COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 42 (2018) 2459–2501

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12677

Thinking About the Opposite of What Is Said: Counterfactual Conditionals and Symbolic or Alternate Simulations of Negation

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 31 January 2017; received in revised form 16 April 2018; accepted 20 July 2018

Abstract

When people understand a counterfactual such as "if the flowers had been roses, the trees would have been orange trees," they think about the conjecture, "there were roses and orange trees," and they also think about its opposite, the presupposed facts. We test whether people think about the opposite by representing alternates, for example, "poppies and apple trees," or whether models can contain symbols, for example, "no roses and no orange trees." We report the discovery of an inference-to-alternates effect—a tendency to make an affirmative inference that refers to an alternate even from a negative minor premise, for example, "there were no orange trees, therefore there were poppies." Nine experiments show the inference-to-alternates effect occurs in a binary context, but not a multiple context, and for direct and indirect reference; it can be induced and reduced by prior experience with similar inferences, and it also occurs for indicative conditionals. The results have implications for theories of counterfactual conditionals, and of negation.

Keywords: Counterfactuals; Conditionals; Negation; Mental models; Embodied representations

1. Introduction

How do people understand counterfactual conditionals, such as "if the flowers had been roses, the trees would have been orange trees"? People appear to think about the situation corresponding to the counterfactual conjecture, "the flowers were roses and the trees were orange trees," and they keep track of its epistemic status as contrary to the facts. But they also appear to think about the situation corresponding to the opposite of what was said, the presupposed or known facts, "the flowers were not roses

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

and the trees were not orange trees," and they keep track of its epistemic status too, as corresponding to the facts (see Byrne, 2016, for a review). Our aim is to examine how people mentally represent the presupposed facts of a counterfactual conditional. The relevance of a psychological account of the mental representations and cognitive processes that underpin counterfactuals has been highlighted in contemporary philosophical, logical, and linguistic treatments of counterfactuals (e.g., Kratzer, 2012; Nickerson, 2015; Williamson, 2007) as well as in artificial intelligence programs for inference and other computational simulations (e.g., Ginsberg, 1986; Pearl, 2013). First we briefly summarize the extensive experimental evidence that people think about two possibilities when they understand a counterfactual. Next we sketch two putative theories about how people might represent the presupposed facts when they understand a counterfactual, extrapolated from accounts of how people represent negation. One view is derived from some accounts of how negation may be represented in embodied experiential simulations (e.g., Kaup, Lüdtke, & Zwaan, 2006; Mayo, Schul, & Burnstein, 2004), and the other view is based on accounts of how negation may be represented in mental models that contain propositional-like symbolic tags (e.g., Johnson-Laird & Byrne, 2002; Khemlani, Orenes, & Johnson-Laird, 2012; Orenes, Beltrán, & Santamaría, 2014). We then outline the inference task that allows us to test the predictions of two alternative theories of how people represent the opposite of what is said when they understand a counterfactual conditional and then we report the results of nine experiments that do so.

1.1. How people understand a counterfactual: Dual possibilities

Evidence that people think about both the counterfactual conjecture and the presupposed facts comes from a range of studies that employ a wide variety of methods (e.g., Byrne, 2005, 2017). For example, people tend to judge that someone who utters the counterfactual "if the flowers had been roses, the trees would have been orange trees" means to imply "there were no roses" and "there were no orange trees" (e.g., Thompson & Byrne, 2002), and when they read the counterfactual, they tend to subsequently mistakenly remember that they were given "there were no roses" and "there were no orange trees" (Fillenbaum, 1974). Priming studies show that people tend to read the conjunction "there were no roses and there were no orange trees" more rapidly when they are primed by the counterfactual compared to when they are primed by the corresponding ordinary conditional "if the flowers were roses, the trees were orange trees"; however, they read the conjunction "the flowers were roses and the trees were orange trees" equally quickly whether they are primed by the counterfactual or the ordinary conditional (e.g., Santamaría, Espino, & Byrne, 2005; see also De Vega, Urrutia, & Riffo, 2007). The counterfactual conjecture conflicts with the presupposed facts, and hence eye-tracking studies show an initial brief disruption in the immediate comprehension of a counterfactual (e.g., Ferguson & Sanford, 2008). fMRI studies also show that counterfactuals activate areas of the medial prefrontal cortex related to conflict detection (e.g., Kulakova, Aichhorn, Schurz, Kronbichler, & Perner, 2013; Van Hoeck et al., 2013). ERP studies indicate that

the initial disruption is rapidly resolved (e.g., Ferguson, Sanford, & Leuthold, 2008; see also Nieuwland & Martin, 2012).

Perhaps the most striking evidence for the dual meaning of counterfactuals is the counterfactual inference effect: the discovery that people readily make deductions from counterfactual conditionals that they otherwise find difficult to make from ordinary conditionals (e.g., Byrne & Tasso, 1999). For an ordinary indicative conditional "if there were roses then there were orange trees" people have difficulty making the modus tollens inference from "there were no orange trees" to "there were no roses"; participants make the inference on a little over half of the trials in experiments, and on many other instances they say instead that nothing follows. However, for the counterfactual conditional people very readily make the modus tollens inference; participants make the inference on about twice as many trials from the counterfactual as from the indicative conditional (e.g., Byrne & Tasso, 1999; Thompson & Byrne, 2002). People also readily make the denial of the antecedent inference, from "there were no roses" to "there were no orange trees" from the counterfactual, and more often than they do from the indicative conditional. The results indicate that participants have envisaged the presupposed facts, "there were no roses and there were no orange trees" (e.g., Byrne, 2017). Of equal importance is the observation that participants make the modus ponens inference from "there were roses," to "there were orange trees" readily from the counterfactual, just as often as they do from an indicative conditional (e.g., Byrne & Tasso, 1999; Thompson & Byrne, 2002). They also readily make the affirmation of the consequent inference, from "there were orange trees" to "there were roses," again as often from the counterfactual as from the indicative conditional. The results indicate that participants have also envisaged the counterfactual conjecture, "there were roses and there were orange trees." Participants do not consider the modus ponens premise to be inconsistent with the counterfactual premise, nor do they consider the modus ponens inference to be contradictory and they do not baulk at making it. Instead, they appear to update the epistemic status of the possibility "there were roses and there were orange trees" from corresponding to a counterfactual conjecture to corresponding to the actual facts (e.g., Byrne, 2017). The willingness of participants to make inferences such as modus ponens and modus tollens from counterfactuals has been examined not only for basic concrete content, such as conditionals about people in places (e.g., Byrne & Tasso, 1999) but also for other sorts of content, such as causal and definitional counterfactuals (e.g., Frosch & Byrne, 2012; Thompson & Byrne, 2002), and deontic and inducement counterfactuals (e.g., Egan & Byrne, 2012; Quelhas & Byrne, 2003). They make such inferences for counterfactuals based on various linguistic forms including "there would have been roses *only if* there had been orange trees" and "*even if* there had been roses there would have been orange trees" (e.g., Egan, Garcia-Madruga, & Byrne, 2009; Moreno-Rios, Garcia-Madruga, & Byrne, 2008). The dual meaning of counterfactuals has been examined for counterfactuals about the past, and also subjunctive conditionals about the present, and pre-factual conditionals about the future (e.g., Byrne & Egan, 2004; Byrne & Tasso, 1999).

Hence, the evidence overwhelmingly indicates that people think about both the counterfactual conjecture and the presupposed facts when they understand a counterfactual. Of

course, there may be additional possibilities that are also consistent with a counterfactual conditional, and participants may think about them if prompted by the content or the task (Thompson & Byrne, 2002). Nonetheless, from the outset, people tend to think about what is mentioned in the counterfactual as the conjecture, and the opposite of what is mentioned as the presupposed facts.

Our aim is to examine how people represent the presupposed facts, the opposite of what is mentioned by a counterfactual such as "if there had been roses there would have been orange trees." We test whether people construct a simulation composed of alternates, for example, "poppies and apple trees" or whether they construct models of possibilities that can be annotated using propositional symbols, for example, "no roses and no orange trees." The question concerns how negation is represented when people understand counterfactuals. It is informed by the results of recent studies that indicate that people represent negation by constructing a representation that is iconic but that nonetheless can contain symbols for negation (e.g., Orenes et al., 2014).

1.2. Negation: Alternates or symbols

Recent studies have attempted to distinguish between alternative views of how people represent negation (e.g., Orenes et al., 2014). One contemporary view is that the meaning of a concept, such as roses, is grounded in or dependent on characteristics of sensory or motor processes; that is, its meaning is embodied in a modality-specific representational format that is experiential in nature (e.g., Barsalou, Simmons, Barbey, & Wilson, 2003; van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008), rather than based on linguistic symbols (e.g., Clark & Chase, 1972; Fodor, 1998; Trabasso, Rollins, & Shaughnessy, 1971). Embodied theories of conceptual meaning have been extended to abstract concepts, including negation (e.g., Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999; see also Waskan, 2006). Hence, "words, phrases, and mathematical and logical symbols all become meaningful through how we perceive and interact with the objects and situations those symbols denote" (Glenberg, 2010, p. 587).

How negation is understood in an embodied system of meaning is debated. In some situations, it may be transformed into an affirmation; for example, "Tom is not guilty" is transformed into its affirmative counterpart, "Tom is innocent," which activates information congruent with innocence, at least for negations that have well-defined opposites (Mayo et al., 2004). Hence, negation need not always entail prolonged processing times (e.g., Glenberg et al., 1999). Some negative sentences, for example, "the door is not closed," may allow immediate access to the actual situation, an open door, because their content refers to a binary alternative situation (e.g., Mayo et al., 2004). In such situations, "negation words may function just like any other context rather than requiring the use of logical operators" (Huette & Anderson, 2012, p. 300). Other situations, for example, "the door is not red," may first require a representation of the negated situation, a red door, which is then subsequently rejected (e.g., Mayo et al., 2004). In such multiple alternative situations, the problem is to understand, "what, if anything, it means for an eagle to not be in the sky. For an embodied account to function correctly, the eagle must be

somewhere, nowhere, everywhere, or perhaps a blend of several locations weighted by the frequency in which they have been encountered" (Huette & Anderson, 2012, p. 300). Even understanding negation in a binary situation, such as "the door is not closed" may be a two-step process over time in which people first create a representation of the negated situation, a closed door, and at a later point in the comprehension process shift attention to the actual situation, an open door (Kaup et al., 2006).

Sentential negation is associated with motor inhibition in the brain, at least for negated actions in imperative form (De Vega et al., 2016), which implies that when people understand a negation such as "there are no roses," the concept "roses" becomes inhibited. Negation may result in "disembodiment" or the reduced activation of an embodied simulation, resulting in "storing a concept in affirmative form in semantic memory, since its negative counterpart can be produced by transiently reducing the access to such stored semantic information" (e.g., Bartoli et al., 2013, p. 1782). There may be flexibility in language-induced motor recruitment such that the negation of an action word prevents the usual recruitment of motor structures for processing the action word (e.g., Aravena et al., 2012). On such accounts, "in a non-linguistic experiential representational format it is not possible to represent a linguistic operator such as negation explicitly" (Kaup et al., 2006, p. 1046), and so alternative mechanisms are required. The representation of the negated situation, such as a closed door, may be maintained in an auxiliary representational system that is not integrated with the representation of the actual situation (Kaup et al., 2006). Negated assertions such as "there is not a balloon above a cloud" may be represented by the failure of the attempted simulation of the information, by the absence of any binding between relevant entities (e.g., Barsalou, 1999; see also Hald, Hocking, Vernon, Marshall, & Garnham, 2013). On such views, "sentential negation blocks the mental representation of the information it denies, engendering a subjective experience of absence" (e.g., Bartoli et al., 2013, p. 1782).

Of course, if the polarity of an utterance directly reflected the nature of the underlying representation of the to-be-described state of affairs, then the mere fact that negation is often used in verbal communication would already be a counterargument against the idea that speakers create non-symbolic representations of negations. However, negation in everyday conversation is often used to deny, to communicate deviations from expectancies, to counter-argue, and so on (e.g., Kaup et al., 2006; Khemlani et al., 2012; Mayo et al., 2004; Wason & Jones, 1963).

A second contemporary view is a pluralist one that the meaning of concepts is captured in a simulation that is iconic but that can nonetheless contain symbols if necessary (e.g., Dove, 2009). Hence, the mental representation of negation may be as iconic as possible but may include symbolic annotations, such as a propositional-like tag, "no" or some other symbol to capture negation (e.g., Byrne & Johnson-Laird, 2009; Johnson-Laird & Byrne, 2002; Johnson-Laird, Byrne, & Schaeken, 1992; Khemlani, Orenes, & Johnson-Laird, 2014; Khemlani et al., 2012). For example, when people understand "Tom is not guilty," they may first think about the negated state of affairs, "Tom is guilty" which activates information congruent with "guilt," and only then do they attach a negation marker (e.g., Johnson-Laird & Tridgell, 1972; see also Clark & Chase, 1972;

Wason & Jones, 1963; Wason & Johnson-Laird, 1972). On this view, people construct iconic representations that nonetheless contain symbols, that is, annotated models (Orenes et al., 2014).

Both sorts of representations, specific alternates and explicit negation, could be captured in symbolic representations. Nonetheless, to discriminate between the two views, in one study participants were given information such as "the figure can be red or green," and they were shown an array in a visual world paradigm consisting of, for example, a red circle, a green triangle, a yellow square, and a blue diamond (Orenes et al., 2014). Eye-tracking measures examined where participants looked in the array when they were told "the figure is not red." In this binary context, the two theories make the same prediction: People will look at the green figure, and the results corroborated this prediction. The two theories make different predictions for a multiple context, that is, when participants are told, "the figure can be red or green or yellow or blue." The specific-alternates view predicts that participants will look at the green or yellow or blue figures; that is, people will scan the three figures other than the red one, or different individuals will look arbitrarily at one of the three figures other than the red one and so overall, the results for the group of participants will be distributed across the three figures other than the red one. The annotated models view makes the opposite, and somewhat counterintuitive, prediction. It predicts that when people are told the figure is not red, they will look at the red figure. They will look at it because they will modify it with some symbol, whether a propositional tag such as "no," or an image such as an X through the figure, or some other such operation (e.g., Johnson-Laird et al., 1992). The experimental results showed that people looked at the red figure. The findings support the pluralist view that in simple binary situations people construct iconic representations comprised of specific alternates, perhaps akin to embodied experiential simulations, but in more complex multiple situations, they can construct iconic representations that nonetheless contain symbols, that is, annotated models (Orenes et al., 2014). The experiments we report extend the logic of this approach to test alternative views of how people represent the presupposed facts for a counterfactual.

1.3. Counterfactuals: Alternates or symbols

The idea that the meaning of counterfactuals is embodied in a mental simulation based on sensory and motor processes is supported by the evidence that counterfactuals show an "action sentence compatibility" effect (e.g., De Vega & Urrutia, 2011). When participants read sentences such as "I *passed* the ball to him" and responded with a hand movement, such as a lever press, away from their body or toward their body, their response times differed depending on whether the direction of movement was compatible or incompatible with the direction of movement described in the sentence (e.g., Glenberg & Kaschak, 2002). The effect is interpreted as showing that the mental simulation of the action described in the sentence recruits the same resources as the physical action itself. An action sentence compatibility effect has been shown for counterfactuals such as "If I had been far away from the basket *I would have passed* the ball to another player" (e.g.,

De Vega & Urrutia, 2011). The length of time to respond is affected by the match or mismatch between the movement word and the movement response, just as for causal assertions.

Our experiments are designed to examine whether people understand a counterfactual by thinking about the presupposed facts by constructing a mental representation based on alternates, consistent with many embodied simulation accounts, or whether they construct a mental representation that represents negation explicitly, consistent with the idea of mental models annotated with symbols. We examined counterfactuals such as,

> If the flowers had been roses, the trees would have been orange trees.

We ensured that the counterfactuals contained concepts that were chosen to be concrete, and to potentially activate the sensory system, for example, "roses" is a concept which can activate the visual and olfactory sensory system (see Appendix A), although we do not intend to investigate representations that are directly dependent on sensory or motor processing. Given that our interest is in the inferences that reasoners make from counterfactual conditionals, our method does not rely on motor responses or brain activation measures, unlike studies of negation carried out to directly test embodied simulations. Nonetheless, we suggest an examination of counterfactual inferences has the potential to contribute converging evidence of relevance to current debates about how negation is represented. To distinguish the predictions of the specific alternates account from the annotated-models account we presented the counterfactuals in either a binary context, for example,

> At the botanical gardens the flowers were roses or poppies and the trees were orange trees or apple trees. If the flowers had been roses, the trees would have been orange trees.

or else in a multiple context, such as a multiple antecedent context, for example,

> At the botanical gardens the flowers were roses or poppies or lilies and the trees were orange trees or apple trees. If the flowers had been roses, the trees would have been orange trees.

or a multiple consequent context, for example,

> At the botanical gardens the flowers were roses or poppies and the trees were orange trees or apple trees or pear trees. If the flowers had been roses, the trees would have been orange trees.

Both the specific alternates and the annotated models views are similar in their implications for how people will represent a counterfactual in a binary context. People will construct a simulation in which the presupposed Facts are represented as alternates:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples

The specific alternates and annotated models views differ in their implications for how people represent a counterfactual in a multiple context. The specific alternates view suggests that people will attempt to enumerate possible alternates; for example, for the multiple antecedent context, they will think about each of the possible alternate antecedents:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples
Presupposed Facts:	Lilies	Apples

In contrast, the annotated models view proposes that because a multiple context requires multiple alternates to be represented, which places high demands on working memory, people can instead represent the information parsimoniously with symbols:

Counterfactual:	Roses	Oranges
Presupposed Facts:	No roses	No oranges

Hence, the annotated models view is a pluralist account that proposes that when the possibilities begin to exceed limited cognitive resources, people can avail of symbols to reduce load. We test the predictions of the two theories for inferences from counterfactual conditionals.

1.4. Inferences in binary and multiple contexts

We tested four sorts of inferences, which have been examined extensively with counterfactuals, and which participants make readily from them (e.g., Byrne & Tasso, 1999; Thompson & Byrne, 2002). One was the simple modus ponens inference, for which participants were told "The flowers were roses" and they were asked to say what, if anything, follows, by completing the sentence stem: "Therefore _____." For the binary context,

> At the botanical gardens the flowers were roses or poppies and the trees were orange trees or apple trees. If the flowers had been roses, the trees would have been orange trees.

both theories propose that people will think about the presupposed facts as alternates:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples

Hence, both theories predict that participants will tend to complete the modus ponens inference by making an affirmative conclusion, "Therefore the trees were orange trees." For example, a simple computer program would construct the models above and, given

the modus ponens premise: "the flowers were roses," it would match the information to the first model, eliminate the second model, update the epistemic status of the first model to correspond to facts, and conclude "the trees are orange trees." We expect that participants who make the inference are likely to construct an affirmative conclusion of this sort. Similarly for the affirmation of the consequent inference, "the trees were orange trees," participants who make the inference are likely to construct an affirmative conclusion, "Therefore the flowers were roses."

In addition to these two inferences with affirmative minor premises, we also gave participants two inferences with negative minor premises. One was the denial of the antecedent inference, for which they were told "the flowers were not roses." The novel prediction from both theories is that participants will tend to make an inference to the alternate and construct the affirmative conclusion, "Therefore the trees were apple trees," rather than the negative conclusion, "Therefore the trees were not orange trees." For example, a simple computer simulation that has constructed models of what is and is not in the botantic gardens as outlined earlier:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples

would integrate the information, "the flowers were not roses,"

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies (*no roses*)	Apples

and then eliminate the first model and conclude "the trees were apple trees." Similarly, for the modus tollens inference, "The trees were not orange trees," both theories make the novel prediction that participants will tend to make an inference to the alternate and construct an affirmative conclusion, "Therefore the flowers were poppies" rather than the negative conclusion, "Therefore the flowers were not roses." Hence, the representation of the presupposed facts as alternates will override any tendency to "match" the negation in the minor premise by constructing a conclusion that also contains a negation. The first experiment tests the novel prediction that such an inference-to-alternates effect should occur for inferences from counterfactuals in a binary context.

The two theories make different predictions about the presence of the inference-to-alternates effect in a multiple context:

> At the botanical gardens the flowers were roses or poppies and the trees were orange trees or apple trees or pear trees. If there had been roses there would have been orange trees.

The specific alternates view proposes that participants represent the counterfactual by thinking about multiple alternates:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples
Presupposed Facts:	Poppies	Pears

In contrast, the annotated models view proposes that participants represent the counterfactual by relying on symbols:

Counterfactual:	Roses	Oranges
Presupposed Facts:	No roses	No oranges

Both theories predict affirmative conclusions for the inferences with affirmative minor premises; for example, from the modus ponens minor premise "the flowers were roses," participants will make the affirmative conclusion, "Therefore the trees were orange trees." But the two theories make different predictions for the inferences with negative minor premises. For the denial of the antecedent minor premise, for example, "there were no roses," participants could construct an affirmative conclusion, "there were apple trees or pear trees" or a negative conclusion, "there were no orange trees." If participants have represented the presupposed facts as alternates, they will construct the affirmative conclusion. But if they have represented the presupposed facts by annotated models with symbols, they will construct the negative conclusion. For example, a simple program to simulate the specific alternates view would construct models with multiple alternates in the multiple context:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples
Presupposed Facts:	Poppies	Pears

Given the modus ponens premise: "the flowers were roses," it would match the information to the first model, eliminate the second and third, update the epistemic status to be facts, and conclude "the trees were orange trees." Given the denial of the antecedent premise, "the flowers were not roses," it would match the information to the second model and the third model, eliminate the first, and conclude, "the trees were apple trees or pear trees." In contrast, a simple program to simulate the annotated-models view would construct models with symbols in the multiple context:

Counterfactual:	Roses	Oranges
Presupposed Facts:	No roses	No oranges

Given the modus ponens premise: "the flowers were roses," it would match the information to the first model, eliminate the second, update the epistemic status to be facts, and conclude "the trees were orange trees." Given the denial of the antecedent premise, "the flowers were not roses," it would match the information to the second model, eliminate the first, and conclude, "the trees were not oranges." Hence, the two theories differ in their predictions for the multiple context—the specific alternates view predicts that the inference-to-alternates effect will be observed in the multiple context, whereas the

annotated-models theory predicts that the inference-to-alternates effect will not be observed in the multiple context. According to the annotated-models view, since participants can construct annotated-models the inference-to-alternates effect will be reduced, that is, participants will make at least as many negative as affirmative conclusions.

The logic of our methodological rationale rests on the assumption that the form of the verbalized conclusion, affirmative or negative, reflects the nature of the mental representation that is used to derive the conclusion, a mental representation of alternates or one with symbols. This assumption is comparable to other assumptions in studies of reasoning, for example, the assumption that the possibilities that people list as consistent with a conditional correspond to the possibilities that they have mentally represented (for a recent review see Johnson-Laird, Khemlani, & Goodwin, 2015). Nonetheless, we return to this assumption in later experimental tests. We test the predictions of the two theories for binary context in Experiment 1 and for multiple context in Experiment 2, and then we test the effects of binary and multiple context on each other.

2. Experiment 1: Binary context

The aim of the experiment was to test the novel prediction of an inference-to-alternates effect for counterfactuals; that is, that participants will tend to construct affirmative conclusions that refer to alternates, even for negative minor premises, in a binary context. Given a binary context, the specific alternates view and the annotated-models view both propose that people think about the presupposed facts as alternates, as Table 1 shows. Hence, both theories predict that participants will produce affirmative conclusions to negative minor premises, such as modus tollens, "there are no orange trees therefore there are poppies" and denial of the antecedent, "there are no roses therefore there are apple trees." Both theories also predict that participants will produce affirmative conclusions to affirmative minor premises, of course, such as modus ponens, "there are roses therefore there are orange trees" and affirmation of the consequent, "there are orange trees therefore there are roses."

To control for the presence of negatives in the minor premises, we also manipulated whether the minor premise made direct or indirect reference to the objects in the counterfactual. For example, in the binary context,

> At the botanic gardens the flowers were roses or poppies and the trees were orange trees or apple trees. If the flowers had been roses, the trees would have been orange trees

modus ponens with *direct* reference to the counterfactual's terms contains the minor premise, "the flowers were roses," whereas modus ponens with *indirect* reference contains the minor premise, "the flowers were not poppies." Modus tollens with direct reference to the counterfactual's terms contains the minor premise, "The trees were not orange trees," whereas modus tollens with indirect reference contains the minor premise, "The trees were apples trees," as Table 1 shows. We wished to control for the

Table 1
Binary context: The form of the inferences used in the experiments for a binary context is illustrated for the counterfactual "if there had been A, there would have been D." The four minor premises for direct reference are presented first and for indirect reference second. The form of the conclusions that were scored as affirmative or negative (with the predicted conclusion in bold) is indicated

Counterfactual If there had been A, there would have been D	Binary context: A or B and D or E If there had been A, there would have been D	
Example	*The flowers were roses or poppies and the trees were orange trees or apple trees* *If there had been roses, there would have been orange trees*	
Models for binary context and counterfactual given direct or indirect reference minor premise	*Counterfactual: roses oranges* *Presupposed facts: poppies apples*	
	Minor Premise	Affirmative Conclusion	Negative Conclusion
Direct reference inference			
Modus ponens	A	D	Not-E
	Roses	**Oranges**	Not-apples
Deny antecedent	Not-A	E	Not-D
	Not-roses	**Apples**	Not-oranges
Affirm consequent	D	A	Not-B
	Oranges	**Roses**	Not-poppies
Modus tollens	Not-D	B	Not-A
	Not-oranges	**Poppies**	Not-roses
Indirect reference inference			
Modus ponens	Not-B	D	Not-E
	Not-poppies	**Oranges**	Not-apples
Deny antecedent	B	E	Not-D
	Poppies	**Apples**	Not-oranges
Affirm consequent	Not-E	A	Not-B
	Not-Apples	**Roses**	Not-poppies
Modus tollens	E	B	Not-A
	Apples	**Poppies**	Not-roses

presence of negatives so that we could examine whether an inference-to-alternates effect is observed for all inference types, that is, not only for the inferences that are usually considered negative—modus tollens and denial of the antecedent with direct reference—but also for inferences that are usually considered affirmative—modus ponens and affirmation of the consequent, which can be rendered negative with indirect reference. Both theories make the same predictions for indirect reference as for direct reference in a binary context.

For example, a simple computer simulation of annotated-models would construct the models:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples

It would integrate the information about what is not in the botanic gardens in the minor premise for modus ponens with indirect reference, "the flowers were not poppies" with the information in its models about what is in the botanic gardens (roses), eliminate the second model, update the epistemic status of the first model to correspond to facts, and conclude "therefore there are orange trees." Likewise, it would integrate the information in the modus tollens minor premise with indirect reference, "The trees were apples trees," by matching it to the second model, it would eliminate the first model, and conclude, "the flowers were poppies." Hence, participants will construct affirmative conclusions even to negative minor premises, for indirect reference as well as for direct reference, as Table 1 shows.

2.1. Method

2.1.1. Materials and design

Participants acted as their own controls in a within-participants design with two variables, that is, a 4 (inference: modus ponens, modus tollens, denial of the antecedent, affirmation of the consequent) × 2 (reference: direct or indirect) design. Each problem consisted of a disjunctive premise describing the binary context, followed by the counterfactual premise, and the minor categorical premise. Participants completed 16 inferences comprised of 2 instances of each of the 8 conditions. We constructed 16 different contents, which are provided in Appendix A; they were assigned at random to the inferences, and the order of the terms in the antecedent and consequent of the problems was counterbalanced.

2.1.2. Procedure

Participants were tested in small groups of about 10 participants in each group. Each participant received a booklet containing the task instructions, illustrated by an example, which informed them that the experiment aimed to examine how people reason with different logical rules. They were told that their task was to draw a conclusion that necessarily followed from the three premises, that is, a conclusion that is true given that the premises are true. They were also told that if they considered that there was no necessary conclusion, they were to write "there is no conclusion." They were asked to work on each problem in the order in which it appeared in the booklet and not to return to any problem or change any answer when they had completed it.

2.1.3. Participants

The 32 participants who took part in the experiment were undergraduate students in psychology at the University of La Laguna, Tenerife, Spain, and hence 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.

2.2. Results and discussion

We recorded whether participants made the standard endorsement of the inference as an affirmative or negative conclusion. For example, for the denial of the antecedent, participants received problems of the form: "A or B, and D or E. If A then D. Not-A." We recorded participants' conclusions as a negative endorsement if it corresponded to "therefore not-D," and as an affirmative endorsement if it corresponded to "therefore E" (see Table 1). The data files for each of the nine experiments are available at https://reasoningandimagination.wordpress.com/data-archive/ and also at the Open Science Framework at https://osf.io/8wg2v/

As Fig. 1 shows, participants showed a robust inference-to-alternates effect; that is, they drew more affirmative conclusions than negative conclusions, not only for the modus ponens and affirmation of the consequent affirmative minor premises, but also for the modus tollens and denial of the antecedent negative minor premises. Participants generated more affirmative conclusions than negative ones for each of the four inferences that used direct reference, binomial sign test $p < .0001$ for each of the four comparisons, and for each of them with indirect reference, binomial sign test $p < .0001$ for each of the four comparisons.

Table B1 in the Supplemental Materials lists the percentages of the main types of responses made by participants. We carried out a 4 (inference: modus ponens, denial of the antecedent, affirmation of the consequent, modus tollens) × 2 (reference: direct or indirect) repeated measures ANOVA on the affirmative conclusions, with a Greenhouse–Geisser correction for the violation of sphericity assumption. There was no main effect of inference, $F(1.751, 54.285) = 1.07$, $MSE = .11$, $p < .34$, or reference, $F < 1$, but the two variables interacted, $F(3, 93) = 3.63$, $MSE = .05$, $p < .02$, $\eta_p^2 = .11$. We decomposed the interaction to test for potential differences between direct and indirect inferences, using a Bonferroni corrected alpha of .0125 for the four comparisons, and found that there were

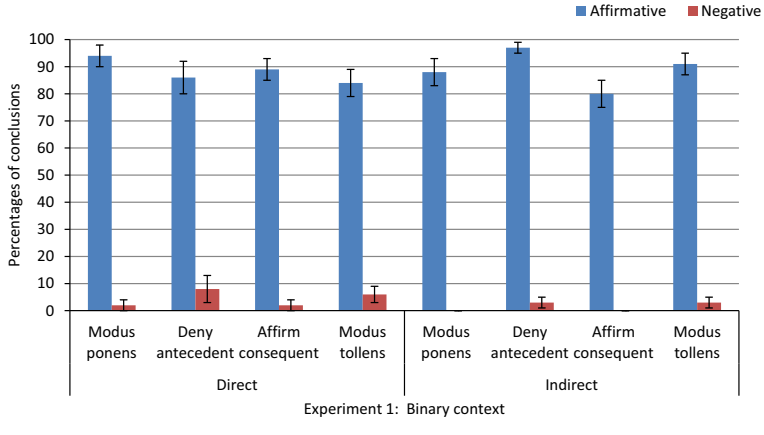

Fig. 1. Percentages of affirmative and negative conclusions in Experiment 1 for counterfactuals in a binary context. Error bars are standard error of the mean.

no differences between direct and indirect reference for any of the inferences—modus ponens, 94% versus 86%; $t(31) = 1.43$, $p = .16$, modus tollens, 84% versus 91%; $t(31) = 1.27$, $p = .21$, denial of the antecedent, 86% versus 97%; $t(31) = 1.87$, $p = .07$, or affirmation of the consequent, 90% versus 80%; $t(31) = 1.89$, $p = .06$. (The interaction arises from differences between the inferences, participants made fewer affirmative conclusions to affirmation of the consequent than denial of the antecedent inferences for indirect minor premises, 80% versus 97%; $t(31) = 3.23$, $p < .004$, $d = .57$, and no other difference between inferences was significant on the corrected alpha, largest t test = 2.03, smallest $p = .051$.)

The results reveal an inference-to-alternates effect, that is, a tendency to draw an affirmative conclusion that refers to an alternate even from a negative minor premise, in a binary context. For example, given the binary context, "At the botanic gardens there were roses or poppies, and there were orange trees or apple trees," with the counterfactual, "If there had been roses, then there would have been orange trees," and the negative minor premise "there were no roses," participants construct an affirmative conclusion, "there were apple trees" rather than a negative conclusion, "there were no orange trees." The result is consistent with both the specific alternates view and the annotated-models view that participants represent the opposite of what is mentioned in the counterfactual, that is, the presupposed facts, by envisaging alternates; that is, they envisage the conjectured possibility "roses and orange trees" and the presupposed facts as "poppies and apple trees." In the next experiment we examine multiple context, for which the two theories make different predictions.

3. Experiment 2: Multiple context

The aim of the experiment was to test the predictions of the specific alternates view and the annotated-models view for a multiple context, for example,

> At the botanical gardens the flowers were roses or poppies or lilies and the trees were orange trees or apple trees. If there had been roses there would have been orange trees.

The two theories propose that people construct different sorts of representations for the multiple context. The specific alternates view proposes that they represent the counterfactual by thinking about multiple alternates:

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples
	Lilies	Apples

Hence, it predicts an inference-to-alternates effect for multiple context; that is, participants will construct affirmative conclusions even to negative minor premises. For example, when participants are given the modus tollens minor premise, "there were no orange trees" they will conclude "there were poppies or lilies," in the multiple context. In

contrast, the annotated-models view proposes that participants can represent the counterfactual by relying on symbols:

Counterfactual:	Roses	Oranges
Presupposed Facts:	No roses	No oranges

Working memory constraints ensure that participants may not attempt to enumerate the multiple alternates for the presupposed facts in the multiple context, and instead they can switch to representing the information using symbols. Accordingly the annotated-models view predicts that participants will construct as many negative conclusions as affirmative conclusions to negative minor premises. For example, when they are given "there were no orange trees" they may conclude "there were no roses." The annotated-models view predicts that the inference-to-alternates effect will be at least reduced in a multiple context, as Table 2 shows. Hence, the two theories make different predictions: The specific alternates view predicts an inference-to-alternates effect for multiple context, that is, participants will make affirmative conclusions to negative premises, whereas the annotated-models view predicts the inference-to-alternates effect will be reduced, that is, participants will make as many negative conclusions as affirmative conclusions to negative premises. The aim of Experiment 2 was to test these predictions.

We used two sorts of multiple context. We used a multiple antecedent context,

> At the botanical gardens the flowers were roses or poppies or lilies and the trees were orange trees or apple trees

for the affirmation of the consequent and modus tollens inferences that affirm or deny the consequent, for example, "there were no orange trees," so that the conclusion that participants had to construct was about the multiple alternatives, for example, "there were poppies or lilies" or "there were no roses." We used a multiple consequent context, for example,

> At the botanical gardens the flowers were roses or poppies and the trees were orange trees or apple trees or pear trees

for the modus ponens and denial of the antecedent inferences that affirm or deny the antecedent, for example, "there were no roses," so that the conclusion that participants had to construct was about the multiple alternatives, for example, "there were apple trees or pear trees" or "there were no orange trees."

Once again we included minor premises that made indirect reference to the counterfactual's terms, as well as minor premises that made direct reference. This time the annotated-models view makes a different prediction for indirect reference compared to direct reference in the multiple context case. Given a multiple context,

Table 2
Multiple context: The form of the inferences used in the experiments for a multiple context is illustrated for the counterfactual "if there had been A, there would have been D." The four minor premises for direct reference are presented first and for indirect reference second. The form of the conclusions that were scored as affirmative or negative (with the predicted conclusion in bold) is indicated

Counterfactual If A had been, D would have		Multiple (antecedent): A or B or C and D or E If there had been A, there would have been D		Multiple (consequent): A or B and D or E or F If there had been A, there would have been D	
Abbreviated Example		*Roses or poppies or lilies and oranges or apples. If there had been roses, there would have been oranges*		*Roses or poppies and oranges or apples or pears. If there had been roses, there would have been oranges*	
Annotated models		*CF: roses oranges* *F: no roses no oranges*		*CF: roses oranges* *F: no roses no oranges*	
	Minor Premise	Affirmative Conclusion	Negative Conclusion	Affirmative Conclusion	Negative Conclusion
Direct reference inference					
Modus ponens	A Roses			D **Oranges**	Not E or F *Not apples or pears*
Deny antecedent	Not-A Not-roses			E or F *Apples or pears*	Not-D **Not-oranges**
Affirm consequent	D Oranges	A **Roses**	Not B or C *Not poppies or lillies*		
Modus tollens	Not-D Not-oranges	B or C *Poppies or lilies*	Not-A **Not-roses**		
Models		*CF: roses oranges* *F: poppies apples* *F: lilies apples*		*CF: roses oranges* *F: poppies apples* *F: poppies pears*	
	Minor Premise	Affirmative Conclusion	Negative Conclusion	Affirmative Conclusion	Negative Conclusion
Indirect reference inference					
Modus ponens	Not-B Not-poppies			D **Oranges**	Not E or F *Not apples or pears*
Deny antecedent	B Poppies			E or F **Apples or pears**	Not-D *Not-oranges*
Affirm consequent	Not-E Not-Apples	A **Roses**	Not B or C *Not poppies or lilies*		
Modus tollens	E Apples	B or C **Poppies or lilies**	Not-A *Not-roses*		

Key: Multiple antecedent context was used for the affirmation of the consequent and modus tollens inferences which draw a conclusion about the antecedent; multiple consequent context was used for the modus ponens and denial of the antecedent inferences which draw a conclusion about the consequent.

At the botanic gardens the flowers were roses or poppies or lilies and the trees were orange trees or apple trees. If there had been roses, then there would have been orange trees

the annotated-models view proposes that participants construct models with symbols:

Counterfactual:	Roses	Oranges
Presupposed Facts:	No roses	No oranges

When they are given a minor premise with indirect reference, for example, for modus tollens, "there were apple trees," the information does not match any information in their models. Unlike any of the inferences encountered thus far, there are no referents in common between the models and the minor premise (apple trees). Hence, they must add to their models further information to represent the presupposed facts as alternates instead, "poppies and apple trees" and "lilies and apple trees." Hence, they will construct the affirmative conclusion "there were poppies or lilies." Similarly, when they are given the indirect minor premise for the affirmation of the consequent "there were no apple trees," the information does not match any information in their models. They must add to their models further information to represent the presupposed facts as alternates instead and so they will produce the affirmative conclusion "there were roses." Hence, the annotated-models view predicts that when participants are given a multiple context but the minor premise has indirect reference, they will exhibit the inference-to-alternates effect and construct affirmative conclusions even to negative minor premises, as Table 2 shows.

In summary, the specific alternates view predicts an inference-to-alternates effect in a multiple context for both direct and indirect reference; the annotated-models view predicts the inference-to-alternates effect will be reduced in a multiple context for direct reference but not for indirect reference.

3.1. Method

3.1.1. Materials, design, and procedure

The same materials, design, and procedure were used as the previous experiment, with the exception that the disjunctive premise presented a multiple context. We gave participants a multiple antecedent context, for example, "At the botanical gardens the flowers were roses or poppies or lilies and the trees were orange trees or apple trees" for the affirmation of the consequent and modus tollens inferences which draw a conclusion about the antecedent; we gave participants a multiple consequent context, for example, "At the botanical gardens the flowers were roses or poppies and the trees were orange trees or apple trees or pear trees" for the modus ponens and denial of the antecedent inferences which draw a conclusion about the consequent. We did not use both a multiple antecedent and multiple consequent disjunction in the context because we believed it might exceed the limits of comprehension for our participants.

3.1.2. Participants

The 32 participants who took part in the experiment were undergraduate students at the University of La Laguna, Tenerife, Spain. None had formal training in logic. They had not taken part in the previous experiment.

3.2. Results and discussion

We used the same scoring of the standard endorsement of the inference as an affirmative or negative conclusion as in the previous experiment. For example, in the multiple context for the denial of the antecedent, participants received problems of the form: "A or B, and C or D or E. If A then C. Not-A." We recorded a participant's conclusion as a negative endorsement if it corresponded to "therefore not-C," and as an affirmative complete endorsement if it corresponded in meaning to "therefore D or E" and an affirmative partial endorsement if it corresponded to "therefore D" or "therefore E" (see Table 2).

As Fig. 2 shows, the inference-to-alternates effect is reduced for multiple context inferences, for direct reference but not for indirect reference. Participants drew more affirmative conclusions than negative conclusions, to the modus ponens and affirmation of the consequent premises, binomial sign test $p < .0001$ for each of the four comparisons for direct and indirect reference, as both theories predict; however, they drew as many negative conclusions as affirmative ones to the modus tollens and denial of the antecedent inferences, in the direct reference condition (denial of the antecedent $p = .61$, modus tollens $p = 1.0$); they drew more affirmative than negative conclusions in the indirect reference condition, binomial sign test $p < .0001$ for both comparisons.

The results show that given a multiple context, such as "At the botanic gardens, there were roses or poppies and there were orange trees or apple trees or pear trees" with the

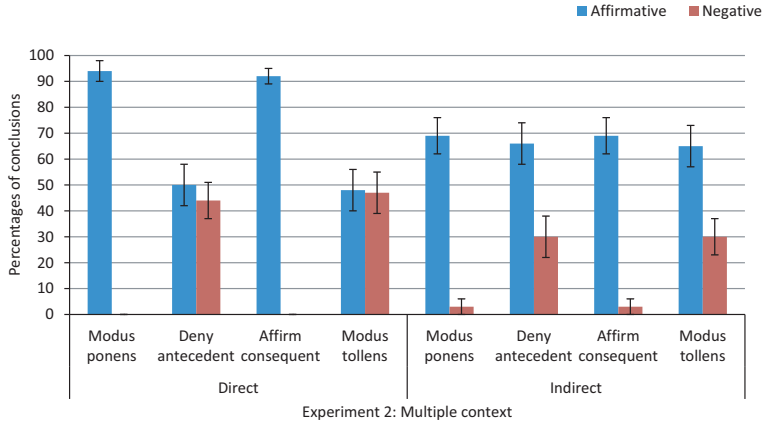

Fig. 2. Percentages of affirmative and negative conclusions in Experiment 2 for counterfactuals in a multiple context. Error bars are standard error of the mean.

counterfactual, "If there had been roses then there would have been orange trees," and a negative minor premise, such as the denial of the antecedent one, "There were no roses," participants construct a negative conclusion, "there were no orange trees" as often as an affirmative one, "there were apple trees or pear trees," in this direct reference condition. The result corroborates the prediction of the annotated-models view that people can switch to representing the presupposed facts by relying on symbols, for example, "no roses and no orange trees," when the multiple possibilities exceed working memory constraints. The inference-to-alternates effect is reduced although not reversed in a multiple context, confirming that the tendency to represent the information as alternates persists even for multiple contexts in some instances.

An ANOVA of the same design as the previous experiment on the affirmative conclusions showed a main effect of inference, $F(1.364, 42.299) = 7.34$, $MSE = .35$, $p < .007$, $\eta_p^2 = .19$, no main effect of reference, $F(1, 31) = 1.26$, $MSE = .07$, $p = .27$, and an interaction between the two variables, $F(1.599, 49.583) = 10.11$, $MSE = .16$, $p < .002$, $\eta_p^2 = .25$ (see Table B2 in the Supplemental Materials). We decomposed the interaction to test for the expected differences between direct and indirect inferences, using a Bonferroni corrected alpha of .0125 for the four comparisons, and found that participants drew more affirmative conclusions in the direct reference condition than the indirect one, for modus ponens, 94% versus 69%; $t(31) = 3.36$, $p < .003$, $d = .59$, and affirmation of the consequent inferences, 92% versus 69%; $t(31) = 2.90$, $p < .009$, $d = .51$. In contrast, they drew somewhat more affirmative conclusions in the indirect reference condition than the direct one for the denial of the antecedent, 66% versus 50%; $t(31) = 2.39$, $p < .025$, $d = .42$ and modus tollens inference, 66% versus 48%; $t(31) = 2.47$, $p < .02$, $d = .43$, although the differences are marginal on the corrected alpha. The results are consistent with the elimination of the inference-to-alternates effect for the direct reference inferences, and its continued exhibition for the indirect reference inferences. (There were also differences between some of the inferences: modus ponens versus denial of antecedent, $t(31) = 2.68$, $p < .013$, $d = .47$, and versus modus tollens, $t(31) = 2.77$, $p < .01$, $d = .49$, and affirmation of the consequent versus denial of the antecedent, $t(31) = 2.90$, $p < .008$, $d = .51$, and versus modus tollens, $t(31) = 2.99$, $p < .006$, $d = .53$.)

Experiment 2 showed that in a multiple context, the inference-to-alternates effect is reduced and participants construct as many negative conclusions as affirmative conclusions to negative minor premises, and they construct affirmative conclusions to affirmative minor premises, for direct reference minor premises. For indirect reference minor premises, the inference-to-alternates effect was observed. The results are consistent with the predictions of the annotated-models view, which proposes that participants can represent the presupposed facts by alternates, but when the number of possibilities exceeds working memory limitations, they can instead represent them by symbols. This flexibility to represent by symbols allows them to make negative conclusions to the negative minor premises readily for the direct minor premises, but for the indirect minor premises it results in a mismatch between the representation and the minor premise that needs to be further remedied.

The experiment shows that participants can in some cases switch from mentally representing the presupposed facts as alternates to representing them with symbols. It is noteworthy, however, that the inference-to-alternates effect is reduced rather than reversed in the multiple context direct reference condition. Participants make as many negative conclusions as affirmative ones, rather than switching entirely to making more negative conclusions than affirmative ones. The result suggests that the switch to representing by symbols occurs for some participants on some trials, rather than uniformly for all participants on all trials. About half of the participants, 44%, provided only one sort of conclusion, either affirmative or negative, to all 8 trials of the modus tollens and denial of the antecedent inferences; that is, 11 participants produced an affirmative conclusion to the 8 modus tollens and denial of the antecedent inferences and 3 participants produced a negative conclusion to them. The other approximately half of the participants, 53%, produced different sorts of conclusions; that is, these 17 participants produced affirmative conclusions to some modus tollens and denial of the antecedent inferences and negative conclusions to others (and 1 participant produced "no valid conclusion" to them). The result may reflect the influence of different experiences or capacities, and the next experiment tests predictions about how experience with binary and multiple context influences the inference-to-alternates effect. The discovery that participants can in some cases switch from alternates to symbols in their mental representation of the presupposed facts of a counterfactual conditional implies that their mental representation by alternates or symbols is not invariant. Accordingly, the aim of the next experiment was to test predictions about the effects of making inferences in a binary context on subsequently making inferences in a multiple context.

4. Experiment 3: Binary context first, multiple context second

The aim of the experiment was to examine whether the inference-to-alternates effect, the tendency to generate affirmative conclusions from negative minor premises, is influenced by experience. To do so, we adopted the logic of the "mental set" paradigm (e.g., Luchins, 1942; Ollinger, Jones, & Knoblich, 2008). "Mental set" is the observation that when participants are presented with a set of similar problems that induces one strategy, and then they are presented with a set of different problems which they usually tend to solve by a second strategy, they nonetheless continue to rely on the first strategy even if the second strategy would be simpler. Hence, we provided participants with a set of inferences in a binary context, followed by a set in a multiple context. Our aim was to examine whether participants' experience of representing the counterfactual by alternates in a binary context would influence them to continue to represent the counterfactual by alternates even in a multiple context. If representation by alternates or symbols is a strategy rather than an invariant process, we expect that participants' experience with inferences in a binary context will influence their responses to inferences in a multiple context, and hence that they will exhibit the inference-to-alternates effect not only in binary contexts but also in multiple contexts.

4.1. Method

4.1.1. Design, materials, and procedure

The design was a within-participants one with two variables: 4 (inference: modus ponens, modus tollens, denial of the antecedent, affirmation of the consequent) × 2 (context: binary or multiple). The materials and procedure were the same as the previous experiments, with two exceptions: one was that participants received two booklets, the first booklet contained all binary context problems and the second booklet contained all multiple context problems; the second exception was that all of the problems contained direct reference rather than indirect reference. Hence, participants received 16 inferences, 8 binary context inferences with direct reference and then 8 multiple context inferences with direct reference.

4.1.2. Participants

The 31 participants who took part in the experiment were undergraduate students at the University of La Laguna, Tenerife, Spain. None had formal training in logic. They had not taken part in the previous experiments.

4.2. Results and discussion

As Fig. 3 shows, participants showed a robust inference-to-alternates effect; that is, they drew more affirmative conclusions than negative conclusions, for each of the four inferences in the binary context, binomial sign test $p < .0001$ for each of the four comparisons. This time they also showed a robust inference-to-alternates effect for the multiple context inferences; that is, they also drew more affirmative conclusions than negative conclusions, for each of the four inferences in the multiple context, even the modus

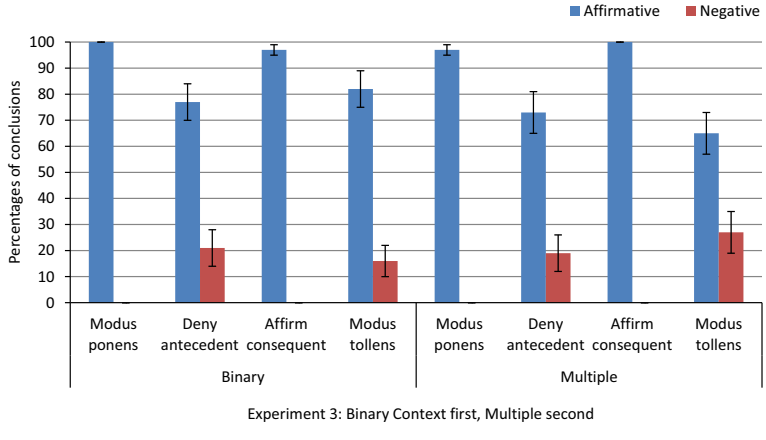

Fig. 3. Percentages of affirmative and negative conclusions in Experiment 3 for counterfactuals in a binary context first, multiple context second, for direct reference. Error bars are standard error of the mean.

tollens and denial of the antecedent inferences, binomial sign test $p < .0001$ for each of the four comparisons. The result shows that participants' experience of representing the counterfactual by alternates in a binary context influences them to continue to represent the counterfactual by alternates even for a multiple context. The discovery supports the idea that representation by alternates or symbols is a strategy rather than an invariant process.

We carried out a 4 (inference: modus ponens, denial of the antecedent, affirmation of the consequent, modus tollens) × 2 (context: binary or multiple) repeated measures ANOVA on the affirmative conclusions, with a Greenhouse–Geisser correction for the violation of sphericity assumption. It showed a main effect for inferences, $F(3, 90) = 15.20$, $MSE = .08$, $p < .001$, $\eta_p^2 = .34$, due to more affirmative conclusions to modus ponens and affirmation of the consequent inferences (98% in each case) than denial of the antecedent and modus tollens inferences (75% and 73%). There was no main effect of context, $F(1, 30) = 1.98$, $MSE = .10$, $p < .16$, as participants showed the inference-to-alternates effect to both the binary and the multiple context, and there was no interaction, $F(2.02, 60.602) = 2.17$, $MSE = .08$, $p = .12$ (see Table B3 in the Supplemental Materials). We note that the percentages for negative conclusions for denial of the antecedent and modus tollens appear somewhat increased in Fig. 3 in binary contexts in comparison to the direct reference condition of Experiment 1, and we conjecture that the apparent increase may reflect random variation.

The results show that when participants gain experience with the binary context inferences first, they subsequently respond to the multiple context inferences in the same way as they respond to the binary context inferences, by generating affirmative conclusions rather than negative ones, thus exhibiting a robust inference-to-alternates effect. The experiment shows that participants' experience of representing the counterfactual by alternates in a binary context influenced them to continue to represent the counterfactual by alternates even for a multiple context. In the next experiment we impose a more stringent test on the hypothesis that representation by alternates or symbols is a strategy: We examine whether experience with inferences in a multiple context affects inferences in a binary context.

5. Experiment 4: Multiple context first, binary context second

We provided participants with a set of inferences with a multiple context first, followed by a set with a binary context. Once again, if representation by alternates or symbols is a strategy rather than an invariant process, we expect that participants' experience with multiple context will influence their inferences in the binary context. If it does, we expect that the inference-to-alternates effect will be reduced even for binary context, that is, participants will draw affirmative conclusions to affirmative minor premises, but they will draw negative conclusions as often as affirmative ones to negative minor premises, not only to multiple context inferences but also to binary context ones. Hence, the aim of the experiment was to test the idea that participants' experience of representing the

presupposed facts in models with symbols in a multiple context would influence them to continue to represent them with symbols even for the simpler binary context.

5.1. Method

5.1.1. Design, materials, and procedure

The design, materials, and procedure were the same as the previous experiment, with the exception that participants started with the multiple context inferences first and then they received the binary context inferences second.

5.1.2. Participants

The 32 participants who took part in the experiment were undergraduate students at the University of La Laguna, Tenerife, Spain. None had formal training in logic. They had not taken part in the previous experiments.

5.2. Results and discussion

As Fig. 4 shows, the inference-to-alternates effect was reduced not only for the multiple context inferences but also for the binary context ones. Participants made more affirmative than negative conclusions to the modus ponens and affirmation of the consequent inferences, in both the binary and the multiple context, binomial sign test, $p < .0001$ for each of the four comparisons. They did not make more affirmative conclusions than negative conclusions to modus tollens or the denial of the antecedent inferences in the multiple context (denial of the antecedent $p = .11$, modus tollens $p = .60$), and this time, neither did they make more affirmative than negative conclusions for these inferences in the binary context (denial of the antecedent $p = .12$, modus tollens $p = .11$). The results show that when participants gain experience with the multiple context inferences first, they subsequently respond to the

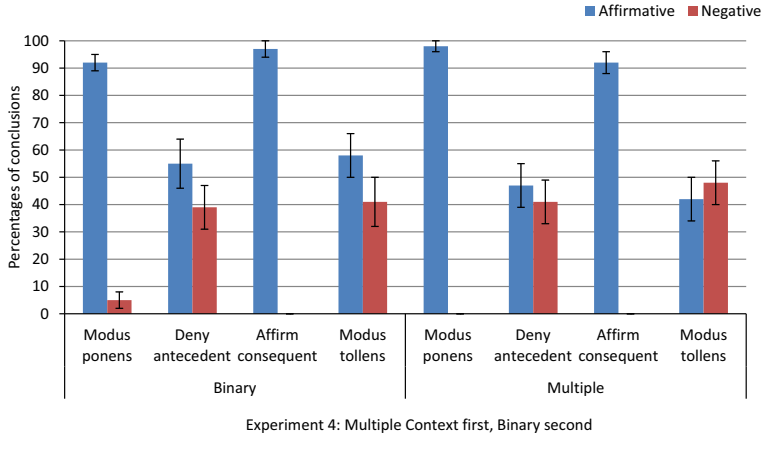

Fig. 4. Percentages of affirmative and negative conclusions in Experiment 4 for counterfactuals in a multiple context first, binary context second, for direct reference. Error bars are standard error of the mean.

binary context inferences in the same way as they respond to the multiple context inferences, by constructing negative conclusions as often as affirmative ones for the modus tollens and denial of the antecedent inferences. The results show that their experience representing the counterfactual by symbols in a multiple context influences them to continue to represent it by symbols even for a binary context.

An ANOVA of the same design as the previous experiment on the affirmative conclusions showed a main effect for type of inferences, $F(1.527, 47.331) = 29.44$, $MSE = .28$, $p < .001$, $\eta_p^2 = .49$, as participants made more affirmative conclusions to modus ponens and the affirmation of the consequent inferences (95% in each case) than denial of the antecedent and modus tollens inferences (51% and 50%, respectively). There was no main effect for type of context, $F(1, 31) = 2.50$, $MSE = .08$, $p < .12$, as the inference-to-alternates effect was reduced for both the binary and the multiple context, and the two variables did not interact, $F(2.266, 70.258) = 2.21$, $MSE = .08$, $p < .11$ (see Table B4 in the Supplemental Materials).

The experiment provides a strong test of the hypothesis that the tendency to represent the presupposed facts of a counterfactual by alternates or symbols is a strategy rather than an invariant process: It shows that the inference-to-alternates effect is reduced for binary context when participants gain experience with the multiple context inferences first; that is, participants respond to the binary context inferences in the same way as they respond to the multiple context ones, by generating negative conclusions as often as affirmative ones to the negative minor premise inferences.

One possibility is that the results of Experiments 3 and 4 may arise simply because of a type of syntactic priming, that is, that processing an utterance of a particular form facilitates processing a subsequent utterance with a similar form in a similar way (e.g., Pickering & Branigan, 1999). In the next experiment we examine mixed binary and multiple contexts. If the effects observed in Experiments 3 and 4 are merely syntactic priming then the inference-to-alternates effect in binary context should be diluted by the presentation of binary context inferences mixed at random with multiple context inferences, and similarly, the reduction in the inference-to-alternates effect for multiple context inferences should also be diluted. If instead, the effects observed in Experiments 3 and 4 arise from a mental set, then when this set is removed by the mixed presentation of binary and multiple context inferences, the observation of the inference-to-alternates effect for binary context inferences should be similar to its observation in Experiment 1, and the observation of the reduction in the inference-to-alternates effect in multiple context inferences should be similar to its observation in Experiment 2.

6. Experiment 5: Mixed binary and multiple contexts

The aim of the experiment was to examine mixed binary and multiple contexts, to test how participants respond to inferences in binary and multiple contexts when they are presented at random and so their experience of them is entirely mixed. The experiment allows us to investigate how robust the inference-to-alternates effect is when participants

have varied experiences. If the inference-to-alternates effect is merely the result of a superficial strategy that participants induce during the course of an experiment, or some sort of syntactic priming, then we expect it to be diluted when participants receive binary and multiple context inferences in a random order. However, if the inference-to-alternates effect reflects a more enduring strategic approach to representational constraints, then we expect it to occur for the binary context inferences, and to be reduced for the multiple context inferences, even in a mixed design. We expect that participants will respond to the binary context inferences by constructing affirmative conclusions even to negative minor premises, in the way they did in Experiment 1, and that they will respond to the multiple context inferences by constructing negative conclusions to negative minor premises as often as affirmative ones, in the way they did in Experiment 2.

6.1. Method

6.1.1. Design, materials, and procedure

The design, materials, and procedure were the same as the previous experiment, with the exception that participants received the binary and multiple context inferences in random order.

6.1.2. Participants

The 32 participants who took part in the experiment were undergraduate students at the University of La Laguna, Tenerife, Spain. None had formal training in logic. They had not taken part in the previous experiments.

6.2. Results and discussion

As Fig. 5 shows, participants exhibited an inference-to-alternates effect for binary context inferences, as they did in Experiment 1, and the effect was reduced for multiple context inferences, as it was in Experiment 2. Participants drew more affirmative than negative conclusions to modus ponens and affirmation of the consequent inferences, for both binary and multiple context, binomial sign test, $p < .0001$ for each of the four comparisons. They drew more affirmative than negative conclusions to modus tollens and the denial of the antecedent inferences for binary context (modus tollens $p < .04$; denial of the antecedent $p < .053$), but not for multiple context (modus tollens $p = .17$; denial of the antecedent $p = .91$). The results replicate the results of Experiment 1 for binary context and Experiment 2 for multiple context.

It is important to note, as Fig. 5 shows, that although participants exhibit the inference-to-alternates effect robustly in the binary context, constructing significantly more affirmative conclusions than negative ones even to negative minor premises, the effect is nonetheless somewhat diminished, and instead there is a small proportion of negative conclusions to negative minor premises. The observation may indicate that the tendency to provide negative conclusions to negative minor premises is a more dominant strategy, which may in turn support the idea that constructing models with symbols rather than by alternates is a more common strategy.

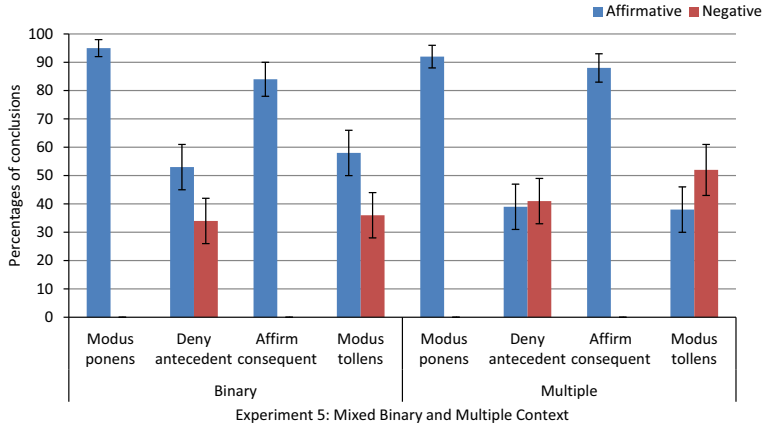

Fig. 5. Percentages of affirmative and negative conclusions in Experiment 5 for counterfactuals in a mixed binary and multiple context, for direct reference. Error bars are standard error of the mean.

An ANOVA of the same design as the previous experiment showed a main effect of inferences, $F(1.505, 46.647) = 21.24$, $MSE = .38$, $p < .001$, $\eta_p^2 = .41$, and context $F(1, 31) = 7.01$, $MSE = .07$, $p < .015$, $\eta_p^2 = .18$, and an interaction of the two, $F(2.281, 70.720) = 6.25$, $MSE = .04$, $p < .003$, $\eta_p^2 = .17$. We decomposed the interaction to test for potential differences between binary and multiple context, using a Bonferroni corrected alpha of .0125 for the four comparisons, and found that participants made more affirmative conclusions to inferences in the binary than multiple context for modus tollens, 58% versus 38%, $t(31) = 3.22$, $p < .004$, $d = .57$, and the denial of the antecedent, 53% versus 39%, $t(31) = 2.33$, $p < .03$, $d = .41$, although the latter effect is marginal on the corrected alpha. There were no differences for modus ponens, 95% versus 92%; $t(31) = 1.0$, $p = .33$, or affirmation of the consequent, 84% versus 88%, $t(31) = 1.0$, $p = .33$ (see Table B5 in the Supplemental Materials).

The results replicate the results observed in Experiments 1 and 2 and extend them to the situation where the binary and multiple contexts are mixed. They corroborate the suggestion that participants adopt different strategies for the binary and multiple context problems when they are presented at random. The experiment indicates that the inference-to-alternates effect is the result of an enduring strategy even when participants have varied experiences, and it is unlikely to be the result of mere syntactic priming or a superficial strategy that participants induce during the course of an experiment. The next experiment examines mixed binary and multiple contexts for inferences with indirect reference.

7. Experiment 6: Mixed binary and multiple contexts with indirect reference

The aim of the experiment was to examine mixed binary and multiple contexts for minor premises that contained indirect reference. We expect to replicate the findings of Experiments 1 and 2 for indirect reference, that is, to observe the inference-to-alternates

effect for indirect reference inferences not only in a binary context but also in a multiple context. The inference-to-alternates effect may be even more robust in the multiple context in the mixed design than it is in the design of Experiment 2, since experience with the ready match between the alternates in the mental representation of a binary context and the terms in the indirect minor premise may lead participants to represent the multiple context by alternates from the outset.

7.1. Method

7.1.1. Design, materials, and procedure

The design, materials, and procedure were the same as the previous experiment, with the exception that the minor premise contained indirect reference to the counterfactual, as in the indirect reference conditions in Experiments 1 and 2.

7.1.2. Participants

The 32 participants who took part in the experiment were undergraduate students at the University of La Laguna, Tenerife, Spain. None had formal training in logic. They had not taken part in the previous experiments.

7.2. Results and discussion

As Fig. 6 shows, participants exhibited an inference-to-alternates effect for both binary and multiple context when the minor premise contained indirect reference. Participants drew more affirmative than negative conclusions to modus ponens and affirmation of the consequent inferences, for both binary and multiple context, binomial sign test, $p < .0001$ for each of the four comparisons; they also made more affirmative conclusions to modus tollens and denial of the antecedent inferences, for both binary and multiple context, binomial sign test, $p < .0001$ for each of the four comparisons.

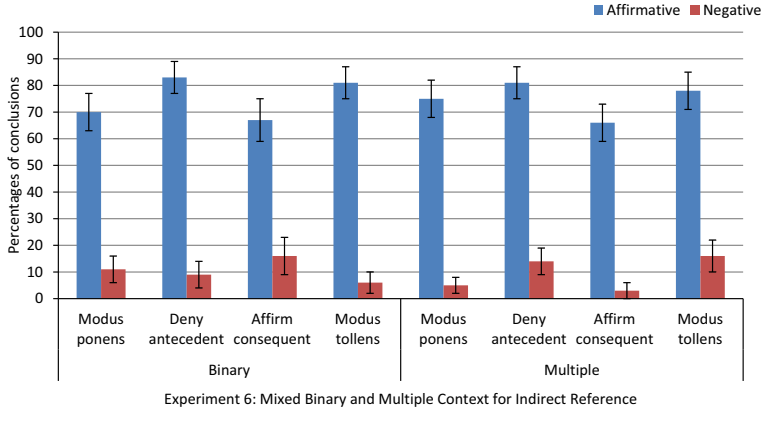

Fig. 6. Percentages of affirmative and negative conclusions in Experiment 6 for counterfactuals in a mixed binary and multiple context, for indirect reference. Error bars are standard error of the mean.

An ANOVA of the same design as the previous experiment on the affirmative conclusions showed no main effects of inference or context, $F < 1$ in both cases, and no interaction, $F(1.703, 52.788) = 1.71$, $MSE = .33$, $p = .19$ (see Table B6 in the Supplemental Materials). The results replicate the results observed in Experiment 1 and 2 for indirect reference and extend them to mixed binary and multiple contexts. As a comparison of Fig. 2 and Fig. 6 shows, the inference-to-alternates effect is very strongly observed for indirect reference in this experiment.

We have assumed thus far that the form of the verbalized conclusion, as affirmative or negative, reflects the nature of the mental representation that is used to derive the conclusion, that is, specific alternates or annotated models. An alternative explanation is that the results may instead reflect a short-cut strategy, based on the verbal response itself rather than the mental representation. People may represent the presupposed facts of a counterfactual conditional such as "if there had been roses there would have been orange trees" in terms of alternates, and they may incorporate a minor premise such as "there are no orange trees" and make the inference that there are poppies or lilies. However, they may be prompted by the information in the context to verbalize their conclusion in negative terms, "there are no roses." Consider the steps that would need to be taken in a computer program to simulate this proposed verbal response strategy. First, the program would represent the premises, "if there had been roses there would have been orange trees" in the multiple context "the flowers are roses or poppies or lilies and the trees are orange trees or apple trees," as alternates,

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples
	Lilies	Apples

Second, the program would incorporate the minor premise, "there are no orange trees," by eliminating the first possibility and retaining the other two possibilities,

Poppies	Apples
Lilies	Apples

Third, it must construct a verbal conclusion that parsimoniously captures the information in these possibilities. We have assumed thus far that the verbal conclusion that captures this representation parsimoniously is "there are poppies or lilies." However, a program could construct a verbal conclusion by computing a fourth step, by matching these possibilities to the multiple context, "the flowers are roses or poppies or lilies and the trees are orange trees or apple trees," and so produce the negative conclusion, "there are no roses." Although such a verbal response strategy is possible, we suggest there is little a priori justification for the fourth step, that is, to suppose that the description of a mental representation requires affirmative possibilities to be translated to negative conclusions. We consider instead that the proposal of such a fourth step violates Occam's razor—the simpler explanation is that when participants represent the information in terms of

alternates, for example, "poppies or lilies" to make an inference, they continue to rely on their representation of alternates to verbalise their conclusion, for example, "there are poppies or lilies." Similarly, when participants represent the information using symbols, for example, "no roses" to make an inference, they continue to rely on their representation of symbols to verbalise their conclusion using a negative referent, for example, "there are no roses." And in any case, the same argument of a verbal response strategy could be used to claim that people represent the presupposed facts of the counterfactual conditional in terms of symbols (e.g., no roses and no orange trees) but when verbalizing the conclusion that there are no roses, they may use the additional information given in the context (e.g., roses or poppies or lilies) to transform their negative conclusion (e.g., no roses) into an affirmative conclusion (e.g., poppies or lilies). We consider either verbal strategy—affirmative representations translated to negative conclusions, or negative representations translated to affirmative conclusions—to be unlikely, since they both predict that participants should make more affirmative than negative conclusions not only in the binary context but also in the multiple context, as participants could apply either verbal response strategy given either context. For example, when participants verbalize the conclusion that there are no roses, they may use the additional information given in the multiple context (e.g., roses or poppies or lilies) to transform their negative conclusion (e.g., no roses) into an affirmative conclusion (e.g., poppies or lilies). Participants tend not to do so, and so a verbal response strategy is unlikely to account for the inference-to-alternates effect in binary contexts and its reduction in multiple contexts. However, it could be the case that people employ a short-cut verbal response strategy only in a binary context and not in a multiple context, perhaps because the binary context is simpler or results in a categorical conclusion, or the multiple context exceeds working memory limitations. Hence, we test the alternative explanation of a verbal response strategy in the next set of three experiments, which rely on indicative conditionals.

8. Experiments 7a, 7b, and 7c: Indicative conditionals

The aim of this set of three experiments was to test the inferences participants make for ordinary indicative conditionals, such as "if there were roses then there were orange trees." We test their inferences given a binary context in Experiment 7a and a multiple context in Experiment 7b, and we directly compare their inferences in binary and multiple contexts in Experiment 7c. We make the novel prediction that an inference-to-alternates effect will be observed in binary contexts for indicative conditionals. Most important, we also predict that an inference-to-alternates effect will be observed not only in binary contexts but also in multiple contexts for indicative conditionals, unlike for counterfactual conditionals. The prediction arises because of the differences in the mental representations of indicative and counterfactual conditionals. When people understand an ordinary conditional in the indicative mood, they tend to envisage just a single possibility at the outset, for example,

Roses Oranges
...

(Johnson-Laird & Byrne, 1995, 2002). Their initial representation parsimoniously corresponds to a single alternative (e.g., Barrouillet & Lecas, 1999; Byrne, 1989; Johnson-Laird et al., 1992). Their interpretation is not merely akin to a conjunction and the three dots in the diagram indicate they may be aware that there are other possibilities which they have not yet thought about explicitly. They can "flesh out" their representation to consider further alternatives (e.g., Hinterecker, Knauff, & Johnson-Laird, 2016; Johnson-Laird & Byrne, 2002). But unlike counterfactual conditionals, they do not represent from the outset an additional possibility corresponding to the presupposed facts, for example,

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples

Hence, for indicative conditionals, it is only when participants receive a subsequent minor premise, for example, "there were no roses," which is incompatible with the single possibility they have considered at the outset, that they must consider alternative possibilities. The minor premise rules out the initial possibility they have envisaged and so participants must consider what other possibilities to flesh out. For a binary context, they can readily flesh out their models to include the additional possibility:

Poppies Apples

and they can conclude "there were apples." Equally, for a multiple context they can also readily flesh out their models to include the additional possibilities:

Poppies Apples
Poppies Pears

because their initial models of the indicative conditional contain only a single explicit possibility, unlike their initial models for the counterfactual conditional which contain from the outset two possibilities. Hence, consideration of alternates in the multiple context for an indicative conditional does not exceed their working memory capacity. As a result, we expect that participants will show an inference-to-alternates effect; that is, they will provide affirmative conclusions even to negative minor premises, not only for a binary context but also for a multiple context.

In contrast, the alternative explanation that the inference-to-alternates effect arises from a short-cut verbal response strategy makes a different prediction. It predicts that the inference-to-alternates effect will be observed in binary contexts, but it will be reduced in multiple contexts, just as it is for counterfactuals, since the strategy can be applied equally regardless of whether the conditional is a counterfactual or an indicative one. Experiments 7a, 7b, and 7c test these predictions. The set of three experiments is designed to examine inferences from indicative conditionals in a binary and multiple context for direct and indirect reference, as well as in a mixed binary and multiple context.

8.1. Method

8.1.1. Design, materials, and procedure

The design, materials, and procedure were the same as the previous experiments. Experiment 7a (binary context) was the same as Experiment 1, Experiment 7b (multiple context) was the same as Experiment 2, and Experiment 7c (mixed binary and multiple contexts) was the same as Experiment 5, with the exception throughout that the conditionals were in the indicative mood, for example, "if the flowers were roses then the trees were orange trees."

8.1.2. Participants

There were 32 participants in Experiment 7a, a different set of 32 participants in Experiment 7b, and a different set of 32 participants in Experiment 7c. The 96 participants were undergraduate students at the University of La Laguna, Tenerife, Spain. None had formal training in logic. They had not taken part in the previous experiments.

8.2. Results and discussion

Participants showed an inference-to-alternates effect for indicative conditionals in a binary context, as Fig. 7a shows, in a multiple context as Fig. 7b shows, and in a mixed binary and multiple context, as Fig. 7c shows.

In Experiment 7a with binary context, participants drew more affirmative conclusions than negative ones to modus ponens and affirmation of the consequent inferences, for both direct and indirect reference, binomial sign test, $p < .0001$ for each of the four comparisons as Fig. 7a shows. They also made more affirmative conclusions than negative ones to modus tollens and denial of the antecedent inferences, for both direct and indirect reference, binomial sign test, $p < .0001$ for each of the four comparisons. The results replicate the results of Experiment 1 for counterfactuals and extend them to indicative conditionals. An ANOVA of the same design as the first experiment on the affirmative conclusions showed a main effect for inference, $F(1.311, 40.363) = 9.75$, $MSE = .29$, $p < .001$, $\eta_p^2 = .24$, as participants made more affirmative conclusions to denial of the antecedent and modus tollens inferences (94% and 94%) than modus ponens and affirmation of the consequent inferences (70% and 70%); there was no main effect for reference, $F > 1$, and no interaction, $F(2.116, 65.597) = 2.58$, $MSE = .06$, $p = .09$ (see Table B7a in the Supplemental Materials). The experiment demonstrates for the first time an inference-to-alternates effect for indicative conditionals, that is, a tendency to produce affirmative conclusions to negative minor premises in a binary context.

In Experiment 7b with multiple context, participants also showed an inference-to-alternates effect, that is, they made more affirmative conclusions to modus ponens and affirmation of the consequent inferences, for both direct and indirect reference, binomial sign test, $p < .0001$ for each of the four comparisons; they also made more affirmative conclusions to modus tollens and denial of the antecedent inferences, for both direct and indirect reference, binomial sign test, $p < .0001$ for each of the four comparisons, as Fig. 7b

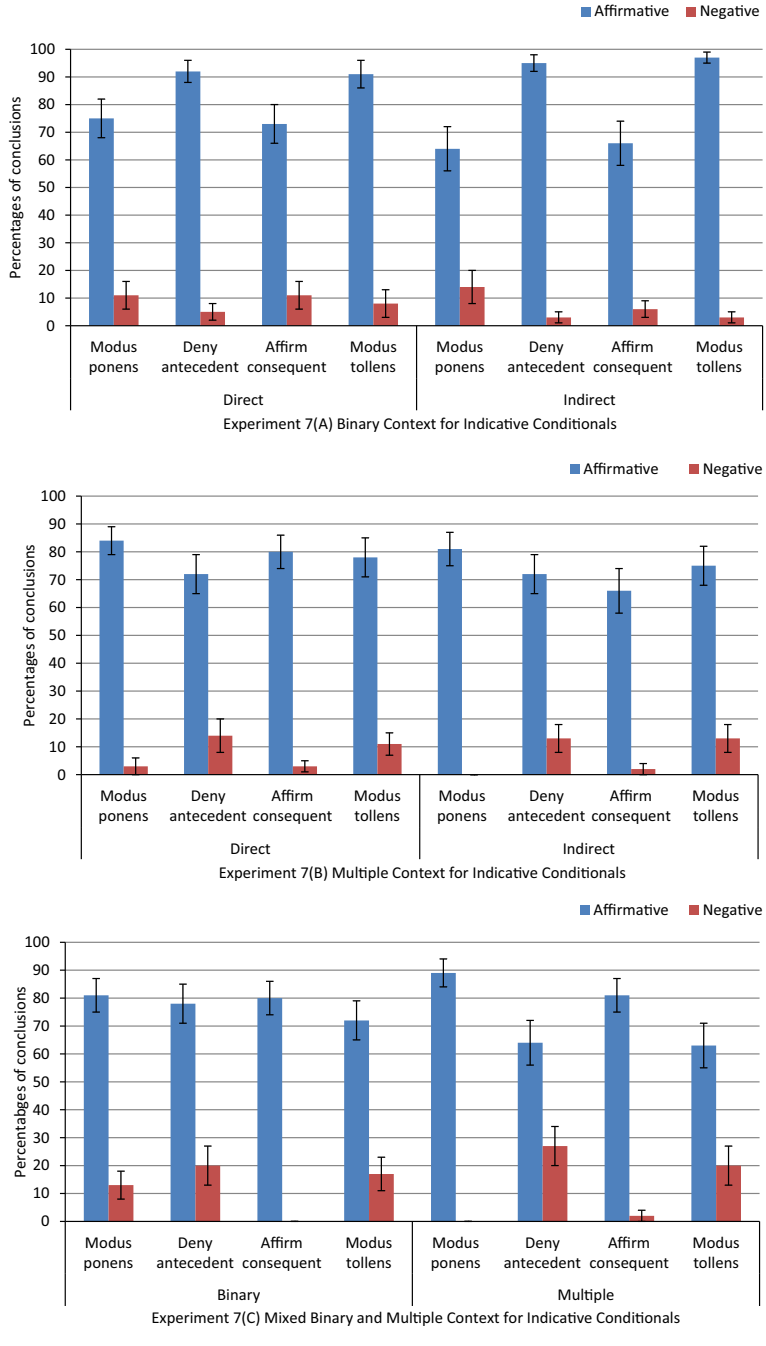

Fig. 7. (a) Percentages of affirmative and negative conclusions in Experiment 7a for indicative conditionals in a binary context. (b) Percentages of affirmative and negative conclusions in Experiment 7B for indicative conditionals in a multiple context. (c) Percentages of affirmative and negative conclusions in Experiment 7c for indicative conditionals in mixed binary and multiple contexts. Error bars are standard error of the mean.

shows. Hence, as we predicted, the results are different from those for counterfactual conditionals, for example, in Experiment 2. The results indicate that for ordinary indicative conditionals, participants are not prompted by the representation of any presupposed facts to represent the information by annotated-models instead of by alternates. An ANOVA of the same design as the previous experiment showed no main effect for inference, $F(2.340, 72.549) = 1.44$, $MSE = .14$, $p = 24$, nor for reference, $F(1, 31) = 2.10$, $MSE = .08$, $p = .15$, and no interaction, $F < 1$ (see Table B7b in the Supplemental Materials).

The results of Experiment 7B also help to rule out an alternative explanation for indirect reference with counterfactuals. For indirect premises, we have suggested that participants need to transform their symbolic representation involving explicit negation into a representation with alternates when reading the indirect premise. An alternative possibility is that participants reformulate the model of the indirect premise as a negation to allow it to be integrated with the model of the presupposed facts. However, the results for indirect premises for indicative conditionals in multiple contexts are the same as those for counterfactual conditionals, and yet participants have not represented any presupposed facts with which to integrate such a reformulated model of the indirect premise, and hence this alternative explanation seems unlikely.

In Experiment 7c with mixed binary and multiple contexts, participants showed an inference-to-alternates effect for both binary and multiple context; that is, they made more affirmative conclusions to modus ponens and affirmation of the consequent inferences, for both binary and multiple context, binomial sign test, $p < .0001$ for each of the four comparisons, as Fig. 7c shows. They also made more affirmative conclusions to modus tollens and denial of the antecedent inferences, for both binary and multiple context, binomial sign test, $p < .0001$ for each of the four comparisons. An ANOVA of the same design as Experiment 5 on the affirmative conclusions showed no main effect for inference, $F(1.815, 56.255) = 2.09$, $MSE = .35$, $p = .14$, or context $F(1, 31) = 1.49$, $MSE = .05$, $p = .23$, but an interaction of the two, $F(3, 93) = 4.04$, $MSE = .04$, $p < .01$, $\eta_p^2 = .11$ (see Table B7c in the Supplemental Materials). The interaction showed that participants made more affirmative conclusions in a binary context than a multiple context for the denial of the antecedent inference, 78% versus 64%, $t(31) = 2.32$, $p < .03$, $d = .15$, and fewer affirmative conclusions in a binary context than a multiple context for modus ponens, 81% versus 90%, $t(31) = 2.40$, $p < .03$, $d = .16$; there were no differences in modus tollens, 72% versus 63%, $t(31) = 1.64$, $p = .11$, or the affirmation of the consequent, 80% versus 81%, $t < 1$.

The results of the three experiments show that for ordinary indicative conditionals, such as "if there were roses then there were orange trees," an inference-to-alternates effect is observed. The effect is observed not only for binary contexts but also for multiple contexts. The result is consistent with the proposal that the inference-to-alternates effect for counterfactuals in a binary context, and its reduction in a multiple context, reflect changes in the nature of the mental representation of the presupposed facts of a counterfactual as alternates or in annotated models. In contrast, people can represent alternates even in a multiple context for an indicative conditional because their initial

representation contains only a single explicit possibility, unlike their initial representation for the counterfactual conditional, which from the outset contains two possibilities. Hence, consideration of alternates for an indicative conditional in a multiple context does not exceed working memory capacity. The results are contrary to the predictions of a short-cut verbal response strategy. A verbal response strategy could be applied regardless of whether the conditional is counterfactual or indicative. Hence, such an account predicts a reduction in the inference-to-alternates effect for indicative conditionals in a multiple context, just as it does for counterfactuals. The results go against its predictions. A defense of the verbal short cut strategy could be that participants rely on a verbal response strategy that translates affirmative representations into negative conclusions, only in complex situations, such as when working memory limitations are reached. In that case, participants would translate their affirmative representations into negative conclusions in multiple contexts but not binary ones for counterfactuals (which require the mental representation of two possibilities), and they would translate their affirmative representations into negative conclusions in both binary and multiple contexts for indicative conditionals (which require the mental representation of just a single possibility at the outset). Once again, however, there appears to be little a priori justification to suppose that the description of a mental representation requires affirmative possibilities to be translated to negative conclusions, and moreover, that this additional step is invoked only in situations of complexity, such as counterfactuals in a multiple context.

9. General discussion

The nine experiments we report reveal two key discoveries. The main discovery is that the experiments reveal for the first time a robust inference-to-alternates effect for counterfactuals such as "If there had been roses there would have been orange trees" when they are presented in a binary context, such as "At the botanic gardens the flowers were roses or poppies and the trees were orange trees or apple trees." Participants draw affirmative conclusions, for example, "there were poppies," even to negative minor premises, for example, "there were no oranges," as Experiment 1 shows. The second discovery is that the inference-to-alternates effect is reduced for counterfactuals presented in a multiple context, such as, "At the botanic gardens the flowers were roses or poppies or lilies and the trees were orange trees or apple trees." Participants draw negative conclusions, for example, "there were no roses," to negative minor premises, for example, "there were no oranges," in the multiple context, as often as affirmative ones, as Experiment 2 shows. The results show that the presence of multiple alternatives reduces the dominance of affirmative conclusions.

The two discoveries allow us to distinguish between the predictions of two alternative accounts of the mental representation of the presupposed facts for counterfactuals. One account is that people represent the presupposed facts of a counterfactual by specific alternates,

Counterfactual:	Roses	Oranges
Presupposed Facts:	Poppies	Apples
Presupposed Facts:	Lilies	Apples

We derived this specific alternates view from some of the proposals made in some studies of negation carried out from an experiential embodied perspective (e.g., Kaup et al., 2006; Mayo et al. 2004; but see also Huette & Anderson, 2012; Aravena et al., 2012; Bartoli et al., 2013; Hald et al., 2013). An alternative account is that people construct mental models that can contain propositional-like symbols,

Counterfactual:	Roses	Oranges
Presupposed Facts:	No roses	No oranges

On this account, annotated-models are as iconic as possible but can contain symbols if necessary (e.g., Johnson-Laird & Byrne, 2002; Orenes et al., 2014). The discovery of the inference-to-alternates effect in a binary context is consistent with both the specific alternates and the annotated-models accounts. Its reduction in the multiple context corroborates the prediction of the annotated-models view that people can in some cases switch to representing the presupposed facts by relying on symbols, such as when the multiple context exceeds working memory constraints. The finding is difficult to explain on a specific alternates view.

The experiments show that people prefer to infer affirmative conclusions in many situations. Importantly, the experiments show that experience affects the inference-to-alternates effect. When participants were given a set of inferences in the binary context first, they exhibited the inference-to-alternates effect, drawing affirmative conclusions to negative minor premises, and when they were then given a set of inferences in a multiple context, they continued to exhibit the inference-to-alternates effect, as Experiment 3 shows. The result indicates that participants' experience of representing the counterfactual by alternates in a binary context inculcates a "mental set" that influences them to continue to represent the counterfactual by alternates even for a multiple context (e.g., Luchins, 1942; Ollinger et al., 2008). Conversely, when participants were given a set of inferences in the multiple context first, they showed a reduced inference-to-alternates effect—they drew negative conclusions as often as affirmative ones to the negative minor premises, and when they were then given a set of inferences in a binary context, they continued to show a reduced inference-to-alternates effect, and to respond to the binary context inferences in the same way as they responded to the multiple context inferences, by drawing negative conclusions as often as affirmative ones to negative minor premises, as Experiment 4 shows. The result indicates that participants' experience of representing the counterfactual by symbols in a multiple context influenced them to continue to represent it by symbols even for a binary context. The discovery of a "mental set" for binary and multiple context inferences corroborates the idea that representation by alternates or symbols is a strategic process rather than an invariant one. Moreover, when participants gain experience with the binary context inferences and the multiple context inferences at random in a mixed design, they show an inference-to-

alternates effect for the binary context inferences, and the effect is reduced for the multiple context inferences, as Experiment 5 shows. The result corroborates the idea that the inference-to-alternates effect is not a superficial artifact induced during the experimental exposure to problems of particular sorts, or from syntactic priming, but instead is sufficiently robust to occur even in mixed presentations.

A noteworthy result is that the experiments show an inference-to-alternates effect for indirect reference in minor premises. When participants are given a counterfactual, such as "If there had been roses, there would have been orange trees" in the binary context, "At the botanic gardens the flowers were roses or poppies and the trees were orange trees or apple trees" and they are given the minor premise for modus ponens with indirect reference such as "There were no poppies," they draw affirmative conclusions, just as they do for direct reference negative minor premises, as Experiment 1 showed. However, the inference-to-alternates effect is observed even for multiple context, for such indirect reference minor premises, as Experiment 2 showed. They exhibit the inference-to-alternates effect for binary and multiple contexts for indirect reference minor premises when both sorts of contexts are presented together in a mixed design, as Experiment 6 shows. The findings corroborate the prediction of the annotated-models account that even though participants can represent the presupposed facts by relying on symbols in the multiple context, they must add to their models information about alternates when they are given an indirect reference minor premise, which refers only to alternates, because it does not refer to anything that they have represented otherwise.

The inference-to-alternates effect occurs for indicative conditionals not only in binary contexts but also in multiple contexts, as Experiments 7a, 7b, and 7c show. The reduced representational demands of indicative conditionals, for which participants generally envisage a single possibility at the outset, facilitate the consideration of alternates even in multiple contexts. The result suggests that the inference-to-alternates effect arises from representational constraints rather than from a verbal response strategy. In particular, a verbal response strategy has difficulty in accounting convincingly for the complete pattern of results for the inference-to-alternates effect as it occurs for counterfactuals in binary or multiple contexts, direct or indirect reference, blocked or mixed presentation, and for indicative conditionals.

In all of the experiments, participants made the modus ponens and modus tollens inferences very readily, and they also made the affirmation of the consequent and denial of the antecedent inference readily. The pattern of inferences suggests that they tend to consider just two possibilities, the conjecture and the presupposed facts, rather than considering any further possibilities, for the sorts of content used in the experiments.

The inference-to-alternates effect shows that people tend to envisage what is assumed to be present in a scene, such as roses or poppies in a botanic garden, rather than tending to envisage what is not present, such as roses or no roses. They tend to infer what is present in a scene in response to information about what is present, "there are roses therefore there are orange trees," but also in response to information about what is not present, "there are no roses therefore there are apple trees," at least in a simple binary context. In contrast, in a multiple context, they infer, in at least some instances, what is not present

in a scene from information about what is not present, "there are no roses therefore there are no orange trees."

Affirmative conclusions were generally preferred in most of the conditions we studied, and negative conclusions were much more rare in each of the experiments. Participants clearly rely on representation by alternates in many circumstances, but nonetheless they can make use of symbolic representations in at least some circumstances, such as when they are given a multiple context with direct reference premises. The materials we used concerned real-world settings. Some real-world situations may have well defined possibilities, akin to the binary context or multiple context in our experiments, but in some real-world situations, people may not have information about the set of possibilities. In such situations, it may be even more crucial for them to rely on annotated models that can use symbols for negation rather than alternates, because otherwise the number of possibilities to envisage would far exceed working memory limitations. The results thus have implications not only for how people understand counterfactuals, but also how they construct them (e.g., Segura, Berrocal, & Byrne, 2002; Walsh & Byrne 2007).

It has been known for some time that people think about two possibilities when they understand a counterfactual such as "if there had been roses, there would have been orange trees." They imagine the conjecture, that there are roses and orange trees, and they also recover the known or presupposed facts, that there are no roses and there are no orange trees. The experiments reported in this paper uncover in further depth how people represent the presupposed facts. Overall, the findings support a pluralist view that people construct representations of specific alternates in simple binary situations, and these representations are consistent with some suggestions from an embodied simulation perspective; however, in more complex multiple situations, people can also rely on annotated models that represent the presupposed facts of a counterfactual in simulations that are iconic but that can nonetheless contain symbols.

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 2014, at the Annual Conference of the Association for Psychological Science in New York in 2015, and at the International Thinking Conference in Rhode Island in 2016. We thank Sabrina Haimovici for helpful comments.

References

Aravena, P., Delevoye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V., & Nazir, T. (2012). Grip force reveals the context sensitivity of language-induced motor activity during "action words" processing: Evidence from sentential negation. *PLoS ONE*, *7*(12), e50287.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, *5*(4), 289–302.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(4), 637–660.

Barsalou, L., Simmons, W. K., Barbey, A., & Wilson, C. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science*, *7*, 84–91.

Bartoli, E., Tettamanti, A., Farronato, P., Caporizzo, A., Moro, A., Gatti, R., Perani, D., & Tettamanti, M. (2013). The disembodiment effect of negation: Negating action-related sentences attenuates their interference on congruent upper limb movements. *Journal of Neurophysiology*, *109*(7), 1782–1792.

Byrne, R. M. (1989). Everyday reasoning with conditional sequences. *The Quarterly Journal of Experimental Psychology Section A*, *41*(1), 141–166.

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. (2017). Counterfactual thinking: From logic to morality. *Current Directions in Psychological Science*, *26*(4), 314–322.

Byrne, R. M. J., & Egan, S. M. (2004). Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology*, *58*, 113–120.

Byrne, R. M. J., & Johnson-Laird, P. N. (2009). "If" and the problems of conditional reasoning. *Trends in Cognitive Sciences*, *13*, 282–287.

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

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517.

van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, *32*, 579–590.

De Vega, M., Morera, Y., León, I., Beltrán, D., Casado, P., & Martín-Loeches, M. (2016). Sentential negation might share neurophysiological mechanisms with action inhibition. Evidence from frontal theta rhythm. *Journal of Neuroscience*, *36*(22), 6002–6010.

De Vega, M., & Urrutia, M. (2011). Counterfactual sentences activate embodied meaning. *Journal of Cognitive Psychology*, *23*, 962–973.

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

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*, 412–443.

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.

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.

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

Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. Oxford, UK: Oxford University Press.

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

Ginsberg, M. L. (1986). Counterfactuals. *Artificial Intelligence*, *30*(1), 35–79.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 586–596.

Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, *1*(1), 19–33.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*, 231–262.

Hald, L., Hocking, I., Vernon, D., Marshall, J. A., & Garnham, A. (2013). Exploring modality switching effects in negated sentences: Further evidence for grounded representations. *Frontiers in Psychology*, *4*, 93.

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1606–1620.

Huette, S., & Anderson, S. (2012). Negation without symbols: The importance of recurrence and context in linguistic negation. *Journal of Integrative Neuroscience*, *11*(03), 295–312.

Johnson-Laird, P. N., & Byrne, R. M. J. (1995). A model point of view. *Thinking and Reasoning*, *1*, 339–350.

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., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, *19*(4), 201–214.

Johnson-Laird, P. N., & Tridgell, J. M. (1972). When negation is easier than affirmation. *Quarterly Journal of Experimental Psychology*, *24*, 87–91.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology*, *60*(7), 976–990.

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, *24*(5), 541–559.

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2014). The negations of conjunctions, conditionals, and disjunctions. *Acta Psychologica*, *151*, 1–7.

Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives*. New York: Oxford 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.

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*(6), i.

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*(4), 433–449.

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

Nickerson, R. (2015). *Conditional reasoning*. Oxford, UK: 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.

Ollinger, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, *55*(4), 269–282.

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, *74*, 36–45.

Pearl, J. (2013). Structural counterfactuals: A brief introduction. *Cognitive Science*, *37*(6), 977–985.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*(4), 136–141.

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470.

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

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.

Segura, S., Fernandez-Berrocal, P., & Byrne, R. M. (2002). Temporal and causal order effects in thinking about what might have been. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(4), 1295–1305.

Thompson, V., & Byrne, R. M. J. (2002). Reasoning counterfactually: Making inferences about things that didn's happen. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 1154–1170.

Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive Psychology*, *2*, 239–289.

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.

Walsh, C. R., & Byrne, R. M. (2007). How people think if only… about reasons for actions. *Thinking & Reasoning*, *13*(4), 461–483.

Waskan, J. A. (2006). *Models and cognition: Prediction and explanation in everyday life and in science*. Cambridge, MA: The MIT Press.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Wason, P. C., & Jones, S. (1963). Negatives: Denotation and connation. *British Journal of Psychology*, *54*, 299–307.

Williamson, T. (2007). Philosophical knowledge and knowledge of counterfactuals. *Grazer Philosophical Studies*, *74*(1), 89–124.

Winer, B. J., Brown, D. R., & Michels, K. M. (1971). *Statistical principles in experimental design* (vol. 2). New York: McGraw-Hill.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table B1. Experiment 1: Binary context with direct and indirect reference.

Table B2. Experiment 2: Multiple context with direct and indirect reference.

Table B3. Experiment 3: Binary context first, multiple context second.

Table B4. Experiment 4: Multiple context first, binary context second.

Table B5. Experiment 5: Mixed multiple and binary context.

Table B6. Experiment 6: Mixed multiple and binary context for indirect reference.

Table B7A. Experiment 7A: Binary context for indicative conditionals.

Table B7B. Experiment 7B: Multiple context for indicative conditionals.

Table B7C. Experiment 7C: Mixed binary and multiple context for indicative conditionals.

Appendix A: The materials for the experiments

Materials: The 16 contents used in the experiments. The binary context is provided first with the multiple context additions in parentheses. We provide the English translation of the original Spanish.

1. At the Botanical Gardens, the flowers were roses or poppies (or carnations) and the trees were orange trees or apple trees (or pear trees). If the flowers had been roses, the trees would have been orange trees.

2. In the store that day the fruit was grapes or watermelons (or melons) and the vegetables were potatoes or peppers (or onions). If fruit had been grapes, the vegetables would have been potatoes.

3. In the restaurant, the first dish was gazpacho or soup (or salad) and the main course was rice or pasta (or seafood). If the first dish had been gazpacho, the main course would have been rice.

4. On the farm, the livestock was sheep or cattle (or goats) and the crop was oats or wheat (or barley). If the cattle had been sheep, the crop would have been oats.

5. At the party, the drink was rum or whisky (or brandy) and the appetizer was olives or peanuts (or pistachios). If the drink had been rum, the appetizer would have been olives.

6. In the kitchen, the utensils were kettles or bowls (or frying pans) and the cutlery were spoons or forks (or knives). If the utensils had been kettles, the cutlery would have been spoons.

7. In the aquarium, the fish were barracudas or trout (or mackerel) and the molluscs were octopus or squid (or limpets). If the fish had been barracudas, the molluscs would have been octopuses.

8. At the museum, the vessels were Phoenician or Etruscan (or Egyptian) and the columns were Ionic or Doric (or Corinthian). If the vessels had been Phoenician, the columns have been Ionian.

9. In the sweetshop, the cakes were nougat or truffle (or lemon) and the ice-cream was lime or coconut (or vanilla). If the cake had been nougat, the ice cream would have been lime.

10. In the zoo, the cats were tigers or lions (or cheetahs) and the cetaceans were grey whales or orcas (or dolphins). If the cats had been tigers, the cetaceans would have been gray whales.

11. In the toy shop, the push toy was a bear or a dog (or a cat), and the board game was goose or parcheesi (or monopoly). If the push toy had been a bear, the board game would have been a goose.

12. On Miguel's birthday, the drink was fanta or coca-cola (or seven-up) and the food was tortilla or sandwiches (or hamburgers). If the drink had been fanta, the food would have been tortilla.

13. Mary's gift was a skirt or a shirt (or a scarf) and John's gift was a sweater or socks (or trousers). If Mary's gift had been a skirt, John's gift would have been a sweater.

14. In the pet shop, the animals were cats or dogs (or hamsters) and the birds were cockatoos or parrots (or macaw). If the animals had been cats, the birds would have been cockatoos.

15. At the wedding, the wedding dress was white or beige (or green) and the bouquet was orchids or jasmines (or lilies). If the dress had been white, the bouquet would have been orchids.

16. In the purse, the lipstick was red or brown (or green) and the eye shadow was black or blue (or grey). If the lipstick had been red, the eye shadow would have been black.