Acta Psychologica 141 (2012) 54–66

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

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

# Causal conditionals and counterfactuals

Caren A. Frosch [a,*], Ruth M.J. Byrne [b]

[a] University of Leicester, UK
[b] Trinity College Dublin, University of Dublin, Ireland

ARTICLE INFO

*Article history:*
Received 11 November 2011
Received in revised form 28 June 2012
Accepted 2 July 2012
Available online 2 August 2012

*PsycINFO classification:*
2300 Human Experimental Psychology
2340 Cognitive Processes

*Keywords:*
Conditional reasoning
Counterfactuals
Causality
Enablers
Mental models

ABSTRACT

Causal counterfactuals e.g., 'if the ignition key *had been* turned then the car *would have started*' and causal conditionals e.g., 'if the ignition key was turned then the car started' are understood by thinking about multiple possibilities of different sorts, as shown in six experiments using converging evidence from three different types of measures. Experiments 1a and 1b showed that conditionals that comprise enabling causes, e.g., 'if the ignition key was turned then the car started' primed people to read quickly conjunctions referring to the possibility of the enabler occurring without the outcome, e.g., 'the ignition key was turned and the car did not start'. Experiments 2a and 2b showed that people paraphrased causal conditionals by using causal or temporal connectives (because, when), whereas they paraphrased causal counterfactuals by using subjunctive constructions (had…would have). Experiments 3a and 3b showed that people made different inferences from counterfactuals presented with enabling conditions compared to none. The implications of the results for alternative theories of conditionals are discussed.



## 1. Introduction

Our primary research question is, what possibilities do people envisage when they understand a causal counterfactual, e.g., 'if the ignition key *had been* turned the car *would have* started'? The causal counterfactual appears to convey something very different from its conditional counterpart, e.g., 'if the ignition key was turned the car started' (e.g., Lewis, 1973; Stalnaker, 1968). People create counterfactual alternatives to reality frequently in everyday life, when they think about how events in the past could have turned out differently, 'if only' (Kahneman & Tversky, 1982; Roese & Olson, 1995). The counterfactual conjecture may help them to work out the various causes of an outcome, and to prepare for the future (e.g., Mandel, Hilton, & Catellani, 2005; Markman, Klein, & Suhr, 2009). Counterfactual thoughts tend to focus on background conditions, that is, enabling causes, rather than on direct, strong causes (Byrne, 2005). For example, participants who read a story about a drunk driver who crashed into an individual driving home by an unusual route identified the cause of the accident as the drunk driver, but they created counterfactual alternatives such as 'if only he had driven home by his usual route' (Mandel & Lehman, 1996; N'gbala & Branscombe, 1995). They tend to focus on enabling conditions rather than strong causes, perhaps because the removal of an enabler within their control effectively prevents a bad outcome even when the cause is outside their control (Byrne, 2007; Egan, Frosch, & Hancock, 2008).

And so our second question is, what possibilities do people envisage when they understand a causal conditional that refers to an enabling cause such as 'if the ignition key was turned the car started'? The enabler is a necessary cause to bring about the outcome but it is not sufficient, that is, the outcome requires other causes to be fulfilled as well, e.g., there is petrol in the car, the battery is charged, and so on (e.g., De Neys, Schaeken, & d'Ydewalle, 2005; Markovits, Lortie Forgues, & Brunet, 2010). We report six experiments to answer these two research questions, by converging evidence from three different methods – causal conditionals as primes, paraphrases of causal conditionals and counterfactuals, and inferences from causal conditionals and counterfactuals. The experiments show that people keep in mind multiple possibilities when they think about counterfactuals, and when they think about enabling causes. First we outline how people understand and reason from ordinary conditionals, then causal conditionals, and then counterfactuals.

### 1.1. Ordinary conditionals

How do people understand and reason from conditionals? In fact, there is as yet no consensus (e.g., Byrne & Johnson-Laird, 2009). One view is that people understand an 'ordinary' or *indicative* conditional, 'if there is a triangle on the blackboard then there is a circle' (if A then B) by thinking about rules of inference, either abstract (Braine & O'Brien, 1998; Rips, 1994) or domain specific (Fiddick, Cosmides, & Tooby, 2000; Holyoak & Cheng, 1995). Another view is that they understand it by thinking about probabilities: they assume the truth of the antecedent, A, and assess whether B or not-B is more probable (Evans & Over, 2004; see also Oaksford & Chater, 2007). A third view is that they understand

* Corresponding author at: School of Psychology, University of Leicester, Henry Wellcome Building, Lancaster Road, Leicester, LE1 9HN, UK. Tel.: +44 116 229 7189.
*E-mail address:* cf159@leicester.ac.uk (C.A. Frosch).

0001-6918/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2012.07.001

it by thinking about possibilities (Johnson-Laird & Byrne, 2002). A principle of truth ensures that they think about only the true possibilities that are consistent with the conditional: a triangle and a circle, no triangle and no circle, and no triangle and a circle; and they do not think about false possibilities that are ruled out by the conditional — a triangle and no circle (Espino & Byrne, 2012; Espino, Santamaria, & Byrne, 2009; Johnson-Laird & Byrne, 2002). Because of the constraints of working memory they also tend to think about few possibilities (Johnson-Laird, Byrne, & Schaeken, 1992), and so they understand the conditional by envisaging initially just a single model, a triangle and a circle (A and B), as Table 1 outlines.

On this account, people can readily make the *modus ponens* inference (A therefore B) because it matches the initial possibility they have kept in mind. They have more difficulty with the *modus tollens* inference (not-B therefore not-A) because they must think about some of the other true possibilities, e.g., not-A and not-B, in order to make it. They tend to make the *affirmation of the consequent* inference (B therefore A), whenever they keep in mind the initial possibility and fail to think of other true possibilities, e.g., not-A and B. They make the *denial of the antecedent* inference (not-A therefore not-B) when they have thought about some of the alternative possibilities (not-A and not-B) but not others (not-A and B). The interpretation of a basic conditional can be modulated by its content and context (Johnson-Laird & Byrne, 2002), as illustrated by conditionals with causal content, in the next section.

### 1.2. Causal conditionals

How do people understand and reason from *causal* conditionals? Causal conditionals can refer to different sorts of causes (e.g., Goldvarg & Johnson-Laird, 2001). They can express a strong cause, e.g., heating water to 100° causes it to boil, which is both necessary and sufficient for the outcome. They can express one of several alternative weak causes, e.g., arson caused the Australian bushfires, or accidental sparks from campfires caused them, any one of which is sufficient but not necessary. Or they can express one of several joint enabling conditions, e.g., arson caused the bushfires, enabled by the presence of dry vegetation, any one of which is necessary but not sufficient.

Alternative views exist about whether causes and enabling relations differ in terms of their meaning or logic, or in terms of characteristics such as normality, conversational relevance, constancy and covariation (e.g., Cheng & Novick, 1992; Einhorn & Hogarth, 1986; Hilton & Erb, 1996; Sloman, 2005; Turnbull & Slugoski, 1988). The interpretation of causality is controversial. One view is that people may think about different possibilities to mentally represent different sorts of causes (e.g., Frosch & Johnson-Laird, 2011; Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Byrne, 1991).

Our focus is on enabling causes, and the possibilities that people consider for enabling causes. Most people consider that the enabling conditional 'if the ignition key was turned then the car started' is consistent with the possibility, the key was turned and the car started (A and B), and with the possibility, the key was not turned and the car did not start (not-A and not-B). But the full interpretation of the causal conditional depends on the retrieval of counterexamples (De Neys, 2011; De Neys et al., 2005; Markovits et al., 2010; see also Geiger & Oberauer, 2007). In this case people appear to think readily about disablers, e.g., the key was turned and the car did not start, perhaps because the battery was dead (A and not-B), that is, they judge the cause to be consistent with a third possibility. They do not tend to think of alternative causes, that is, possibilities consistent with the key not being turned and the car starting anyway. Their interpretation of the conditional as an enabling causal relation rules out as false the possibility that the key was not turned and the car started (not-A and B). People make different inferences from different causal relations because of the availability of counterexamples (e.g., Byrne, 1989; Byrne, Espino, & Santamaria, 1999). As a result, for an enabling cause, they make the affirmation of the consequent (B therefore A) and denial of the antecedent (not-A therefore not-B) inferences only, and they resist the modus ponens (A therefore B) and modus tollens (not-B therefore not-A) inferences, because they can retrieve a disabler — the battery being flat caused the car not to start (e.g., Cummins, Lubart, Alksnis, & Rist, 1991; De Neys et al., 2005; Markovits & Potvin, 2001).

An enabling cause can be contrasted with other sorts of causes, such as a weak cause. For example, most people judge that a cause such as 'if the apples were ripe then they fell from the tree' is consistent with the possibility, the apples were ripe and they fell from the tree (A and B), and with the possibility, the apples were not ripe and they did not fall from the tree (not-A and not-B). In this case people appear to think readily about counterexamples based on alternative causes, that is, they judge that the cause is consistent with a third possibility, that the apples were not ripe and they fell from the tree anyway, perhaps because of strong winds (not-A and B). They do not tend to think readily of disablers in this case, that is, possibilities consistent with the apples being ripe and not falling from the tree, and so this possibility is ruled out as false. Hence the interpretation of the conditional is as a *weak* causal relation. For a weak causal relation, they make the modus ponens (A therefore B) and modus tollens (not-B therefore not-A) inferences but they resist the affirmation of the consequent (B therefore A) and denial of the antecedent (not-A therefore not-B) inferences.

For a third sort of causal relation, a strong cause, such as 'if Joe cut his finger then it bled' (A causes B), people tend to think of just two possibilities: he cut his finger and it bled (A and B) and he did not cut his finger and it did not bleed (not-A and not-B), as Table 1 shows. Most people do not tend to think readily of disablers, that is, possibilities consistent with Joe cutting his finger and it not bleeding, and they do not tend to think of alternative causes, that is, possibilities consistent with Joe not cutting his finger and it bleeding — even if such possibilities exist (e.g., Cummins et al., 1991; De Neys et al., 2005). Hence they come to an interpretation of the causal relation as a strong cause, which rules out as false two possibilities: he cut his finger and it did not bleed (A and not-B) and he did not cut his finger and it bled (not-A and B). As a result, people make all four inferences from a strong cause. Enabling causes tend to be focused on when people create counterfactual conditionals, and so we turn now to a consideration of counterfactuals.

**Table 1**
The consistent possibilities for indicative and counterfactual conditionals expressing basic content and enabling causal relations; with information on strong and weak causes for comparison.

| | Indicative | Counterfactual | |
|---|---|---|---|
| | *If A then B* | *If A had been then B would have been* | |
| Basic | A and B | A and B | (Conjecture) |
| | Not-A and not-B | Not-A and not-B | (Presupposed facts) |
| | Not-A and B | Not-A and B | |
| Enabler | A and B | A and B | (Conjecture) |
| | Not-A and not-B | Not-A and not-B | (Presupposed facts) |
| | A and not-B | A and not-B | |
| Strong Cause | A and B | A and B | (Conjecture) |
| | Not-A and not-B | Not-A and not-B | (Presupposed facts) |
| Weak Cause | A and B | A and B | (Conjecture) |
| | Not-A and not-B | Not-A and not-B | (Presupposed facts) |
| | Not-A and B | Not-A and B | |

### 1.3. Counterfactual conditionals

Counterfactual conditionals often express causal claims (e.g., Thompson & Byrne, 2002), and the relation between counterfactuals and causal assertions has long been of interest to philosophers and psychologists (e.g., Byrne, 2011; Chisholm, 1946; Hoerl, McCormack, & Beck, 2011). Even with non-causal content, a counterfactual conditional in the subjunctive mood, e.g., 'if there had been a triangle then there would have been a circle' seems to mean something very different from an indicative one, 'if there was a triangle then there was a circle' (Lewis, 1973; Stalnaker, 1968). People tend to judge that someone