Indicative and counterfactual 'only if' conditionals

Suzanne M. Egan [a,*], Juan A. García-Madruga [b], Ruth M.J. Byrne [c]

[a] Department of Psychology, Mary Immaculate College, University of Limerick, South Circular Road, Limerick, Ireland
[b] Department of Psychology, UNED, Juan del Rosal, No. 10, Ciudad Universitaria, Madrid 28040, Spain
[c] Trinity College Dublin, University of Dublin, Ireland

ARTICLE INFO

ABSTRACT

We report three experiments to test the possibilities reasoners think about when they understand a conditional of the form 'A only if B' compared to 'if A then B'. The experiments examine conditionals in the indicative mood (e.g., A occurred only if B occurred) and counterfactuals in the subjunctive mood (A would have occurred only if B had occurred). The first experiment examines the conjunctions of events that reasoners judge to be consistent with conditionals, e.g., A and B, not-A and not-B. It shows that people think about one possibility to understand 'if' and two possibilities to understand 'only if'; they think about two possibilities to understand counterfactual 'if' and 'only if'. The second experiment shows that the possibilities people think about when they understand 'only if' are in a different temporal order (e.g., B and A) to the possibilities they think about for 'if' (A and B). The third experiment shows that people make different inferences from 'only if' and 'if' conditionals and counterfactuals. The implications of the results for theories of counterfactual conditionals are considered.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

People often imagine counterfactual alternatives about what might have been different in the past, particularly after bad outcomes, e.g., 'if John had felt well he would have passed the examination' (Kahneman & Tversky, 1982; Mandel, Hilton, & Catellini, 2005; Markman, Klein, & Suhr, 2009; Roese & Olson, 1995). People understand the counterfactual conditional to mean something different from an indicative one, e.g., 'if John felt well he passed the examination' (Byrne, 2002, 2005, 2007). Moreover, they understand an indicative conditional based on 'if' to mean something different from one based on 'only if', e.g., 'John felt well only if he passed the examination' (Johnson-Laird & Byrne, 1989). How do people understand counterfactual conditionals based on 'only if', such as 'John would have felt well only if he had passed the examination'? The answer is not known and our aim is to test a new account of what people think about when they understand 'only if' conditionals, both indicative and counterfactual.

We provide a novel account of how people understand and reason from indicative and counterfactual 'only if' conditionals. First we sketch the view that people envisage possibilities to understand indicative 'if' conditionals (Johnson-Laird & Byrne, 2002), and we consider evidence that they think about more possibilities to understand counterfactual 'if' compared to indicative 'if'. We

outline our new account in Sections 2 and 3, we report three experiments that test predictions derived from this account by comparing indicative and counterfactual 'if' and 'only if' conditionals.

2. Indicative and counterfactual 'if'

People may understand a conditional such as 'if Mary went to the meeting then she received the documentation' (if A then B) by thinking about possibilities, such as 'Mary went to the meeting and she received the documentation' (A and B) (Johnson-Laird & Byrne, 2002). The possibilities they think about may be constrained by principles. For example, people think about true possibilities but not false possibilities, such as 'Mary went to the meeting and she did not receive the documentation' (A and not-B) (Espino, Santamaría, & Byrne, 2009; Johnson-Laird & Byrne, 2002). They tend to think about few possibilities, perhaps because of their limited working memories, and so they do not think about all of the true possibilities, e.g., they do not think initially about the other true possibilities, 'Mary did not go to the meeting and she did not receive the documentation' (not-A and not-B), or 'Mary did not go to the meeting and she received the documentation' (not-A and B) (Johnson-Laird & Byrne, 1991). These principles of truth and parsimony constrain the number of mental models of the possibilities that they construct (see Byrne and Johnson-Laird (2009) for a review).

The model theory explains why people make some inferences readily, e.g., given 'Mary went to the meeting', most people make

* Corresponding author. Tel.: +353 61 204333.
E-mail address: suzanne.egan@mic.ul.i.e (S.M. Egan).

the *modus ponens* inference, 'she received the documentation' (Johnson-Laird, 2006). The inference corresponds to the single possibility they have thought about from the outset. They do not make other inferences readily, e.g., given 'Mary did not receive the documentation' they do not make the *modus tollens* inference, 'she did not go to the meeting' (Evans, Newstead, & Byrne, 1993). Instead some of them infer erroneously that nothing follows. The information 'Mary did not receive the documentation' does not correspond to the possibility they have thought about from the outset, Mary went to the meeting and she received the documentation. They have to think about other true possibilities to make the inference. Likewise given, 'Mary received the documentation' people sometimes make an *affirmation of the consequent* inference, 'she went to the meeting'. To resist the inference they must think about other true possibilities, e.g., Mary did not go to the meeting and she received the documentation. Finally, given, 'Mary did not go to the meeting', they sometimes make the *denial of the antecedent inference* 'she did not receive the documentation'. They do not make the inference if they have thought about just a single possibility, or if they have thought about all of the true possibilities.

There is considerable evidence to support the mental model view of reasoning (see Byrne and Johnson-Laird (2009) for a review). Nonetheless, the cognitive processes that underlie conditional reasoning continue to be a matter of debate. An alternative view is that reasoners rely on rules of inference, either abstract rules that operate in virtue of their form (e.g., Braine & O'Brien, 1998; Rips, 1994), or domain-specific rules that are sensitive to some content (e.g., Fiddick, Cosmides, & Tooby, 2000; Holyoak & Cheng, 1995). Another alternative view is that reasoners rely on judgements of the probability of the conditional, either by supposing the antecedent to be true and considering whether the consequent is true or false (e.g., Evans & Over, 2004; Evans, Over, & Handley, 2005), or by judging the likelihood of the antecedent and consequent (e.g., Oaksford & Chater, 2007). The view that reasoners imagine possibilities (Johnson-Laird & Byrne, 2002) has offered the only empirically corroborated account of reasoning with counterfactual conditionals, and so we will outline it further here and return to alternative accounts later.

A counterfactual conditional, e.g., 'if Mary had gone to the meeting then she would have received the documentation' can seem to mean something quite different from the indicative conditional. Logically, counterfactuals have false antecedents and so on a truth functional account they must be true (Quine, 1954), and an account of their logic led to the development of 'possible world' semantics (Lewis, 1973; Stalnaker, 1968). Psychologically, the subjunctive mood of the counterfactual, 'if Mary *had gone* to the meeting then she *would have* received the documentation' can help convey the presupposition that in fact, Mary did not go to the meeting, and in fact, she did not receive the documentation (Fillenbaum, 1974). According to the model theory, people understand a counterfactual conditional by thinking about two possibilities, the conjecture, 'Mary went to the meeting and she received the documentation', and the presupposed facts, 'Mary did not go to the meeting and she did not receive the documentation'. They keep track of the epistemic status of these possibilities as imagined, or as corresponding to the facts (Johnson-Laird & Byrne, 1991, pp. 68–69).

Evidence from inference and comprehension studies corroborates the idea that reasoners think about a single possibility to understand indicative 'if' whereas they think about two possibilities to understand counterfactual 'if'. First, people make more of the inferences that require access to the negative possibility (Mary did not go to the meeting and she did not receive the documentation), that is, the *modus tollens* and *denial of the antecedent* inferences, from the counterfactual because they have access to the negative possibility for the counterfactual but not for the indicative

conditional (Byrne & Egan, 2004; Byrne & Tasso, 1999). They make the same frequency of the affirmative inferences, *modus ponens* and *affirmation of the consequent*, because they have access to the affirmative possibility for both the counterfactual and the indicative. Second, people judge that someone uttering the counterfactual means to imply 'Mary did not go to the meeting' and 'Mary did not receive the documentation' (Thompson & Byrne, 2002). Third, when readers first read a counterfactual they are then primed to read more quickly a subsequent negative conjunction 'Mary did not go to the meeting and she did not receive the documentation' compared to when they are primed by an indicative conditional (Santamaría, Espino, & Byrne, 2005). They read the affirmative conjunction equally quickly whether it is primed by a counterfactual or an indicative (see also De Vega, Urrutia, & Riffo, 2007; Stewart, Haigh, & Kidd, 2009). These three strands of evidence converge to support the view that people think about two possibilities when they understand counterfactual 'if', and only one possibility when they understand indicative 'if'. The evidence from priming studies supports the view that when people understand a counterfactual conditional, they construct two possibilities from the outset during their comprehension of the counterfactual, that is, they think about dual possibilities in order to comprehend the meaning of the counterfactual assertion, independently of, and prior to, the task demands of inferential, consistency, or truth judgement tasks (Byrne, 2005). Of course there may be a time course to their construction of each possibility one after the other, and different individuals may differ in which possibility they tend to construct first. In the next section, we present a novel account of the possibilities that people envisage to understand indicative and counterfactual 'only if'.

3. Indicative and counterfactual 'only if'

A conditional of the form 'if Mary goes to the meeting then she receives the documentation' (if A then B) is of course logically equivalent to one of the forms 'Mary goes to the meeting only if she receives the documentation' (A only if B) (Jeffrey, 1981). But their equivalence is not immediately obvious to most people. Nonetheless, people can appreciate that both conditionals are false in the same situation, that is, Mary goes to the meeting and she does not receive the documentation (A and not-B). The psychological interpretations of 'only if' have been debated (Evans, 1977; Keenan, 1971). An 'only if' assertion seems to work best when its second component (B) precedes in time its first component (A) (Cheng & Holyoak, 1985; Evans & Beck, 1981; Ormerod, Manktelow, & Jones, 1993; Rips & Marcus, 1977; Thompson & Mann, 1995). 'Only if' often contains a precondition in its consequent, for example, 'you can go out to play only if you tidy your room' (Girotto, Mazzocco, & Cherubini, 1992).

Our view is that reasoners keep in mind different possibilities to understand 'if' and 'only if'. We suggest that reasoners think about two possibilities from the outset to understand 'only if' (Johnson-Laird & Byrne, 1989). As a result, they can readily make both the *modus ponens* (A therefore B) inference and the *modus tollens* (not-B therefore not-A) inference from 'only if', unlike from 'if'. Previous research has shown that the difference between *modus ponens* and *modus tollens* inferences from 'if' disappears with 'only if', and reasoners can make both inferences readily (Evans & Beck, 1981; Roberge, 1978). Moreover, we suggest that the possibilities people keep in mind preserve the temporal order of occurrence of events in the world (Byrne, 2005; Byrne, Segura, Culhane, Tasso, & Berrocal, 2000). Reasoners understand 'A only if B' by thinking in the direction B to A (Carriedo, García-Madruga, Moreno, & Gutiérrez, 1999; Evans & Newstead, 1977; see also García-Madruga, Gutiérrez, Carriedo, Moreno, & Johnson-Laird, 2002). Hence they

make more readily the *affirmation of the consequent* (B therefore A) and *modus tollens* inferences (not-B therefore not-A) from 'only if' than from 'if'. These two suggestions combine to result in the novel proposal that people understand 'A only if B' by thinking about two possibilities, in the direction B to A, that is, they think about 'B and A' and 'not-B and not-A'. Existing evidence that supports this account lies in the observation that reasoners are faster to make the backward 'B to A' inferences than the forward 'A to B' inferences from 'only if' (García-Madruga, Carriedo, Moreno, Gutiérrez, & Schaeken, 2008; Grosset & Barrouillet, 2003; Santamaría & Espino, 2002). Moreover, when reasoners read 'A only if B', they are primed to understand quickly 'B and A', and also 'not-B and not-A' (Santamaría & Espino, 2002, Experiment 3). Participants read 'not-B and not-A' reliably faster when primed by 'A only if B' than when primed by 'if A then B'.

On our account, people understand counterfactual 'only if', e.g., 'Mary would have gone to the meeting only if she had received the documentation' and indicative 'only if', e.g., 'Mary went to the meeting only if she received the documentation' by thinking about very similar possibilities, as Table 1 shows. For indicative 'only if' they think about two true possibilities, whereas for counterfactual 'only if' one of the possibilities is understood to correspond to the counterfactual conjecture (B and A), and the other to the presupposed facts (not-B and not-A). Nonetheless, because people envisage these two possibilities for both indicative and counterfactual 'only if', our account makes novel predictions about the possibilities people will judge to be consistent with the conditionals and about the inferences they will draw from them. We test this new hypothesis about the mental representations and cognitive processing of 'only if' by comparing indicative and counterfactual 'if' and 'only if' conditionals.

4. Experiment 1

The aim of the first experiment was to examine the possibilities that reasoners judge to be consistent with indicative and counterfactual 'only if', and to compare them to indicative and counterfactual 'if'. We gave participants problems that consisted of a conditional premise, e.g., 'Joe would have been in Meath only if Ann had been in Dublin'. They were given four possibilities: 'Joe was in Meath and Ann was in Dublin', 'Joe was not in Meath and Ann was not in Dublin', 'Joe was not in Meath and Ann was in Dublin', and 'Joe was in Meath and Ann was not in Dublin'. They were asked to judge whether each possibility was 'consistent', 'inconsistent', or 'irrelevant' with regard to the conditional. We examined judgements of consistency and inconsistency, in line with previous research on counterfactuals, because of the unsuitability of judgements of truth (e.g., true, false, irrelevant) for counterfactual conditionals (Thompson & Byrne, 2002; see also Schroyens, 2008).

We derived the following set of predictions from our novel account of indicative and counterfactual 'only if':

Table 1
A summary of the possibilities people envisage initially for indicative and counterfactual 'only if' and 'if then'.

Form	Mood	
	Indicative	Counterfactual
If A then B	A and B	A and B (conjecture)
	. . .	Not-A and not-B (presupposed facts)
		. . .
A only if B	B and A	B and A (conjecture)
	Not-B and not-A	Not-B and not-A (presupposed facts)
	. . .	. . .

Note: the ellipsis (. . .) indicates that there are other true possibilities consistent with the conditional which reasoners may have fleshed out to be more explicit but which are not mentally represented in the initial models.

(1) Indicative versus counterfactual 'if': in line with earlier research, we predicted people would tend to judge that the negative possibility (not-A and not-B) is consistent with a counterfactual more often than they would make this judgement for an indicative conditional (Thompson & Byrne, 2002). People keep in mind the negative possibility from the outset for counterfactual 'if' but not for indicative 'if'. We did not expect a difference in judgements about the affirmative possibility (A and B) for factual and counterfactual 'if'.

(2) Indicative versus counterfactual 'only if': in contrast to our predictions for 'if', for 'only if' we expected that judgements about the consistency of possibilities for 'only if' would show no differences between indicative and counterfactual conditionals. People keep in mind two possibilities for both indicative and counterfactual 'only if'.

(3) 'Only if' versus 'if' indicative conditionals: our account commits us to the prediction of a difference in consistency judgements for indicative 'only if' and indicative 'if'. We expect that reasoners will judge the 'not-A and not-B' possibility to be consistent more often following indicative 'only if' than indicative 'if', and that there will be no difference for the affirmative 'A and B' possibility. Earlier findings suggest that there may be no differences in judgements of the truth of such cases (e.g., Evans & Newstead, 1977). However, on our account reasoners think about the 'A and B' possibility for both indicative 'if' and indicative 'only if', but they think about the negative possibility only for indicative 'only if'.

(4) 'Only if' versus 'if' counterfactuals: we expect that there should be no differences in judgements of consistency for counterfactual 'only if' and counterfactual 'if'. On our account, both counterfactuals are understood by thinking about two possibilities.

We confined our predictions to the 'A and B' and 'not-A and not-B' cases and refrained from making predictions of any differences between the linguistic forms ('if' and 'only if') or the grammatical moods (subjunctive and indicative) for the possibilities 'not-A and B' and 'A and not-B'. People do not often judge these possibilities to be consistent in such tasks: for example, they judge the 'A and not-B' possibility to be consistent or true about 5% of the time, and the 'not-A and B' possibility to be consistent or true about 15% of the time, as a recent meta-analysis shows (Schroyens, 2008).

4.1. Method

4.1.1. Materials and design

We gave one group of participants a set of problems based on indicative conditionals (e.g., 'if Joe was in Meath then Ann was in Dublin'), and the other group received counterfactual conditionals (e.g., 'if Joe had been in Meath then Ann would have been in Dublin'). There were six problems in total, three 'if A then B' and three 'A only if B' problems which were presented in blocks and counterbalanced across participants in a within-participant design. Each problem consisted of a conditional followed by four possibilities presented in random order ('A and B', 'not-A and B', 'A and not-B' and 'not-A and not-B'). We used three sorts of neutral content: locations, ingredients (e.g., if David used basil then Jodie used parsley), and actions (e.g., if Mary walked then Fred jumped). To control for content effects, the contents were assigned to the problems at random twice to make two different sets of problems.

4.1.2. Procedure

The participants were tested in groups and were asked to sign a consent form prior to participating. The experimenter read the

Table 2
Percentages of possibilities judged as 'consistent' in Experiment 1.

		A and B	Not-A and not-B	Not-A and B	A and not-B
If	Indicative	99	58	3	4
	Counterfactual	100	90	0	0
Only if	Indicative	97	70	10	3
	Counterfactual	100	89	13	0

instructions aloud which asked participants to complete the problems in the order they were presented and not to change any responses once they had made them. The experimenter answered any questions the participants had and participants were advised that they could take as long as they needed to complete the task.

4.1.3. Participants

The participants were 44 undergraduate psychology students from Dublin Business School who took part voluntarily. They had not received any prior training in logic nor had they taken part in a reasoning study before. There were 31 women and 13 men and their average age was 25 years (with an age range from 19 to 40 years). They were assigned at random to the indicative ($n = 23$) or counterfactual ($n = 21$) groups.

4.2. Results and discussion

We carried out a $2 \times 2 \times 4$ ANOVA on the frequency with which participants judged a possibility to be consistent. The factors were linguistic mood (indicative, subjunctive), linguistic form (if, only if) and possibility (A and B, not-A and not-B, not-A and B, A and not-B), with repeated measures on the last two factors. There were main effects of linguistic form, $F(1, 42) = 4.72$, Mse = .90, $p = .02$; mood, $F(1, 42) = 5.67$, Mse = 2.92, $p = .01$; and possibility, $F(1, 65) = 325.67$, Mse = 366.09, $p < .01$.[1] Form and mood did not interact, $F(1, 42) = .15$, Mse = .29, $p = .35$, but each interacted with possibility: form and possibility, $F(2, 85) = 2.71$, Mse = .77, $p = .04$; and mood and possibility, $F(3, 126) = 6.08$, Mse = 3.57, $p < .01$. The three-way interaction was marginal, $F(3, 126) = 1.72$, Mse = .33, $p = .08$. To test our predictions we conducted a series of planned comparisons on the options 'A and B' and 'not-A and not-B'. As Table 2 shows, 'not-A and B' and 'A and not-B' were selected as consistent infrequently ($\leqslant 13\%$).

(1) Indicative versus counterfactual 'if': as we predicted, participants judged 'not-A and not-B' to be consistent more often for counterfactual 'if' than indicative 'if', 90% versus 58%, $t(39) = 2.88$, $p = .01$, equal variances not assumed, and there was no difference for 'A and B', 99% versus 100%, $t(42) = .96$, $p = .17$.[2] These findings support the idea that people think about the negative possibility 'not-A and not-B' for a counterfactual more than an indicative conditional. The result is consistent with previous findings (e.g., Byrne & Tasso, 1999; Thompson & Byrne, 2002).

(2) Indicative versus counterfactual 'only if': as we expected, there were no differences between indicative and counterfactual 'only if' for 'A and B', 97% versus 100%, $t(42) = .96$, $p = .17$. Unexpectedly, participants judged the 'not-A and

not-B' possibility to be consistent more often for counterfactual 'only if' than indicative 'only if', 89% versus 70%, $t(42) = 2.25$, $p = .01$. However, participants judged the 'not-A and not-B' possibility to be consistent with indicative 'only if' (70%), reliably more so than indicative 'if' as we will see below, supporting our suggestion that they think about two possibilities to understand indicative 'only if'. Nonetheless, they appear to think about the negative possibility even more readily for counterfactual 'only if'. We return to this result in the next experiment.

(3) 'Only if' versus 'if' indicative conditionals: participants judged 'not-A and not-B' to be consistent more often for indicative 'only if' than indicative 'if' as we predicted, 70% versus 58%, $t(22) = 2.01$, $p = .03$. There were no differences for 'A and B', 97% versus 99%, $t(22) = 1.00$, $p = .17$, again as expected. These results support our hypothesis that reasoners initially think about just one possibility for indicative 'if then' conditionals, but two for indicative 'only if' conditionals.

(4) 'Only if' versus 'if' counterfactuals: there were no differences between 'if' and 'only if' counterfactual conditionals for 'not-A and not-B', 90% versus 89%, $t(20) = .25$, $p = .40$, and 'A and B', 100% for each, $t(22) = .00$, $p = .5$. This finding is consistent with our suggestion that people keep in mind two possibilities for both sorts of counterfactual.

The results of the experiment support our hypothesis that people understand 'A only if B' by thinking about two possibilities. The results are difficult to explain on the view that people understand indicative 'A only if B', by keeping in mind a single possibility (Evans, 1993; Grosset & Barrouillet, 2003; Santamaría & Espino, 2002). Such an account cannot explain the finding that participants judge 'not-A and not-B' to be consistent more often for indicative 'only if' than for indicative 'if'. The single-possibility view proposes that for 'A only if B' people think about a single possibility in the temporal order 'B and A' and they have a processing preference for making inferences in this direction, i.e., from B to A, whereas for 'if A then B' they think about a single possibility in the temporal order, 'A and B', and they have a processing preference for making inferences in this direction, i.e., from A to B. As a result, the backward inferences, *affirmation of the consequent* (B therefore A) and *modus tollens* (not-B therefore not-A) are made more often from 'A only if B' (e.g., Evans, 1993; Evans, Clibbens, & Rood, 1995, 1996). In fact, the evidence that 'if' favours the forward inferences, *modus ponens* (A therefore B) and *denial of the antecedent* (not-A therefore not-B) is not clear-cut (see Evans, 1977; Evans & Beck, 1981; Johnson-Laird & Byrne, 1989).

The result that participants readily judge both 'not-A and not-B' and 'A and B' to be consistent for indicative 'only if' is problematical for the view that reasoners tend to think about just a single possibility, in line with a 'singularity principle' (Evans, 2007; Evans & Over, 2004; Evans et al., 2005). So too is the result that reasoners think about these two possibilities for counterfactual 'if' and counterfactual 'only if'). A singularity principle in the suppositional theory constrains individuals to think about only a single possibility (Evans & Over, 2004). On the suppositional account, the supposition of an antecedent and the assessment of the believability of the consequent occur in a single mental simulation. The single mental model may contain information added by pragmatic implicatures as well as information stated in the premise. Individuals assign a subjective probability to a conditional represented by a number corresponding to their strength of belief (Evans, 2007), e.g., 'if Paul wins the lottery he buys a Ferrari' is represented as:

• Paul wins the lottery → .8 Paul buys a Ferrari

Even a counterfactual conditional 'if Paul had won the lottery he would have bought a Ferrari' is represented in a single possibility, but in this case, according to the suppositional theory, the single mental model contains alternative components. Each component is annotated to include numbers corresponding to the strength of belief in the conditional, and the strength of belief in the individual components (Evans, 2007, p.74):

- Paul had won the lottery → .8 Paul would have bought a Ferrari
- [Paul did not win the lottery (.999); Paul did not buy a Ferrari (.999)].

On our account, people imagine alternative possibilities to correspond to these alternatives, the conjecture 'Paul won the lottery and he bought a Ferrari' and the presupposed facts, 'Paul did not win the lottery and he did not buy a Ferrari'. The experimental evidence supports the conclusion that people think about these two distinct alternative possibilities when they understand indicative 'A only if B', counterfactual 'if A then B' and counterfactual 'A only if B'.

Our experimental results support our predictions with one exception: participants judged 'not-A and not-B' to be consistent more often for counterfactual 'only if' (89%) than for indicative 'only if' (70%), whereas we expected no difference. Their judgements that 'not-A and not-B' is consistent with indicative 'only if' are reliably higher than for indicative 'if' (58%) and so it cannot be the case that reasoners are thinking about a single possibility in each case (pace Evans & Over, 2004; Evans et al., 2005). We conjecture that the judgements about 'not-A and not-B' may have been somewhat suppressed in the indicative 'only if' condition (70%) compared to the counterfactual 'only if' condition (89%), because of the presentation of the response options in the direction of A to B, rather than B to A. The order in which the information enters working memory (i.e., A followed by B) could impact on the judgements (Espino & Hernandez, 2009; Grosset & Barrouillet, 2003). It may be that the subjunctive mood of the counterfactual provides an additional cue to encourage the representation of the negative possibility, and so the impact of the order in which the information enters working memory is observed more strongly for indicative 'only if'. On our account, reasoners think about two possibilities to understand 'A only if B', and both possibilities are in the direction B to A, that is, they think about 'B and A' and about 'not-B and not-A' (see Table 1). Accordingly, our next experiment examines judgements of consistency for possibilities phrased in the direction of B to A.

5. Experiment 2

The aim of the second experiment was once again to examine the possibilities that reasoners judge to be consistent with indicative and counterfactual 'A only if B', and to compare them to indicative and counterfactual 'if A then B', this time for four possibilities in the B to A order: B and A, not-B and not-A, not-B and A, and B and not-A. We made the same predictions as in the previous experiment. We aimed to test the prediction that people would judge 'not-B and not-A' to be consistent with counterfactual 'only if' and indicative 'only if' equally often when the possibilities were presented in the B to A direction.

5.1. Method

5.1.1. Materials, design and procedure

The materials used were the same as the previous experiment with the exception that the possibilities were presented in the order B to A, rather than A to B. The procedure was the same as in the previous experiment. We used the same general design and in this experiment both mood and linguistic form were between-participant variables, hence there were four groups of participants: indicative 'if', indicative 'only if', counterfactual 'if' and counterfactual 'only if'.

5.1.2. Participants

The participants were 95 undergraduate psychology students from Mary Immaculate College, University of Limerick, who took part voluntarily. They had not received any prior training in logic or taken part in a reasoning study before. There were 82 women and 13 men and their average age was 22 years (with an age ranging from 17 to 52 years). They were assigned at random to one of four groups, indicative 'if' ($n = 25$), indicative 'only if' ($n = 23$), counterfactual 'if' ($n = 20$) or counterfactual 'only if' ($n = 27$).

5.2. Results and discussion

We analysed the results in an ANOVA carried out on the possibilities that participants selected as 'consistent'. The factors were linguistic mood (indicative, counterfactual), linguistic form (if A then B, A only if B) and possibility (B and A, not-B and not-A, not-B and A, B and not-A), with repeated measures on the last factor. There was a main effect of linguistic form, $F(1, 91) = 10.09$, Mse = 5.75, $p < .01$, and possibility, $F(2, 273) = 202.92$, Mse = 193.6, $p < .01$, but not of mood, $F(1, 91) = .23$, Mse = .13, $p = .32$. Possibility interacted with form, $F(3, 273) = 4.97$, Mse = 3.51, $p < .01$, but not with mood, $F(3, 273) = 5.15$, Mse = 3.64, $p < .01$, and mood and form did not interact, $F(1, 91) = .23$, Mse = .13, $p = .32$. The three-way interaction was not reliable, $F(3, 273) = 1.54$, Mse = 1.08, $p = .1$. To test our predictions we conducted a series of planned comparisons on the non-significant three-way interaction (see Winer (1971) for the legitimacy of such comparisons) to test judgements about the possibilities 'B and A' and 'not-B and not-A':

(1) Indicative versus counterfactual 'if': as we predicted, participants judged 'not-B and not-A' to be consistent more often for counterfactual 'if' than indicative 'if', 72% versus 40%, $t(43) = 2.59$, $p = .01$, and there was no difference for 'B and A', 88% versus 82%, $t(43) = .73$, $p = .24$. This finding replicates the results of Experiment 1 and extends them to possibilities in the B to A direction.

(2) Indicative versus counterfactual 'only if': as expected, there were no differences between indicative and counterfactual 'only if' for 'B and A', 93% versus 94%, $t(48) = .23$, $p = .41$, as Table 3 shows. Also as expected, participants judged 'not-B and not-A' to be consistent as often for counterfactual 'only if' as indicative 'only if', 75% versus 85%, $t(37.02) = .99$, $p = .16$, equal variances not assumed. The result corroborates the hypothesis that the direction of the response options in the previous experiment may have contributed to the unexpected difference for 'not-A and not-B' judgements between counterfactual and indicative 'only if'.

(3) 'Only if' versus 'if' indicative: as expected, participants judged 'not-B and not-A' to be consistent more often for indicative 'only if' than for 'if', 75% versus 40%, $t(46) = 2.97$,

Table 3
Percentages of possibilities judged as 'consistent' in Experiment 2.

		B and A	Not-B and not-A	B and not-A	Not-B and A
If	Indicative	88	40	9	16
	Counterfactual	82	72	5	5
Only if	Indicative	93	75	19	4
	Counterfactual	94	85	10	3

p = .01, and there were no differences for 'B and A', 93% versus 88%, $t(46)$ = .76, p = .23. The result replicates Experiment 1, for responses in the direction B to A.

(4) 'Only if' versus 'if' counterfactuals: as expected, there were no differences between 'if' and 'only if' counterfactual conditionals for 'not-B and not-A', 85% versus 72%, $t(31.46)$ = 1.33, p = .10, or 'B and A', 94% versus 82%, $t(25.6)$ = 1.51, p = .08. The result replicates the first experiment.

Experiment 2 replicates the findings of Experiment 1 and extends them to possibilities phrased in the B to A direction. The results are consistent with our proposal that people understand 'only if' conditionals by thinking about two possibilities in the B to A direction, 'B and A' and 'not-B and not-A': the direction of the response options in Experiment 1 led to an unexpected difference for 'not-A and not-B' judgements between counterfactual and indicative 'only if' and this difference was eliminated in Experiment 2 for 'not-B and not-A' judgements. However, a comparison of the results of Experiments 1 and 2 shows that participants appear to judge 'A and B' and 'B and A' to be consistent with indicative 'A only if B' equally often (97% and 93%, respectively), and with counterfactual 'A only if B' equally often (100% and 94%, see Tables 2 and 3). Likewise, they judge 'not-A and not-B' and 'not-B and not-A' to be consistent with indicative 'A only if B' equally often (70% and 75%) and with counterfactual 'A only if B' equally often (89% versus 85%). Hence, we test our proposal further in the next experiment by comparing forward inferences (e.g., not-A therefore not-B) and backward inferences (e.g., not-B therefore not-A) for 'if A then B' and 'A only if B'. We predict systematic differences between 'if A then B' and 'A only if B' to arise not only because of the number of possibilities to be kept in mind from the outset, but also because of the direction of the possibilities.

Experiment 2 provides further evidence that people tend to think about two possibilities to understand indicative 'A only if B' conditionals whereas they think about a single possibility initially to understand indicative 'if A then B' conditionals. According to the suppositional theory: 'The proposal that two models are used to represent the 'only if' conditional is inconsistent with the singularity principle of our own hypothetical thinking theory.' (Evans & Over, 2004, p. 70). As a result, the data from Experiment 2 present difficulties for the idea that individuals think about just one possibility for indicative 'A only if B'.

People understand indicative and counterfactual 'A only if B' conditionals by keeping in mind the same two possibilities, 'B and A' and 'not-B and not-A'. In contrast, people understand indicative and counterfactual 'if A then B' conditionals by thinking about different possibilities initially: they think about a single possibility to understand indicative 'if', 'A and B', and they think also about a second possibility to understand counterfactual 'if', 'not-A and not-B'. The differences between indicative and counterfactual 'if' and 'only if' lead us to predict that different inferences will be made from the different sorts of conditionals and our final experiment tests the inferences people make from counterfactual 'only if' compared to indicative 'only if'.

6. Experiment 3

The aim of the experiment was to examine the inferences that reasoners make from indicative and counterfactual 'if' and 'only if'. Participants were given a conditional such as, 'Joe would have been in Meath only if Ann had been in Dublin'. They were given a second premise such as 'Ann was in Dublin'. They were asked to say what if anything follows from the premises, and they chose their conclusion from a set, 'therefore, (a) Joe was in Meath, (b) Joe was not in Meath, (c) Joe may or may not have been in Meath'.

We derived a novel set of predictions about the expected frequency of inferences based on our proposal that people think about two possibilities in the direction of B to A for 'A only if B' indicative and counterfactual conditionals whereas they think about one possibility in the direction of A to B for 'if A then B' indicative conditionals and two possibilities in the direction A to B for 'if A then B' counterfactuals:

(1) Indicative versus counterfactual 'if A then B': on our account people think about one possibility initially for indicative 'if A then B' (A and B), and two possibilities for counterfactual 'if A then B' (A and B, not-A and not-B). Accordingly, we expected that reasoners would make more of the negative inferences, *modus tollens* (not-B therefore not-A) and *denial of the antecedent* (not-A therefore not-B), from counterfactual 'if A then B' compared to indicative 'if A then B'. We expected that reasoners would make the same frequency of the affirmative inferences, *modus ponens* (A therefore B) and *affirmation of the consequent* (B therefore A). We expected to replicate the findings of previous research for these inferences from indicative and counterfactual 'if A then B' (Byrne & Tasso, 1999).

(2) Indicative versus counterfactual 'A only if B': on our account people think about two possibilities for indicative and counterfactual 'A only if B' (B and A, not-B and not-A) and so we expected that reasoners would make the same frequency of inferences from both indicative and counterfactual 'A only if B', unlike indicative and counterfactual 'if A then B'.

(3) 'A only if B' versus 'if A then B' indicative conditionals: on our account reasoners think about two possibilities for indicative 'A only if B', in the direction B to A; they envisage initially one possibility for 'if A then B' in the direction A to B. On this account reasoners should make more *modus tollens* (not-B therefore not-A) and *affirmation of the consequent* (B therefore A) inferences from 'A only if B' compared to 'if A then B', because the possibilities they think about for 'A only if B' are in the same B to A direction as these inferences. On this account, reasoners should also make more of the negative inferences, *modus tollens* (not-B therefore not-A) and *denial of the antecedent* (not-A therefore not-B), from 'A only if B' because they think about the negative possibility from the outset. Hence we predict more *modus tollens*, *affirmation of the consequent*, and *denial of the antecedent* inferences from indicative 'only if' compared to indicative 'if'.

(4) 'A only if B' versus 'if A then B' counterfactuals: on our account reasoners think about two possibilities for both sorts of counterfactuals, but in a different direction, B to A for 'only if' and A to B for 'if'. Accordingly, we predicted that reasoners would make more *modus tollens* and *affirmation of the consequent* inferences from 'only if' counterfactuals compared to 'if' counterfactuals, because the possibilities they think about for 'only if' counterfactuals are in the same B to A direction as these inferences. We also predicted they would make more *modus ponens* and *denial of the antecedent* inferences from 'if' counterfactuals for the same reason – the possibilities they think about for 'if' counterfactuals are in the same A to B direction as these inferences.

6.1. Method

6.1.1. Materials and design

We constructed two sets of problems, an indicative set and a counterfactual set and all the assertions were in the past tense. Each set contained 24 problems, 12 for the linguistic form 'if' and 12 for 'only if'. We used three sorts of neutral content as in the pre-

vious experiment: locations, actions and ingredients (e.g., Joe would have been in Meath only if Ann had been in Dublin). Each problem consisted of a conditional premise and a categorical premise corresponding to *modus ponens* (Joe was in Meath), *modus tollens* (Ann was not in Dublin), *denial of the antecedent* (Joe was not in Meath) and *affirmation of the consequent* (Ann was in Dublin) and participants had a choice of three conclusions to select from as shown in the example earlier. Each type of inference was presented once for each content and for each linguistic form (i.e., 4 inferences × 3 contents × 2 linguistic forms = 24 problems). To control for content effects, the contents were assigned to the problems at random twice to make two different sets of problems. The problems were presented in two blocks ('if' problems and 'only if' problems) and these blocks were counterbalanced across all participants. The 12 problems within each block were presented in a different random order for each participant.

6.1.2. Procedure

Participants were tested individually or in groups of two or three participants. They first signed a consent form and then the 24 problems were presented on Macintosh computers using Super-Lab 1.75. The instructions were presented on the computer and included an example problem and three practice problems (based on conjunctions and disjunctions of shapes) to familiarise participants with the task presentation and keyboard response options. Participants were advised that they could take as long as they needed to complete the task. They pressed the space bar to view each new piece of information (the conditional, the minor premise, the conclusion set), and each remained on screen to be joined by the subsequent information. The participants pressed one of the keys labelled 'a', 'b', or 'c' to select a conclusion. These keys were in the centre of the keyboard and corresponded to the T, G, and B keys, respectively.

6.1.3. Participants

The participants were 40 undergraduate psychology students from Trinity College, Dublin, who participated voluntarily for course credits. There were 28 women and 12 men and their average age was 22 years, ranging from 18 to 45 years. They had not been trained in logic nor had they participated in any previous reasoning study. They were assigned at random to the indicative group ($n = 19$) or the counterfactual group ($n = 21$).

6.2. Results and discussion

We analysed the results in an ANOVA on the endorsements of conclusions with the factors of mood (indicative, counterfactual), form (if then, only if), and inference (modus ponens, tollens, denial of the antecedent and affirmation of the consequent), with repeated measures on the last two factors. It showed a main effect of form, $F(1, 38) = 3.98$, Mse = 3.41, $p = .03$, and inference, $F(2, 89) = 12.62$, Mse = 13.09, $p < .01$, but not of mood, $F(1, 38) = .05$, Mse = .09, $p = .41$. Form interacted with inference, $F(2, 94) = 4.16$, Mse = 2.38, $p = .01$, but not mood, $F(1, 38) = 1.70$, Mse = 1.46, $p = .10$, and mood and inference did not interact, $F(3, 114) = .67$, Mse = .55, $p = .29$. The three-way interaction was reliable, $F(3, 114) = 2.61$, Mse = 1.24, $p = .03$. To test our predictions we carried out a series of planned comparisons:

(1) Indicative versus counterfactual 'if A then B': we replicated the findings of previous research that reasoners tend to make more negative inferences from counterfactual 'if' compared to indicative 'if', and the difference was reliable for the *denial of the antecedent*, 84% versus 63%, $t(38) = 1.79$, $p = .04$, although not for *modus tollens*, 71% versus 79%, $t(38) = .66$, $p = 0.26$, perhaps because we found a particularly high rate

of *modus tollens* for indicative 'if' in this experiment (79%). As expected there were no differences in the frequency of the affirmative inferences, *modus ponens*, 100% in each case, $t(38) = .00$, $p = .5$, and *affirmation of the consequent*, 67% in each case $t(38) = .00$, $p = .5$, as Table 4 shows.

(2) Indicative versus counterfactual 'A only if B': as expected, reasoners made the same frequency of inferences from both indicative and counterfactual 'only if' conditionals, unlike indicative and counterfactual 'if'. There were no differences in the frequency of negative inferences from counterfactual and indicative 'only if', for the *denial of the antecedent*, 68% versus 79%, $t(38) = 1.00$, $p = .17$, and *modus tollens*, 92% versus 96%, $t(38) = 1.09$, $p = .14$. Likewise, there were no differences in the affirmative inferences from counterfactual and indicative 'only if', for *modus ponens*, 98% and 95%, $t(38) = .92$, $p = .18$, and *affirmation of the consequent*, 73% and 84%, $t(38) = 1.03$, $p = .15$. The result is consistent with our hypothesis that people think about two possibilities for both indicative and counterfactual 'only if' (B and A, not-B and not-A).

(3) 'A only if B' versus 'if A then B' indicative conditionals: as we predicted, reasoners made more of the backward inferences from 'A only if B': significantly so for *modus tollens*, 96% versus 79%, $t(18) = 2.04$, $p = .03$, and marginally so for *affirmation of the consequent* inferences, 84% versus 67%, $t(18) = 1.56$, $p = .07$, from 'A only if B' compared to 'if A then B'. This result supports the suggestion that the possibilities they think about for 'only if' are in the same B to A direction as these inferences. Also as we expected, they made more of the negative inferences, *modus tollens* (as reported above) and marginally more *denial of the antecedent*, 79% versus 63%, $t(18) = 1.63$, $p = .06$, from 'only if' and the result supports the suggestion that they think about the negative possibility 'not-B and not-A' as well as the affirmative one 'B and A'. As expected there were no differences for *modus ponens*, 95% versus 100%, $t(18) = 1.37$, $p = .09$. The results for *modus tollens* and *affirmation of the consequent* are compatible with previous research (e.g., Evans et al., 1995). The predicted result for the *denial of the antecedent* inference corroborates our suggestion that reasoners think about the negative possibility, and the magnitude of the difference between 'only if' and 'if' for this inference (79% versus 63%) is impressive given that for 'only if' it is a 'backward' inference (i.e., in the direction A to B whereas the mental representation is in the direction B to A). In fact earlier studies also show some small differences in the predicted direction between 'only if' and 'if' for *denial of the antecedent* (e.g., Evans et al., 1995), despite it being a negative and backward (for only if) inference.

(4) 'Only if' versus 'if' counterfactuals: as we expected reasoners made more of the B to A inferences from 'A only if B' counterfactuals compared to 'if A then B' counterfactuals, and the difference was reliable for *modus tollens*, 92% versus 71%, $t(20) = 2.44$, $p = .01$, although not for *affirmation of the consequent*, 73% versus 67%, $t(20) = .68$, $p = .25$. The results

Table 4
Percentages of conclusions endorsed by participants in Experiment 3.

		Modus ponens	Affirmation of the consequent	Modus tollens	Denial of the antecedent
If A then B	Indicative	100	67	79	63
	Counterfactual	100	67	71	84
A only if B	Indicative	95	84	96	79
	Counterfactual	98	73	92	68

provide some support for the suggestion that the possibilities people think about for 'A only if B' counterfactuals are in the same B to A direction as these inferences. Also as we expected, they made more of the A to B inferences from 'if A then B' counterfactuals compared to 'A only if B' counterfactuals and the difference was reliable for the *denial of the antecedent* inferences, 84% versus 68%, $t(20) = 2.35$, $p = .02$, although not for *modus ponens* which was at ceiling, 98% versus 100%, $t(20) = 1.00$, $p = .17$. The result provides some support for the suggestion that the possibilities people think about for 'if A then B' counterfactuals are in the same A to B direction as these inferences. The result supports our suggestion that reasoners keep two possibilities in mind for both sorts of counterfactuals, but in a different direction, B to A for 'A only if B' and A to B for 'if A then B'.

The results for inferences in this experiment are consistent with the results for consistency judgements in Experiments 1 and 2. The results are difficult to explain on the view that 'A only if B' is paraphrased to capture a double negative similar to 'if not-B then not-A' (e.g., Braine, 1978). On this account, the *modus tollens* inference (not-B therefore not-A) is made readily because it is in effect a *modus ponens* one (A therefore B) from 'if not-B then not-A' (Braine, 1978). But on such an account *modus ponens* should be difficult (it is in effect *modus tollens* from 'if not-B then not-A'), and as the result show, it is not (see also Johnson-Laird & Byrne, 1989).

The results are also difficult to explain on the view that 'A only if B' is interpreted as 'A, if B', in other words, 'if B, A'. Note that 'if B, A' is false in the situation 'B and not-A'; it is true in the situations 'B and A', 'not-B and not-A', and 'not-B and A'. But if individuals interpreted 'A only if B' to mean 'if B then A', they should not make the *modus ponens* inference ('A only if B', and 'A', therefore 'B'). Nor should they make the *modus tollens* inference. Yet they make these inferences readily as previous studies have shown (Evans, 1977; Evans & Beck, 1981; Johnson-Laird & Byrne, 1989). Moreover, as Table 3 shows, only 19% of participants in Experiment 2 judged the 'B and not-A' situation to be consistent for 'A only if B'; 9% made this judgement for 'if A then B'.

The results replicate the findings of previous experiments on inferences from indicative and counterfactual 'if A then B'. They provide the first comparison of inferences from indicative and counterfactual 'A only if B'. They support the novel proposal that reasoners understand and reason from indicative and counterfactual 'A only if B' by thinking about two possibilities in the direction B to A (B and A, and not-B and not-A).

7. General discussion

We have provided a new account of the mental representations and cognitive processes underlying 'A only if B'. An indicative conditional, 'if A then B' is understood by keeping in mind initially the possibility 'A and B' (Johnson-Laird & Byrne, 2002). In contrast, 'A only if B' is understood by thinking about two possibilities from the outset, in the B to A direction: 'B and A' and 'not-B and not-A'. We have also provided the first account of the mental representations and cognitive processes underlying counterfactual 'A only if B'. The counterfactual 'if A had been then B would have been' is understood by envisaging initially two possibilities, the conjecture 'A and B' and the presupposed facts 'not-A and not-B' (Byrne & Tasso, 1999; Thompson & Byrne, 2002). The counterfactual 'A would have been only if B had been' is understood initially by thinking about the same two possibilities as for the indicative 'A only if B', the conjecture, 'B and A', and the presupposed facts 'not-B and not-A'.

The results of the three experiments support this new account of indicative and counterfactual 'A only if B' and corroborate existing accounts of indicative and counterfactual 'if A then B':

(1) For indicative and counterfactual 'if A then B', people judge the possibility 'not-A and not-B' to be consistent with the counterfactual more than the indicative conditional, as Experiments 1 and 2 showed, and they make more of the negative inferences from the counterfactual than the indicative conditional, as Experiment 3 showed. The results replicate the findings of previous research, supporting the view that people think about a single possibility for indicative 'if' (A and B), and two possibilities for counterfactual 'if' (A and B, not-A and not-B) (Byrne & Tasso, 1999; Thompson & Byrne, 2002).

(2) For indicative and counterfactual 'only if', people judge the possibilities 'B and A' and 'not-B and not-A' to be consistent for both conditionals, as Experiment 2 showed, and they make the same frequency of inferences from the conditionals, as Experiment 3 showed. The results support the view that people think about two possibilities for indicative and counterfactual 'only if' in the direction B to A (B and A, not-B and not-A).

(3) For 'only if' and 'if' indicative conditionals, people judge the 'not-B and not-A' possibility to be consistent more often for indicative 'only if' than indicative 'if', and there is no difference for 'B and A', as Experiments 1 and 2 show. They make more of the B to A inferences, *modus tollens* and *affirmation of the consequent*, and more of the negative inferences, *modus tollens* and *denial of the antecedent*, from 'only if' than from 'if', as Experiment 3 shows.

(4) For 'only if' versus 'if' counterfactuals, people judge the 'B and A' and the 'not-B and not-A' possibility to be consistent equally often for both sorts of conditionals, as Experiments 1 and 2 show. However, the direction of the possibilities has an influence on the inferences reasoners make, as Experiment 3 shows. Reasoners made more of the backwards *modus tollens* inference and fewer of the forwards *denial of the antecedent* inferences from 'only if' than from 'if'.

We suggest that reasoners think initially about two possibilities to understand 'only if', 'B and A', and 'not-B and not-A'. Our suggestion is distinct from the view that 'only if' is a biconditional ('if and only if A then B'). On a biconditional account both the denial of the antecedent and the affirmation of the consequent inference should be made more often from 'only if' than from 'if'. On our account, reasoners may tend to make more of these inferences, because they have constructed two possibilities, but prudent reasoners can flesh out these possibilities to think about a third one, 'B and not-A'. Previous studies have shown that the affirmation of the consequent inference is made more readily and more quickly from 'only if' (Evans, 1977; Evans & Beck, 1981; Johnson-Laird & Byrne, 1989; Santamaría & Espino, 2002), and our results are consistent with them. But, the denial of the antecedent inference is unstable. Most studies of 'only if' have been based on truth table tasks (e.g., Evans et al., 1996) or the selection task (e.g., Evans, Legrenzi, & Girotto, 1999), but the few inference studies show that the denial of the antecedent inference is sometimes made more often, sometimes less often, and sometimes the same (e.g., Evans, 1977; Evans & Beck, 1981; Johnson-Laird & Byrne, 1989). In Experiment 3, reasoners made the denial of the antecedent inference more often from 'only if' than from 'if', as we predicted, corroborating our suggestion that they thought about the negative possibility (not-B and not-A). Nonetheless, reasoners can flesh out their initial understanding of 'only if' to be more explicit if necessary, and they can appreciate that a third possibility is consistent, 'B and not-A'. Just

as counterfactual 'if' is not necessarily interpreted as a biconditional 'if and only if' (Thompson & Byrne, 2002), so too indicative and counterfactual 'only if' are not necessarily interpreted as biconditionals. For example, in Experiment 1 the 'not-A and B' was judged consistent with indicative 'only if' (10%) and counterfactual 'only if' (13%) more so than indicative or counterfactual 'if' (3% and 0%, respectively); likewise in Experiment 2 the 'B and not-A' possibility was judged consistent with indicative 'only if' (19%) and counterfactual 'only if' (10%) more so than indicative or counterfactual 'if' (9% and 5%, respectively).

The results are difficult to explain if reasoners tend to construct only a single mental model (Evans, 2007; Evans & Over, 2004). First, the results go against the idea that people keep in mind only a single possibility for 'A only if B' even if in the direction B to A (Evans, 1993; Grosset & Barrouillet, 2003; Santamaría & Espino, 2002). The observation of no difference between indicative and counterfactual 'only if' is difficult to explain if a single possibility is envisaged for indicative 'only if'. In fact, previous observations that reasoners read 'not-B and not-A' faster when they are primed by 'A only if B' than 'if A then B' (Santamaría & Espino, 2002) also go against the idea that people keep a single possibility in mind. Second, the results also go against the idea that people construct only a single mental model to understand counterfactual conditionals (Evans, 2007; Evans & Over, 2004; Evans et al., 2005). The experiments reported here show that participants judge the possibility 'not-A and not-B' to be consistent with counterfactual 'if' more than indicative 'if' and they make more of the negative inferences from counterfactual 'if' than indicative 'if'. Both of these lines of evidence, from indicative 'only if' and from counterfactuals, are incompatible with the singularity principle of the suppositional theory (Evans & Over, 2004; Evans et al., 2005).

When people think about a counterfactual, e.g., 'if there had been a triangle on the blackboard there would have been a circle' they think about two possibilities, for counterfactuals based on 'if' (Byrne & Tasso, 1999; Santamaría et al., 2005), 'even if' (Moreno-Rios, García-Madruga, & Byrne, 2008), and 'only if', as we have shown. One possibility corresponds to the conjecture temporarily supposed to be true, 'there was a triangle and there was a circle' and the other to the presupposed facts 'there was no circle and there was no triangle'.

The generation of counterfactual thoughts is very common in everyday life, especially after bad outcomes (e.g., McEleney & Byrne, 2006; Roese, 1997; Walsh & Byrne, 2007). People express their counterfactual thoughts in many different ways. The assertion 'if I had married a millionaire I would have owned a string of horses', conveys subtly different information from the assertion, 'I would have married a millionaire only if I had owned a string of horses'. We suggest that reasoners keep in mind initially two possibilities to understand both, the conjecture 'I married a millionaire and owned a string of horses', and the presupposed facts 'I did not marry a millionaire and did not own a string of horses'. But the possibilities differ in their temporal direction, A to B for the former ('I married a millionaire and owned a string of horses') and B to A for the latter ('I owned a string of horses and married a millionaire'). The rich complexity of language allows counterfactual thoughts of varying nuances to be conveyed readily, as our exploration of 'only if' counterfactuals suggests.

References

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1–21.

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

Byrne, R. M. J. (2002). Mental models and counterfactual thinking. *Trends in Cognitive Sciences, 6*, 405–445.

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

Byrne, R. M. J. (2007). Precis of the rational imagination. *Behavioral and Brain Sciences, 30*, 439–480.

Byrne, R. M. J., & Johnson-Laird, P. N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*(7), 282–287.

Byrne, R. M. J., & Egan, S. M. (2004). Counterfactual and pre-factual conditionals. *Canadian Journal of Experimental Psychology, 58*(2), 113–120.

Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory and Cognition, 28*, 264–281.

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

Carriedo, N., García-Madruga, J., Moreno, S., & Gutiérrez, F. (1999). How does content affect 'unless' conditional reasoning? In *Proceedings of the European conference on cognitive science* (pp. 271–277). Sienna, Italy.

Cheng, P., & Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391–416.

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

Espino, O., Santamaría, C., & Byrne, R. M. J. (2009). People think about what is true for conditionals, not what is false. *Quarterly Journal of Experimental Psychology*, in press.

Espino, O., & Hernandez, E. (2009). Effect of directionality on conditionals. *Psicologica, 30*(1), 41–57.

Evans, J. Bt. B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29*, 297–306.

Evans, J. Bt. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition, 48*, 1–20.

Evans, J. St. B. T. (2007). *Hypothetical thinking*. Hove: Psychology Press.

Evans, J. Bt. B. T., & Beck, M. A. (1981). Directionality and temporal factors in conditional reasoning. *Current Psychological Research, 1*, 111–120.

Evans, J. Bt. B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*, 644–670.

Evans, J. Bt. B. T., Clibbens, J., & Rood, B. (1996). The role of explicit and implicit negation in conditional reasoning. *Journal of Memory and Language, 35*, 392–409.

Evans, J. Bt. B. T., Legrenzi, P., & Girotto, V. (1999). The influence of linguistic form on reasoning: The case of matching bias. *Quarterly Journal of Experimental Psychology, 52*, 185–216.

Evans, J. Bt. B. T., & Newstead, S. E. (1977). Language and reasoning: A study of temporal factors. *Cognition, 5*, 265–283.

Evans, J. Bt. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum Associates Inc..

Evans, J. Bt. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.

Evans, J. Bt. B. T., Over, D. E., & Handley, S. J. (2005). Supposition, extensionality and conditionals: A critique of Johnson-Laird & Byrne (2002). *Psychological Review, 112*, 1040–1052.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition, 77*, 1–79.

Fillenbaum, S. (1974). Information amplified: Memory for counterfactual conditionals. *Journal of Experimental Psychology, 102*, 44–49.

García-Madruga, J. A., Carriedo, N., Moreno, S., Gutiérrez, F., & Schaeken, W. (2008). Unless reasoning. *The Spanish Journal of Psychology, 11*(2), 386–399.

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Moreno, S., & Johnson-Laird, P. N. (2002). Mental models in deductive reasoning. *The Spanish Journal of Psychology, 5*, 125–140.

Girotto, V., Mazzocco, A., & Cherubini, P. (1992). Judgements of deontic relevance in reasoning: A reply to Jackson and Griggs. *Quarterly Journal of Experimental Psychology, 45A*, 547–574.

Grosset, N., & Barrouillet, P. (2003). On the nature of mental models of conditional: The case of if, if then and if and only if. *Thinking and Reasoning, 9*, 289–306.

Holyoak, K. J., & Cheng, P. (1995). Pragmatic reasoning with a point of view: A response. *Thinking and Reasoning, 1*, 373–388.

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

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1989). Only reasoning. *Journal of Memory and Language, 28*, 313–330.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Erlbaum.

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

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

Keenan, E. (1971). Quantifier structures in English. *Foundations of Language, 7*, 255–284.

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

Mandel, D. R., Hilton, D. J., & Catellani, P. (2005). *The psychology of counterfactual thinking*. New York: Routledge.

Markman, K. D., Klein, W. M. P., & Suhr, J. A. (2009). *The handbook of imagination and mental simulation*. New York: Psychology Press.

McEleney, A., & Byrne, R. M. J. (2006). Spontaneous causal and counterfactual thoughts. *Thinking and Reasoning, 12*, 235–255.

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

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

Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology, 46A*, 653–677.

Quine, W. V. O. (1954). *The ways of paradox and other essays*. Cambridge, MA: Harvard University Press.

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

Rips, L. J., & Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale: Erlbaum.

Roberge, J. J. (1978). Linguistic and psychometric factors in propositional reasoning. *Quarterly Journal of Experimental Psychology, 30*, 705–716.

Roese, N. J., & Olson, J. M. (Eds.). (1995). *What might have been: The social psychology of counterfactual thinking*. Hillsdale: Erlbaum.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133–149.

Santamaría, C., & Espino, O. (2002). Conditionals and directionality: On the meaning of 'if' vs. 'only if'. *Quarterly Journal of Experimental Psychology, 55A*(1), 41–57.

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

Schroyens, W. (2008). A critical review of thinking about what is true, possible and irrelevant in reasoning from or reasoning about conditional propositions. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th annual cognitive science society* (pp. 2116–2121). Washington, DC: Cognitive Science Society.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*. Oxford: Basil Blackwell.

Stewart, A. J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Quarterly Journal of Experimental Psychology*, in press.

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

Thompson, V. A., & Mann, J. M. (1995). Perceived necessity explains the dissociation between logic and meaning: The case of 'only if'. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1554–1567.

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

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