The development of counterfactual reasoning about doubly-determined events

Teresa McCormack*, Margaret Ho, Charlene Gribben, Eimear O'Connor, Christoph Hoerl

Queen's University Belfast, UK, Christoph Hoerl, University of Warwick, UK

ARTICLE INFO

ABSTRACT

Previous studies of children's counterfactual reasoning have focused on scenarios in which a single causal event yielded an outcome. However, there are also cases in which an outcome would have occurred even in the absence of its actual cause, because of the presence of a further potential cause. In this study, 128 to 152 children aged 4–9 years reasoned counterfactually about such scenarios, in which there were 'doubly-determined' outcomes. The task involved dropping two metal discs down separate runways, each of which was sufficient to knock over a toy pig. One of the runways was shorter than the other, meaning that one of the discs actually knocked over the pig whereas the other always arrived too late to do so. Children were asked whether the pig would have been knocked over in the absence of the first metal disc descending the runway. We found that children could accurately answer such counterfactual questions by 6–7 years.

1. Introduction

Numerous studies have examined the development of counterfactual reasoning (e.g., Beck, Riggs, & Gorniak, 2009; Beck, Riggs & Gorniak, 2010; Beck, Riggs, & Burns, 2011; German & Nichols, 2003; McCormack, Butterfill, Hoerl, & Burns, 2009; Perner, Sprung, & Steinkogler, 2004; Riggs, Peterson, Robinson, & Mitchell, 1998). In these studies, children were typically either shown or told about an event sequence and asked to judge what would have happened if an aspect of the world had been different. For example, Riggs et al. (1998) told children about a fireman, Peter, who is at home in bed because he does not feel very well, but then goes to the Post Office after getting a phone call asking him to go to help put out a fire there. In the counterfactual reasoning task, children had to judge where Peter would be if there had been no fire, with the correct answer being that he would be in bed. Children who are 4–5 years can typically answer these sorts of questions correctly. Children's ability to correctly answer such counterfactual questions has been shown to be related to other cognitive skills, in particular false belief understanding (Perner et al., 2004; Riggs et al., 1998) and aspects of executive functions (Beck et al., 2009; Drayton, Turley-Ames, & Guajardo, 2011).

All of the studies mentioned have examined what can be termed reasoning involving "real-world" counterfactuals: reasoning about alternatives to events as they happened at a particular past time (e.g., in Rigg's et al.'s study, what would have been the case if the Post Office fire had not occurred). Reasoning about real-world counterfactuals has been distinguished from reasoning about other types of non-actual scenarios, including hypothetical, pretend, or fictitious scenarios (Beck, 2016; Hoerl, McCormack & Beck, 2011). Whilst this latter type of reasoning also involves considering scenarios that do not match reality, it does not require mentally undoing an aspect of an actual past event sequence that has happened. Theorists differ in terms of whether they think that the term

counterfactual reasoning should be extended to cover this type of reasoning as well (Beck, 2016; Hoerl et al., 2011; Perner & Rafetseder, 2011; Weisberg & Gopnik, 2013; Woodward, 2011). For example, Weisberg and Gopnik (2013, 2016) suggest that while it might be possible to distinguish between reasoning about real-world counterfactuals and reasoning about other non-actual scenarios, the latter is also a type of counterfactual thought that relies on the very same cognitive structures. Their motivation for this claim stems from the idea that both sorts of thought can be modelled within a Causal Bayes Net framework by assuming that they both involve an imaginary intervention on a variable in a graphical causal model to fix its value and then calculating the effects of this intervention (Weisberg & Gopnik, 2013; for a detailed description of such modelling, see Hagmayer, Sloman, Lagnado, & Waldmann, 2007). However, other developmental psychologists argue that the ability to reason using real-world counterfactuals is quite different from the ability to reason about pretend or hypothetical scenarios, and that the former is particularly significant for both theoretical and practical reasons (Beck, 2016; Beck et al., 2011; Perner & Rafetseder, 2011; Rafetseder & Perner, 2014).

From a practical perspective, real-world counterfactual reasoning is assumed to be important for a range of aspects of judgment, learning, and decision making, such as learning not to repeat one's mistakes as a result of experiencing regret, according responsibility for past actions, and making moral judgments (Byrne, 2016; Epstude & Reese, 2008). From a theoretical perspective, reasoning using real-world counterfactuals has been taken to be indicative of a distinctive type of cognitive skill. Specifically, it is assumed to involve (i) the ability to construct and hold in mind simultaneously two representations of the world (what actually happened, and the counterfactual scenario; Byrne, 2007, 2016) and (ii) an understanding of how these two representations are related to each other (Beck et al., 2011; Perner & Rafetseder, 2011). In holding both representations in mind and understanding the relation between them, children are assumed to grasp that the counterfactual scenario involves a negation or undoing of a specific aspect of the actual sequence of events specified in the antecedent of the counterfactual (e.g., in the above example, there not having been a fire in the Post Office). Because children must consider them to be alternative representations of the very same past time, there is a sense in which there is a conflict between the two representations that does not exist in the case of entertaining representations of pretend or hypothetical scenarios (Beck, 2016; Beck et al., 2010).

This way of characterizing what is involved in reasoning using real-world counterfactuals has important implications. First, Beck, Robinson, Carroll, and Apperly (2006) suggest that grasping that there is "a common past that unites counterfactual and real worlds" (Beck, 2016, p. 254) is part of understanding a fundamental feature of the world, namely that at any given point in time multiple possible events could occur. Such a claim forges a close link between this type of counterfactual reasoning and an understanding of the nature of time itself (McCormack, 2015; McCormack & Hoerl, 2008). Second, this connection between the two representations places important constraints on the construction of the counterfactual alternative: the counterfactual alternative should mirror what actually happened in the real world, only differing in terms of the impact of negating the specific event mentioned in the antecedent. We can see how this principle might be put to work in the case of reasoning about Peter and the Post Office fire: it is judged that Peter would be in bed if the Post Office had not caught fire because it is assumed that all other things are equal in that counterfactual scenario (e.g., that Peter did not separately get phoned and asked to help put out a fire in the hospital). Following Lewis (1973a), Rafetseder, Schwitalla, and Perner (2013) refer to this as applying the Nearest Possible World Constraint.

1.1. When does real-world counterfactual reasoning develop?

If the distinction between real-world counterfactual reasoning and reasoning about other types of non-actual scenarios is a meaningful one, then it might be predicted that these two types of reasoning show different developmental trajectories. It is well-established that children are capable of reasoning about pretend and fictitious scenarios early in development, with 2-year-olds able to make certain types of judgments about pretend events (Harris, Kavanaugh, & Meredith, 1994; Harris & Kavanaugh, 1993). However, it is plausible that, because of its cognitive demands, real-world counterfactual reasoning does not emerge until later. The empirical challenge is to try to devise tasks that distinguish between this specific type of counterfactual reasoning and other types of reasoning. This challenge arises because of the possibility that children may give the correct answers in counterfactual reasoning tasks without actually engaging in real-world counterfactual reasoning (Beck et al., 2010; Beck et al., 2011; Rafetseder, Cristi-Vargas, & Perner, 2010; Rafetseder et al., 2013).

An example from a task used in Harris, German, and Mills' (1996) influential study of counterfactual reasoning can be used to illustrate this possibility. In Harris et al.'s study, children observed short sequences of events, such as a doll called Carol walking across the floor and leaving dirty footprints, because she was wearing muddy boots. The majority of 3-year-olds were able to answer such counterfactual questions as "What if Carol had taken her shoes off, would the floor be dirty?" On the face of it, it looks like children were engaging in real-world counterfactual reasoning, in which they mentally changed a specific aspect of the past event of Carol arriving home and walking across the floor, namely whether she kept her shoes on. However, Rafetseder et al. (2013; see also Rafetseder et al., 2010; Rafetseder & Perner, 2010) argued that children can answer these sorts of questions without thinking counterfactually at all. They suggested that instead such questions can be answered correctly using what they term basic conditional reasoning (BCR). BCR is assumed to involve reasoning using conditionals that apply to general states of affairs: e.g., "If someone takes their shoes off, the floor stays clean."; "If someone doesn't get a phone call, they will stay in bed."

The general states of affairs that such conditionals describe can sometimes differ from what has actually obtained: for example, children may, using BCR, correctly apply the principle "If someone takes their shoes off, the floor stays clean" to answer a question in circumstances in which a person did not actually take their shoes off and the floor is dirty. That does not mean, though, that under such circumstances children need to be engaging in real-world counterfactual reasoning in order to do so. The claim is that children can correctly answer questions that may be expressed as counterfactuals by applying only tenseless general if-then principles rather