COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12827

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12827

The Suppression of Inferences From Counterfactual Conditionals

Orlando Espino,[a] Ruth M. J. Byrne[b]

[a]*Department of Psychology, University of La Laguna*
[b]*School of Psychology and Institute of Neuroscience, Trinity College Dublin,University of Dublin*

Received 18 November 2018; received in revised form 30 January 2020; accepted 19 February 2020

Abstract

We examine two competing effects of beliefs on conditional inferences. The suppression effect occurs for conditionals, for example, "if she watered the plants they bloomed," when beliefs about additional background conditions, for example, "if the sun shone they bloomed" decrease the frequency of inferences such as *modus tollens* (from "the plants did not bloom" to "therefore she did not water them"). In contrast, the counterfactual elevation effect occurs for counterfactual conditionals, for example, "if she *had* watered the plants they *would have* bloomed," when beliefs about the known or presupposed facts, "she did not water the plants and they did not bloom" increase the frequency of inferences such as *modus tollens*. We report six experiments that show that beliefs about additional conditions take precedence over beliefs about presupposed facts for counterfactuals. The *modus tollens* inference is suppressed for counterfactuals that contain additional conditions (Experiments 1a and 1b). The *denial of the antecedent* inference (from "she did not water the plants" to "therefore they did not bloom") is suppressed for counterfactuals that contain alternatives (Experiments 2a and 2b). We report a new "switched-suppression" effect for conditionals with negated components, for example, "if she had not watered the plants they would not have bloomed": *modus tollens* is suppressed by alternatives and *denial of the antecedent* by additional conditions, rather than vice versa (Experiments 3a and 3b). We discuss the implications of the results for alternative theories of conditional reasoning.

Keywords: Counterfactuals; Conditional reasoning; Suppression; Beliefs

1. Introduction

Knowledge affects the inferences that people make in everyday hypothetical reasoning, and different sorts of knowledge have different effects. Consider, for example, the *modus tollens* inference: When participants in experiments are given a conditional in the

Correspondence should be sent to Ruth Byrne, School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin 2, Ireland. E-mail: rmbyrne@tcd.ie

indicative mood, often called a *factual* conditional, such as "if she watered the plants they bloomed," and they are told, "the plants did not bloom," about two-thirds of them make the inference, "therefore she did not water them" (see Nickerson, 2015, for a review). One striking effect of knowledge is that it can *suppress* such inferences. For example, when participants know of an additional condition such as "if the sun shone on the plants they bloomed," the frequency with which they make the inference *decreases* dramatically (Byrne, 1989). An equally striking effect of knowledge is that it can *elevate* such inferences. For example, when participants are given a conditional in the subjunctive mood, often called a *counterfactual* conditional, such as, "If she *had* watered the plants they *would have* bloomed," they recover the known or presupposed facts, "she did not water the plants and they did not bloom" and the frequency with which they make the *modus tollens* inference *increases* dramatically (Byrne & Tasso, 1999). Our aim is to examine the conflicting influence of these two different sorts of beliefs on conditional inferences. We report six experiments that examine whether participants make inferences such as *modus tollens* for counterfactuals, which convey presupposed facts, when they also know about additional background conditions.

Our question is: Which will take precedence in reasoners' inferences, their beliefs about additional background conditions, for example, that sun is needed as well as water for the plants to bloom, or their beliefs about the presupposed facts, for example, that in fact, she did not water the plants and they did not bloom? There are at least three potential answers: (a) Knowledge about background conditions will over-ride knowledge about the presupposed facts so that inferences such as *modus tollens* will be suppressed, compared to situations without such knowledge; (b) Knowledge about presupposed facts will over-ride knowledge about background conditions so that inferences such as *modus tollens* will be elevated; or (c) Knowledge about the presupposed facts will compete equally with knowledge about background conditions, and since they pull in opposite directions, inferences such as *modus tollens* will be made at the same frequency in situations with both sorts of knowledge as in situations without any such knowledge.

Different answers are suggested to this question by different theories of human reasoning. We first outline the predictions derived from theories of reasoning based on the idea that people simulate possibilities, which predict answer (a): Knowledge of additional conditions will take precedence over knowledge of the presupposed facts. Next, we report the results of the six experiments that test these predictions. Then, we consider alternative theories, including those based on the idea that people assess likelihood from prior beliefs, which predict answer (b): Knowledge of the presupposed facts will take precedence over knowledge of additional conditions. To begin, we describe the experimental evidence for the two different effects of knowledge on conditional inferences in the next section.

2. Conditional suppression and counterfactual elevation

For a conditional such as "if she watered the plants they bloomed," almost all participants in experiments make the inference from "she watered the plants" to "they bloomed," called *modus ponens* by logicians (e.g., Jeffrey, 1981). About two-thirds of them make the inference from "the plants did not bloom" to "she did not water them," called *modus tollens* in standard propositional logic (see Table 1). Similar frequencies are observed for the *denial of the antecedent* inference from "she did not water the plants" to "they did not bloom" and the *affirmation of the consequent* inference from "the plants bloomed" to "she watered them" (e.g., Nickerson, 2015).

2.1. The suppression effect

Knowledge of additional background conditions *suppresses* inferences such as *modus ponens* and *modus tollens*. For a pair of conditionals that contain an *additional* condition (sometimes called an enabler), such as "if she watered the plants they bloomed, if the sun shone they bloomed," the frequency with which participants make the inferences is suppressed. When participants are told, "she watered the plants," they no longer tend to make the *modus ponens* inference "therefore they bloomed," and when they are told "the plants did not bloom," they no longer tend to make the *modus tollens* inference "therefore she did not water them" (Byrne, 1989). The additional condition has no effect on the other two inferences (see Table 1). In contrast, knowledge of alternative conditions

Table 1
Four inferences from arguments based on a single conditional, the pattern of suppression for an additional condition and an alternative condition, and the pattern of elevation for a counterfactual

	Categorical Premise	Conclusion
Conditional premise	*If she watered the plants they bloomed*	
Modus ponens	She watered the plants	They bloomed
Modus tollens	The plants did not bloom	She did not water them
Affirm antecedent	The plants bloomed	She watered them
Deny consequent	She did not water the plants	They did not bloom
Conditional premise	*If she watered the plants they bloomed*	
Additional condition	*If the sun shone on the plants they bloomed*	
Modus ponens	She watered the plants	They bloomed—Suppressed
Modus tollens	The plants did not bloom	She did not water them—Suppressed
Affirm antecedent	The plants bloomed	She watered them
Deny consequent	She did not water the plants	They did not bloom
Conditional premise	*If she watered the plants they bloomed*	
Alternative condition	*If it rained on the plants they bloomed*	
Modus ponens	She watered the plants	They bloomed
Modus tollens	The plants did not bloom	She did not water them
Affirm antecedent	The plants bloomed	She watered them—Suppressed
Deny consequent	She did not water the plants	They did not bloom—Suppressed
Counterfactual premise	*If she had watered the plants they would have bloomed*	
Modus ponens	She watered the plants	They bloomed
Modus tollens	The plants did not bloom	She did not water them—Elevated
Affirm antecedent	The plants bloomed	She watered them
Deny consequent	She did not water the plants	They did not bloom—Elevated

suppresses inferences such as the *denial of the antecedent* and *affirmation of the consequent*. For a pair of conditionals that contain an *alternative* condition, such as "if she watered the plants they bloomed, if it rained on the plants they bloomed," the frequency with which participants make such inferences is suppressed (e.g., Rumain, Connell, & Braine, 1983). When participants are told, "she did not water the plants," they no longer tend to make the *denial of the antecedent* inference "therefore they did not bloom," and when they hear "the plants bloomed," they no longer tend to make the *affirmation of the consequent* inference "therefore she watered them." The alternative condition has no effect on the other two inferences (see Table 1).

There have been many hundreds of experiments carried out on the suppression effect since its first demonstration, to explore the factors that influence its exhibition in adults, including its implications in practical situations such as consumer choices or legal reasoning (e.g., Chandon & Janiszewski, 2008; Gazzo Castañeda & Knauff, 2016), its early emergence in children (e.g., de Chantal & Markovits, 2017; De Neys & Everaerts, 2008; Janveau-Brennan & Markovits, 1999), and its activation of various areas of the brain such as those associated with expectations (e.g., Bonnefond, Kaliuzhna, Van der Henst, & De Neys, 2014; see also Pijnacker, Geurts, Van Lambalgen, Buitelaar, & Hagoort, 2011). Its disrupted pattern in autistic people is characterized by suppression of inferences by alternatives, but not by additional conditions, perhaps because of differences in synthesizing context to interpret exceptions (e.g., Pijnacker et al., 2009; Pijnacker, Geurts, Van Lambalgen, Buitelaar, & Hagoort, 2010; see also McKenzie, Evans, & Handley, 2011; Stenning & Van Lambalgen, 2019), and it may contribute to delusion formation in schizophrenia (e.g., Sellen, Oaksford, & Gray, 2005; see also Phillips, Howard, & David, 1997). It exemplifies the flexibility of human reasoning to engage in belief revision to make non-monotonic or defeasible inferences (e.g., Elio, 1997; Pelletier & Elio, 1997), of relevance to the development of various artificial intelligence and machine learning algorithms (e.g., Dietz, Holldobler, & Ragni, 2012; Holldobler & Ramli, 2009; Ragni, Eichhorn, & Kern-Isberner, 2016; see also Kowalski, 2011). It occurs whether participants are explicitly told about additional or alternative conditions, or instead retrieve them from memory (e.g., Cummins, Lubart, Alksnis, & Rist, 1991; De Neys, Schaeken, & D'Ydewalle, 2002, 2003; Dieusseurt, Schaeken, Schroyens, & d'Ydewalle, 2000; Verschueren, Schaeken, & d'Ydewalle, 2005), whether they make a categorical judgment about what follows, or a judgment about their degree of certainty in the conclusion (e.g., Geiger & Oberauer, 2007; Markovits, Brisson, & de Chantal, 2015; Markovits, Brunet, & Lortie Forges, 2010), and whether the additional condition is presented as an enabler, for example, "if the sun shone the plants bloomed" or a disabler, for example, "if the sun did not shine the plants did not bloom" (e.g., Markovits et al., 2010; see also Markovits & Potvin, 2001). In many everyday situations, causal outcomes are over-determined or depend on multiple joint causes (e.g., Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015; Rehder, 2014; Strickland, Silver, & Keil, 2017), and suppression occurs not only for inferences about causal relations but also for relations based on intentions (e.g., Juhos, Quelhas, & Byrne, 2015), inducements (e.g., Couto, Quelhas, & Byrne, 2017), and polite discourse (e.g., Bonnefon & Hilton, 2002, 2004; Demeure, Bonnefon, & Raufaste, 2009; see also Chan & Chua, 1994). Despite the extensive research on the suppression effect, no studies have examined how it interacts with other beliefs, or with other reasoning effects, such as the counterfactual elevation effect, to which we now turn.

2.2. The counterfactual elevation effect

Knowledge about presupposed or known facts *elevates* the *modus tollens* and *denial of the antecedent* inferences. When participants understand a counterfactual conditional such as "if she *had* watered the plants they *would have* bloomed," the frequency with which they make these inferences is increased. For example, when they are told, "the plants did not bloom," they make the *modus tollens* inference "she did not water them," more often than they do from a factual conditional. When they are told "she did not water the plants," they make the *denial of the antecedent* inference "they did not bloom," more often than they do from a factual conditional. The counterfactual has no effect on the frequency of the other two inferences—when participants are told that "she watered the plants," they still make the *modus ponens* inference "they bloomed," as often as they do from a factual conditional, and when they are told, "the plants bloomed" they still make the *affirmation of the consequent* inference "she watered them" (Byrne & Tasso, 1999).

The tendency to make inferences such as *modus tollens* readily from counterfactuals is robust for various contents such as causes and definitions (Thompson & Byrne, 2002; see also Frosch & Byrne, 2012; Lucas & Kemp, 2015; Sloman & Lagnado, 2005). It has also been examined for deontic content such as obligations and inducement content such as promises and threats (Egan & Byrne, 2012; Quelhas & Byrne, 2003). It has been tested for different sorts of linguistic forms, such as *only if* and *even if* (Egan, Garcia-Madruga, & Byrne, 2009; Moreno-Rios, Garcia-Madruga, & Byrne, 2008), and for subjunctive conditionals about the past and present as well as about the future, that is, pre-factuals (Byrne & Egan, 2004; Byrne & Tasso, 1999).

People create counterfactuals very often in their everyday lives (e.g., De Brigard & Parikh, 2019; Kahneman & Tversky, 1982; Phillips, Morris, & Cushman, 2019; Roese & Epstude, 2017). A counterfactual such as "if she *had* watered the plants they *would have* bloomed" can seem to mean something very different from its factual counterpart, "if she watered the plants they bloomed." The counterfactual conveys the known or presupposed facts that its antecedent is not the case, "she did not water the plants," and its consequent is not the case either, "the plants did not bloom." Determining the truth of a counterfactual presents a logical challenge for a truth-functional account of conditionals—according to which a conditional is true if its antecedent is false or its consequent is true—because on such an account every counterfactual must be true, given that its antecedent is false (see Nickerson, 2015; see also Chisholm, 1946). Alternative analyses of counterfactuals based on possible-worlds semantics address the challenge in different ways (e.g., Lewis, 1973; Stalnaker, 1968). The differences in the inferences people endorse from factual and counterfactual conditionals seem to arise because people detect that counterfactuals convey the presupposition that the facts are quite different from what is mentioned (Byrne, 2005). For example, when participants read a counterfactual such as, "if she had watered the plants then they would have bloomed," and then are given a surprise memory test, they tend to mistakenly believe they were given "she did not water the plants" and "the plants did not bloom" (Fillenbaum, 1974). Participants also judged that someone who uttered the counterfactual meant to imply "she did not water the plants" and "the plants did not bloom" (Thompson & Byrne, 2002). They are *primed* to read the conjunction "she did not water the plants and the plants did not bloom" more quickly when they have read the counterfactual than the factual conditional (e.g., Santamaría, Espino, & Byrne, 2005; see also De Vega, Urritia, & Riffo, 2007; Stewart, Haigh, & Kidd, 2009).

Nonetheless, participants also read a conjunction "she watered the plants and the plants bloomed" equally quickly whether they are primed by a counterfactual or a factual conditional (Santamaría et al., 2005). Eye-tracking studies also indicate they have detected both possibilities (e.g., Ferguson & Sanford, 2008; Ferguson, Sanford, & Leuthold, 2008; see also Nieuwland & Martin, 2012). For example, they look at an image corresponding to the presupposed facts more often when they hear the counterfactual than a factual conditional, and they look at an image corresponding to the conjecture equally often for both sorts of conditionals (e.g., Orenes, Garcia-Madruga, Gomez-Veiga, Espino, & Byrne, 2019). Accordingly, brain imaging fMRI studies show that counterfactuals activate areas related to conflict detection such as the medial prefrontal cortex (e.g., Kulakova, Aichhorn, Schurz, Kronbichler, & Perner, 2013; Van Hoeck et al., 2013).

Our aim in this paper is to examine the competing impact of these two sorts of knowledge: on the one hand, knowledge about additional conditions, which decreases the frequency with which people make inferences such as *modus tollens*, and on the other hand, knowledge about presupposed facts, which increases the frequency with which people make inferences such as *modus tollens*. Consider a pair of counterfactuals that contain an additional condition, "If the plants had been watered they would have bloomed. If it had been sunny they would have bloomed." Given the *modus tollens* premise, "the plants did not bloom," three different patterns could potentially be observed:

1. Inferences such as *modus tollens* may be suppressed from counterfactuals that contain additional conditions, compared to single counterfactuals, just as they are suppressed from factual conditionals that contain additional conditions, compared to single factual conditionals. Knowledge about background conditions may override knowledge about the presupposed facts.
2. Inferences such as *modus tollens* may be elevated from counterfactuals, compared to factual conditionals, even when the counterfactuals contain an additional background condition. Knowledge about the presupposed facts may override knowledge about background conditions.
3. Inferences such as *modus tollens* may show no elevation or suppression effects, if the two tendencies exert an equally strong effect in opposite directions, and hence inferences will be made at the same frequency from counterfactuals as from factual conditionals, even when the counterfactuals contain an additional background condition. Knowledge about the presupposed facts and knowledge about background conditions may apply equal, opposing effects.

We turn now to consider cognitive processes in the suppression of inferences by additional conditions, and the elevation of inferences by counterfactuals.

3. Cognitive processes in counterfactual suppression

The suppression of conditional inferences has been explained in competing ways by psychological theories based on formal inference rules (e.g., Politzer & Braine, 1991; Rumain et al., 1983), mental models (e.g., Johnson-Laird & Byrne, 1991, 2002), probabilist suppositions (e.g., Over, 2017; Over & Cruz, 2017; Stevenson & Over, 1995), causal Bayes nets (e.g., Oaksford & Chater, 2007, 2017), logic programming (e.g., Stenning & Van Lambalgen, 2005, 2016), machine learning (e.g., Dietz, Holldobler, & Wernhard, 2014), and formal linguistics (e.g., Cariani & Rips, 2017). One key distinguishing feature of these theories is that some of them propose that suppression arises because people conjoin the antecedent of the first conditional, for example, "she watered the plants" and the additional condition, for example, "the sun shone," to create a combined interpretation, for example, "if she watered the plants AND the sun shone then they bloomed." This feature is central to the account of suppression provided by theories based on mental models (e.g., Byrne, Espino, & Santamaría, 1999), and it is shared by other accounts such as those based on logic programming (e.g., Stenning & Van Lambalgen, 2007, 2012), machine learning (e.g., Dietz et al., 2014), and formal linguistics (e.g., Cariani & Rips, 2017). In contrast, other theories propose that suppression arises because the additional condition, for example, "the sun shone" reduces the degree of belief in the first conditional "if she watered the plants they bloomed"—and this feature is common to theories based on formal inference rules (e.g., George, 1995, 1997; O'Brien, 1993; Politzer & Braine, 1991), probabilist suppositions (e.g., Over, 2017; Stevenson & Over, 1995), and causal Bayes nets (e.g., Oaksford & Chater, 2004, 2017). The inspiration for our studies came from the idea that people simulate possibilities in mental models, and we outline first the predictions we derived from this theory; we return to consider how alternative theories could account for our findings after we report the experimental results.

We illustrate the idea that the suppression of inferences arises because people construct a combined interpretation, for example, "if she watered the plants AND the sun shone then they bloomed," with reference to the mental model theory. It proposes that people understand a factual conditional such as "if she watered the plants they bloomed," by simulating the possibilities in an iconic model of how the world would be if the assertion was true (e.g., Johnson-Laird & Byrne, 2002). They construct an initial set of models that are parsimonious because of the limitations of working memory and so they think about only the mentioned events at the outset (Johnson-Laird, Byrne & Schaeken, 1992),

watered bloomed
. . .

They make a mental note that there could be other possibilities (captured by the ellipsis in the diagram), which can be "fleshed out" to be explicit (see Table 2). Accordingly, they make the *modus ponens* inference readily because the information in its premise

Table 2
Initial and explicit mental models for a factual conditional and a counterfactual (for a conditional interpretation)

	Factual Conditional		Counterfactual Conditional		
	If she watered the plants they bloomed		*If she had watered the plants they would have bloomed*		
Initial models	watered	bloomed	*Counterfactual*:	watered	bloomed
	. . .		*Facts*:	not watered	not bloomed
				. . .	
Explicit models	watered	bloomed	*Counterfactual*:	watered	bloomed
	not watered	not bloomed	*Facts*:	not watered	not bloomed
	not watered	bloomed		not watered	bloomed

"she watered the plants" corresponds to information explicitly captured in the initial representation. But to make the *modus tollens* inference, they must consider other possibilities because the information in its premise "the plants did not bloom" does not correspond to the information in the initial representation.

The "fleshed out" set of models can correspond to many different interpretations of *if* (e.g., Johnson-Laird & Byrne, 2002). One interpretation is consistent with three possibilities, as Table 2 shows, which are conjunctive—it is possible that she watered the plants and they bloomed, *and* it is possible that she did not water the plants and they did not bloom, *and* it is possible that she did not water the plants and they bloomed anyway (e.g., Khemlani, Byrne, & Johnson-Laird, 2018). They do not think about the possibility ruled out as false, that is, she watered the plants and they did not bloom (e.g., Johnson-Laird & Byrne, 2002; Espino, Santamaria, & Byrne, 2009). If they flesh out their models to consider the possibility that she did not water the plants and they did not bloom, they can make the *modus tollens* inference from the information "the plants did not bloom"—it leads to the elimination of the possibilities in which the plants bloomed, and from the remaining possibility, participants can infer "she did not water the plants." Another interpretation of *if* corresponds to a *biconditional*, consistent with just the first two possibilities in Table 2. A third interpretation corresponds to an enabling interpretation, in which the true possibilities include the first two possibilities in Table 2, and the possibility that she watered the plants and they did not bloom. Knowledge modulates the possibilities that people consider, leading to as many as 10 different interpretations of *if* (e.g., Johnson-Laird & Byrne, 2002).

For a pair of conditionals that contain an additional condition, such as "if she watered the plants they bloomed, if the sun shone they bloomed," participants can construct an initial set of models in which it is possible that she watered the plants and the sun shone and they bloomed (e.g., Byrne et al., 1999):

watered   sun   bloomed
. . .

Their knowledge ensures that they combine the antecedents conjunctively. Given that the premises refer to two conditions, people can consider the possibilities in which both conditions occur or only one occurs with the outcome,

watered	sun	bloomed
watered	not sun	
not watered	sun	
. . .		

that is, it is possible that the plants were watered and it was sunny and they bloomed, but it is possible that the plants were watered and it was not sunny, and it is possible that the plants were not watered and it was sunny. They can simulate several such possibilities but because of the constraints of working memory, their initial models may contain only some information explicitly (e.g., Barres & Johnson-Laird, 2003; Johnson-Laird, Lotstein, & Byrne, 2012),

watered	sun	bloomed
watered		
	sun	
. . .		

The implicit information can be fleshed out if needed (e.g., Johnson-Laird et al., 1992). They consider the outcomes in these different possibilities and since the additional premise conveys that the plants must have sun as well as water, they represent the plants as blooming only in the situation in which both of the conditions occur,

watered	sun	bloomed
watered		not bloomed
	sun	not bloomed
	. . .	

As a result, because the second conditional makes explicit that there is an additional background condition, they consider whether it is possible that she watered the plants but the sun did not shine and so they did not bloom.

The suppression of the *modus ponens* and *modus tollens* inferences arises because people can readily access this ready-made counterexample to the inferences (e.g., De Neys, Schaeken, & D'Ydewalle, 2005a, 2005b; Markovits et al., 2010). The additional condition is jointly necessary with the first antecedent, and so the first antecedent by itself is not sufficient for the consequent (see Table 3). Hence, people make the inferences when both

Table 3
Fully explicit models for pairs of conditionals that contain additional or alternative conditions

Additional condition					
If she watered the plants they bloomed, if the sun shone they bloomed. (If she watered the plants AND the sun shone, they bloomed).			*p q*, if *r q*		
watered	**sun**	**bloomed**	**p**	**r**	**q**
watered	not sun	**not bloomed**	**p**	not-r	**not-q**
not watered	**sun**	**not bloomed**	not-p	**r**	**not-q**
not watered	not sun	not bloomed	not-p	not-r	not-q
watered	not sun	bloomed	p	not-r	q
not watered	sun	bloomed	not-p	r	q
not watered	not sun	bloomed	not-p	not-r	q
Alternative condition					
If she watered the plants they bloomed, if it rained they bloomed. (If she watered the plants OR it rained, they bloomed).			*p q*, if *r q*		
watered	**rain**	**bloomed**	**p**	**r**	**q**
watered	not rain	**bloomed**	**p**	not-r	**q**
not watered	**rain**	**bloomed**	not-p	**r**	**q**
not watered	not rain	not bloomed	not-p	not-r	not-q
not watered	not rain	bloomed	not-p	not-r	q

Note. The initial mental models that people construct are in bold. For the additional condition, the first possibility (in which the joint antecedent is true and the outcome occurs) and the next three (in which the joint antecedent is not true and the outcome does not occur) correspond to a "biconditional" interpretation of "if p and r then q"; the addition of the final three possibilities (in which the joint antecedent is not true and yet the outcome occurs) corresponds to a conditional interpretation. For the alternative condition, the first three possibilities (in which the disjunctive antecedent is true and the outcome occurs) and the fourth possibility (in which the disjunctive antecedent is false and the outcome does not occur) correspond to a "biconditional" interpretation of "if p or r then q," the addition of the final possibility (in which the disjunctive antecedent is false and the outcome occurs) corresponds to a conditional interpretation.

antecedents are jointly presented, and they construct conclusions that sometimes refer to the other antecedent (e.g., Byrne, 1989, 1991; Byrne et al., 1999).

For completeness, Table 3 illustrates the fully fleshed-out explicit set of models which of course contains many more possibilities than the initial models. Participants may consider some of these possibilities as preferred models (e.g., Ragni & Knauff, 2013; see also Stenning & van Lambalgen, 2005), although it is unlikely they consider very many of them given working memory constraints (Espino & Byrne, 2013), and individuals differ in their ability to do so (e.g., De Neys et al., 2005a, 2005b; see also Murray & Byrne, 2005). Instead because of the limitations of working memory, participants tend to consider initially just the possibilities corresponding to the occurrence of the two conditions, or one or other of them, and the outcomes in these possibilities (highlighted in bold in the table).

An analogous explanation can be provided for a pair of conditionals that contain an alternative condition, such as "if she watered the plants they bloomed, if it rained they bloomed." Participants' knowledge ensures that they combine the antecedents disjunctively, "If she watered the plants OR it rained they bloomed" (e.g., Byrne et al., 1999). Once again, they consider the possibilities in which both conditions occur or only one occurs,

watered	rain	bloomed
watered		
	rain	
...		

Now when they consider the outcomes in these different possibilities, the alternative condition conveys that the plants can have rain instead of being watered, and so this time they represent the plants as blooming in the situations in which either or both of the conditions occur,

watered	rain	bloomed
watered		bloomed
	rain	bloomed
	...	

Accordingly, they have ready access to the possibility in which she did not water the plants but it rained and so the plants bloomed anyway.

The suppression of the *denial of the antecedent* and *affirmation of the consequent* inferences arises because people can readily access this counterexample (e.g., Dieusseurt, Schaeken, & d'Ydewalle, 2002; Markovits et al., 2010; Verschueren, Schaeken, De Neys, & d'Ydewalle, 2004). The alternative is individually sufficient for the consequent, and so the first antecedent is not necessary for the consequent (see Table 3). Accordingly, the model theory explains the suppression of inferences by additional background conditions or alternatives because the combined antecedents provide *counterexamples* to the inferences. Several other theories also emphasize the combination of the antecedents as central to the explanation of suppression (e.g., Cariani & Rips, 2017; Dietz et al., 2014; Stenning & Van Lambalgen, 2012).

Consider now the elevation of inferences for counterfactuals. The model theory proposes that people understand the counterfactual by envisaging dual possibilities (e.g., Byrne, 2005). They simulate not only the conjecture, "she watered the plants and they bloomed" but also the known or presupposed facts, "she did not water the plants and they did not bloom" (e.g., Espino & Byrne, 2018). They keep track of the epistemic status of these models, one as a conjecture that is contrary to the facts, that was once possible but is so no longer, and the other as corresponding to the presupposed or known facts (e.g., Johnson-Laird & Byrne, 1991),

Counterfactual	watered	bloomed
Fact	not watered	not bloomed
	...	

From the outset, they construct a more explicit set of initial models for a counterfactual, compared to a factual conditional, as Table 2 shows. Counterfactuals can seem to mean something very different than factual conditionals, for example, most people probably agree that if Oswald did not shoot Kennedy, then someone else did, whereas they may disagree that if Oswald had not shot Kennedy, then someone else would have (Adams, 1975). Such apparent peculiarities have led logicians to consider how the meanings of factual and counterfactual conditionals could parallel one another, and whether counterfactuals require a semantics based on "possible worlds" (e.g., Lewis, 1973; see also Stalnaker, 1968). Yet, although factual and counterfactual conditionals can seem to mean very different things, a systematic mapping can be made between them (Byrne & Johnson-Laird, 2019), as Table 2 illustrates. The models can be fleshed out to be explicit, and they result in the same set of models for a counterfactual as for a factual conditional, albeit annotated to keep track of their epistemic status.

The more explicit representation of a counterfactual ensures that the *modus tollens* and *denial of the antecedent* inferences are made more readily from the counterfactual than the factual conditional. For example, when participants are given the *modus tollens* information "the plants did not bloom," it can be integrated into the models for the counterfactual—it leads to the elimination of the first model, and from the second model participants can infer "she did not water the plants." When they are given the *modus ponens* information, "she watered the plants," it can also be integrated into the models—it leads to the elimination of the second model, the updating of the epistemic status of the first model to correspond to the facts, and from this model participants can infer "they bloomed."

This account of suppression and counterfactual elevation makes a unique prediction: Knowledge about background conditions will take precedence over knowledge about presupposed facts. Hence, inferences such as *modus tollens*, which are usually elevated for counterfactuals, will instead be suppressed from counterfactuals with additional conditions. When people understand a counterfactual such as "if she had watered the plants they would have bloomed," they construct an initial set of models that make the presupposed facts explicit, "she did not water the plants and they did not bloom,"

Counterfactual	watered	bloomed
Fact	not watered	not bloomed
	...	

But when they understand a pair of counterfactuals with additional conditions, "if she had watered the plants they would have bloomed, if the sun had shone they would have bloomed," they combine the counterfactuals conjunctively, "If she had watered the plants AND the sun had shone they would have bloomed." They construct an initial set of models that capture not only the presupposed facts but also the counterexample possibility, “she watered the plants and the sun did not shine and the plants did not bloom,” just as they do for factual conditionals,

Counterfactual	watered	sunny	bloomed
	watered		not bloomed
		sunny	not bloomed
Facts	not watered	not sunny	not bloomed
		. . .	

Given the information that the plants did not bloom, participants have access to a ready-made counterexample to the *modus tollens* conclusion that she did not water the plants (i.e., she watered the plants but it was not sunny), and so they will refrain from endorsing the conclusion, just as they do for a pair of factual conditionals. Hence, the theory’s prediction corresponds to pattern (a) outlined earlier: Knowledge about background conditions overrides knowledge about the presupposed facts. We test the theory’s predictions that inferences such as the *modus tollens* inference that are usually elevated for counterfactuals will instead be suppressed for counterfactuals with additional conditions, in the six experiments we report. We consider other theories such as those based instead on probabilities, after we report the results.

Experiments 1a and 1b test the frequency of *modus tollens* inferences in arguments based on counterfactual conditionals that contain additional conditions. Experiments 2a and 2b test the frequency of *denial of the antecedent* inferences in arguments based on counterfactual conditionals that instead contain alternative conditions. Experiments 3a and 3b examine *switched-suppression* by testing for the first time the suppression effect for conditionals that contain negation, such “if she did not water the plants they did not bloom,” for factual conditionals and counterfactuals.

4. Experiment 1a: Suppression of *modus tollens* for counterfactuals

The aim of the experiment was to test whether *modus tollens* is suppressed for counterfactuals that contain additional conditions. Our interest is in arguments based on counterfactuals with additional conditions, for example,

If Lisa had had an essay to write she would have studied late in the library.
If the library had stayed open she would have studied late in the library.
Lisa did not study late in the library.
Therefore:
(a) Lisa had an essay to write.
(b) Lisa did not have an essay to write.
(c) Lisa may or may not have had an essay to write.

As control comparisons, we compared them to arguments based on a single counterfactual, for example,

If Lisa had had an essay to write she would have studied late in the library.
Lisa did not study late in the library.

and also to arguments based on factual conditionals with additional conditions, for example,

If Lisa had an essay to write she studied late in the library.
If the library stayed open she studied late in the library.
Lisa did not study late in the library.

and to arguments based on a single factual conditional, for example,

If Lisa had an essay to write she studied late in the library.
Lisa did not study late in the library.

The control comparisons allow us to test that we replicate (a) the standard suppression effect for factual conditionals, that is, participants will make fewer *modus tollens* inferences from factual conditionals with additional conditions compared to single factual conditionals; and (b) the standard counterfactual inference effect, that is, participants will make more *modus tollens* inferences from single counterfactuals compared to single factual conditionals. The main aim of the experiment was to test the model theory's prediction that the *modus tollens* inference will be suppressed for counterfactuals with additional conditions. Our focus is on the *modus tollens* inference because its frequency is affected differently by counterfactuals and by additional conditions.

4.1. Methods

4.1.1. Participants

The 120 participants who took part in the experiment were undergraduate students from the University of La Laguna, Tenerife, Spain and they were drawn from a sample of students comprised in general of two-thirds women and one-third men within the age range of 18–24 years. None of them had formal training in logic nor had they taken part in an experiment on reasoning before. For all the studies, the participants gave their informed consent, and we report all of our manipulations and measures, and each study's sample size was determined prior to data collection based on effect sizes in the literature. The participants were randomly assigned to one of four groups: counterfactuals with additional conditions, single counterfactuals, factual conditionals with additional conditions, and single factual conditionals ($n = 30$ in each group).

4.1.2. Design and materials

The design was a between-participants one with four groups of participants who received four different sorts of arguments, as described earlier: single factual conditional arguments, single counterfactual arguments, additional condition factual conditional arguments, and additional condition counterfactual arguments. Participants in all groups carried out three instances of the *modus tollens* inference, based on three different contents (adapted from Byrne et al., 1999; see Supplemental Materials).

For completeness and to ensure participants attempted to reason about each inference, they completed not only the *modus tollens* inference but also the three other sorts of conditional inferences, *modus ponens*, *denial of the antecedent*, and *affirmation of the consequent*. We included three instances of each inference based on the three contents, that is, 12 in total. We also included eight fillers based on a second conditional that contained an alternative rather than an additional condition (in the additional condition group), with two instances of each of the four sorts of inferences; we matched these fillers in the single arguments conditions with two arguments based on different content (see Supplemental Materials). The 20 arguments were presented in a different randomized order to each participant. They were presented to the participants in their native language of Spanish. The participants' task was to choose a conclusion from a set of three conclusions, for example, "(a) She had an essay to write, (b), She did not have an essay to write, (c) She may or may not have had an essay to write."

4.1.3. Procedure

Participants were tested in small groups, and they were given the arguments in a single booklet. The instructions printed on the front page explained the task with reference to a single indicative conditional as an example. In line with previous studies of suppression, participants were asked to assume that the premises were true and to "choose one of the conclusions, whichever one you think follows from the premises." Participants were asked to read each item carefully and to work from beginning to end at their own pace without changing any responses or skipping any items.

4.2. Results and discussion

The dataset for this experiment and the subsequent ones is available at https://osf.io/z4nj9/ and www.reasoningandimagination.com. The results show that the *modus tollens* inference was suppressed for counterfactuals that contained additional conditions, as Fig. 1 illustrates.

We carried out a 2 (argument type: single vs. additional) $\times$ 2 (conditional type: factual vs. counterfactual) $\times$ 4 (inference type: *modus tollens, modus ponens, denial of the*

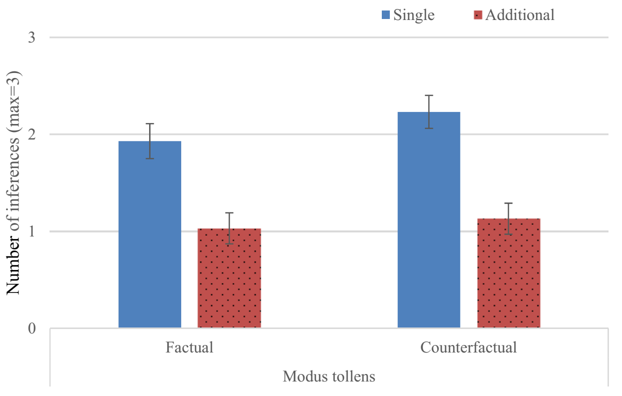


Fig. 1. The mean frequencies of *modus tollens* inferences for factual and counterfactual conditionals in single and additional condition arguments in Experiment 1a. Error bars are standard error of the mean.

antecedent, and *affirmation of the consequent*) ANOVA with repeated measures on the third factor, on the mean frequencies of the inferential conclusions that participants chose (see Table 4), using a Greenhouse–Geisser correction for the violation of sphericity assumption. There was a main effect of inference type indicating differences in the frequency of *modus tollens, modus ponens, denial of the antecedent,* and *affirmation of the consequent,* $F(2.489, 288.675) = 13.14$, $MSE = 0.72$, $p < .001$, $\eta_p^2 = .10$. There was a main effect of argument type, as participants made more inferences to the single arguments than the additional ones, $F(1, 116) = 28.41$, $MSE = 1.65$, $p < .001$, $\eta_p^2 = .20$; and no main effect of conditional type, $F < 1$. Inference interacted with argument type, $F(2.489, 288.675) = 4.3$, $MSE = 0.72$, $p < .01$, $\eta_p^2 = .35$, and conditional type, $F(2.489, 288.675) = 4.2$, $MSE = 0.72$, $p < .01$, $\eta_p^2 = .35$; argument and conditional type did not interact and the three-way interaction was not reliable, $F < 1$ in both cases (see Supplemental Materials for more details).

To test our hypotheses about *modus tollens*, we decomposed the nonsignificant three-way interaction, with a Bonferroni corrected alpha of $p < .0167$ for three comparisons (see Winer, 1971, for a defense of this strategy in such circumstances). *Modus tollens* was suppressed for counterfactuals with additional conditions compared to single counterfactuals, $F(1, 58) = 22.47$, $MSE = 0.81$, $p < .001$, $\eta_p^2 = .28$, as Fig. 1 shows. It was suppressed for factual conditionals with additional conditions compared to single factual conditionals, $F(1, 58) = 13.86$, $MSE = 0.88$, $p < .001$, $\eta_p^2 = .19$, replicating previous findings of a suppression effect for factual conditionals. Given the nature of the competing tendencies, it is unsurprising that suppression for counterfactuals is no greater than suppression for factual conditionals. *Modus tollens* was somewhat elevated for single counterfactuals compared to single factual conditionals, although the difference did not reach significance, $F(1, 58) = 1.47$, $MSE = 0.92$, $p = .23$, $\eta_p^2 = .03$, perhaps because of its already high rate for single factual conditionals. For completeness, we also computed such comparisons for the other three sorts of inference (for which there is no conflict between additional conditions and presupposed facts); see Supplemental Materials.

Table 4
The frequencies of inferences (with *SE*s in parentheses) in Experiments 1a, 2a and 2b, and 3a and 3b

	Modus Tollens	Deny Antecedent	Modus Ponens	Affirm Consequent
Premises	*If A, C*	*If A, C*	*If A, C*	*If A, C*
	(if B, C)	*(If B, C)*	*(if B, C)*	*(If B, C)*
Minor premise:	*not-C*	*not-A*	*A*	*C*
Conclusions:	*not-A*	*not-C*	*C*	*A*
Experiment 1a				
Factual Single	1.93 (.18)	2.30 (.17)	2.53 (.12)	2.3 (.19)
Counterfactual Single	2.23 (.17)	2.37 (.15)	1.97 (.19)	2.33 (.18)
Factual Additional	1.03 (.16)	1.83 (.17)	1.73 (.19)	1.90 (.17)
Counterfactual Additional	1.13 (.16)	1.83 (.16)	1.4 (.16)	2.1 (.18)
Experiment 2a				
Counterfactual Single	2.71 (.13)	2.67 (.12)	2.63 (.16)	2.46 (.17)
Counterfactual Additional	1.58 (.18)	2.29 (.22)	1.71 (.2)	1.83 (.21)
Counterfactual Alternative	2.04 (.19)	1.0 (.24)	2.75 (.11)	0.96 (.19)
Experiment 2b				
Counterfactual Single	2.1 (.19)	2.33 (.17)	2.57 (.09)	2.23 (.19)
Counterfactual Additional	1.3 (.17)	2.13 (.15)	1.5 (.21)	2.13 (.17)
Counterfactual Alternative	1.9 (.19)	.73 (.18)	2.83 (.11)	1.03 (.22)
Premises	*If not-A, not-C*	*If not-A, not-C*	*If not-A, not-C*	*If not-A, not-C*
	(if not-B, not-C)	*(if not-B, not-C)*	*(if not-B, not-C)*	*(if not-B, not-C))*
Minor premise:	*C*	*A*	*not-A*	*not-C*
Conclusions:	*A*	*C*	*not-C*	*not-A*
Experiment 3a				
Factual Single	1.88 (.22)	1.75 (.23)	2.5 (.13)	2.17 (.17)
Factual Additional	2.58 (.15)	1.25 (.24)	2.25 (.17)	1.33 (.21)
Factual Alternative	1.41 (.23)	2.33 (.18)	1.67 (.24)	2.29 (.19)
Experiment 3b				
Counterfactual Single	2.75 (.09)	2.42 (.19)	2.63 (.16)	2.38 (.21)
Counterfactual Additional	2.13 (.17)	1.63 (.21)	1.75 (.21)	1.42 (.23)
Counterfactual Alternative	1.21 (.23)	2.67 (.12)	.92 (.18)	2.13 (.15)

The experiment shows that *modus tollens* is suppressed for counterfactuals with additional conditions compared to single counterfactuals. The finding shows that knowledge about additional background conditions takes precedence over knowledge about presupposed facts. The result corroborates the proposal that people mentally simulate the possibilities that the conditionals refer to, in a conjunctive representation. In the next experiment, we aim to replicate the effect in a more demanding task to rule out the possibility that the results merely reflect participants carrying out the task in a superficial manner.

5. Experiment 1b. Counterfactual suppression with inference generation

The aim of the experiment was to replicate the suppression of *modus tollens* inferences for counterfactuals with additional conditions compared to single counterfactuals, this time using the more exacting construction task in which participants generate their own conclusion (adapted from Byrne et al., 1999). We gave them the same premises but asked them to say what, if anything, followed from them, for example,

If Lisa had had essay to write she would have studied late in the library.
If the library had stayed open she would have studied late in the library.
Lisa did not study late in the library.
What, if anything, follows?

We focused on the comparison of counterfactuals with additional conditions to single counterfactuals in this experiment.

5.1. Methods

5.1.1. Participants and procedure

The 48 participants who took part in the experiment were a new set of undergraduate students from the University of La Laguna, Tenerife, Spain and they had not previously participated in any experiments of this sort. They were randomly assigned to one of two groups, counterfactuals with additional conditions and single counterfactuals ($n = 24$ in each group). The procedure was the same as the previous experiment with the exception that the participants generated their own conclusions rather than selecting a conclusion from a presented set.

5.1.2. Design and materials

We used two of the argument types from the previous experiment, counterfactuals with additional conditions and single counterfactuals, in a between-participants design. The participants' task this time was to construct a conclusion in response to the question "What, if anything, follows?" They carried out a *modus tollens* inference based on a minor premise that denied the consequent shared by both counterfactuals (e.g., "Lisa did not study late in the library"). As fillers, they also completed two sorts of *modus ponens* inferences, *modus ponens-1* based on a minor premise that affirmed the antecedent of the first counterfactual, for example, "Lisa had an essay to write," and *modus ponens-2* based on a minor premise that affirmed the antecedent of the second counterfactual, "the library stayed open." They carried out three instances of each inference and so they completed nine inferences in total. We used the same materials as the previous experiment and assigned the three contents at random to each of the three inference forms to construct a different set of nine problems for each participant (see Supplemental Materials). The matched inferences for the single counterfactual were three *modus tollens* inferences and six *modus ponens* inferences. The problems were presented to the participants in their

Table 5
Mean frequencies of different sorts of conclusions to the *modus tollens* inference for counterfactuals with additional conditions and single counterfactuals in Experiment 1b

	Additional	Single
Premises	If A, C	If A, C
Given:	(if B, C)	not-C
	not-C	
Categorical endorsements		
not-A	**.42 (.12)**	**2.88 (.07)**
not-B	**.46 (.12)**	—
A	.04 (.04)	—
not-C	.33 (.17)	—
Modal endorsements		
C can or C cannot	—	.12 (.07)
Answers referring to more than one component		
Cannot A or cannot-B	.46 (.19)	—
Not A or not-B	.75 (.23)	—
Not-A and not-B	.42 (.16)	—
A and not-B	.04 (.04)	—
Adding further information	.08 (.06)	—
Total	*3*	*3*

Note. The categorical inferential endorsements are in bold; see also Table S1 in Supplemental Materials.

native Spanish. We gave each participant the nine arguments in a different randomized order.

5.2. Results and discussion

Participants produced a rich variety of conclusions to the counterfactuals with additional conditions, as Table 5 shows. The classification of their conclusions was carried out according to some simple principles (adapted from Byrne et al., 1999). Conclusions were categorized according to (a) which component was mentioned, (b) whether it contained modal auxiliaries such as “may,” “might,” and “maybe” to qualify the conclusion, and (c) whether it mentioned a single component or more than one and, if so, which connectives were used, and (d) whether the participant concluded there was not enough information or added other information, as Table 5 shows. Some conclusions were categorical endorsements of the *modus tollens* inference from counterfactuals with additional conditions, for example, given “Lisa did not study late in the library,” participants concluded “Lisa did not have an essay to write” or “the library did not stay open.” Most of the remaining conclusions were answers referring to both antecedents (for further information, see Table S1 in the Supplemental Materials).

As predicted, the *modus tollens* inference was suppressed from counterfactuals with additional conditions even when participants constructed their own conclusions. We carried out a 2 (counterfactuals with additional conditions vs. single counterfactuals) × 3 (*modus tollens, modus ponens-1, modus ponens-2*) ANOVA with repeated measures on the second factor, on the frequency of categorical inferential responses (e.g., for *modus tollens*, given, "if A then C, if B then C," and "not-C," a categorical inferential conclusion is "not-A," or "not-B"). It showed a main effect of argument type, $F(1, 46) = 133.174$, $MSE = 0.92$, $p < .001$, $\eta_p^2 = .74$, indicating a difference in the frequency of inferences made from counterfactuals with additional conditions and single counterfactuals; no main effect of type of inference, $F(2, 92) = 2.63$, $MSE = 0.90$, $p = .08$, $\eta_p^2 = .05$, indicating no differences in the overall frequency of *modus tollens*, *modus ponens-1*, or *modus ponens −2*, and a significant interaction, $F(2, 92) = 4.19$, $MSE = 0.50$, $p < .02$, $\eta_p^2 = .08$.

To test our hypotheses, we decomposed the interaction using a single pairwise comparison, which showed that *modus tollens* inferences were suppressed from counterfactuals with additional conditions compared to single counterfactuals, $F(1, 46) = 87.44$, $MSE = 0.55$, $p < .001$, $\eta_p^2 = .66$. Of the 48 participants, 30 gave mostly categorical answers (i.e., to six or more of the nine trials), 7 gave mostly answers referring to more than one component, and the remaining 11 gave a mix of categorical and compound answers.

The experiment shows that *modus tollens* is suppressed from counterfactuals with additional conditions even when participants construct their own conclusions, and so its suppression cannot be attributed to participants engaging in a superficial way with the task. We turn now to examine the suppression of the *denial of the antecedent* inference in counterfactuals with alternative conditions.

6. Experiment 2a: Suppression of *denial of antecedent* for counterfactual alternatives

The aim of the experiment was to test whether the *denial of the antecedent* inference is suppressed for counterfactuals with *alternative* conditions. Our focus is on the *denial of the antecedent* inference because it is the only inference whose frequency is increased by counterfactuals but decreased by *alternative* conditions. When participants are given a single counterfactual argument, such as,

If Lisa had had an essay to write she would have studied late in the library.
Lisa did not have an essay to write.

most of them tend to make the *denial of the antecedent* inference, "Lisa did not study late in the library," more often than they do for a single factual conditional (e.g., Byrne & Tasso, 1999). In contrast, they make *fewer* of these inferences from factual conditionals that contain *alternatives* compared to single conditionals (e.g., Rumain et al., 1983). We examined whether the inference is suppressed for counterfactuals with *alternatives*,

If Lisa had had an essay to write she would have studied late in the library.
If Lisa had had a textbook to read she would have studied late in the library.
Lisa did not have an essay to write.
Therefore:
(a) Lisa studied late in the library.
(b) Lisa did not study late in the library.
(c) Lisa may or may not have studied late in the library.

We compared the *denial of the antecedent* inference in counterfactuals with alternatives, to single counterfactual arguments. However, given the high rate of all inferences to the single counterfactual in the first experiment, we considered that an extra, and more stringent, test of suppression was to compare the frequency of the inference for counterfactuals with alternatives to counterfactuals with additional conditions, which we also included.

6.1. Methods

6.1.1. Participants and procedure

The 72 participants who took part in the experiment were a new set of undergraduate students from the University of La Laguna, Tenerife, Spain, who had not participated in previous experiments of this sort. They were randomly assigned to one of three groups (n = 24 participants in each). The procedure was the same as the previous experiments.

6.1.2. Design and materials

We gave participants arguments based on single counterfactuals, and on counterfactuals with additional conditions similar to the previous experiments, and also arguments based on counterfactuals with alternatives. Participants received only one sort of argument in a between-participant design and so there were three groups, and all of the arguments were based on counterfactuals. The materials were similar to Experiment 1a and their task was to choose a conclusion from a set of three conclusions. Even though our focus is on the *denial of the antecedent* inference, we once again ensured that participants also completed the *modus tollens, modus ponens*, and *affirmation of the consequent* inferences as fillers. We used the same materials as the previous experiments. Participants carried out three instances with three different contents for each of the four sorts of inferences and so they completed 12 inferences in total, along with 8 other filler arguments in each condition, as before. They completed the 20 inferences in a different randomized order for each participant. The problems were presented to the participants in their native Spanish.

6.2. Results and discussion

The *denial of the antecedent* inference was suppressed for counterfactuals with alternative conditions, as Fig. 2 shows. We carried out a 3 (argument type: single

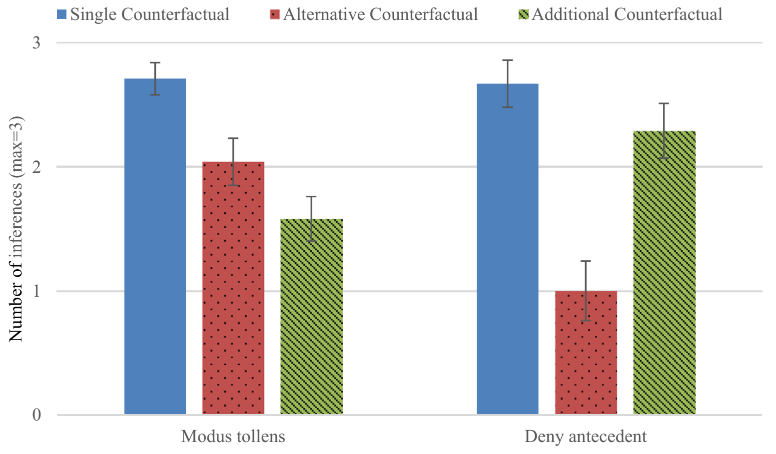


Fig. 2. The mean frequencies of *modus tollens* and *denial of the antecedent* inferences for counterfactual conditionals in single, alternative, or additional conditions arguments in Experiment 2a. Error bars are standard error of the mean.

counterfactual, counterfactuals with alternative conditions, counterfactuals with additional conditions) × 4 (inference type: *denial of the antecedent, modus tollens, modus ponens,* and *affirmation of the consequent*) ANOVA with repeated measures on the second factor, on the conclusions that participants selected, using a Greenhouse–Geisser correction for the violation of sphericity assumption. There was a main effect of argument type, indicating differences in the frequency of inferences made to single, additional, and alternative arguments, $F(2, 69) = 16.92$, $MSE = 1.38$, $p < .001$, $\eta_p^2 = .32$, as Table 4 shows; a main effect of inference type indicating differences in the frequency of *denial of the antecedent*, *modus tollens*, *modus ponens*, and *affirmation of the consequent*, $F(2.501, 172.479) = 8.06$, $MSE = 0.70$, $p < .001$, $\eta_p^2 = .10$; and the two variables interacted, $F(4.999, 172.560) = 13.80$, $MSE = 0.70$, $p < .001$, $\eta_p^2 = .28$.

We decomposed the interaction to test our hypotheses about the *denial of the antecedent* inference using pairwise comparisons with a Bonferroni corrected alpha of $p < .0167$ for three comparisons. Participants made fewer *denial of the antecedent* inferences from counterfactuals with alternatives compared to single counterfactuals, $F(1, 46) = 38.99$, $MSE = 0.86$, $p < .001$, $\eta_p^2 = .46$, and also for the more stringent comparison to counterfactuals with additional conditions, $F(1, 46) = 15.62$, $MSE = 1.28$, $p < .001$, $\eta_p^2 = .25$, as Fig. 2 shows. As expected, there was no difference between *denial of the antecedent* inferences from counterfactuals with additional conditions and single counterfactuals, $F(1, 46) = 2.26$, $MSE = 0.75$, $p = .13$, $\eta_p^2 = .05$. For completeness, we also computed such comparisons for the other three inferences (see Supplemental Materials). We note here that participants made fewer *modus tollens* inferences from counterfactuals with additional conditions compared to single counterfactuals, replicating the previous experiments, $F(1, 46) = 26.08$, $MSE = 0.58$, $p < .001$, $\eta_p^2 = .36$, as Fig. 2 shows. They also made fewer *modus tollens* inferences even from counterfactuals with alternative conditions compared to single counterfactuals, $F(1, 46) = 8.79$, $MSE = 0.61$, $p < .006$, $\eta_p^2 = .16$, perhaps because the provision of an alternative can in some instances lead participants to generate additional conditions too (see Markovits & Potvin, 2001).

The experiment shows that the *denial of the antecedent* inference is suppressed for counterfactuals with alternatives compared to single counterfactuals. In the next experiment, we aimed to replicate the finding in arguments that contained subjunctive minor premises and subjunctive conclusions to rule out the possibility that the results merely reflect some confusion about the counterfactual nature of the information.

7. Experiment 2b: Counterfactual suppression and subjunctive conclusions

The aim of the experiment was to test whether the inferences are suppressed for counterfactuals, even when the minor premise and conclusion are phrased as counterfactual. We gave participants *denial of the antecedent* inferences for counterfactuals with alternatives, this time with subjunctive minor premises and subjunctive conclusions, such as,

If Lisa had had an essay to write she would have studied late in the library.
If Lisa had had a textbook to read she would have studied late in the library.
Suppose Lisa had not had an essay to write.
Therefore:
(a) Lisa would have studied late in the library.
(b) Lisa would not have studied late in the library.
(c) It is not possible to know whether Lisa would have studied late in the library or would not have studied late in the library.

We also gave them *modus tollens* inferences for counterfactuals with additional conditions, and subjunctive minor premises and conclusions. The use of the subjunctive mood in the minor premise and conclusion allows us to rule out the possibility that participants are responding merely on the basis of prior beliefs about the likely facts, rather than on the basis of their combination of the antecedents for the pair of counterfactuals, and their presupposed facts.

7.1. Methods

7.1.1. Participants and procedure

The 90 participants who took part in the experiment were undergraduate students from the University of La Laguna, Tenerife, Spain, and they had not previously participated in any experiments of this sort. They were randomly assigned to one of three groups ($n = 30$ in each group). The procedure was the same as the previous experiments.

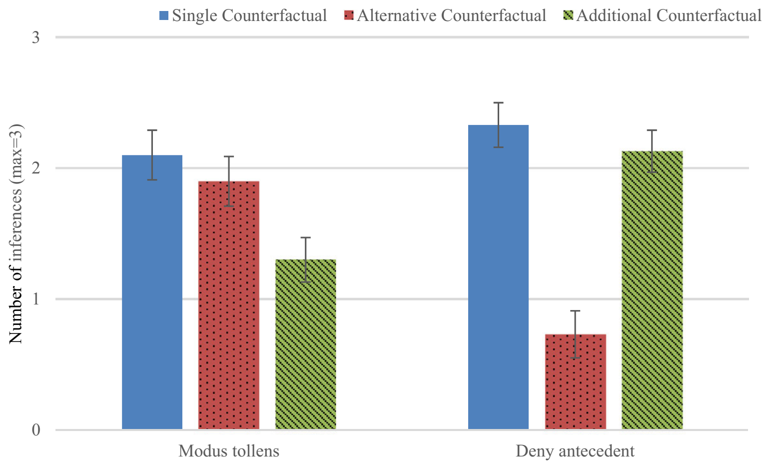


Fig. 3. The mean frequencies of *modus tollens* and *denial of the antecedent* inferences for counterfactual conditionals in single, alternative, or additional conditions arguments in Experiment 2b. Error bars are standard error of the mean.

7.1.2. Design and materials

The design and materials were the same as the previous experiment, with the exception that the arguments contained a subjunctive minor premise and a subjunctive conclusion.

7.2. Results and discussion

The *denial of the antecedent* inference was suppressed by counterfactuals with alternatives and the *modus tollens* inference was suppressed by counterfactuals with additional conditions, even for arguments with minor premises and conclusions that were phrased as counterfactual, as Fig. 3 shows. We carried out the same sort of analysis as in the previous experiment. There was a main effect of argument type, indicating differences in the frequency of inferences made to the single, additional, and alternative arguments, $F(2, 87) = 10.55$, $MSE = 1.48$, $p < .001$, $\eta_p^2 = .19$; a main effect of inference type indicating differences in the frequency of the *denial of the antecedent*, *modus tollens, modus ponens,* and *affirmation of the consequent* inferences, $F(2.675, 232.729) = 8.98$, $MSE = 0.80$, $p < .001$, $\eta_p^2 = .09$, and an interaction of the two variables, $F(5.350, 232.729) = 18.81$, $MSE = 0.80$, $p < .001$, $\eta_p^2 = .30$.

We decomposed the interaction to test our hypotheses about the *denial of the antecedent* and *modus tollens* inferences using pairwise comparisons with a Bonferroni corrected alpha of $p < .008$ for six comparisons. For the *denial of the antecedent*, participants made fewer inferences from counterfactuals with alternatives compared to single counterfactuals, $F(1, 59) = 42.39$, $MSE = 0.90$, $p < .001$, $\eta_p^2 = .42$, and on the more stringent comparison, to counterfactuals with additional conditions, $F(1, 59) = 34.56$, $MSE = 0.85$, $p < .001$, $\eta_p^2 = .37$, as Fig. 3 shows. As expected, there was no difference for *denial of the antecedent* inferences from counterfactuals with additional conditions and single counterfactuals, $F(1, 59) = .75$, $MSE = 0.79$, $p = .39$, $\eta_p^2 = .01$.

Participants made fewer *modus tollens* inferences from counterfactuals with additional conditions compared to single counterfactuals, $F(1, 59) = 9.43$, $MSE = 1.01$, $p < .004$, $\eta_p^2 = .14$, and on the more stringent comparison, to counterfactuals with alternatives, $F(1, 59) = 5.30$, $MSE = 1.01$, $p < .03$, $\eta_p^2 = .08$, although the latter difference is not significant on the adjusted alpha. As expected (and unlike the previous experiment), there was no difference for *modus tollens* inferences from counterfactuals with alternative conditions and single counterfactuals, $F(1, 59) = .53$, $MSE = 1.12$, $p = .46$, $\eta_p^2 = .01$. (For further information, see Supplemental Materials.)

The experiment shows that the inferences are suppressed, even for arguments with minor premises and conclusions phrased as counterfactual. It shows that the results do not reflect some confusion about the counterfactual status of the information, nor do they merely reflect inferences made on the basis of the likely facts. The next two experiments extend the findings of counterfactual suppression to the interesting case of conditionals with negated components.

8. Experiment 3a: Reversal of suppression in negated conditionals

The aim of the experiment was to examine a novel case in which the model theory predicts that the typical pattern of suppression should be switched. We examined pairs of conditionals that contained negated alternatives, such as,

If Lisa did not have an essay to write she did not study late in the library
if Lisa did not have a textbook to read she did not study late in the library

and pairs that contained negated additional conditions, such as,

Table 6
Examples of structurally equivalent inferences from affirmative and negative conditionals

	Affirmative	Negative
Antecedent	A: Lisa had an essay to write	A: Lisa did not have an essay to write
Consequent	C: She studied late in the library	C: She did not study late in the library
Conditional		
If A	If Lisa had an essay to write	If Lisa did not have an essay to write
then C	then she studied late in the library	then she did not study late in the library
Deny antecedent		
not-A	Lisa did not have an essay to write	Lisa had an essay to write
therefore not-C	She did not study late in the library	She studied late in the library
Modus tollens		
not-C	Lisa did not study late in the library	Lisa studied late in the library
therefore not-A	She did not have an essay to write	She had an essay to write

If Lisa did not have an essay to write she did not study late in the library
if the library did not stay open then she did not study late in the library.

We note first that the presence of negation in the premises does not change the underlying logical structure of an inference (e.g., Evans, 2007). For an affirmative conditional "if Lisa had an essay to write…," denying its antecedent is "Lisa did not have an essay to write"; but for a negative conditional, "if Lisa did not have an essay to write…," denying its antecedent is "Lisa had an essay to write." The inference remains the *denial of the antecedent* one regardless of the polarity of the assertions, as Table 6 shows.

Affirmative alternatives such as "if Lisa had an essay to write she studied late in the library, if she had a textbook to read she studied late in the library" suppress the *denial of the antecedent* inference, from "she did not have an essay to write" to "she did not study late in the library." The affirmative antecedents are independently sufficient for the outcome, and we have proposed that people mentally simulate them in a disjunctive combination, "if Lisa had an essay to write OR she had a book to read then she studied late in the library" (see, e.g., Byrne et al., 1999). Our new proposal is that in contrast, negated alternatives are jointly necessary for the outcome, and so people will mentally simulate them in a *conjunctive* combination, "if Lisa did not have an essay to write AND she did not have a textbook to read then she did not study late in the library." The proposal is somewhat analogous to De Morgen's theorem, that is, the negation of a disjunction (it is not the case that *p* or *q*) results in the conjunction of its negated components (not-*p* and not-*q*). As a result, we make the novel prediction that conditionals with *negated alternatives* will suppress the *modus tollens* inference (from "she studied late in the library" to "she had an essay to write"), rather than the *denial of the antecedent* inference (from "she had an essay to write" to "she studied late in the library").

Similarly, affirmative additional conditions such as "if Lisa had an essay to write she studied late in the library, if the library stayed open she studied late in the library" suppress the *modus tollens* inference, from "she did not study late in the library" to "she did not have an essay to write." The affirmative additional conditions are jointly necessary for the outcome, and we have proposed that people mentally simulate them in a conjunctive combination, "if Lisa had an essay to write AND the library stayed open then she studied late in the library." Our new proposal is that in contrast, negated additional conditions are independently sufficient for the outcome and so people will mentally simulate them in a *disjunctive* combination, "if Lisa did not have an essay to write OR the library did not stay open then she did not study late in the library." As a result, we make the novel prediction that conditionals with *negated additional conditions* will suppress the *denial of the antecedent* inference, rather than the *modus tollens* inference. We test the predictions of a "switched-suppression" effect for negated factual conditionals in Experiment 3a, and then for counterfactuals in Experiment 3b.

8.1. Methods

8.1.1. Participants and procedure

The 72 participants who took part in the experiment were undergraduate students from the University of La Laguna, Tenerife, Spain, and they had not previously participated in any experiments of this sort. They were randomly assigned to one of three groups (n = 24 in each group). The procedure was the same as the previous experiments.

8.1.2. Design and materials

The design and materials were the same as the previous experiments, with the difference that the arguments contained negated factual conditionals (and the minor premise and conclusion were in the indicative mood).

8.1.3. Results and discussion

As predicted, *modus tollens* inferences were suppressed for factual conditionals with negated alternatives, and *denial of the antecedent* inferences were suppressed for factual conditionals with negated additional conditions, as Fig. 4 shows. We carried out the same sort of analysis as in the previous experiment. There was no main effect of argument type, $F(2, 69) = .73$, $MSE = 1.61$, $p = .48$, $\eta_p^2 = .02$, showing no difference between negated single factual conditionals, negated additional conditions, or negated alternative conditions; nor of inference type, $F(3, 207) = 2.22$, $MSE = .70$, $p = .09$, $\eta_p^2 = .03$, showing no difference between *denial of the antecedent*, *modus tollens*, *modus ponens*, or *affirmation of the consequent*, but the two factors interacted, $F(6, 207) = 11.79$, $MSE = 0.70$, $p < .001$, $\eta_p^2 = .25$.

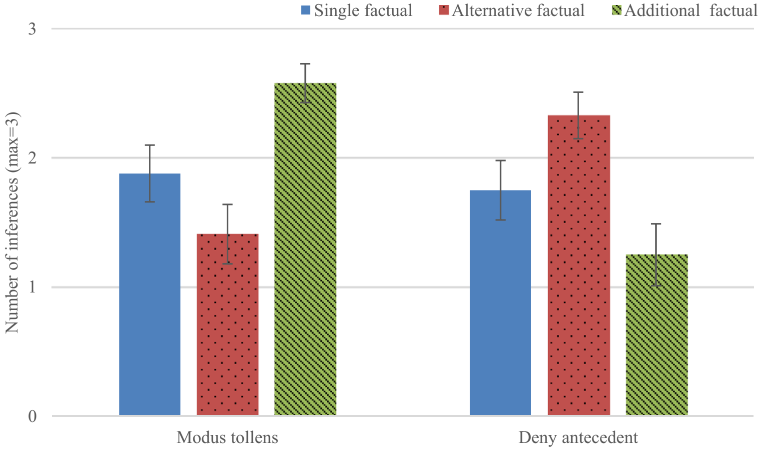


Fig. 4. The mean frequencies of *denial of the antecedent* and *modus tollens* inferences for factual negated conditionals in single, alternative, or additional condition arguments in Experiment 3a. Error bars are standard error of the mean.

We decomposed the interaction to test our hypotheses about the *modus tollens* and *denial of the antecedent* inferences using pairwise comparisons with a Bonferroni corrected alpha of $p < .008$ for six comparisons. Participants made fewer *modus tollens* inferences from factual conditionals with negated *alternative* conditions compared to factual conditionals with negated additional conditions, $F(1, 46) = 18.94$, $MSE = 0.86$, $p < .001$, $\eta_p^2 = .29$, although the comparison to single factual conditionals was not significant, $F(1, 46) = 2.12$, $MSE = 1.18$, $p = .15$, $\eta_p^2 = .05$. As expected, there was no difference between single factual conditionals and factual conditionals with negated additional conditions, $F(1, 46) = 7.20$, $MSE = 0.84$, $p < .015$, $\eta_p^2 = .14$, on the adjusted alpha. It is important to note that a similar pattern of suppression was observed for the *modus ponens* inference (see Supplemental Materials).

In contrast, participants made fewer *denial of the antecedent* inferences from factual conditionals with negated *additional* conditions compared to factual conditionals with negated alternatives, $F(1, 46) = 13.54$, $MSE = 1.04$, $p < .001$, $\eta_p^2 = .23$, although the comparison to single factual conditionals, was not significant, $F(1, 46) = 2.34$, $MSE = 1.28$, $p = .13$, $\eta_p^2 = .05$. As expected, there was no difference between single factuals and factual conditionals with negated alternatives, $F(1, 46) = 4.09$, $MSE = 0.99$, $p < .05$, $\eta_p^2 = .08$, on the adjusted alpha. A similar pattern of suppression was observed for the *affirmation of the consequent* inference (see Supplemental Materials).

The experiment reveals a novel "switched-suppression" effect for negated factual conditionals. Pairs of conditionals with negated alternatives suppress the *modus tollens* inferences, and those with negated additional conditions suppress the *denial of the antecedent* inferences, the reverse of what is observed for affirmative conditionals. In the next experiment, we examine the suppression of inferences for negated counterfactuals.

9. Experiment 3b: Suppression in negated counterfactuals

The aim of the experiment was to test the suppression of inferences in negated counterfactuals, for example, for the *modus tollens* inference from negated alternatives,

If Lisa had not had an essay to write she would not have studied late in the library.
If Lisa had not had a textbook to read she would not have studied late in the library.
Lisa studied late in the library.
Therefore:
(a) Lisa had an essay to write.
(b) Lisa did not have an essay to write.
(c) Lisa may or may not have had an essay to write.

and the *denial of the antecedent* inference for negated additional conditions,

If Lisa had not had an essay to write she would not have studied late in the library.
If the library had not stayed open she would not have studied late in the library.
Lisa had an essay to write.

We expect to replicate the switched-suppression effect of the previous experiment but this time with negated counterfactuals.

9.1. Methods

9.1.1. Participants and procedure

The 72 participants who took part in the experiment were undergraduate students from the University of La Laguna, Tenerife, Spain, and they had not previously participated in any experiments of this sort. They were randomly assigned to one of the three groups ($n = 24$ in each group). The procedure was the same as the previous experiments.

9.1.2. Design and materials

The design and materials were the same as the previous experiment, with the only difference that the arguments contained negated counterfactuals.

9.1.3. Results and discussion

Modus tollens inferences were suppressed by negated counterfactuals with *alternative* conditions, and *denial of the antecedent* inferences were suppressed by negated counterfactuals with *additional* conditions, as Fig. 5 shows. We carried out the same sort of analysis as in the previous experiment. There was a main effect of argument type, indicating differences in the frequency of inferences made to negated single counterfactuals,

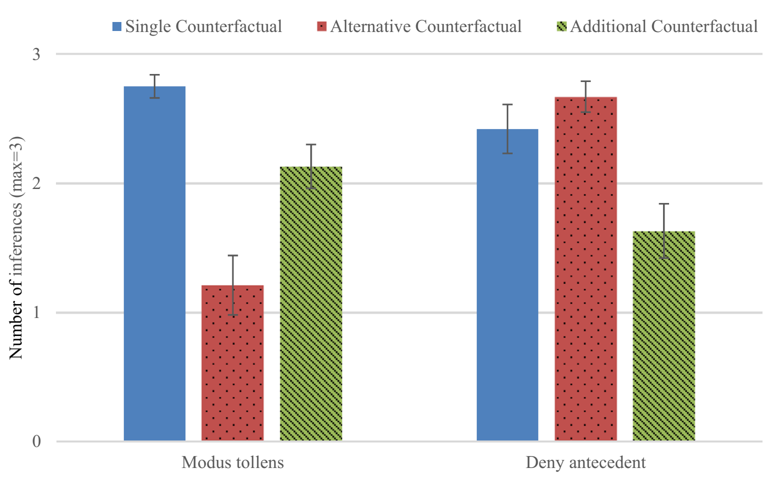


Fig. 5. The mean frequencies of *modus tollens* and *denial of the antecedent* inferences for negated counterfactual conditionals in single, alternative, or additional conditions arguments in Experiment 3b. Error bars are standard error of the mean.

negated additional conditions, and negated alternatives, $F(2, 69) = 19.90$, $MSE = 1.06$, $p < .001$, $\eta_p^2 = .36$; a main effect of inference type indicating differences in the frequency of *modus tollens, denial of the antecedent, modus ponens,* and *affirmation of consequent,* $F(2.424, 167.233) = 3.72$, $MSE = 0.90$, $p < .02$, $\eta_p^2 = .05$, and the two variables interacted, $F(4.847, 167.233) = 10.92$, $MSE = 0.90$, $p < .001$, $\eta_p^2 = .24$, as Fig. 5 shows.

We decomposed the interaction to test our hypotheses about the *modus tollens* and *denial of the antecedent* inferences using pairwise comparisons with a Bonferroni corrected alpha of $p < .008$ for six comparisons. For *modus tollens*, participants made fewer inferences from counterfactuals with negated alternatives compared to single counterfactuals, $F(1, 46) = 38.07$, $MSE = 0.75$, $p < .001$, $\eta_p^2 = .45$, and compared to counterfactuals with negated additional conditions, $F(1, 46) = 9.95$, $MSE = 1.01$, $p < .004$, $\eta_p^2 = .18$. They also made fewer inferences from counterfactuals with negated additional conditions than single counterfactuals, $F(1, 46) = 10.20$, $MSE = 0.46$, $p < .004$, $\eta_p^2 = .18$. A similar pattern of suppression was observed for the *modus ponens* inference (see Supplemental Materials).

For the *denial of the antecedent*, participants made fewer inferences from counterfactuals with negated additional conditions compared to single counterfactuals, $F(1, 46) = 7.96$, $MSE = 0.95$, $p < .008$, $\eta_p^2 = .15$, and compared to negated alternative conditions, $F(1, 46) = 19.34$, $MSE = 1.01$, $p < .001$, $\eta_p^2 = .30$. There was no difference between counterfactuals with negated alternatives and single counterfactuals, $F(1, 46) = 1.27$, $MSE = .59$, $p = .26$, $\eta_p^2 = .02$. A similar pattern of suppression was observed for the *affirmation of the consequent* inference (see Supplemental Materials).

It is also notable that the frequency of *modus tollens* and *denial of the antecedent* from the single negated counterfactuals was high, showing the standard elevation effect, for example, compared to the frequency of *modus tollens* or *denial of the antecedent* from the single factual conditionals in the previous experiment, as a glance at Figs. 4 and 5 confirms. The experiment shows that the switched-suppression effect occurs for counterfactuals. Counterfactuals with negated *alternatives* suppress the *modus tollens* inference, and counterfactuals with negated *additional* conditions suppress the *denial of the antecedent* inference.

10. General discussion

We began with the question of which sort of knowledge takes precedence in reasoners' inferences, their beliefs about additional background conditions, for example, that sun is needed as well as water for plants to bloom, or their beliefs about the known or presupposed facts of a situation, for example, that in this case, there was no sun and the plants did not bloom in fact. We have discovered that knowledge about additional background conditions takes priority over knowledge about presupposed facts.

The six experiments show that inferences are suppressed from counterfactual conditionals. *Modus tollens* is suppressed for affirmative counterfactuals with additional conditions compared to single counterfactuals, as Experiment 1a showed. The finding indicates that its suppression by additional conditions takes precedence over its elevation by counterfactuals. *Modus tollens* was also suppressed from counterfactuals with additional conditions even when participants constructed their own conclusions, and so its suppression cannot be attributed to participants engaging in a superficial way with the task, as Experiment 1b showed.

Similarly, the *denial of the antecedent* inference was suppressed for counterfactuals with alternatives compared to single counterfactuals, as Experiment 2a showed. The finding indicates that its suppression by alternatives takes precedence over its elevation by counterfactuals. The *denial of the antecedent* and *modus tollens* inferences were suppressed for counterfactuals even for arguments with minor premises and conclusions phrased as counterfactual, as Experiment 2b showed. The result shows that the suppression does not merely reflect some confusion about the counterfactual status of the information.

The results corroborate the idea that when people understand counterfactuals with additional conditions, such as “if she had watered the plants they would have bloomed, if the sun had shone they would have bloomed,” they combine the counterfactuals conjunctively, “if she had watered the plants AND the sun had shone they would have bloomed.” Even though they have also envisaged the presupposed facts, “she did not water the plants and the sun did not shine and they did not bloom,” they nonetheless keep in mind the counterexample, “she watered the plants and the sun did not shine and the plants did not bloom,” just as they do for factual conditionals.

Although conditionals with additional conditions typically suppress inferences such as *modus tollens*, and conditionals with alternatives typically suppress inferences such as the *denial of the antecedent*, this pattern is switched for negated conditionals. Conditionals that contain negated alternatives, such as, “if Lisa did not have an essay to write she did not study late in the library, if Lisa did not have a book to read she did not study late in the library” suppress *modus tollens* inferences (and *modus ponens*), and those that contain negated additional conditions, such as, “if Lisa did not have an essay to write she did not study late in the library, if the library did not stay open she did not study late in the library” suppress the *denial of the antecedent* inference (and the *affirmation of the consequent*). The switched-suppression effect corroborates the proposal that people combine a pair of conditionals with negated alternatives in a *conjunction*, “If Lisa did not have an essay to write AND she did not have a book to read then she did not study late in the library,” and a pair of conditionals with negated additional conditions in a *disjunction*, “If Lisa did not have an essay to write OR the library did not stay open then she did not study late in the library.” The switched-suppression effect occurs for negated conditionals, as Experiment 3a shows, and for negated counterfactuals, as Experiment 3b shows.

Overall, the results are compatible with theories that propose that people create a combined interpretation of the antecedents, for example, for affirmative additional conditions, “if she watered the plants AND the sun shone then they bloomed.” Hence, they are consistent with theories of suppression based on mental models (e.g., Byrne et al., 1999), and also those based on logic programming (e.g., Stenning & Van Lambalgen, 2007, 2012), machine learning (e.g., Dietz et al., 2014), and formal linguistics (e.g., Cariani &

Rips, 2017). They present something of a challenge to theories that do not highlight the combination of the antecedents, to which we now turn.

10.1. Alternative theories

The suppression of the *denial of the antecedent* and *affirmation of the consequent* inferences was initially demonstrated by theorists who advocated a theory of reasoning based on abstract inference rules (e.g., Braine & O'Brien, 1998; see also Rips, 1994). They proposed that the "mental logic" consisted of elementary rules, including only those considered valid in propositional logic, such as *modus ponens*. Inferences such as the *affirmation of the consequent,* considered a fallacy on the material implication interpretation of "if," arose from interpretational errors, they proposed. The demonstration that such inferences could be suppressed by the provision of extra information was taken to indicate that there were no rules corresponding to these inferences in the mental logic (Rumain et al., 1983). But the demonstration of the suppression of the *modus ponens* and *modus tollens* inferences called this explanation into question (Byrne, 1989). A defense of the formal inference rule account is that suppression arises because the additional condition, for example, "the sun shone" reduces the degree of belief in the first conditional "if she watered the plants they bloomed" (e.g., George, 1995, 1997; O'Brien, 1993; Politzer & Braine, 1991). This view is shared by theories based on probabilistic logic and suppositions (e.g., Over, 2017; Stevenson & Over, 1995), and causal Bayes nets (e.g., Oaksford & Chater, 2004, 2018).

However, if the suppression of inferences arises because the additional condition reduces people's belief in the first conditional, then knowledge of the presupposed facts should take precedence over knowledge of additional conditions. The "suppositional conditional" theory proposes that people understand a conditional "if p then q" by adding the antecedent, "p," to their beliefs and calculating the probability of the consequent, "q" (Evans, 2007; Evans & Over, 2004). They think about true antecedents only (p) and not about their belief in the negated antecedent (not-p), or whether or not the consequent (q) follows in such cases (e.g., Handley, Evans, & Thompson,2006; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007). They understand a conditional by estimating their degree of belief in it, based on assessing the prior likelihood of the consequent given the antecedent, $P\ (q|p)$, that is, its conditional probability (e.g., Over, 2017; Over & Cruz, 2017). An individual who thinks it is very likely that the plants bloom when the person waters them will assign a high number to this conditional probability (for illustration, we chose .75 to capture "very likely"),

If she watered the plants they bloomed $P(q|p) = .75$

An individual who does not think it is very likely that the plants bloom when the person waters them will assign a lower number. On this view, people tolerate exceptions to a conditional in their consideration of the likelihood of conclusions from it and so the probability of the conditional does not have to be 1 for it to be considered true (e.g., Geiger & Oberauer, 2007; see also Oaksford & Chater, 2007, 2018; Singmann, Klauer, & Beller, 2016; but for evidence against this claim, see Byrne & Johnson-Laird, 2019). A person who assigned a high degree of certainty to the conditional will make the *modus ponens* inference with a high degree of certainty too. A probabilistically valid inference is one for which the uncertainty of its conclusion does not exceed the sum of the uncertainties of its premises. The *modus tollens* inference requires additional steps. One strategy is akin to a *reductio ad absurdum*—assume *p* and conclude *q* by *modus ponens*, and since the conclusion *q* and the premise not-*q* are a contradiction, conclude not-*p* (e.g., Evans, 2007; Evans & Over, 2004).

On this theory, the suppression of inferences by an additional condition arises because the second conditional leads people to no longer believe in the certainty that the first conditional is true (e.g., George, 1995; O'Brien, 1993; Politzer & Braine, 1991; Stevenson & Over, 1995). For example, an individual who assigned a high probability to the first conditional,

If she watered the plants they bloomed $P(q|p) = .75$

may assign an even higher probability to an additional condition, if they believe it is a crucial condition, for example,

If the sun shone the plants bloomed $P(q|r) = .95$

And their high belief in the second conditional will lead them to doubt the truth of the first conditional, which they will now revise downwards, for example,

If she watered the plants they bloomed $P(q|p) = .25$

The *modus ponens* inference is thus suppressed because the degree of belief in the first conditional is reduced through this revision. We note this account appears to struggle to explain the observation that people make inferences such as *modus ponens* with equal frequency from each conditional (Cariani & Rips, 2017), and *modus ponens* is suppressed equally whether individuals are given the minor premise corresponding to the antecedent of the first conditional or the second one (Byrne et al., 1999).

The suppositional account proposes that the probability of a counterfactual such as "if she had watered the plants they would have bloomed" at the present time is the same as the probability of a factual conditional such as "if she watered the plants they bloomed" at a previous time (e.g., Evans, 2007; Evans & Over, 2004; Over et al., 2007). Given a high degree of belief in the factual conditional at a previous time, If she watered the plants they bloomed $P(q|p) = .75$

the probability of a counterfactual at the present time will be similar,

If she had watered the plants they would have bloomed $P(q|p) = .75$

People think about only one possibility according to a singularity principle (Evans, 2007, p. 74), and numerical values represent both their strength of belief in the counterfactual consequent given the counterfactual antecedent, and strengths of belief in each of the implicated facts—that she did not water the plants, and that they did not bloom. On this view, *"Johnson-Laird and Byrne (2002) propose two models rather than one for the initial representation of counterfactual conditionals. The suppositional theory assumes that implicature may be readily added to the representation of all conditionals, and hence there is nothing intrinsically different in adding the -p -q case here"* (Evans, Over, & Handley, 2005, p. 1049). Inferences such as *modus ponens* are made from a counterfactual by assessing prior beliefs about its conditional probability,

If she had watered the plants they would have bloomed $P(q|p) = .75$

whereas inferences such as *modus tollens* are made by assessing beliefs about the implied facts:

She did not water the plants P (not-p) = .95
The plants did not bloom P (not-q) = .95

As a result, one way for the theory to provide an *ad hoc* explanation for why *modus tollens* is made more frequently from a counterfactual than a factual conditional is to assume that the implied facts are assigned higher degrees of belief than the counterfactual itself (e.g., we assign .95 to them above for illustration). A challenge for this approach, however, is that the elevation is observed even for neutral or abstract counterfactuals, for which participants can have no prior beliefs about the facts (e.g., Byrne & Tasso, 1999).

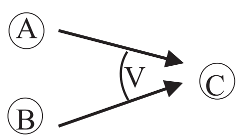
If the account proposes that counterfactual elevation occurs because the *modus tollens* inference is made by reference to the known or presupposed facts, then it must predict that *modus tollens* will continue to be elevated even when additional conditions are known about. Suppose someone says to you about a game of cards, "if I had had two kings I would have won." You may believe the counterfactual to be highly likely, and you may believe the presupposed facts to be highly likely too—she did not have two kings, she did not win. But suppose you then discover that her opponent had two aces. You will revise your belief in the counterfactual—it now seems highly unlikely. Instead, you may believe if she had had two kings—and her opponent had not had two aces—she would have won. But what of the facts—is there any need to revise your belief in them? It is still highly likely that she did not have two kings, and that she did not win. So, there is no need to revise your degrees of belief in the facts just because new information leads you to revise your degrees of belief in the counterfactual. Hence, *modus tollens* from a counterfactual, which is based on the facts according to the suppositional theory, should continue to be elevated, even when additional conditions are known about. The account must predict that knowledge of the known or presupposed facts will take precedence over knowledge of additional conditions. A putative way to provide an *ad hoc* explanation for counterfactual suppression is to propose that *modus tollens* is suppressed for counterfactuals with additional conditions because new information (about the additional condition) leads you to revise not only your belief in the first counterfactual but also your belief in the counterfactual's presupposed facts. But this explanation requires a systematic account of the circumstances under which degrees of belief in the facts should be revised because of revision to degrees of belief in the counterfactual. As the cards example shows, revision of belief in a counterfactual need not necessarily affect belief in the facts. The results of Experiments 1a and 1b do not support the idea that people make inferences such as *modus tollens* by relying on the likelihood of the presupposed facts, separately from the likelihood of the counterfactuals.

Another putative explanation is that a conditional with an additional background condition, such as "if the library stays open she will study late in the library" casts doubt on the certainty of the first conditional because the background condition fundamentally undermines its assumptions (George, 1995; Politzer & Braine, 1991; Stevenson & Over, 1995). On this view, the default interpretation of the first conditional is that "all things being equal" the antecedent leads to the consequent, but the conditional with the background condition calls this "no abnormal conditions" assumption into question (see also Oaksford & Chater, 2007; Stenning & van Lambalgen, 2012). But the results of Experiment 3a indicate that the suppression of *modus tollens* does not depend on a background additional condition that calls assumptions of normality into question to undermine the certainty of the first conditional, since it occurs for negated *alternatives*. The "abnormal conditions" idea cannot explain the switched-suppression observed in Experiment 3a. An *alternative* about having a book to read does not have the quality of undermining "all things being equal." So this account must predict that belief in the likelihood of a conditional, "if Lisa did not have an essay to write she did not study late in the library" is not affected in such a way by a conditional with an *alternative*, "if Lisa did not have a book to read she did not study late in the library." The alternative in the second conditional does not undermine the truth of the first conditional or call its default assumptions into question. The suppositional conditional version of probabilist theories of reasoning cannot explain why *modus tollens* is suppressed by conditionals with negated *alternatives*.

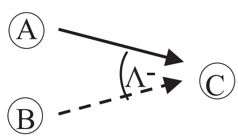
A different probabilist theory is based on a structural theory of causation, causal Bayes nets (Pearl, 2011, 2013). The theory proposes that people construct causal models, for example,

in which nodes represent causes and effects, arrows represent causal direction from cause to effect, and a conditional probability table at each node gives the probability that a node is present or absent conditional on its causes (e.g., Lucas & Kemp, 2015; Sloman & Lagnado, 2005). A reasoner can make changes to the model by "intervention" on a node to reflect a counterfactual assumption, for example, the intervention of imagining that B is absent deletes links into B, and inferences can be made about the effect of such an action on A, and on C (e.g., Pearl, 2011, 2013). A counterfactual's probability is determined by conditional probability information about how effects depend on causes in the causal model, and it relies on the "do" intervention operator, $P(c \mid b.\ do\ b)$, that is, the conditional probability of *c* given that *b* was observed, but counterfactually removed. The theory has been adapted as an account of human reasoning (e.g., Sloman & Lagnado, 2005) and can accommodate causal relations such as over-determination (Lagnado, Gerstenberg, & Zultan, 2013). It may require additions to incorporate norms (e.g., Dehghani, Iliev, & Kaufmann, 2010) and heuristics to identify nodes as causes or effects of a target node (e.g., Meder, Gerstenberg, Hagmayer, & Waldmann, 2010).

To explain the suppression of inferences when there are several causes, integration rules determine how the parents of a node combine, for example, an OR rule gives probability 1 unless both causes are absent (e.g., Fernbach & Erb, 2013; Oaksford & Chater, 2018),



An AND-NOT rule gives probability 1 if a generative cause, for example, she waters the plants, is present and a preventative cause, for example, the sun does not shine, is absent,



People endorse an inference in direct proportion to the conditional probability of the conclusion given the categorical premise (e.g., Oaksford & Chater, 2018). The *modus ponens* inference is made with a probability corresponding to the degrees of belief in the conditional probability of the conditional, $P\ (q|p)$, that is, $1 - P(not\text{-}q|p)$ (Oaksford & Chater, 2018). People consider the *modus tollens* premise as a falsifying, *p, not-q*, counterexample to an *if p then q* conditional (Oaksford & Chater, 2018). It requires assessing belief in the likelihood of the non-occurrence of the antecedent, given the non-occurrence of the consequent, $P\ (not\text{-}p|not\text{-}q)$, that is (Oaksford & Chater, 2018, p. 884):

$$1-P(q)-P(p)P(not-q|p)/1-P(q).$$

When there are multiple causes, people may make the *modus tollens* inference by assuming that an explicit disabler, B (e.g., the sun not shining), occurs, that is, *P(not-a|not-c, b)* (Oaksford & Chater, 2017, 2018). A probability, P0, before information is learned, can be changed given new evidence to a different probability, P1; P1 can be lower than P0 given non-monotonicity (Oaksford & Chater, 2004; Oaksford & Chater, 2018).

On this account, dynamical reasoning could explain why additional information suppresses inferences from counterfactual conditionals. For example, for the *modus tollens* inference from "if she watered the plants they bloomed, if the sun shone they bloomed," the conditional with the additional condition, for example, "if the sun shone the plants bloomed" lowers the conditional probability, *P(not-p|not-q)*, that is, the probability she did not water the plants given that they did not bloom because it makes clear that the plants may not have bloomed because the sun did not shine. Reasoners can compare the probability of the plants not blooming because they were not watered, and the probability of them not blooming because there was no sun. But the evidence from Experiments 3a and 3b that the *modus tollens* inference is *not* suppressed for a pair of negated conditionals with additional conditions goes against this view. For negated conditionals with additionals, such as "if she did not water the plants they did not bloom, if the sun did not shine they did not bloom," the conditional with the additional condition, "if the sun did not shine the plants did not bloom" should also lower the conditional probability, *P(not-p|not-q)*, this time, the probability she watered the plants given that they bloomed because it makes clear that the plants may have bloomed because the sun shone. Reasoners can compare the probability of the plants blooming because they were watered, and the probability of them blooming because the sun shone. The *modus tollens* inference should thereby be suppressed for negated conditionals with additional conditions on this view.

The evidence from Experiments 3a and 3b that the *modus tollens* inference is suppressed instead for a pair of negated conditionals with *alternatives* is also challenging. A putative defense could be that reasoners re-interpret the *modus tollens* inference from a negated conditional, such as "if she did not water the plants they did not bloom" to be instead an *affirmation of the consequent* inference from its affirmative counterpart, "if she watered the plants they bloomed." The probability of endorsing an inference depends on the conditional probability of the conclusion, given the minor premise (e.g., Oaksford & Chater, 2018). The *modus tollens* inference from the pair of negated conditionals with alternatives, "if she did not water the plants they did not bloom, if it did not rain they did not bloom" depends on calculating the conditional probability of the conclusion "she watered the plants" given the minor premise "the plants bloomed," that is, P (*watered* | *bloomed*). Similarly, the *affirmation of the consequent* inference from the corresponding pair of affirmative conditionals, "if she watered the plants they bloomed, if it rained they bloomed" depends on calculating the conditional probability of the conclusion "she watered the plants" given the minor premise "the plants bloomed," that is, P (*watered* | *bloomed*). Hence, the suppression occurs because people translate the given negative information to be affirmative and calculate only the probability that she watered the plants given that they bloomed. This defense requires the assumption that reasoners do not reason about the plants not blooming if she did not water them, but instead draw conclusions based only on their beliefs about her watering the plants given that they bloomed. Yet it appears overly restrictive to claim that people cannot think directly about negative conditionals.

The causal Bayes net theory does not explain the counterfactual elevation effect. Instead, tests of the theory have focused on the tendency to *refrain* from *modus tollens* when counterfactuals describe the prevention of a cause (e.g., Sloman & Lagnado, 2005; see also Lucas & Kemp, 2015; Meder, Hagmayer, & Waldmann, 2009). For example, when participants were told that movement of A causes B to move, and both A and B were moving, and they were asked, “suppose B were observed to not be moving, would A still be moving?” most of them said no, that is, they made the *modus tollens* inference. But when they were asked instead, “suppose B were prevented from moving, would A still be moving?” most of them said yes (Sloman & Lagnado, 2005). They made fewer *modus tollens* inferences from counterfactuals that described interventions compared to ones that described observations, for counterfactual interventions for which the real state was known, and hypothetical ones for which the real state was not known (e.g., Lucas & Kemp, 2015; Meder et al., 2009). A further challenge to this account is the observation that in a causal system in which A or B causes C, an intervention on C minimally prunes the causal model to remove links into C, which supports the counterfactual, “If C hadn’t occurred, A would still have occurred.” But people tend to infer instead that when C did not occur, neither A *nor* B could have occurred (e.g., Rips 2010; see also Rips & Edwards, 2013). Another difficulty for the causal Bayes nets approach is that both the suppression effect and the counterfactual elevation effect have been observed for conditionals that convey a wide variety of relations between the antecedent and consequent, including not only causal content but also many other sorts of contents, such as intentions, inducements, definitions, social regulations, and even arbitrary relations. Notwithstanding applications of the probabilistic logic approach to abstract conditionals (e.g., Oaksford & Chater, 2004, 2018), it remains unspecified how the causal Bayes nets approach extends to non-causal contents, nor indeed how it can explain the findings of differences in counterfactual thoughts about causal sequences from those about other sorts of relations, such as intention sequences (e.g., Walsh & Byrne, 2007; see also Walsh & Byrne, 2004; for a review see Byrne, 2016).

11. Conclusion

What is the best way to combine beliefs about background conditions and beliefs about presupposed facts? The question is a crucial epistemological one. The results of the six experiments reported here indicate that most people give precedence to beliefs about additional conditions over beliefs about presupposed facts. The results were predicted by the theory that people combine these beliefs by relying on the simulation of the possibilities that are consistent with them (e.g., Byrne & Johnson-Laird, 2019; Khemlani et al., 2018).

Understanding how different sorts of conflicting beliefs are resolved is crucial for understanding human reasoning in many applied settings such as reasoning by juries (e.g., Moreno-Rios & Byrne, 2018). Another important application is in eXplainable AI

(XAI), to explain the otherwise unintelligible decisions of complex AI systems, such as automated neural networks trained on vast amounts of data (e.g., Biran & Cotton, 2017; Hoffman, Miller, Mueller, Klein, & Clancey, 2018). The use of counterfactuals in XAI has many benefits (e.g., Byrne, 2019; Miller, 2019). For example, an explanation system for an autonomous vehicle could provide an explanation for a decision to swerve onto the pavement, in the form of a counterfactual such as, “if the car had braked, the passenger would have been injured.” It could provide further elaboration by referring to additional background conditions, “if a vehicle detected close behind the car had rear-ended it, the passenger would have been injured.” Our results indicate that beliefs about additional background conditions—for example, that the car braking and a vehicle close behind rear-ending it would lead to a passenger being injured—take precedence over beliefs about the presupposed facts of a situation—for example, that in this case, the car did not brakes and the vehicle behind it did not rear-end it in fact. The discovery has important consequences for the sorts of counterfactual explanations that human users may find helpful to ensure they trust the decisions of complex AI systems. Overall, the suppression of inferences by background conditions, and the elevation of inferences by presupposed facts, and the way people integrate these two sorts of knowledge, have many practical implications, as well as widespread consequences for understanding human reasoning.

Acknowledgments

The research was supported by the Research Project PSI2015-70771-P from the Spanish Ministry of Science and Innovation (Ministerio de Ciencia e Innovación). Some of the results were presented at the London Reasoning Workshop in 2017. We thank Mark Keane and Robert Mackiewicz for comments on an earlier draft.

References

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. Dordrecht, The Netherlands: Reidel.

Barres, P., & Johnson-Laird, P. N. (2003). On imagining what is true (and what is false). *Thinking & Reasoning*, *9*(1), 1–42.

Biran, O., & Cotton, C. (2017). Explanation and justification in machine learning: A survey. IJCAI-17 Workshop on Explainable AI (XAI), *8*, 8–13.

Bonnefon, J. F., & Hilton, D. J. (2002). The suppression of modus ponens as a case of pragmatic preconditional reasoning. *Thinking & Reasoning*, *8*(1), 21–40.

Bonnefon, J. F., & Hilton, D. J. (2004). Consequential conditionals: invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 28.

Bonnefond, M., Kaliuzhna, M., Van der Henst, J. B., & De Neys, W. (2014). Disabling conditional inferences: an EEG study. *Neuropsychologia*, *56*, 255–262.

Braine, M., & D. P. O’Brien (Eds.) (1998). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 61–83.

Byrne, R. M. (1991). Can valid inferences be suppressed? *Cognition*, *39*(1), 71–78.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press.

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, *67*, 135–157.

Byrne, R. M. J. (2019). Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning (pp. 6276–6282). Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. IJCAI. https://doi.org/10.24963/ijcai.2019/876.

Byrne, R. M., & Egan, S. M. (2004). Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *58*(2), 113.

Byrne, R. M. J., Espino, O., & Santamaría, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory & Language*, *40*, 347–373.

Byrne, R. M. J., & Johnson-Laird, P. N. (2019). 'If' and 'or': Real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory & Cognition*. [Epub ahead of print]. https://doi.org/10.1037/xlm0000756

Byrne, R. M. J., & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition*, *27*, 726–740.

Cariani, F., & Rips, L. J. (2017). Conditionals, context, and the suppression effect. *Cognitive Science*, *41*(3), 540–589.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, *53*(3), 217–238.

Chandon, E., & Janiszewski, C. (2008). The influence of causal conditional reasoning on the acceptance of product claims. *Journal of Consumer Research*, *35*(6), 1003–1011.

Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, *55*(220), 289–307.

Couto, C., Quelhas, A. C., & Byrne, R. M. J. (2017). Advice conditionals about tips and warnings: Interpretations and inferences. *Journal of Cognitive Psychology*, *29*(3), 364–380.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274–282.

De Brigard, F., & Parikh, N. (2019). Episodic counterfactual thinking. *Current Directions in Psychological Science*, *28*, 59–66.

de Chantal, P. L., & Markovits, H. (2017). The capacity to generate alternative ideas is more important than inhibition for logical reasoning in preschool-age children. *Memory & Cognition*, *45*(2), 208–220.

De Neys, W., & Everaerts, D. (2008). Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay. *Journal of Experimental Child Psychology*, *100*(4), 252–263.

De Neys, W., Schaeken, W., & D'ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition*, *30*(6), 908–920.

De Neys, W., Schaeken, W., & D'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, *31*(4), 581–595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*(4), 349–381.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning*, *11*(2), 123–150.

De Vega, M., Urrutia, M., & Riffo, B. (2007). Cancelling updating in the comprehension of counterfactuals embedded in narratives. *Memory and Cognition*, *35*, 1410–1421.

Dehghani, M., Iliev, R., & Kaufmann, S. (2010). Causal explanation and fact mutability in counterfactual reasoning. *Mind & Language*, *27*, 55–85.

Demeure, V., Bonnefon, J. F., & Raufaste, E. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 260.

Dietz, E. A., Holldobler, S., & Ragni, M. (2012). A computational logic approach to the suppression task. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (*34*, pp. 1500–1505).

Dietz, E. A., Hölldobler, S., & Wernhard, C. (2014). Modeling the suppression task under weak completion and well-founded semantics. *Journal of Applied Non-Classical Logics*, *24*(1–2), 61–85.

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology*, *49*(3), 181.

Dieussaert, K., Schaeken, W., Schroyens, W., & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking & Reasoning*, *6*(2), 125–160.

Egan, S., & Byrne, R. M. J. (2012). Inferences from counterfactual threats and promises. *Experimental Psychology*, *59*(4), 227–235.

Egan, S., Garcia-Madruga, J., & Byrne, R. M. J. (2009). Indicative and counterfactual 'only if' conditionals. *Acta Psychologica*, *132*(3), 240–249.

Elio, R. (1997). What to believe when inferences are contradicted. The impact of knowledge type and inference rule. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 211–216).

Espino, O., & Byrne, R. M. J. (2013). The compatibility heuristic in non categorical hypothetical reasoning: Inferences between conditionals and disjunctions. *Cognitive Psychology*, *67*(3), 98–129.

Espino, O., & Byrne, R. M. J. (2018). Thinking about the opposite of what is said: Counterfactual conditionals and symbolic or alternate simulations of negation. *Cognitive Science*, *42*(8), 2459–2501.

Espino, O., Santamaría, C., & Byrne, R. M. J. (2009). People think about what is true for conditionals, not what is false: Only true possibilities prime the comprehension of 'if'. *Quarterly Journal of Experimental Psychology*, *62*, 1072–1078.

Evans, J. S. B. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, UK: Psychology Press.

Evans, J. S. BT. & Over, D. E. (2004). *If*. Oxford: Oxford University Press.

Evans, J. S. B., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, *112*(4), 1040–1052.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, *58*, 609–626.

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research*, *1236*, 113–125.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1327.

Fillenbaum, S. (1974). Information amplified: Memory for counterfactual conditionals. *Journal of Experimental Psychology*, *102*(1), 44.

Frosch, C., & Byrne, R. M. J. (2012). Causal conditionals and counterfactuals. *Acta Psychologica*, *141*(1), 54–66.

Gazzo Castañeda, L. E., & Knauff, M. (2016). Defeasible reasoning with legal conditionals. *Memory & Cognition*, *44*(3), 499–517.

Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition*, *35*(8), 2060–2074.

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology*, *86*(1), 93–111.

George, C. (1997). Reasoning from uncertain premises. *Thinking & Reasoning*, *3*(3), 161–189.

Handley, S. J., Evans, J. S. B., & Thompson, V. A. (2006). The negated conditional: A litmus test for the suppositional conditional? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 559–569.

Hoffman, R., Miller, T., Mueller, S. T., Klein, G., & Clancey, W. J. (2018). Explaining explanation, part 4: A deep dive on deep nets. *IEEE Intelligent Systems*, *33*(3), 87–95.

Hölldobler, S., & Ramli, C. D. P. K. (2009). Logic programs under three-valued Łukasiewicz semantics. In *International Conference on Logic Programming* (pp. 464–478). Berlin, Heidelberg: Springer.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, *35*(4), 904.

Jeffrey, R. (1981). *Formal logic: Its scope and limits* (2nd ed.). New York: McGraw-Hill.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, *99*, 418–439.

Johnson-Laird, P. N., Lotstein, M., & Byrne, R. M. J. (2012). The consistency of disjunctive assertions. *Memory and Cognition*, *40*, 769–778.

Juhos, C., Quelhas, A. C., & Byrne, R. M. J. (2015). Reasoning about intentions: Counterexamples to reasons for actions. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *41*(1), 55–76.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Khemlani, S. S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model theory of sentential reasoning. *Cognitive Science*, *42*(6), 1887–1924.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, *137*, 196–209.

Kowalski, R. (2011). *Computational logic and human thinking: How to be artificially intelligent*. Cambridge, UK: Cambridge University Press.

Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., & Perner, J. (2013). Processing counterfactual and hypothetical conditionals: An fMRI investigation. *NeuroImage*, *72*, 265–271.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *37*, 1036–1073.

Lewis, D. (1973). *Counterfactuals*. Oxford, England: Blackwell.

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*, *122*, 700–734.

Markovits, H., Brisson, J., & de Chantal, P. L. (2015). Deductive updating is not Bayesian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 949.

Markovits, H., Forgues, H. L., & Brunet, M. L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition*, *38*(4), 485–492.

Markovits, H., & Potvin, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition*, *29*(5), 736–744.

McKenzie, R., St. B. T. Evans, J., & Handley, S. J. (2011). Autism and performance on the suppression task: reasoning, context and complexity. *Thinking & Reasoning*, *17*(2), 182–196.

Meder, B., Gerstenberg, T., Hagmayer, Y., & Waldmann, M. R. (2010). Observing and intervening: Rational and heuristic models of causal decision making. *The Open Psychology Journal*, *3*, 119–135.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & Cognition*, *37*, 249–264.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38.

Moreno-Rios, S., & Byrne, R. M. J. (2018). Inferences from disclosures about the truth and falsity of expert testimony. *Thinking and Reasoning*, *24*(1), 41–78.

Moreno-Rios, S., Garcia-Madruga, J., & Byrne, R. M. J. (2008). Semifactual 'even if' reasoning. *Acta Psychologica*, *128*, 197–209.

Murray, M. A., & Byrne, R. M. J. (2005). Attention and working memory in insight problem solving. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1571–1576). Hillsdale, NJ: Erlbaum.

Nickerson, R. (2015). *Conditional reasoning*. Oxford, England: Oxford University Press.

Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, *122*, 102–109.

O'Brien, D. P. (1993). Mental logic and human irrationality: We can put a man on the moon so why can't we solve those logical-reasoning problems? In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 110–135). London: Routledge.

Oaksford, M., & Chater, N. (2004). Probabilities and pragmatics in conditional inference: Suppression and order effects. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 95–122). Chichester, UK: Wiley.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Oaksford, M., & Chater, N. (2017). Causal models and conditional reasoning. In M. Waldman (Ed.), *Oxford handbook of causal reasoning* (pp. 327–346). Oxford, UK: Oxford University Press.

Oaksford, M., & Chater, N. (2018). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, *19*(3–4), 346–379.

Orenes, I., Garcia-Madruga, J., Gomez-Veiga, I., Espino, O., & Byrne, R. M. J. (2019). The comprehension of counterfactual conditionals: Evidence from eye-tracking in the visual world paradigm. *Frontiers in Psychology*, *10*, 1172.

Over, D. E. (2017). Causation and the probability of causal conditionals. In M. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 307–325). Oxford, UK: Oxford University Press.

Over, D. E., & Cruz, N. (2017). Probabilistic accounts of conditional reasoning. In L. Ball & V. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 434–450). London, UK: Routledge.

Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*(1), 62–97.

Pearl, J. (2011). The algorithmization of counterfactuals. *Annals of Mathematics and Artificial Intelligence*, *61*, 29–39.

Pearl, J. (2013). Structural counterfactuals: A brief introduction. *Cognitive Science*, *37*, 977–985.

Pelletier, F. J., & Elio, R. (1997). What should default reasoning be, by default? *Computational Intelligence*, *13*(2), 165–187.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040.

Phillips, M. L., Howard, R., & David, A. S. (1997). A cognitive neuropsychological approach to the study of delusions in late-onset schizophrenia. *International Journal of Geriatric Psychiatry*, *12*(9), 892–901.

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940–2951.

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2011). Reasoning with exceptions: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, *23*(2), 471–480.

Pijnacker, J., Geurts, B., Van Lambalgen, M., Kan, C. C., Buitelaar, J. K., & Hagoort, P. (2009). Defeasible reasoning in high-functioning adults with autism: Evidence for impaired exception-handling. *Neuropsychologia*, *47*(3), 644–651.

Politzer, G., & Braine, M. D. S. (1991). Responses to inconsistent premisses cannot count as suppression of valid inferences. *Cognition*, *38*(1), 103–108.

Quelhas, A. C., & Byrne, R. M. J. (2003). Reasoning with deontic and counterfactual conditionals. *Thinking and Reasoning*, *9*, 43–66.

Ragni, M., Eichhorn, C., & Kern-Isberner, G. (2016). Simulating human inferences in the light of new information: A formal analysis. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (pp. 2604–2610). IJCAI.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred models. *Psychological Review*, *120*, 561–588.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, *72*, 54–107.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.

Rips, L. J. (2010). Two causal theories of counterfactual conditionals. *Cognitive Science*, *34*(2), 175–221.

Rips, L. J., & Edwards, B. J. (2013). Inference and explanation in counterfactual reasoning. *Cognitive Science*, *37*(6), 1107–1135.

Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. In *Advances in Experimental Social Psychology* (Vol. 56, pp. 1–79). Cambridge, MA: Academic Press.

Rumain, B., Connell, J., & Braine, M. D. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psychology*, *19* (4), 471.

Santamaría, C., Espino, O., & Byrne, R. M. J. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1149–1154.

Sellen, J. L., Oaksford, M., & Gray, N. S. (2005). Schizotypy and conditional reasoning. *Schizophrenia Bulletin*, *31*(1), 105–116.

Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology*, *88*, 61–87.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science*, *29*, 5–39.

Stalnaker, R. (1968). A theory of conditionals. *American Philosophical Quarterly Monograph Series*, *2*, 98–112.

Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*(6), 919–960.

Stenning, K., & van Lambalgen, M. (2007). Logic in the study of psychiatric disorders: Executive function and rule-following. *Topoi*, *26*(1), 97–114.

Stenning, K., & Van Lambalgen, M. (2012). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Stenning, K., & van Lambalgen, M. (2016). Logic programming, probability, and two-system accounts of reasoning: A rejoinder to Oaksford and Chater (2014). *Thinking & Reasoning*, *22*(3), 355–368.

Stenning, K., & Van Lambalgen, M. (2019). Reasoning and discourse coherence in autism spectrum disorder. In K. Morsanyi & R. M. J. Byrne (Eds.), *Thinking, reasoning, and decision making in autism* (pp. 135–155). London: Routledge.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology*, *48*(3), 613–643.

Stewart, A. J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2113–2125.

Strickland, B., Silver, I., & Keil, F. C. (2017). The texture of causal construals: Domain-specific biases shape causal inferences from discourse. *Memory & Cognition*, *45*(3), 442–455.

Thompson, V., & Byrne, R. M. J. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology Learning Memory and Cognition*, *28*, 1154–1170.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., & Van Overwalle, F. (2013). Counterfactual thinking: An fMRI study on changing the past for a better future. *Social Cognitive and Affective Neuroscience*, *8*, 556–564.

Verschueren, N., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2004). The difference between generating counterexamples and using them during reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *57*(7), 1285–1308.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*(3), 239–278.

Walsh, C. R., & Byrne, R. M. J. (2004). Counterfactual thinking: The temporal order effect. *Memory & Cognition*, *32*, 369–378.

Walsh, C. R., & Byrne, R. M. J. (2007). The effects of reasons for acting on counterfactual thinking. *Thinking and Reasoning*, *13*, 461–483.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Supplemental Materials. The materials used in the experiments and additional statistical information for each experiment.