REPORT

Children’s causal reasoning: counterfactual thinking occurs for ‘negative’ outcomes only

Tim P. German

University of Essex, UK

Abstract

Harris, German and Mills (Children’s use of counterfactual thinking in causal reasoning. Cognition, 61 (1996), 223–259) following Mackie, argue that children make explicit use of counterfactual thinking in arriving at causal judgments. They showed that children as young as 3, in explaining simple mishap events, made reference to courses of action that a protagonist had rejected, when that course of action would have prevented the observed outcome. It is hypothesized here that such counterfactual thinking might have been invoked by the ‘negative’ mishaps rather than as part of the causal reasoning process. Although the generation of counterfactuals in explanation was replicated using mishap outcomes such as those used by Harris et al., counterfactual thinking was not evident in children’s explanations of ‘positive’ outcomes. These results undermine the view that a counterfactual thinking process, as indexed by reference to possible actions rejected by a protagonist, is necessary for causal reasoning. Alternative characterizations of the relationship between causals and counterfactuals are discussed.

The nature of the relationship between causal reasoning and counterfactual thinking has been of interest not only to philosophers and psychologists (e.g. Mackie, 1974; Jackson, 1977; Roese & Olson, 1995a; Harris, German & Mills, 1996; Mandel & Lehman, 1996) but also to areas of enquiry as diverse as history, politics and the law (e.g. Hart & Honoré, 1959/1985; Fearon, 1991).

In a recent paper exploring children’s causal reasoning, Harris et al. (1996) advanced evidence in support of Mackie’s (1974) contention that causal reasoning depends on counterfactual thinking. Mackie proposed that in reaching the conclusion that event \( x \) caused event \( y \) we first consider a counterfactual situation where \( x \) did not occur and imagine what outcome would follow. If our simulation reveals that \( y \) would also not have occurred, then a causal relationship between the two events is inferred. For example, a lighted match goes out after a window is opened and we consider whether the match would have continued to burn had the window not been opened; if our thought experiment reveals that the match would have continued to burn, we are likely to conclude that opening the window caused the match to be extinguished (see Harris et al., 1996, p. 234).

Harris et al. (1996, Experiment 3) presented preschool children with stories modeled on stories presented to adult subjects by Wells and Gavanski (1989), in which a protagonist makes a choice between two courses of action. In experimental stories the choice was important in determining the outcome which followed; in control stories it was irrelevant to the outcome. For example, one story involved a woman dying as a result of an allergic reaction to a dish ordered by her boss at a celebratory meal. In the experimental story the boss wavered between two dishes of which one, unknown to him, would provoke the reaction. In the control story either dish would precipitate the reaction. Subjects rated the choice as more causally important in the experimental condition. Wells and Gavanski interpreted this result in terms of subjects bringing to mind counterfactual situations in reaching their causal conclusions.

Harris et al. (1996) assessed such counterfactual thinking in children by requiring them to explain similar simplified stories. For example, an experimental story involved Sally choosing some white chocolate for a snack instead of a sandwich and ending up hungry, while the control story involved an irrelevant choice – Sally chose white chocolate over brown chocolate and, again, ended...
up hungry. Children were probed with explanation questions (e.g. ‘why did Sally get all hungry?’) and then prevention questions (e.g. ‘what should Sally have done instead so she wasn’t hungry?’) for each of six stories (three experimental, three control). Following Wells and Gavanski (1989), Harris et al. looked at whether children’s explanations made reference to the point of choice. In particular, they reasoned that if children made reference, in their explanations, to the rejected course of action, then this would show that they considered events that might have happened, but that did not in fact happen (i.e. counterfactual events) to be causally relevant. For the purposes of the study to be reported here, and following Harris et al., reference to actions that were not chosen is assumed to count as counterfactual thinking. Harris et al. found that children’s explanations in experimental stories were indeed more likely to make reference to the rejected course of action (e.g., ‘she was hungry because she didn’t pick the sandwich’). For the control stories, children were more likely to generate alternative counterfactual actions that would have prevented the outcome (e.g., ‘she should have had more food’). This difference was more pronounced for the answers to the prevention question. Harris et al. interpreted these data as evidence that counterfactual thinking plays a key role in causal reasoning.

An issue that arises for Harris et al.’s conclusion that children engage in counterfactual thinking in order to reach a causal conclusion is how well the evidence for references to counterfactual states of affairs generalizes to different events. In particular, Harris et al. made exclusive use of scenarios where the outcome was a minor mishap. In an overview of the counterfactual literature, Roese (1997) defends the thesis that counterfactual thinking, though producing consequences that might be regarded as both beneficial and aversive to an individual, has a net beneficial effect for the individual that is functional. Part of Roese’s claim is based on evidence that counterfactual thinking is elicited more by negative events than by positive or neutral events. That is, when events go wrong, thoughts of how things might have been better are provoked (e.g. Johnson, 1986; Boninger, Gleicher & Stratham, 1994). Moreover, these so-called upward counterfactuals, where comparison results in an outcome better than that observed in actuality, are generated spontaneously more frequently than downward counterfactuals, which are generated only very rarely (Roese & Olson, 1995b).

In view of this evidence, an alternative explanation is possible for the counterfactual thinking observed by Harris et al. (1996). Rather than counterfactual thinking playing a key role in causal reasoning, this explanation pattern may in fact reflect spontaneous upward counterfactual thinking provoked by negative events; children’s explanations might be contaminated by issues of prevention rather than cause. Recall that Harris et al. required children to answer explicit questions about prevention, in alternation with their explanation questions. Perhaps interleaving explanation with prevention questions in this way, across the course of six stories, amplified the tendency to invoke counterfactual thinking in explanation. Analysis of children’s answers to the explanation question on the first story, which ought to have been uncontaminated by later prevention questions, showed that only six children (just under 20%) made reference to the rejected option (Harris et al., 1996, p. 250). This result is consistent with the view that thoughts of prevention, rather than causal reasoning, invoke counterfactual thinking. Relatedly, Mandel and Lehman (1996) required adult subjects to judge what caused an outcome, or to judge how an outcome might have been prevented or to imagine counterfactual events (i.e. completions of ‘if only...’ prompts). They showed ascriptions of cause typically made reference to different events than did either judgments of preventability or counterfactual judgments, leading them to suggest that causal ascriptions are based on different criteria than are counterfactual and preventability judgments (see also N’gbalà & Branscombe, 1995).

The prevention hypothesis is tested in the experiment to be presented here; if mention of rejected options in explanation is evidence that counterfactual judgment forms a critical part of causal reasoning, then evidence of such thinking should emerge when children explain all kinds of events, not just mishaps. However, if the mention of rejected options in explanation reflects counterfactual thinking provoked by negative outcomes, we should find no evidence of such counterfactual thought when the events to be explained are not mishaps.

**Experiment**

In this experiment, children were asked to explain the outcomes of stories modeled on those used in Harris et al. (1996), some of which involved negative outcomes, others of which involved positive outcomes. For example, in one story Sally is offered a choice between chocolate and a sandwich for a snack. She either picks chocolate and ends up hungry (negative outcome) or picks the sandwich and ends up full (positive outcome).

**Method**

**Subjects**

Forty 5-year-old children participated, divided into a ‘decisive choice’ condition (12 boys, 8 girls, mean age 5–

© Blackwell Publishers Ltd. 1999
7) and an ‘irrelevant choice’ condition (12 boys, 8 girls, mean age 5–7). The subjects were drawn from schools in Colchester, Essex, UK, which served a wide variety of socioeconomic backgrounds. All participants had English as their first language.

Stories and design

In the decisive choice condition children received stories where had the protagonist chosen the alternative option the outcome would not have occurred. In the irrelevant choice condition the outcome would have occurred whichever option the protagonist had chosen. The outcome to be explained was either a minor mishap (termed here negative outcomes) or the opposite (positive outcomes). For example the decisive choice story with a negative outcome was as follows:

Sally is going on a trip. Her mother asks her if she wants to take the white chocolate or the cheese sandwich. Sally chooses the white chocolate. At the end of her trip, she is all hungry.

In the positive outcome story, Sally chooses the sandwich and ends up all full. For the irrelevant choice stories the choices were between white chocolate and brown chocolate (negative outcome) or between a cheese sandwich and a ham sandwich (positive outcome). The details of the other three stories appear in tables in the Appendix.

The design was a $2^2$ mixed factorial design. Story type was manipulated between subjects, outcome type was manipulated within subjects; children in each condition received two negative outcome and two positive outcome stories. Assignment to either the decisive choice or irrelevant choice condition was pseudo-random, under the constraint that age and sex were as far as possible balanced across conditions. Stories were presented in a random order.

Materials and procedure

Two pictures accompanied each story. One picture showed the protagonist (drawn on white card, 10 cm × 16 cm), and the other depicted the two options drawn on each half of an A4 sheet of paper and colored.

Children sat opposite the experimenter in a quiet corner of the classroom. A picture of the first protagonist was placed on the table and the first story commenced. The second picture was placed below the first at the point in the story where the protagonist made her choice. The experimenter indicated the selected option during the story. The pictures remained on the table until the child had responded to the explanation question.

Results

First, children’s explanations were examined for references to the protagonist’s rejected action in the story, the dependent variable adopted by Harris et al. (1996). The percentages of children in each condition making such a reference on either of their two stories with a given outcome appear in Table 1.

With negative outcomes, children were more likely to refer to the rejected action for decisive choice than irrelevant choice stories (50% versus 15%: Upton’s $\chi^2$ (DF = 1, N = 40) = 5.6, $p < 0.01$, one-tailed). However, there was no such difference for positive outcomes (15% decisive choice, 10% irrelevant choice: Upton’s $\chi^2$, ns). Moreover, within the decisive choice condition, significantly more children mentioned the rejected option for negative stories and failed to do so for positive stories than showed the opposite pattern (McNemar, binomial: $N = 12$, $k = 2$, $p = 0.019$). This pattern was not evident for irrelevant choice stories (McNemar, binomial: ns).

Second, children’s explanations were examined for references to the selected action. The results for this dependent variable appear in Table 1. From these results it can be seen that there is no evidence that children were more likely to offer such explanations for the decisive choice stories than they were for the irrelevant choice stories, either for positive or negative outcomes (Upton’s $\chi^2$, ns).

Children’s explanations were also examined for references to alternative actions that would have prevented the outcome. These results appear in Table 2. With negative outcomes, children generated no more alternative pre-

<table>
<thead>
<tr>
<th>References to rejected option</th>
<th>References to selected option</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ outcome</td>
<td>- outcome</td>
</tr>
<tr>
<td>Decisive choice condition</td>
<td>15</td>
</tr>
<tr>
<td>Irrelevant choice condition</td>
<td>10</td>
</tr>
</tbody>
</table>

© Blackwell Publishers Ltd. 1999
ventative actions for irrelevant choice stories than for decisive choice stories (25% versus 35% respectively: Upton’s $\chi^2$, ns). For positive outcomes, there were fewer examples of such explanations overall, and no effect of condition (5% for decisive choice, 10% for irrelevant choice: Upton’s $\chi^2$, ns). However, in the decisive choice condition an effect of outcome was apparent; significantly more children generated an alternative outcome for negative stories and failed to do so for positive stories than showed the opposite pattern (McNemar, binomial: $N = 8, k = 1, p = 0.035$). In the irrelevant choice condition there was no such effect (McNemar, binomial: ns).

Finally, to assess the possibility that counterfactual thinking might be manifested explicitly (i.e. references to the rejected action) in answer to stories with negative outcomes and implicitly (through reference to the selected action) for positive outcomes, a more lenient criterion for counterfactual thinking was adopted. Children’s answers to explanation questions were examined for references to the point of choice (i.e. reference to either the selected or the rejected action) across their two stories of each outcome type. Table 3 shows these data. It can be seen that children in the decisive choice condition were more likely than those in the irrelevant choice condition to refer to the point of choice for the stories with negative outcomes (90% versus 60%: Upton’s $\chi^2$ (DF = 1, $N = 40$) = 5.06, $p < 0.05$) while stories with positive outcomes yielded no such difference (95% versus 80%: Upton’s $\chi^2$, ns).

Thus, even adopting a more lenient criterion for counterfactual thinking, there is no evidence that children are any more likely to engage in this kind of thinking in stories where the point of choice makes a difference than in cases where the choice is irrelevant.

**Table 2** Percentage of children generating alternative courses of action for either story as a function of condition and outcome

<table>
<thead>
<tr>
<th>References to alternative actions</th>
<th>+ outcome</th>
<th>– outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decisive choice condition</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>Irrelevant choice condition</td>
<td>05</td>
<td>35</td>
</tr>
</tbody>
</table>

**Table 3** Percentage of children making reference to the ‘point of choice’ as a function of condition and outcome

<table>
<thead>
<tr>
<th>References to ‘point of choice’</th>
<th>+ outcome</th>
<th>– outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decisive choice condition</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Irrelevant choice condition</td>
<td>80</td>
<td>60</td>
</tr>
</tbody>
</table>

Discussion

Children in this study offered explanations for events with both positive and negative outcomes. Counterfactual thinking, as indexed by references to rejected actions in decisive choice stories, was evident when events to be explained involved negative outcomes, replicating the results of Harris *et al.* (1996). Moreover, unlike the Harris *et al.* procedure, the current procedure employed no explicit prevention questions in combination with the explanation questions. It is therefore unlikely that the counterfactual thinking observed by Harris *et al.* was an artifact of their questioning procedure.

However, the results also revealed that references to the rejected course of action dropped dramatically in children’s explanations of positive events, undermining the claim that counterfactual thinking, as indexed by this explanation pattern, is required in causal reasoning. If reference to rejected actions during explanation is to be taken as evidence that counterfactual thinking is involved in causal judgments, the pattern should hold for positive and negative outcomes. Even adopting a more lenient criterion for counterfactual thinking, by counting references to either the selected or rejected option (i.e. references to the point of choice), it was only for stories with negative outcomes that an advantage for the decisive choice condition was evident. This measure is of interest because one might argue, on the basis that the choices in these stories involved only two options, that reference to selected actions might be made in reference to ‘implicit’ consideration of the rejected alternative.

The results are consistent with the hypothesis under test, derived from consideration of the counterfactual literature, that counterfactual thinking is invoked in response to negative outcomes (e.g. Boninger *et al.*, 1994); children appear to mutate antecedents in order to undo events that go wrong. The results therefore undermine the view that counterfactual thinking is an essential part of reaching causal judgments, an interpretation in line with recent empirical work on this topic. Both Mandel and Lehman (1996) and N’gala and Branscombe (1995), as indicated in the introduction, found little overlap between events judged by subjects to be causally implicated in an outcome and events that were spontaneously changed by subjects in answer to a counterfactual prompt (‘the outcome would have been different if only...’). These authors conclude that people do not assign causes on the basis of considerations of counterfactual states of affairs:

we characterize counterfactuals not as test results indicating why an outcome occurred but as highly
available mental simulations that often represent how various outcomes could have been prevented. (Mandel & Lehman, 1996, p. 461, italics in original)

Though the data presented here do not support the notion that an explicit counterfactual thinking process is implicated in reaching causal conclusions, it seems intuitive that there is some relationship between counterfactual conditional statements and causal statements. An interesting possibility, discussed but rejected by Harris et al. (1996, p. 250), is that causal judgments might occur prior to and license subsequent counterfactual inferences rather than vice versa. At least one philosophical analysis argues that some counterfactual statements depend for their truth on our appreciation of causal laws (Jackson, 1977). On this view, causal impressions arise first and may, or may not, give rise to counterfactual judgments.

Several strands of evidence favor this possibility. First, there is evidence from studies of the representation of causal and counterfactual information in memory; while subjects sometimes mistake counterfactual conditional statements for related causal statements that they originally encountered, they are less likely to mistake causal statements for related counterfactual conditionals that were originally presented (Fillenbaum, 1974). This result is consistent with the notion that the priority direction of inference is from causal statements to counterfactual information based on them, rather than from counterfactual statements to causal knowledge.

Second, some causal impressions seem unlikely to be mediated by analytic counterfactual thinking processes, such as the launching phenomena discussed by Michotte (1963), especially given that such impressions arise in 6-month-old infants (Leslie & Keeble, 1987).

Finally, the view that causal reasoning is mediated by counterfactual thinking faces a version of the frame problem. That is, given an indefinite number of counterfactual antecedent states of affairs, which should be tested in a simulation of causality? Mandel and Lehman (1996) make the point that counterfactual thinking seems more like confirmation than experiment; subjects focus early on plausible candidate events which rarely fail to change the outcome (Hofstadter, 1979). If counterfactual thinking is a test for causality, people should often test antecedent events that fail to change the outcome.

These problems do not arise if causal impressions arise first, and serve to focus attention on events that might or might not be subject to confirmatory counterfactual thinking processes. Whether counterfactual thinking is actually invoked will depend on various factors such as the type of outcome, as demonstrated here (see also Roese & Olson, 1995b), and other factors such as whether the events are under the control of the protagonist or not (e.g. Girotto, Legrenzi & Rizzo, 1991; Markman, Gavanski, Sherman & McMullen, 1995). The function of counterfactual thinking, rather than being a process for ascribing causes, might instead have to do with issues of prevention and future action planning (Mandel & Lehman, 1996).

---

**Appendix: Further story details**

**Story 2  Billy is going out to fly his kite. His mother asks him what he is going to wear.**

<table>
<thead>
<tr>
<th>Negative outcome</th>
<th>Decisive choice</th>
<th>Irrelevant choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice between woolly jacket and red cardigan.</td>
<td>Choice between blue cardigan and red cardigan.</td>
</tr>
<tr>
<td></td>
<td>Picks red cardigan.</td>
<td>Picks red cardigan.</td>
</tr>
<tr>
<td></td>
<td>Gets cold and has to come home early.</td>
<td>Gets cold and has to come home early.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive outcome</th>
<th>Decisive choice</th>
<th>Irrelevant choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice between woolly jacket and red cardigan.</td>
<td>Choice between woolly jacket and thick coat.</td>
</tr>
<tr>
<td></td>
<td>Picks woolly jacket.</td>
<td>Picks woolly jacket.</td>
</tr>
<tr>
<td></td>
<td>Stays warm and plays all day.</td>
<td>Stays warm and plays all day.</td>
</tr>
</tbody>
</table>

**Story 3  Jenny is going to play in the garden. Her mother asks her what she is going to put on her feet.**

<table>
<thead>
<tr>
<th>Negative outcome</th>
<th>Decisive choice</th>
<th>Irrelevant choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice between black wellies and green trainers.</td>
<td>Choice between black wellies and green trainers.</td>
</tr>
<tr>
<td></td>
<td>Picks green trainers.</td>
<td>Picks black wellies.</td>
</tr>
<tr>
<td></td>
<td>Feet get wet and cold.</td>
<td>Feet get wet and cold.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive outcome</th>
<th>Decisive choice</th>
<th>Irrelevant choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice between black wellies and green trainers.</td>
<td>Choice between black wellies and green wellies.</td>
</tr>
<tr>
<td></td>
<td>Picks black wellies.</td>
<td>Picks black wellies.</td>
</tr>
<tr>
<td></td>
<td>Feet stay dry and warm.</td>
<td>Feet stay dry and warm.</td>
</tr>
</tbody>
</table>
Acknowledgements

The author would like to thank Simon Royston for help in the collection of the data reported here and the staff, children and parents of Cherry Tree County Primary School and Alton Infants School for their time. Thanks also to Paul Bloom, Paul Harris, Shaun Nichols, Sandeep Prasada, Richard Samuels and one anonymous reviewer for comments on an earlier draft of this paper and/or useful discussion of the issues presented here.

References


Received: 8 December 1998
Accepted: 8 February 1999