pre-school age children. In both experiments, participants were presented with two possible causes for a novel event, and were asked to determine the cause of that event. In one condition the two possible causes were independently occurring, while in the other condition the two candidate causes always occurred together. If people are sensitive to confounding they should be able to make a causal attribution in the first condition but not the second.

Methods

These experiments were designed to test whether people are sensitive to the independent occurrence of potential causes of an effect when making judgments of causality. The first experiment was conducted on adults. Even if adults are able to succeed in this task, however, their success might well be due to prior training. The second experiment was therefore conducted on children. Similar materials were used for both experiments.

The second experiment was actually conducted first. Because the data with the children was not very clean, in order to develop a better protocol the materials were piloted with adults. The small sample sizes in both experiments are due to the preliminary nature of the data. Also, the adult pilot was ended when the adult answers became consistent. Both experiments will be re-run with larger sample sizes using the final adult version of the stimuli. Below we describe the methods for both experiments before reporting the results.

Experiment 1

Participants 10 undergraduates at the University of California, Los Angles enrolled in an Introduction to Psychology Course participated in the study. Students received class credit for participating in the study and were recruited using an on-line bulletin board for this course.

Design This experiment had two conditions and utilized a within-subjects design. In one condition, the two possible causes of an unusual event were perfectly correlated (confounded). In the other condition, the same two possible causes occurred independently of one another (unconfounded). Subjects were asked about the causality of the candidate causes in turn. The ordering of the stories, as well as the order in which the subject was asked about each candidate cause, was counterbalanced across subjects.

Materials Two passages of approximately the same length were constructed (one story was 668 words and the other was 681 words). Both passages tell the story of bunny rabbits that went to two different parties.

In both stories, the parties occur at the same time and on the same day. On the day of the party, the bunnies are randomly assigned to a party via a coin toss. Half of the bunnies ate candy before going to the party. At one of the parties the bunnies ate cake, while at the other party they did not. In the confounded condition all the bunnies who ate candy also ate cake, whereas in the unconfounded condition half of the bunnies who ate candy also ate cake, and vice versa. All the bunnies at the cake party grew new pink wings; none of the bunnies at the "no cake" party did. To avoid confusion between the two stories, in one story, the bunnies ate green grass candy and yellow cheesecake; in the other story the bunnies ate blue berry candy and orangey orange cake.

At the end of the story, participants were asked about the causality of each of the causal candidates in the story:

1) Does Yellow Cheese Cake/ Blue Berry Cake all by itself make bunnies grow new pink wings? Yes, No, or Impossible to tell?
2) Does Green Grass Candy/ Orangey Orange Candy all by itself make bunnies grow new pink wings? Yes, No, or Impossible to tell?

The text of the story was accompanied by illustrations. An appropriately colored wedge in the bunnies' stomachs represented the cake, and a candy shaped object in the bunnies' stomach represented the candy.

Because we were attempting to revise the stimuli in order to make the directions clearer for the children, the stimuli underwent slight modification across the 10 subjects. The conditions remained the same, but there were slight changes in wording and pictorial presentation across groups. The most significant wording change was that the story narrative was condensed, leaving only a concise explanation of the meaning of the symbols that represented the outcomes as well as the candidate causes. The most significant pictorial modifications occurred in the confounded condition. In the original stimuli, the bunnies were in two groups (the cake group and the no cake group), with the bunnies who ate candy evenly distributed throughout each group. In the final stimuli, the bunnies were arranged into four groups of bunnies that underwent each treatment (i.e., one group had cake and candy, one group had only candy, one group had only cake, and one group had neither cake nor candy). Both of these changes served to make the experiment easier for the subjects to understand and interpret.

The stories were shown as a power point presentation. The power point presentation was presented on a 15" computer screen.

Procedure Participants were randomly assigned to conditions that differed on the ordering of the stories and assessment questions. Participants were then told that they were going to hear a story about bunny rabbits in two little bunny towns. They were told that something interesting was going to happen to these bunny rabbits, and that it was their job to try to figure out what happened.

Participants looked at the illustrations on the screen as the experimenter read the story aloud. At the end of the story, participants were asked about the causality of each of the causal candidates in the story. The experimenter wrote down their answers on an answer sheet as they progressed through the story.

Experiment 2

Participants Sixteen pre-school children from the Bellagio daycare center at the University of California, Los Angeles participated in the study. Nine male and seven female children between the ages of 4;5 and 5;7, with a mean age of 4;11 participated in the study. One child was excluded from the analysis for answering incorrectly factual questions about the stories presented. The rest of the children answered all of these questions correctly (as explained later).

Design This experiment had the same two conditions as Experiment 1 but utilized a between-subjects design. The order in which children were asked about each candidate cause was counterbalanced across conditions.

Materials The stories presented to the children had the same content as the stories presented to the adults, with three differences. First, the children's protocols did not undergo significant changes. The initial adult protocols (with the distributed confounding variable and long narrative) are the same as the child protocol. Second, in both conditions, children saw green grass candy and yellow cheesecake. (This was possible because subjects only saw one story, which ruled out the possibility of carryover between stories.) The children's assessment procedure also differed from that of the adults.

Children were first asked for their spontaneous attribution. "Do you think that it is possible to figure out why the bunnies grew new pink wings?" If the child answered yes then the following questions were asked:

1) Why do you think these bunnies [pointing to those who went to the cake party] grew new pink wings?
2) Why do you think these bunnies [pointing to those who went to the no cake party] did not grow new pink wings? The ordering of these two questions was counterbalanced across conditions.

Because children sometimes did not give a free response, did not address both of the causal candidates, or did not address the causal candidates in their responses (e.g., "Bunnies grew wings because they wanted to"), additional probes were added, asking about each of the candidate causes separately. Children were told about statements that other children had made while reading this story. Children were asked whether they thought these statements were "definitely right, definitely wrong, or impossible to tell." The statements they were asked to judge were

1) GREEN GRASS candy all by itself makes bunnies grow pink wings.
2) YELLOW CHEESE CAKE all by itself makes bunnies grow pink wings.
3) YELLOW CHEESE CAKE and GREEN GRASS candy together make bunnies grow pink wings.

If the child had previously indicated that the yellow cheesecake was causal, they were not asked about the yellow cheesecake again (and the same for the other candidates).

Procedure Children were randomly assigned to the conditions. They were video taped during the session. In order to accustom children to the camera, they were first introduced to the camera and allowed to see themselves on the LCD screen. As with the adults, children were told that they would hear a story about bunny rabbits, and that it was their job to figure out what happened.

The children looked at the illustrations on the screen as the experimenter read the story aloud. At the end of the story, the children were asked four factual questions to assess whether they understood and remembered the content of the story. The experimenter pointed to a picture of the bunnies with the candy in their tummies and asked, “What did these bunnies eat?”, the correct answer being “candy” (or cake and candy in the confounded condition). The experimenter then pointed to the bunnies without candy in their tummies and asked, “Did these bunnies eat candy?”, the correct answer being “no”. The experimenter then pointed to a picture of the bunnies at the cake party and asked, “What did these bunnies eat at the party?” (this question was omitted in the confounded condition if children answered cake and candy to the first question above), the correct answer being “cake”. The experimenter then pointed to a picture of the bunnies at the no-cake party and asked, “Did these bunnies eat cake?” the correct answer being “no”. Children who did not correctly answer all factual questions were excluded from the study.

Results

Experiment 1

Adult subjects were sensitive to confounding when they made causal judgments. The data are presented in Tables 1 and 2 below, with the number who answered correctly in bold font for each condition.

Table 1: Cake attributions for adults

	Causal Attribution		
Condition	Yes	No	Can't Tell
Confounded	0	0	**10**
Unconfounded	**8**	1	1

Table 2: Candy attributions for adults

	Causal Attribution		
	Yes	No	Can't Tell
Confounded	0	0	**10**
Unconfounded	0	**6**	4

Using McNemar’s test for 2-related samples of categorical data, we see that the pattern of responses differed across conditions for both of the causal candidates. Subjects were more likely to say the cake was causal in the unconfounded condition than in the confounded condition, and conversely, more likely to say it was impossible to assess causality in the confounded condition than in the unconfounded condition ($p < 0.05$, exact statistic, binomial distribution used). Likewise, subjects were more likely to say that the candy was not causal in the unconfounded condition than in the confounded condition, and more likely to say it was impossible to tell in the confounded condition than in the unconfounded condition ($p < 0.05$, exact statistic, binomial distribution used). In fact, as can be seen in Table 1, for both cake and candy in the confounded condition, all subjects said that it was impossible to tell if either candidate was causal. In contrast, in the unconfounded condition, most subjects said that the cake was causal, and none said that the candy was causal.

The order of presentation of the conditions had no effect: Subjects were just as likely to give a correct response whether they received the confounded or unconfounded condition first (n.s.). We were unable to test the effect of the ordering of the questions on the two candidate causes because some information on the ordering was lost. There was no effect of candidate in either condition: subjects were just as likely to make the correct causal judgment for the candy as they were for the cake (n.s.).

Experiment 2

Children were also sensitive to confounding when they made causal judgments, but this data show more variability than the adult data.

Children were first categorized into one of five causal attribution categories: the cake is causal, the candy is causal, both causal (jointly or independently), it is impossible to tell, and other causal attribution (Table 3). Because the focus of this paper is on children's ability to determine whether there is sufficient evidence to attribute causality to an individual cause in the case of confounding, the answers were collapsed into two groups, either assigning causality to individual candidates or not assigning causality to individual candidates (Table 4). If children made a spontaneous causal attribution, this was taken as the value for the measure. If a child did not give a spontaneous response or gave an ambiguous answer, the value for this measure was taken from the child’s answers to follow-up questions about each individual candidate.

Table 3: Children’s causal responses

	Condition	
Child responses	Confounded	Unconfounded
Cake	0	**4**
Candy	0	0
Both	3	2
Can’t Tell	**4**	1
Other	0	1

Table 4: Children’s attributions to individual candidates

	Condition	
Causal Category	Confounded	Unconfounded
Individual Attribution	2	**7**
No Individual Attribution	**5**	1

In the confounded condition, 3 children said both candidates were causal and 4 said it was impossible to tell.

Children in this condition gave no other responses. All of these statements were spontaneous attributions, and their answers in the follow-up questioning were consistent with their spontaneous answers. If children in the "Both" category had meant that the candy and the cake together made the pink wings, but that it was not possible to judge either candidate by itself, this would have been a correct response. Given these children's pattern of responding, for two of the children it is not clear whether they meant the two were independently or conjunctively causal. To score conservatively, against sensitivity to confounding, we counted these as attributions to the individual candidates. The third child indicated that she thought the conjunctive answer was the best answer despite the fact that she answered yes to the probes asking about the causality of each candidate individually. This child was recoded as "no individual attribution" in Table 4.

In the unconfounded condition, 4 children said cake was causal, 2 children said they both were causal, one child gave an alternate attribution, and another said that it was impossible to establish causality. Five of these statements were spontaneous attributions, only one of which was consistent with follow-up questioning. Two children who were coded as making a causal cake-attribution initially began by giving unclear spontaneous responses.

Children were more consistent in their responses across question formats in the confounded condition than in the unconfounded condition (Fisher's exact test, $p < 0.05$).

Even with our small samples, given the scoring as we just explained, individual attribution (see Table 3) was more likely in the unconfounded condition than in the confounded condition (Fisher's exact test, $p < 0.05$). Moreover, the correct response was the modal response in each condition (Table 4). Even so, given our small sample size, in both the confounded and unconfounded conditions children's ability to make a correct causal attribution was not significantly different from chance (Fisher's exact test, n.s.).

Discussion

Like Kuhn et al. (1988), it is likely that this study underestimates children's causal reasoning abilities. The correctness of the children's causal attributions, when they occurred, where not impressive in this experiment. These results may seem at odds with recent literature suggesting that children are often able to make correct causal attributions at an early age (e.g., Gopnik et al., 2001). It is likely that the poor performance in Experiment 2 is due to an overly confusing experimental protocol, particularly the visual presentation of the stimuli. Because the confounding cause was distributed equally across both groups, it was difficult to visually isolate the correct focal set.

Because the stimuli used were different between the child and adult studies, it is impossible to tell what, if anything, developed from childhood to adulthood. Our results do not speak to whether adults are better at causal reasoning than children. But, it is clear that both adults and children can differentiate situations in which it is possible to make a causal determination from those in which it is impossible to make a causal determination.

One puzzling result is that the adults responded unanimously to the confounded condition but less clearly to the unconfounded condition, and seemed to be more confused about the candy. This difference between conditions, however, is likely due to confusion in the protocol. After the assessment, when subjects were asked to explain their answers, the two subjects in the unconfounded condition who did not answer that the cake was causal cited reasons of experimental control. For example, one subject answered, "The bunnies at one party could have drank something that the bunnies at the other party did not." When it was made clear that the bunnies were exactly the same, except for whether they had eaten cake or candy, both subjects gave the correct response. Similarly, the subjects who said "can't tell" for the candy in the unconfounded condition occurred in the earlier portion of the experiment, before the materials were finalized.

One might think that the unconfounded condition could use a simpler, less confusing, design. But, in order to compare the unconditional ΔP model with the conditional ΔP model regarding confounding, the unconfounded condition must present information on all four possible combinations of the two candidate causes: both candidate causes are present, neither is present, and only one or the other of the candidate causes is present. To form the appropriate focal set for each candidate cause, the subject could compare the outcome in a single candidate "cell" to the outcome in the no-candidate-cause cell. Alternatively, the subject could compare the outcome in the combined candidate cell to outcomes in each of the single candidate cause cells.

Let us consider two potential two-cell designs. In the simpler design, there is a single candidate cause. In one cell both the candidate cause and the effect are present, in the other cell neither the candidate cause nor the effect is present. In this case, if children attribute causality to the candidate cause, it could be due purely to association. Furthermore, if fewer children attribute causality in the confounded condition, it could be argued that that condition is simply more difficult because it involves two candidate causes rather than one.

In a more complex two-cell design, there are two candidate causes: In one cell candidate cause A is paired with the effect, in the other cell candidate cause B occurs without the effect. This situation is, in fact, confounded. There is a perfect (inverse) correlation between the occurrence of candidates A and B. The only appropriate focal set would compare the occurrence of the effect before and after the introduction of the candidate causes, essentially adding a no-candidate-cause cell. Even if temporal information is used, this type of design does not rule out the possibility of a third unobserved candidate systematically affecting both candidates as well as the effect. There would be a number of correct responses depending on the extraneous assumptions made by the subjects regarding the influence of the unobserved candidate.

Conclusion

Both children and adults distinguish between confounded and unconfounded candidate causes when making causal inferences. All adults said that it was impossible to tell whether the cake or the candy alone caused the wings in the confounded condition. Because the subjects in the adult experiment were UCLA undergraduates and not less well-educated adults, and because the conditions used in this experiment were consistent with examples used in scientific methodology classes, it is possible that the adults have had prior training that allowed them to say that causal attribution was not possible when the causes were confounded. The same was not true of the children.

The child data suggests that children are able to state, at a much earlier age than documented by previous studies, that it is impossible to infer causality when two candidate causes perfectly covary. Children as young as 5 years old, less than half as old as previously believed, made such judgments. Moreover, these judgments were consistent across response formats.

Despite the variability in the data, and despite the fact that the children were presented with a version of the task that was more difficult than necessary, more children indicated that it was impossible make a causal judgment in the confounded condition than in the unconfounded condition. As predicted by the conditional ΔP model, children (and adults) are sensitive to confounding.

Acknowledgments

The preparation of this paper was supported by an NSF fellowship to the first author and an NIH Grant MH64810 to the second author.

References

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

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99,* 365-382.

Darley, J. M. & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44,* 20-33.

Gelman, S. A. (1996) Concepts and Theories. In T.K.F Au & R. Gelman (Eds), *Perceptual Development* (pp. 117-150). San Diego, CA: Academic Press.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical models in psychology*. Cambridge: MIT Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111 (1),* 3-32.

Gopnik, A., Sobel, D. M., Schulz, L. E. (2001) Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620-629.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*(1, Whole No. 594).

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988) *The development of scientific thinking skills*. New York, NY: Academic Press.

Novick, L.R., & Cheng, P.W. (2004). Assessing interactive causal influence. *Psychological Review*, *111*, 455-485.

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

Schulz, L. E. (2003, April) The play's the thing: Intervention and causal inference. In L. E. Schulz (Chair) *Understanding children's causal knowledge: Exploring the origins of causal inference.* Symposium conducted at the meeting of the Society for Research in Cognitive Development, Tampa, FL.

Siegler, R. S. (1998) C*hildren's Thinking* (3rd ed.). Upper Saddle River, NJ : Prentice Hall.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction and Search* (2nd edition). Boston, MA: MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in Neural Information Processing Systems 13* (pp. 59-65). Cambridge, MA: MIT Press.