Memory & Cognition
1999, 27 (4), 726-740

Deductive reasoning with factual, possible, and counterfactual conditionals

RUTH M. J. BYRNE
University of Dublin, Trinity College, Dublin, Ireland

and

ALESSANDRA TASSO
University of Padua, Padua, Italy

We compared reasoners' inferences from conditionals based on possibilities in the present or the past (e.g., "If Linda had been in Dublin then Cathy would have been in Galway") with their inferences based on facts in the present or the past (e.g., "If Linda was in Dublin then Cathy was in Galway"). We propose that people construct a richer representation of conditionals that deal with possibilities rather than facts: Their models make explicit not only the suppositional case, in which Linda is in Dublin and Cathy is in Galway, but also the presupposed case, in which Linda is not in Dublin and Cathy is not in Galway. We report the results of four experiments that corroborate this model theory. The experiments show that reasoners make more inferences from conditionals based on possibilities rather than on facts when the inferences depend on the presupposed case. The results also show that reasoners generate different situations to verify and falsify conditionals based on possibilities and facts.

Everyday reasoning is concerned not just with inferences about facts but also with inferences about possibilities. Our aim in this paper is to develop and test a psychological theory of reasoning with conditionals based not only on facts but also on possibilities. We will examine conditionals that deal with current facts, such as

If Linda is in Dublin then Cathy is in Galway. (1)

and we will compare them with conditionals that deal with nonfactual or hypothetical states of affairs such as present possibilities (that could happen given the actual state of the world), such as

If Linda were in Dublin then Cathy would be in Galway. (2)

We will also examine conditionals that deal with past facts, such as

If Linda was in Dublin then Cathy was in Galway. (3)

and we will compare them with conditionals that deal with past possibilities (that could have happened given the actual state but did not):

If Linda had been in Dublin then Cathy would have been in Galway. (4)

(See Johnson-Laird & Byrne, 1991.) Conditionals based on past possibilities, such as Example 4, are usually called *counterfactual* conditionals, and they have attracted attention in philosophy (e.g., Jackson, 1991; Lewis, 1973; Stalnaker, 1968) and linguistics (e.g., Dudman, 1988; Isard, 1974), as well as artificial intelligence (e.g., Ginsberg, 1986) and psychology (e.g., Johnson-Laird, 1986; Kahneman & Miller, 1986). Conditionals based on present possibilities, such as Example 2, which we will call *nonfactual* conditionals, have attracted less attention, but our points apply equally to both counterfactual and nonfactual conditionals.

Early psychological interest in counterfactual conditionals focused on aspects of memory and comprehension (e.g., Carpenter, 1973; Fillenbaum, 1974) and the role of linguistic markers such as the subjunctive mood in their usage in different languages (e.g., Au, 1983). As yet, there have been no experiments to examine how people make deductions from counterfactual and nonfactual conditionals, and we do not know the cognitive processes that underlie their evaluations of them as true or false. Accordingly, our aim in the present paper is to provide the first report of these aspects of the basic phenomena of counterfactual and nonfactual deduction.

Progress in understanding counterfactual conditionals has been slow, in part because they seem to mean some-

We thank Vittorio Girotto, Simon Handley, Phil Johnson-Laird, Mark Keane, Mac MacLachlan, Alberto Mazzocco, David O'Brien, Shane O'Mara, David Over, and Valerie Thompson for their helpful comments on the research. We are grateful to Rachel McCloy for collecting and analyzing the data for the second experiment, and to the Dublin University Arts and Social Sciences Benefactions fund for support to do so. The results of some of the experiments were reported at various conferences, including the Sixteenth Annual Conference of the Cognitive Science Society in Atlanta in 1994 and the Fifth International Colloquium on Cognitive Science in San Sebastian, Spain, in 1997. Correspondence should be addressed to R. M. J. Byrne, University of Dublin, Trinity College, Dublin 2, Ireland (e-mail: rmbyrne@tcd.ie, http://www.tcd.ie/psychology/ruth_byrne/).

thing very different from their corresponding factual conditionals (Byrne, 1997). Given the factual conditional

If John wore a seatbelt then his injuries were slight. (5)

a reasoner cannot tell whether or not John wore his belt, nor the extent of his injuries. However, the counterfactual conditional

If John had worn a seatbelt then his injuries would have been slight. (6)

not only hypothesizes a relation between John wearing his seatbelt and the extent of his injuries, but also presupposes that the factual situation is that John did not wear his seatbelt and his injuries were not slight. Unless the content, context, or general knowledge suggests the contrary, conditionals in the subjunctive mood, such as Example 6, convey information about the truth status of their antecedents and consequents, unlike conditionals in the indicative mood, such as Example 5 (Comrie, 1986). People assert some counterfactuals that seem true and others that seem false, and reasoners can distinguish between them (see, e.g., Miyamoto & Dibble, 1986; Miyamoto, Lundell, & Tu, 1989). But, how is it possible to assess whether a counterfactual is true or false if it presupposes the falsity of its antecedent and consequent? A truth-functional semantics can be provided for a factual conditional, such as Example 5 (see, e.g., Jeffrey, 1981). But the problem of counterfactual conditionals is that they do not yield readily to a truth-functional account of their semantics (Goodman, 1973; Lewis, 1973; Stalnaker, 1968). When people understand them, they seem to go beyond an analysis of the truth of their components. When listeners understand a speaker's intentions in the utterance of the counterfactual conditional, they are likely to suppose that the utterance rules out the situations in which John wore his seatbelt. In fact, on a "material implication" interpretation of a conditional, both of the following counterfactuals are true:

If John had worn his seat belt { *his injuries would have been slight.* / *his injuries would* not *have been slight.* (7)

If a general theory of conditional reasoning is to account for both counterfactual and factual conditionals, it needs to explain both their differences and their similarities (Adams, 1970, 1975; Ayers, 1965; Nute, 1980; Rescher, 1973). The vast philosophical literature on the topic suggests that conditionals are often interpreted as being supported by law-like generalizations, and the close connections between the use of counterfactuals and the comprehension of causality have been well documented (e.g., Barwise, 1986; Chisholm, 1946; Goodman, 1973; Mackie, 1973). What, then, does a counterfactual conditional mean? We will attempt to provide one possible answer to this question.

The Model Theory of Counterfactual and Nonfactual Reasoning

Our account of counterfactual deduction[1] locates it within the general domain of suppositional reasoning (see, e.g., Byrne & Handley, 1997; Byrne, Handley, & Johnson-Laird, 1995). Of course, not all suppositions are counterfactual (i.e., suppositions may be true in fact but unknown to the supposer). Counterfactual inferences require the imagination of an alternative to what is currently believed to be the factual situation. We suggest that to make inferences from counterfactuals requires reasoning procedures that are an extension of those used for conditional reasoning. We will sketch the tenets of the model theory of factual conditionals (for details of the theory and the various developing computer algorithms that implement it, see Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992; Johnson-Laird & Savary, 1995).

The first step is to understand a conditional by constructing an initial set of models of it (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). Consider the factual conditional

If Linda was in Dublin then Cathy was in Galway. (8)

Reasoners represent what is true in their models, and the conditional is consistent with three separate situations that capture the way the world would be if it were true:

Linda	Cathy
not-Linda	not-Cathy
not-Linda	Cathy

—where the diagram uses *Linda* to represent *Linda is in Dublin*, *Cathy* to represent *Cathy is in Galway*, and *not-Linda* relies on a propositional-like tag for negation to represent that Linda is not in Dublin (Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1992). Separate models are represented on separate lines: The first model corresponds to the situation where Linda is in Dublin and Cathy is in Galway. The models may be filled with information about who Linda and Cathy are, where they are located, and what the connection between their locations is, but these details are not our concern here; the structure of the models is. The set of models represents a conditional interpretation; if reasoners come to a biconditional interpretation instead, they will construct only the first two models in the set.

Reasoners construct an initial representation that is more economical than the fleshed-out set, because of the limitations of working memory:

Linda	Cathy
...	

—where the three dots represent a model with no explicit content, which captures the idea that alternatives exist that have not been mentally articulated. It may be fleshed out to be explicit if necessary, and it rules out a conjunctive interpretation (Johnson-Laird & Byrne, 1991). Reasoners represent explicitly the case mentioned in the conditional, and they keep track of the possibility that there may be alternatives to it. In fact, a less rudimentary initial representation must record that Linda being in Dublin has been represented *exhaustively* with respect to Cathy being in Galway; that is, it can occur again in the fleshed-out set

only with Cathy in Galway. We captured this notion in our diagrams with square brackets:

[Linda] Cathy

...

(Johnson-Laird & Byrne, 1991), although such mental footnotes may be rapidly forgotten (Johnson-Laird & Savary, 1995; for discussion, see Evans, 1993; Johnson-Laird, Byrne, & Schaeken, 1994; O'Brien, Braine, & Yang, 1994). The theory predicts that inferences that require a single model can be made more readily than inferences that require multiple models, and inferences that can be based on the initial set of models are easier than inferences that require the models to be fleshed out. We have corroborated the primary tenets of the model theory of conditional reasoning experimentally (Byrne, 1989a, 1989b; Byrne & Johnson-Laird, 1992; Johnson-Laird et al., 1992; Schaeken, Johnson-Laird, Byrne, & d'Ydewalle, 1995).

Reasoners engage in a similar process to understand the counterfactual conditional

If Linda had been in Dublin then Cathy would have been in Galway. (9)

(Johnson-Laird & Byrne, 1991). They construct an economical mental representation based on the information hypothesized, and they also represent the presupposed factual situation, insofar as they know it, or can induce it from the cues of mood or content. They keep track of the epistemic status of their models, keeping in mind whether the models correspond to factual or to counterfactual situations, and they tag the models accordingly:

factual:	not-Linda	not-Cathy
counterfactual:	[Linda]	Cathy

...

(Johnson-Laird & Byrne, 1991). The representation of the counterfactual may recruit memories that provide further information about the belief in the actual status of the antecedent, the consequent, and the connection between them. The representation for a nonfactual conditional—

If Linda were in Dublin then Cathy would be in Galway. (10)

—is similar to the representation for the counterfactual:

factual:	not-Linda	not-Cathy
possible:	[Linda]	Cathy

...

According to the model theory, the critical difference between factual and counterfactual conditionals is how much is made explicit in the initial set of models. The more explicit representation of the counterfactual is consistent with Fillenbaum's (1974) findings on memory for counterfactuals. He gave people sentences to read, such as

If he had caught the plane he would have arrived on time. (11)

and then gave them an unexpected memory task. He found that they tended to falsely recognize the negated antecedent—*He did not catch the plane*—and even more so, the negated consequent—*He did not arrive on time*. Thinking about what might be or what might have been is unique in that it requires reasoners to represent what is false, temporarily supposed to be true. Models represent situations that are true possibilities (Johnson-Laird & Byrne, 1991), but in the case of counterfactual thinking, reasoners represent not only the true possibilities but also the contrary-to-fact possibilities, and they "tag" their models to keep track of their epistemic status.

It is well established that inferences based on an initial representation are easier than inferences that require reasoners to flesh out models (Johnson-Laird et al., 1992), as shown in several deductive domains (Byrne & Johnson-Laird, 1989; Girotto, Mazzocco, & Tasso, 1997; Johnson-Laird & Byrne, 1989; Johnson-Laird, Byrne, & Tabossi, 1989). The model theory's proposals about the initial representations of factual and counterfactual conditionals result in a set of novel and unique predictions about their similarities and differences. We will test the proposals indirectly by examining inferences. In the four experiments that we report, we show how the difference in representations leads to differences in the inferences that reasoners make from factual and counterfactual conditionals, and in their generation of instances to verify and falsify them.

EXPERIMENT 1
Conditional Inferences About Present Facts and Possibilities

The model theory predicts that certain inferences will be made more readily from a counterfactual than from a factual conditional. Consider the *modus tollens* (MT) inference: The inferential process leads to a conclusion more readily for a counterfactual, and so more MT inferences should be made from it. The MT inference from the factual conditional

If Linda was in Dublin then Cathy was in Galway.
Cathy was not in Galway. (12)

requires reasoners to construct an initial model of the first premise—

[Linda] Cathy

...

—and a model of the second premise:

not-Cathy

The procedures that combine models may fail at this point because the sets of models appear to contain noth-

ing in common. Indeed, the most common error that reasoners make to this inference is to conclude that nothing follows (Johnson-Laird & Byrne, 1991). A prudent reasoner fleshes out the models to be more explicit:

Linda	Cathy
not-Linda	not-Cathy
not-Linda	Cathy

—and then the combination of the model of the second premise rules out all but the second model:

not-Linda	not-Cathy

The valid conclusion can be made that Linda was not in Dublin (Johnson-Laird & Byrne, 1991). Reasoners find the MT inference difficult because they must flesh out their models and keep several alternatives in mind. The MT inference should be easier from a counterfactual conditional, such as

If Linda had been in Dublin then Cathy would have been in Galway.
Cathy was not in Galway. (13)

Reasoners first construct an initial set of models of the premises:

factual:	not-Linda	not-Cathy
counterfactual:	[Linda]	Cathy

. . .

—and they combine the models with the model for the second premise:

factual:	not-Cathy

The models can be combined with no need to flesh them out further. The procedures that combine models can eliminate the counterfactual models, and the first model only is retained:

factual:	not-Linda	not-Cathy

—which supports the valid conclusion that Linda was not in Dublin. The initial representation of a counterfactual conditional is more explicit than that of the factual conditional. As a result, an MT inference can be made directly without any need to flesh out the set of models. Table 1 summarizes these processes (for ease of exposition, we omit the square brackets notation). Table 1 also illustrates the processes required for the simpler modus ponens (MP) inference. As it shows, the theory proposes that the inferential process for MP inferences is essentially the same for the two sorts of conditionals, and so it predicts no difference in the frequency of this inference. The theory makes a similar set of predictions for denial of the antecedent (DA) and affirmation of the consequent (AC) inferences. These inferences are fallacies on a "material implication" interpretation, but they are valid on a biconditional interpretation of "if" as "if and only if." Reasoners make the inferences when they fail to flesh out their models to the conditional interpretation (Johnson-Laird & Byrne, 1991). As Table 1 shows, DA inferences can be made more readily from the counterfactual than from the factual conditional. As Table 1 also shows, the AC inferences can be made by essentially the same process for the two sorts of conditionals.

In this experiment, we compared inferences based on present facts, such as

If Linda is in Dublin then Cathy is in Galway. (14)

with inferences based on present possibilities, such as

If Linda were in Dublin then Cathy would be in Galway. (15)

In a subsequent experiment we compared these conditionals with conditionals based on past facts and past possibilities. We begin with conditionals based on present possibilities, because they can more readily turn out to be true, whereas past possibilities are strictly speaking no longer possible. Present possibilities may lend themselves more readily to verification by, for example, the information that in fact Linda is in Dublin, and so they are our starting point (Byrne & Tasso, 1994). Our focus is on reasoning from conditionals based on facts and possibilities, and one way to convey facts and possibilities is to use the indicative and subjunctive moods. Our interest is not in these moods per se, but the subjunctive mood allows us syntactically to cue a listener that we wish to consider counterfactual situations. Mood is not a necessary component, since context alone can cue the counterfactuality of a situation (see, e.g., Au, 1983; Dudman, 1988).

In this experiment, we examined the four sorts of inferences—MP, MT, DA, and AC—for a factual and a nonfactual conditional. Our predictions rest on our claims about what is made explicit in the initial representation for different sorts of conditionals, and so we took care to present each participant with just a single inference to avoid any possible interference across the four sorts of inference. For example, a reasoner who has fleshed out his/her models to be explicit to make an MT inference may retain the explicit models when he/she carries out a subsequent inference, such as an AC inference. The reasoner may reach a different conclusion than a reasoner who is presented with the AC inference *ab initio*.

Method

Materials and Design. We constructed eight problems, half based on a factual conditional in the indicative mood and the present tense, and the other half based on a nonfactual conditional in the subjunctive mood and the present tense. Each conditional was accompanied by a second premise that corresponded to the minor premise for an MP, an MT, a DA, or an AC inference. We gave only one problem to each participant, and there were eight groups of participants. The content of the conditionals was based on people-in-places (i.e., Linda in Dublin and Cathy in Galway). The components were negated explicitly (e.g., Linda is not in Dublin).

Procedure. The participants were tested in groups, and they were randomly assigned to one of the eight conditions. The written in-

Table 1
Inferential Steps for the Four Sorts of Inferences for Factual and Counterfactual Conditionals

Factual	Counterfactual
If L then C	If L had been then C would have been
1. Models of first premise:	1. Models of first premise:
L C	factual: not-L not-C
. . .	counterfactual: L C
	. . .
Modus Tollens	
2. Model of second premise: not-C	2. Model of second premise: not-C
3. Combined models:	3. Combined models:
nil	not-L not-C
4. Conclusion: Nothing follows	4. Conclusion: not-L
5. Fleshed-out models:	
L C	
not-L not-C	
not-L C	
6. Combined models:	
not-L not-C	
7. Conclusion: not-L	
Modus Ponens	
2. Model of second premise: L	2. Model of second premise: L
3. Combined models:	3. Combined models:
L C	L C
4. Conclusion: C	4. Conclusion: C
Denial of the Antecedent	
2. Model of second premise: not-L	2. Model of second premise: not-L
3. Combined models:	3. Combined models:
nil	not-L not-C
4. Conclude: Nothing follows	4. Conclude: not-C
5. Fleshed-out models:	5. Fleshed-out models:
L C	factual: not-L not-C
not-L not-C	counterfactual: L C
not-L C	not-L C
6. Combined models:	6. Combined models:
not-L not-C	not-L not-C
not-L C	not-L C
7. Conclude: C may or may not	7. Conclude: C may or may not
Affirmation of the Consequent	
2. Model of second premise: C	2. Model of second premise: C
3. Combined models:	3. Combined models:
L C	L C
4. Conclude: L	4. Conclude: L
5. Fleshed-out models:	5. Fleshed-out models:
L C	factual: not-L not-C
not-L not-C	counterfactual: L C
not-L C	not-L C
6. Combined models:	6. Combined models:
L C	L C
not-L C	not-L C
7. Conclude: L may or may not	7. Conclude: L may or may not

structions explained that the experiment was investigating ordinary reasoning and thus the task was not an intelligence test. We asked them to read the problem carefully before writing their answer, and to take as long as they needed. The problem based on the nonfactual conditional was presented to the participants in the following way:

> Imagine you are given information about the location of different people in different places. You know that:
>
> *If Linda were in Dublin then Cathy would be in Galway*.
>
> Then you are told:
>
> *Linda is in Dublin*.
>
> What conclusion, if anything, follows from these premises?

The problem based on the factual conditional was identical except that the conditional was

> *If Linda is in Dublin then Cathy is in Galway*.

The participants were asked to write their responses on the sheet provided.

Participants. Eighty undergraduates from a variety of departments in Dublin University, Trinity College, took part in the experiment voluntarily. They had no formal training in logic, and they had not previously participated in an experiment on reasoning. The participants were randomly assigned to one of eight groups (n = 10 in each).

Table 2
Percentages of Four Sorts of Inferences for the Different Sorts of Conditionals in Each of Three Experiments

Content	Condition	*n*	MP	AC	MT	DA
		Experiment 1				
	Factual	10	100	30	40	40
	Nonfactual	10	90	50	80*	80*
		Experiment 2				
Overall	Factual	68	96	63	60	47
	Nonfactual	69	90	57	73*	53
Locational	Factual	35	100	71	49	57
	Nonfactual	34	82*	59	71*	59
Referential	Factual	33	91	55	70	36
	Nonfactual	35	97	54	74	46
		Experiment 3				
Present	Factual	35	91	31	40	20
	Nonfactual	32	91	38	56	69*
Past	Factual	31	94	32	42	39
	Counterfactual	32	81	50	66*	59*

Note—MP, modus ponens; AC, affirmation of the consequent; MT, modus tollens; DA, denial of the antecedent. *The percentage of inferences in the cell differs reliably from the percentage of inferences in the cell immediately above it.

Results and Discussion

Participants made more MT inferences from the nonfactual conditional (80%) than from the factual one (40%), and they made more DA inferences from the nonfactual conditional (80%) than from the factual one (40%); both of these differences were reliable (Meddis, 1984,[2] test, $n = 20$, $z = 1.77$, $p < .05$ for each comparison). As Table 2 shows, there were no reliable differences in the frequency with which they made MP inferences from the nonfactual conditional (90%) and the factual one (100%, Meddis test, $n = 20$, $z = 1$, $p > .10$), or in the frequency with which they made the affirmation of the consequent inferences from the nonfactual conditional (50%) and from the factual one (30%) (Meddis test, $n = 20$, $z = 0.88$, $p > .10$). These results corroborate the suggestion that reasoners make more of the two sorts of inferences supported by the more explicit representation, MT and DA, from the nonfactual than from the factual conditional.

Within each conditional, the frequency of the different sorts of inferences also differed in systematic ways. For the factual conditional, participants made more MP inferences (100%) than any of the other inferences: MT (40%), DA (40%) (Meddis test, $n = 20$, $z = 2.66$, $p < .01$ for both), and AC (30%) (Meddis test, $n = 20$, $z = 3.06$, $p < .01$). There were no other differences between the inferences. This pattern replicates a typical pattern found with neutral content: People usually find it easier to make the MP inference than the MT one, whereas the differential difficulty of the DA and AC inferences is a labile phenomenon, with some experiments showing a difference in one direction, some in the opposite direction, and some none at all (see Evans, Newstead, & Byrne, 1993, chap. 2, for a review). For the nonfactual conditional, the only difference between the inferences was that participants made more MP (90%) than AC inferences (50%) (Meddis test, $n = 20$, $z = 1.9$, $p < .05$). The more explicit representation of a nonfactual ensures that MT and DA are made more often, thus eliminating any difference between them and MP.

Perhaps people make more of all sorts of inferences from nonfactual than from factual conditionals? The frequency with which the MP inference is made is at ceiling for factual conditionals (100% of participants make it), and so no increase can be observed for it, but the AC inference was made 20% more often from the nonfactual than from the factual conditional. This difference was not statistically reliable, and it was not as large as the 40% difference for MT and DA inferences from the nonfactual and factual conditionals. However, the power of the experiment (with 10 participants in each condition) may raise doubts about whether the real pattern in the data is an increase in the frequency of all inferences from nonfactual conditionals. We suggest that this conjectured pattern is not the case; the next experiment was designed to shed further light on this point.

This first experiment shows that the inferences reasoners make differ in frequency between nonfactual and factual conditionals. These results lend some support to our proposals about the representation of information in nonfactual and factual conditionals. We suggest that people keep two situations in mind explicitly for the nonfactual conditional, whereas they keep only one situation in mind explicitly for the factual conditional. They can more readily make the valid MT inference from a nonfactual conditional; however, they also more readily make the invalid DA inference. Hence, our suggestion is not that they find it easier to be logical when reasoning nonfactually than factually, nor that the more difficult-to-grasp nonfactual conditional turns out to be an easier conditional from which to reason. Our suggestion is that they make more of those inferences that correspond to what is explicitly represented in their models. This first experiment on the deductive inferences people make from conditionals based on possibilities rather than on facts has established that reasoners make inferences at different rates from the two sorts of conditionals.

Our next experiment aimed to demonstrate the robustness of the central findings of the first experiment by replicating the effect with a greater number of participants, each of whom carried out the four sorts of inferences, and by extending the experiment from the locational content of people-in-places to a referential content of shapes and colors, and from a conclusion-production task to a conclusion-selection task.

EXPERIMENT 2
Inferences About Present Possibilities

The aim of this experiment was to replicate and generalize the finding of the first experiment that people make more MT and DA inferences from conditionals based on present possibilities than from conditionals based on present facts. In this experiment we gave the

participants factual and nonfactual conditionals based on locational relations, such as

If Linda were in Dublin then Cathy would be in Galway. (16)

and referential relations, such as

If the shape were a triangle then it would be red. (17)

Each participant carried out each of the four sorts of inferences: MP, MT, DA, and AC. Instead of giving the participants a conclusion-production task where they were asked to say what, if anything, follows from the premises, we gave them a conclusion-selection task where they were asked to choose one conclusion from a set of three: (1) Cathy is in Galway, (2) Cathy is not in Galway, (3) Cathy may or may not be Galway. We predicted, once again, that people would make more MT and DA inferences from conditionals based on possibilities than from conditionals based on facts, and that they would make the same frequency of MP and AC inferences.

Method

Materials and Design. We constructed two sets of problems. The problems in the factual set were phrased in the indicative mood and the present tense; the problems in the nonfactual set were phrased in the subjunctive mood and the present tense. Each conditional was accompanied by a second premise, which corresponded to the minor premise for an MP, an MT, a DA, or an AC inference. Each participant carried out four inferences (MP, MT, DA, AC) that were presented in a different random order. The participants in each group were given the four inferences based either on a locational content, *If Linda were in Dublin then Cathy would be in Galway*, or on a referential content, *If the shape were a triangle then it would be red*. Participants were required to select a conclusion from a set of three conclusions.

Procedure. The participants were randomly assigned to one of the two conditions, factual or nonfactual. They were given a booklet with each inference presented on a separate page. The instructions on the first page were similar to those in the previous experiment: The participants were informed that the task was a reasoning task. They were given an example (a disjunctive inference) to illustrate the conclusion-selection task, and they were asked to mark their responses on the sheet. We asked them to read each problem carefully before choosing their answer, not to go back over previous answers or change any, and to take as long as they needed.

Participants. The 137 undergraduate students from Dublin University participated in the experiment voluntarily. They had no formal training in logic, and they had not previously participated in an experiment on reasoning. They were randomly assigned to one of two groups (factual, $n = 68$; nonfactual, $n = 69$).

Results and Discussion

As Table 2 shows, the participants made more MT inferences from the nonfactual conditional (73%) than from the factual (60%), and this difference was reliable (Meddis test, $n = 137$, $z = 1.68$, $p < .05$). They made somewhat more DA inferences from the nonfactual (53%) than from the factual (47%) conditional, but this difference did not approach reliability (Meddis test, $n = 137$, $z = 0.60$, $p > .10$). As Table 2 also shows, there were no reliable differences in the frequency with which participants made the MP inferences from the nonfactual (90%) and the factual conditional (96%), nor in the frequency with which they made the AC inferences from the nonfactual (57%) and the factual (63%) conditional (Meddis test, $n = 137$, $z = 1.29$, $p > .05$, and $n = 137$, $z = 0.80$, $p < .10$, respectively).[3]

A similar pattern of results was found for both locational and referential content, as Table 2 shows. Participants made more MT inferences from the nonfactual conditional than from the factual for the locational content (71% vs. 49%; Meddis test, $n = 69$, $z = 1.85$, $p < .05$), although not for the referential content (74% vs. 70%; Meddis test $= n = 68$, $z = 0.42$, $p > .10$) because of the very high rate of making MT inferences from the factual conditional. They made somewhat more DA inferences from the nonfactual than from the factual conditional, but the difference was not reliable for either content—locational (59% vs. 57%; Meddis test, $n = 69$, $z = 0.14$, $p > .10$) or referential (46% vs. 36%; Meddis test, $n = 68$, $z = 0.78$, $p > .10$). The participants did not make more MP inferences from the nonfactual than from the factual conditional for the locational content (82% and 100%; in fact the difference was in the opposite direction; Meddis test, $n = 69$, $z = 2.58$, $p < .01$) or for the referential content (97% and 91%; Meddis test, $n = 68$, $z = 0.02$, $p > .10$). There were no differences in the frequency with which they made AC inferences from the nonfactual and the factual conditional for the locational content (59% and 71%; Meddis test, $n = 69$, $z = 1.09$, $p > .10$) or the referential content (54% and 55%; Meddis test, $n = 68$, $z = 1.08$, $p > .10$).

No systematic order effects were observed.[4] For factual conditionals, people made as many inferences when they were given them first as compared with fourth, for MP (100% and 94%; Meddis test, $n = 30$, $z = 0.82$, $p > .10$), MT (55% and 71%; Meddis test, $n = 34$, $z = 0.96$, $p > .10$), AC (79% and 58%; Meddis test, $n = 36$, $z = 1.34$, $p > .05$), and DA (35% and 53%; Meddis test, $n = 34$, $z = 1.02$, $p > .10$). Likewise, for nonfactual conditionals, people made as many inferences when they were given them first as compared with fourth, for MP (80% and 82%; Meddis test, $n = 32$, $z = 0.17$, $p > .10$), AC (50% and 55%; Meddis test, $n = 36$, $z = 0.29$, $p > .10$), and DA (48% and 63%; Meddis test, $n = 37$, $z = 0.89$, $p > .10$), with the exception of MT, which was made more often first than fourth (94% and 56%; Meddis test, $n = 33$, $z = 2.50$, $p < .01$). A similar pattern is observed whether we consider all four problems completed by each participant or just the first problem that each participant was given (i.e., if we ignore their three subsequent problems and treat problem type as a between-participants variable, as it was in the first experiment). Participants made more MT inferences from the nonfactual conditional than from the factual (94% vs. 55%; Meddis test, $n = 37$, $z = 2.63$, $p < .01$), and more DA inferences from the nonfactual than from the factual (48% vs. 35%), although again the difference did not reach significance (Meddis test, $n = 38$, $z = 0.76$, $p > .10$). Participants made somewhat fewer MP inferences from the nonfac-

tual than from the factual conditional (80% and 100%), although the reliability of the difference was marginal (Meddis test, $n = 27$, $z = 1.61$, $p < .06$), and they made fewer AC inferences from the nonfactual than from the factual conditional (50% and 79%; Meddis test, $n = 35$, $z = 1.77$, $p < .05$).

The experiment replicates the crucial finding that people make more MT inferences from conditionals based on present possibilities than on present facts. This result is reliable whether we consider all inferences that participants were given or just their first one. It is reliable for the locational content, although not for the relational content—because of the very high rate of inferences from the factual conditional. Although people make slightly more DA inferences from conditionals based on present possibilities than present facts, the effect was not reliable in any of the comparisons in this experiment, and we return to examine the frequency of DA inferences in the next experiment. The experiment replicates the finding that there is no overall difference in the frequency of MP inferences from conditionals based on present possibilities than on present facts. Participants made the same number of MP inferences from nonfactual as from factual conditionals overall; however, they made fewer MP inferences from nonfactual conditionals for the locational content but not the referential content, and they tended to make fewer MP inferences from the nonfactual conditional if we consider their first inferences only. Vagaries are also observed in the frequency of AC inferences. The experiment confirms the finding that there is no overall difference in the frequency of AC inferences from conditionals based on present possibilities than on present facts. Participants made the same number of AC inferences from nonfactual as from factual conditionals overall, and this pattern is observed for both contents. However, when we consider their first inferences only, they made fewer AC inferences from the nonfactual conditional (in contrast to the observation in the previous experiment of a tendency, albeit nonreliable, in the opposite direction—that is, *more* AC inferences from the nonfactual).

In this regard, the experiment helps to resolve the question raised by the first experiment about whether people make *more* of all sorts of inferences from nonfactual than from factual conditionals. The frequency with which the MP inference was made from factual conditionals was not wholly at ceiling (96%) and yet no increase was observed for nonfactual conditionals (90%). The AC inference was not made more often from the nonfactual (57%) than from the factual conditional (63%) either. In each case there was a slight decrease instead, and although the decreases were not reliable unless only the first responses are considered, they go against the idea of a general *increase* in the frequency of all sorts of inferences for nonfactual relative to factual conditionals. The experiment has greater power than the previous one (with 68 and 69 participants in the two conditions), yet any differences for these two inferences overall remain statistically unreliable. These weak tendencies to make *fewer* MP and AC inferences from nonfactual conditionals may result from the greater difficulty of processing the multiple-model initial representation for the nonfactual conditional, and we return to this possibility in our third experiment.

The experiment replicates three of the four findings of the first experiment, generalizing the results to referential as well as locational content, and to a conclusion-selection as well as a conclusion-production task. The findings support our suggestion that people construct a more explicit initial set of models for nonfactual conditionals, and they represent the presupposed factual situation as well as the hypothesized situation. The discrepancy lies in the failure to replicate the observed difference between the DA inferences from nonfactual and factual conditionals. The frequency with which the fallacies are made from indicative conditionals appears to be particularly labile, as reviews testify (see Evans et al., 1993, chap. 2). We will return to the frequency with which people make the fallacies in the next experiment, which aimed to extend these findings from nonfactual conditionals to counterfactual conditionals—that is, from present possibilities to past possibilities.

EXPERIMENT 3
Inferences About Present and Past Conditional Possibilities

Our account of the mental representations that people construct applies equally to nonfactual conditionals that deal with present possibilities, such as

If Linda were in Dublin then Cathy would be in Galway. (18)

and to counterfactual conditionals that deal with past possibilities, such as

If Linda had been in Dublin then Cathy would have been in Galway. (19)

We expect that people will also make more MT and DA inferences from counterfactual conditionals compared with factual conditionals and essentially the same frequency of MP and AC inferences, and our next experiment tested this set of predictions. We compared four sorts of conditionals based on past facts and past possibilities and present facts and present possibilities. The aims of the next experiment were to extend the findings of the first two experiments from present facts and possibilities to past facts and past possibilities, and to replicate the findings of the first two experiments for present facts and possibilities.

Method

Materials and Design. We constructed two sets of problems. The problems in the present tense set were similar to those in the previous experiments: They were phrased in the present tense, and half were in the indicative mood and half were in the subjunctive mood. The problems in the past tense set were phrased in the past tense, and half were in the indicative mood and half were in the subjunctive mood. Each conditional was accompanied by a second

premise that corresponded to the minor premise for an MP inference, an MT inference, a DA inference, or an AC inference. Conditional type was a between-participants variable: We gave each participant just one of the four different sorts of conditionals, and hence there were four groups of participants. Inference type was a within-participants variable: Each participant carried out one instance of each of the four sorts of inferences (MP, MT, DA, AC) in a different random order. The content of the conditionals was based on people-in-places (e.g., Alberto in Padua and Vittorio in Trieste), and each inference contained a different content concerning different people and different places, for each participant. The components were negated explicitly (e.g., *Alberto is not in Padua*). The materials were in the participants' native Italian. Participants were required to produce a conclusion in response to the question "What, if anything, follows?"

Procedure. The participants were tested in several groups and they were randomly assigned to one of the four conditions. The written instructions were based on those in the previous experiments. We asked them to read the problem carefully before writing their answer, and to take as long as they needed. They were asked to write their responses on the sheet provided.

Participants. The 141 undergraduates from Padua University participated voluntarily. They had no training in logic, and they had not previously participated in a reasoning experiment. We eliminated 11 participants prior to any data analysis because they failed to give answers to any inferences. The remaining 130 participants were randomly assigned to one of the four groups (present fact $n = 35$; present possibility $n = 32$; past fact $n = 31$; past possibility $n = 32$).

Results and Discussion

The frequency of inferences made from the conditionals that deal with past facts and possibilities followed the same pattern as the frequency of inferences made from the conditionals dealing with present facts and possibilities. As Table 2 shows, the participants made more MT inferences from the counterfactual conditional (66%) than from the factual (42%) and more DA inferences from the counterfactual conditional (59%) than from the factual (39%), and both of these differences were reliable (Meddis test, $n = 63$, $z = 1.87$, $p < .05$, and $n = 63$, $z = 1.63$, $p < .05$, respectively). As Table 2 also shows, there were no reliable differences in the frequency with which participants made the MP inferences from the counterfactual (81%) and from the factual (94%), or in the frequency with which they made AC inferences from the counterfactual (50%) and from the factual (32%) (Meddis test, $n = 63$, $z = 1.45$, $p > .05$, Meddis test, $n = 63$, $z = 1.42$, $p > .05$, respectively). These results generalize the findings of the first and second experiments with conditionals based on present facts and possibilities to conditionals based on past facts and possibilities. Once again the difference between the counterfactual and factual conditionals for the AC fallacy seems large even though it was not statistically reliable, and we will return to this issue shortly.

The participants also made more MT inferences from the nonfactual (56%) than from the factual (40%), and they made more DA inferences from the nonfactual (69%) than from the factual (20%), although only the second of these differences was reliable (Meddis test, $n = 67$, $z = 1.32$, $p > .05$, and $n = 67$, $z = 3.99$, $p < .001$, respectively). The difference between the nonfactual and factual conditionals for the MT inference, although large, was not reliable, and we will return to this issue shortly. There were no reliable differences in the frequency with which participants made MP inferences from the nonfactual and from the factual (91% in each case) or in the frequency with which they made AC inferences from the nonfactual (38%) and from the factual (31%) (Meddis test, $n = 67$, $z = 0.12$, $p > .10$, and $n = 67$, $z = 0.52$, $p > .10$, respectively).

No systematic order effects were observed (of the 16 comparisons, 14 showed no significant differences). A similar pattern is observed whether we consider all four problems completed by each participant or if we consider just the first problem that each participant was given: more MT inferences from the counterfactual than from the factual (75% vs. 25%), and more DA inferences (40% vs. 33%); and the same frequency of both MP inferences (92% and 100%) and AC inferences (17% and 22%). Likewise, more DA inferences from the nonfactual than from the factual (60% vs. 30%), although not more MT inferences (44% vs. 50%); and the same frequency of MP (100% in each case) and AC inferences (27% in each case). However, we must exercise caution because unlike the first two experiments, here the number of participants carrying out each inference when order is introduced as a variable falls below 10 (on average 8 participants, with a range from 4 to 12), and none of these eight comparisons was statistically significant.

Table 2 provides a summary of the frequency of the four sorts of inferences for the factual, nonfactual, and counterfactual conditionals in each of the experiments. As it shows, the MT inference was made more often from nonfactual and counterfactual conditionals than from factual conditionals: a reliable increase of 40% in the first experiment, a reliable increase of 13% in the second, an increase of 16% in the nonfactual comparison in the third experiment that missed reliability, and a reliable increase of 24% in the counterfactual comparison in the third experiment. Given that three of the four increases was reliable, and one was large although it missed reliability, we conclude that MT inferences are made more often from nonfactual and counterfactual conditionals than from factual conditionals. Of course, the conclusion must be tentative given the variability in the data. Likewise, the DA inference was also made more often from nonfactual and counterfactual conditionals than from factual conditionals: a reliable increase of 40% in the first experiment, an increase of 6% in the second that was not reliable, a reliable increase of 49% in the nonfactual comparison in the third experiment, and a reliable increase of 20% in the counterfactual comparison in the third experiment. Once again, given that three of the four increases was reliable, we wish to conclude that DA inferences are made more often from non-

factual and counterfactual conditionals than from factual conditionals, and once again, the conclusion must be a cautious one given the variability of the data.

The MP inference is not made at a different frequency systematically from nonfactual and counterfactual conditionals than from factual conditionals: a nonsignificant decrease of 10% in the first experiment, a nonsignificant decrease of 6% in the second, the same frequency in the nonfactual comparison in the third experiment, and a nonsignificant decrease of 13% in the counterfactual comparison in the third experiment. None of the four differences reached reliability. MP inferences do not seem to be made at a different frequency from nonfactual and counterfactual conditionals than from factual ones, but the data are somewhat variable once again. Likewise, AC inferences were not made at a different frequency systematically from nonfactual and counterfactual conditionals than from factual conditionals: a nonsignificant *increase* of 20% in the first experiment, a nonsignificant *decrease* of 6% in the second, a nonsignificant *increase* of 7% in the nonfactual comparison in the third experiment, and a nonsignificant *increase* of 18% in the counterfactual comparison in the third experiment. None of the four differences reached reliability: two were large, and two were not, and one was in the opposite direction to the others. There seems to be no clear tendency for AC inferences to be made at a different frequency from nonfactual and counterfactual conditionals than from factual ones, although again, the data show variability. The nonsignificant decreases that we observed for MP and the nonsignificant increases that we observed for the AC inferences may have resulted from the greater difficulty of processing the more explicit multiple-model representation of nonfactual and counterfactual conditionals relative to the single explicit model for factual conditionals. The richer representation of counterfactual conditionals may be a double-edged sword: It may make it easier for people to make inferences such as MT, which are difficult from factual conditionals because they require models to be fleshed out to be more explicit, but it may make it harder for them to make inferences such as MP, which are easy from factual conditionals because they require just the single explicit model.

The pattern of inferences made to counterfactual conditionals extends the findings of the first two experiments to counterfactuals. The pattern of inferences made to nonfactual conditionals provides a reasonable replication of the findings of the first two experiments. The experiment shows that when people make inferences from conditionals that deal with either present or past possibilities, they make more MT and DA inferences (and the same frequency of MP and AC inferences) relative to conditionals that deal with either present or past facts. These findings generally support our suggestion that people construct a more explicit initial set of models for nonfactual and counterfactual conditionals: They represent the presupposed facts as well as the contrary-to-fact suppositions. In our next experiment we examined the interpretation of counterfactuals and the situations in which reasoners consider them to be true and false.

EXPERIMENT 4
Verifying and Falsifying Counterfactual Conditionals

Our aim in the fourth experiment was to examine how reasoners interpret conditionals based on possibilities. The inferences people make shed light on their understanding of them, and further light is shed by the instances they judge to render them true or false (see Evans et al., 1993). We examined the situations that they spontaneously constructed to verify a counterfactual and to falsify it. We compared a factual conditional—

If there was a circle on the blackboard, *there was a triangle*. (20)

—to a counterfactual conditional:

If there had been a circle on the blackboard, there would have been a triangle. (21)

—and we asked people what two shapes would best fit the description and what two shapes would definitely go against it. This task was a generation task, and we constrained it by asking for a single answer in each case so that we could examine the instances reasoners generated as the best exemplars of verification and falsification.

In what circumstances do reasoners consider a counterfactual to be true or false? One view of human reasoning that could be considered to be implicit in philosophical investigations (e.g., Lewis, 1973; Stalnaker, 1968) is that reasoners understand a counterfactual by imagining the possible worlds in which the antecedent is true and by assessing whether the consequent is true in at least one of them. On this account, reasoners may consider the counterfactual to be true where the antecedent and consequent are true (a circle and a triangle are on the board), and false where the antecedent is true and the consequent is false (a circle is on the board and no triangle). Indeed, reasoners may assess the plausibility of a counterfactual by comparing these two situations (see, e.g., Miyamoto & Dibble, 1986).

An alternative possibility based on the model theory is that to assess what shapes would best fit the description reasoners construct an initial representation, which for the factual conditional is the following:

O Δ

...

The verifying instance they will generate is: *a circle*, *a triangle*. For the counterfactual conditional, their initial representation is more complex:

factual:	not-O	not-Δ
counterfactual:	O	Δ

...

Reasoners can base their judgments on the explicit model within the counterfactual set, *a circle*, *a triangle*, or equally on the explicit model of the facts: *not a circle*, *not a triangle*. For the factual conditional their initial representation is not fleshed out to include the model *not a circle*, *not a triangle*. We predicted that they would think of *not a circle*, *not a triangle* more often to verify the counterfactual than the factual conditional.

To assess what two shapes definitely go against the description, reasoners must flesh out their models. For the factual conditional, the fleshed-out set of models is as follows:

O	Δ
not-O	not-Δ
not-O	Δ

They can infer the instance that does not fit with any of the models: *a circle*, *not a triangle*. Fleshing out models and constructing the complement set is difficult, requiring the manipulation of multiple models, and so people are likely to make errors in this process. Reasoners may fail to flesh out their initial models:

[O]	Δ
. . .	

and they may construct a model based on the negation of the two explicitly represented elements in the initial set:

not-O	not-Δ

They will conclude: *not a circle*, *not a triangle*. Similar processes have been identified in the negation of compound conjunctions and disjunctions (Handley, 1996; Handley & Byrne, 1999).

For the counterfactual conditional, the fleshed-out set of models is as follows:

factual:	not-O	not-Δ
counterfactual:	O	Δ
	not-O	Δ

Reasoners can infer that the instance that does not fit logically is *a circle*, *not a triangle*, if they construct the complement to the set of models, and so they may generate the same instance for the counterfactual as for the factual conditional. Once again it is plausible that they may fail to flesh out their models, and they may base their answer on their initial representation:

factual:	not-O	not-Δ
counterfactual:	O	Δ
	. . .	

They may negate elements in the initial representation. If they negate the elements in the counterfactual model, they will generate the conclusion *not a circle*, *not a triangle*, just as they may for the factual conditional. If they negate the elements in the factual model, they will generate the conclusion *a circle*, *a triangle*. We predicted that reasoners would generate the instances *a circle*, *not a triangle* and *not a circle*, *not a triangle* equally often to falsify the factual and the counterfactual. However, we also predicted that they would generate more instances of *a circle*, *a triangle* to falsify the counterfactual than the factual conditional.

Method

Materials and Design. We constructed problems based on two conditionals, a factual conditional in the indicative mood and the past tense, and a counterfactual conditional in the subjunctive mood and the past tense. Each conditional was accompanied by two questions that asked what two shapes could have been drawn on the blackboard that would best fit the description, and what two shapes could have been drawn that would definitely go against it. We gave one conditional only to each participant, and each participant received each of the two questions. The content of the conditionals was based on shapes—circles and triangles—drawn on a blackboard.

Procedure. The participants were tested in groups and randomly assigned to one of the two conditions. We instructed them in general terms, with the following written instructions:

> Yesterday there were shapes drawn on a blackboard, chosen from the following selection:
>
> 
>
> You did not see the blackboard. A person who did see it says that the shapes were selected according to the following rule:
>
> *If there was a circle on the blackboard, there was a triangle.*
>
> Imagine what was on the blackboard yesterday. What two shapes could have been drawn on it that would best fit the description? Write your answer here:
>
> Now imagine what two shapes could have been drawn on the blackboard that would definitely go against the description. Write your answer here:

The instructions for the counterfactual conditional were the same except that the conditional presented was a counterfactual conditional in the past tense: "If there had been a circle on the blackboard, there would have been a triangle."

Participants. Thirty-eight undergraduates in Dublin University, Trinity College, took part in the experiment voluntarily. They had no formal training in logic and had not previously participated in a reasoning experiment. The participants were randomly assigned to one of two groups (factual, $n = 18$; counterfactual, $n = 20$).

Results and Discussion

First we compared the instances generated as the best fit for the two conditionals. The participants generated a circle and a triangle as verifying the factual conditional more often (78%) than the counterfactual (50%) (Meddis test, $n = 38$, $z = 1.74$, $p < .05$), and they generated a

Table 3
Percentages of Four Sorts of Verifying and Falsifying Situations Generated for the Two Conditionals in Experiment 4

	Shapes Generated			
Task	o, Δ	not- o, not-Δ	o, not-Δ	not-o, Δ
Verifying				
Factual	78	17	0	0
Counterfactual	50	50	0	0
Falsifying				
Factual	0	44	44	0
Counterfactual	30	30	30	0

shape that was not a circle and one that was not a triangle to verify the counterfactual (50%) more often than the factual (17%) (Meddis test, $n = 38$, $z = 1.85$, $p < .01$).[5] They generated a circle and triangle to verify the factual conditional (78%) more often than they generated a shape that was not a circle and one that was not a triangle (17%); this difference was reliable (binomial test, $n = 18$, $y = 3$, $p < .004$). In contrast, to verify the counterfactual conditional, they generated a circle and triangle, and equally often they generated a shape that was not a circle and one that was not a triangle (50% in each case), as Table 3 shows.

The results for the falsifying task are equally informative. The participants generated a circle with a shape that was not a triangle to falsify the factual conditional (44%), or else they generated two shapes that were neither a circle nor a triangle (44%). They also generated a circle with a shape that was not a triangle to falsify the counterfactual conditional (30%), or else they generated two shapes that were neither a circle nor a triangle (30%). In addition, some participants generated a circle and a triangle to falsify the counterfactual (30%). Participants generated a circle and a triangle to falsify the counterfactual (30%) reliably more often than the factual (0%) (Meddis test, $n = 38$, $z = 2.56$, $p < .01$). They produced a circle with a shape that was not a triangle, and two shapes that were neither a circle nor a triangle, equally often for both conditionals (Meddis test, $n = 38$, $z = 0.9$, $p > .10$), for both comparisons. The results are consistent with the observation that reasoners find it hard to negate a compound conjunction or disjunction, and their errors indicate they have constructed just one of the possible models by negating each of the elements in the compound expression (Handley, 1996). No participant generated a shape that was not a circle with a triangle to falsify the conditionals, which suggests they did not construct a biconditional representation of either.[6]

Perhaps some participants wrote down the shapes they thought would make the conditionals true or false, and other participants wrote down the shapes they thought would have been on the blackboard? The uniformity of the judgments of the participants to the factual conditionals suggests not: Most of them considered it to be verified by a circle and a triangle, and falsified either by a circle with no triangle or else by two shapes that were neither a circle nor a triangle. Moreover, relatively few of the participants judged the counterfactual to be verified by a circle and a triangle and falsified by a circle with no triangle (pace Lewis, 1973; Stalnaker, 1968; see also Miyamoto et al., 1989). Instead, many of them produced answers that referred to the two situations that we suggest they represented explicitly in their initial set of models for the counterfactual, the hypothesized situation (a circle and a triangle), and the factual one (no circle and no triangle). Perhaps the demands of the falsification task result in participants fleshing out their models to be more explicit than they would otherwise be? Task demands should contribute equally to the tasks for the counterfactual and the factual conditionals, and so such demands cannot explain the observed differences between them.

Do the results imply that participants construct an initial representation of a counterfactual that makes explicit only the model of the hypothesized situation (a circle and a triangle) or only the model of the factual situation (no circle and no triangle)? Such a conclusion is not warranted, we believe, and it could not explain their performance on the inference tasks: A participant who represented the counterfactual only by the hypothesized situation (a circle and a triangle) would not readily make MT and DA inferences; likewise, a participant who represented the counterfactual only by the factual situation (no circle and no triangle) would have great difficulty in making MP and AC inferences. Instead it seems that the initial representation of the counterfactual makes explicit both models.

GENERAL DISCUSSION

We suggest that reasoners construct models of conditionals based on possibilities that are similar to but more explicit than their models of conditionals based on facts. The model theory has led to the discovery of a previously unsuspected set of similarities and differences in the inferences people make about possibilities and facts. Reasoners tend to make more MT and DA inferences from a conditional based on present or past possibilities than from a conditional based on present or past facts. They tend to make the same frequency of MP and AC inferences from conditionals based on possibilities and facts, although in some cases they make fewer MP and more AC inferences from a conditional based on possibilities. There was variability in the data in each of the experiments, but the overall pattern provides tentative support for our suggestions. The first experiment demonstrated this phenomenon for conditionals based on present possibilities and facts about a locational content, for which reasoners generated their own conclusion in the relative purity of making a single inference. The second experiment replicated the effect for a locational and a referential content as well, for which reasoners selected a conclusion from a choice of conclusions for each of the four sorts of inferences. The third experiment extended the effect to conditionals dealing with past facts and past possibilities.

The novel and unique predictions about the similarities and differences in the frequency of inferences from counterfactual and nonfactual conditionals relative to factual conditionals were derived a priori from the model theory. It proposes that reasoners understand a counterfactual conditional, such as

If Linda had been in Dublin then Cathy would have been in Galway. (22)

by representing in their initial models not only the hypothesized case—Linda is in Dublin and Cathy is in Gal-

way—but also the factual—Linda is not in Dublin and Cathy is not in Galway:

factual:	not-Linda	not-Cathy
counterfactual:	Linda	Cathy
	...	

In contrast, for the factual conditional, they represent explicitly in their initial models only the hypothesized case—Linda is in Dublin and Cathy is in Galway. The two inferences that draw on the more explicit representation are thus predicted to be made more often from nonfactual and counterfactual conditionals relative to factual conditionals. The results are broadly supportive of this prediction, showing that MT and DA inferences tend to be made more often from the nonfactual and counterfactual conditionals than from the factual conditionals. The results also hint at the possibility that fewer MP and more AC inferences may sometimes be made from nonfactual and counterfactual conditionals, perhaps because the multiple models in the initial representation make these inferences more difficult to process.

The fourth experiment showed that reasoners verify and falsify a counterfactual differently from a factual conditional. To verify a factual conditional, such as

If there was a circle on the blackboard then there was a triangle. (23)

most people think of the hypothesized instance, a circle and a triangle. To verify a counterfactual conditional, such as

If there had been a circle on the blackboard then there would have been a triangle. (24)

they think of the hypothesized case, a circle and a triangle, or the factual case, not a circle and not a triangle. To falsify the factual conditional, they think of the logically prudent case, a circle and not a triangle, or else they think of the instance that negates each proposition mentioned by the assertion, not a circle and not a triangle. To falsify the counterfactual conditional, they likewise think of the logically prudent case, a circle and not a triangle, or else they think of the instance that negates each proposition mentioned by the counterfactual, not a circle and not a triangle. Uniquely, they also sometimes think of the instance that negates each proposition presupposed by the counterfactual, a circle and a triangle.

Of course it may turn out that people represent their beliefs in some sort of representation that is not akin to the kinds of models that we have proposed. However, the discovery of these novel phenomena in reasoning about factual and counterfactual conditionals was made on the basis of predictions of the model theory, derived from its core tenets: (1) The initial representation of conditionals contains some information represented explicitly and some represented implicitly because of the constraints of working memory; (2) inferences that can be based on an initial representation are made more often than inferences that require models to be fleshed out to be explicit; and (3) for counterfactual and nonfactual conditionals, the presupposed factual situation is represented explicitly as well as the temporarily supposed counterfactual situation (Johnson-Laird & Byrne, 1991). Of course, we have examined experimentally only a small and limited set of contents, and the scope of our conclusions is necessarily constrained by this limitation. Our account should generalize to other sorts of contents, too, perhaps especially to conditionals concerned with causality, which maintain strong ties with counterfactuality (e.g., Johnson-Laird & Byrne, 1991). Our account should also generalize to other sorts of thinking, perhaps especially to counterfactual thinking about what might have been (e.g., Byrne, 1996, 1997; Byrne, Culhane, & Tasso, 1995; Byrne & McEleney, 1997). The results also go some way toward providing an empirical resolution of the long-standing philosophical question about whether it is possible to have a general theory of conditionals that encompasses factual, nonfactual, and counterfactual conditionals. On this account, we can also make the further novel prediction that the initial understanding of a counterfactual is more difficult than the initial understanding of a factual conditional, because the counterfactual requires the construction of multiple models. Once this extra work is completed, however—as the results of these experiments have shown—it provides a richer basis for the subsequent tasks of deduction, verification, and falsification.

We have explored one view of how people reason about possibilities, a view similar to one examined in the philosophy of counterfactual conditionals—that a counterfactual is true if the consequent is true in the scenarios constructed by adding the false antecedent to the set of beliefs it recruits about the actual world, and making any necessary adjustments to accommodate the antecedent (e.g., Hansson, 1992; Lewis, 1973; Pollock, 1986; Stalnaker, 1968; and for a psychological adaptation, see Byrne & Tasso, 1994; Johnson-Laird, 1986; Johnson-Laird & Byrne, 1991). Ramsey (1931, p. 248) proposed that conditionals are understood by the following process:

> In general we can say with Mill that "if p then q" means that q is inferable from p, that is, of course, from p together with certain facts and laws not stated, but in some way indicated by the context ... if two people are arguing about "if p, will q?" and are both in doubt as to p, they are adding hypothetically to their stock of knowledge and arguing on that basis about q.

And an alternative view is that a counterfactual is true if the consequent follows from the antecedent taken together with any relevant premises (e.g., Chisholm, 1946; Goodman, 1973; cf. Kvart, 1986); the problem then is to specify the set of relevant premises. For example, counterfactual conditionals may be understood in the same way as factual conditionals, by accessing inference rules (Braine & O'Brien, 1991, p. 183) corresponding to the MP rule:

1. Given if p then q and p, one can infer q.

and to the rule of conditional proof:

2. To derive or evaluate *if p then* . . . , first suppose *p*, for any proposition *q* that follows from the supposition of *p* taken together with other information assumed, one may assert *if p then q*.

Constraints on the application of the conditional proof rule can be designed to ensure, for example, that suppositions are consistent with prior assumptions, and in the case of a counterfactual supposition, that the assumptions are not a record of an actual state of affairs (Braine & O'Brien, 1991). However, such accounts do not predict, nor can they readily explain, the similarities and differences in the frequencies of inferences that reasoners make from factual, nonfactual, and counterfactual conditionals, and the differences in the instances they consider to verify and falsify them. An account of conditional inference based solely on abstract inference rules (e.g., Braine & O'Brien, 1991; Rips, 1994) or domain-specific inference rules (e.g., Cheng & Holyoak, 1985; Cosmides, 1989) cannot account for the data we have reported here.

Our exploration of reasoning about possibilities has illuminated its similarities to reasoning about facts, and its differences. Counterfactual conditionals can seem to mean something very different from their corresponding factual conditionals. But these differences arise, according to the model theory, because counterfactuals are represented in a richer mental representation that captures both the conjectured possibilities and the presupposed facts. The representations and processes underlying counterfactual reasoning are nonetheless based on the same mechanisms as those underlying reasoning about facts.

REFERENCES

Adams, E. W. (1970). Subjunctive and indicative conditionals. *Foundations of Language*, **6**, 89-94.

Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: Reidel.

Au, T. K. (1983). Chinese and English counterfactuals: The Sapir–Whorf hypothesis revisited. *Cognition*, **15**, 155-187.

Ayers, M. R. (1965). Counterfactuals and subjunctive conditionals. *Mind*, pp. 347-364.

Barwise, J. (1986). Conditionals and conditional information. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. Ferguson (Eds.), *On conditionals* (pp. 21-54). Cambridge: Cambridge University Press.

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of IF: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, **98**, 182-203.

Byrne, R. M. J. (1989a). Everyday reasoning with conditional sequences. *Quarterly Journal of Experimental Psychology*, **41A**, 141-166.

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

Byrne, R. M. J. (1996). Towards a model theory of imaginary thinking. In J. Oakhill & A. Garnham (Eds.), *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird* (pp. 155-174). Hove, U.K.: Psychology Press.

Byrne, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 37, pp. 105-154). San Diego: Academic Press.

Byrne, R. M. J., Culhane, R., & Tasso, A. (1995). The temporality effect in thinking about what might have been. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 385-390). Hillsdale, NJ: Erlbaum.

Byrne, R. M. J., & Handley, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition*, **62**, 1-49.

Byrne, R. M. J., Handley, S. J., & Johnson-Laird, P. N. (1995). Reasoning with suppositions. *Quarterly Journal of Experimental Psychology*, **48A**, 915-944.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory & Language*, **28**, 564-575.

Byrne, R. M. J., & Johnson-Laird, P. N. (1992). The spontaneous use of propositional connectives. *Quarterly Journal of Experimental Psychology*, **44A**, 89-110.

Byrne, R. M. J., & McEleney, A. (1997). Cognitive processes in regret for actions and inactions. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 73-78). Hillsdale, NJ: Erlbaum.

Byrne, R. M. J., & Tasso, A. (1994). Counterfactual reasoning: Inferences from hypothetical conditionals. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 124-129). Hillsdale, NJ: Erlbaum.

Carpenter, P. A. (1973). Extracting information from counterfactual clauses. *Journal of Verbal Learning & Verbal Behavior*, **12**, 512-520.

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

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

Comrie, B. (1986). Conditionals: A typology. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. Ferguson (Eds.), *On conditionals* (pp. 77-99). Cambridge: Cambridge University Press.

Cosmides, L. (1989). The logic of social exchange. *Cognition*, **31**, 187-276.

Dudman, V. H. (1988). Indicative and subjunctive. *Analysis*, **48**, 113-122.

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

Evans, J. St. B. T., Newstead, S., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.

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

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

Girotto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition*, **63**, 1-28.

Goodman, N. (1973). *Fact, fiction and forecast* (3rd ed.). New York: Bobbs-Merrill.

Handley, S. J. (1996). *Explicit models, disjunctive alternatives, and deductive reasoning*. Unpublished doctoral dissertation, University of Wales, Cardiff.

Handley, S. J., & Byrne, R. M. J. (1999). *The negation of conjunctions*. Manuscript submitted for publication.

Hansson, S. O. (1992). In defense of the Ramsey test. *Journal of Philosophy*, pp. 522-540.

Isard, S. D. (1974). What would you have done if . . . ? *Journal of Theoretical Linguistics*, **1**, 233-255.

Jackson, F. (1991). *Conditionals*. Oxford: Oxford University Press.

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

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1986). Conditionals and mental models. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. Ferguson (Eds.), *On conditionals* (pp. 55-75). Cambridge: Cambridge University Press.

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

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

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

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1994). Why

models rather than rules give a better account of propositional reasoning: A reply to Bonatti, and to O'Brien, Braine, and Yang. *Psychological Review*, **101**, 734-739.

Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, **96**, 658-673.

Johnson-Laird, P. N., & Savary, F. (1995). How to make the impossible seem probable. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 381-384). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, **93**, 136-153.

Kvart, I. (1986). *A theory of counterfactuals*. Indianapolis: Hackett.

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

Mackie, J. L. (1973). *Truth, probability, and paradox*. Oxford: Oxford University Press, Clarendon Press.

Meddis, R. (1984). *Statistics using ranks*. Oxford: Blackwell.

Miyamoto, J. M., & Dibble, E. (1986). Counterfactual conditionals and the conjunction fallacy. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Miyamoto, J. M., Lundell, J. W., & Tu, S. (1989). Anomalous conditional judgements and Ramsey's thought experiment. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Nute, D. (1980). *Topics in conditional logic*. Dordrecht: Reidel.

O'Brien, D. P., Braine, M. D. S., & Yang, Y. (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review*, **101**, 711-724.

Pollock, J. L. (1986). *Subjunctive reasoning*. Dordrecht: Reidel.

Ramsey, S. P. (1931). *The foundations of mathematics and other logical essays*. London: Kegan Paul.

Rescher, N. (1973). *The coherence theory of truth*. Oxford: Oxford University Press, Clarendon Press.

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

Schaeken, W., Johnson-Laird, P. N., Byrne, R. M. J., & d'Ydewalle, G. (1995). A comparison of conditional and disjunctive inferences: A case study of the mental model theory of reasoning. *Psychologica Belgica*, **35**, 57-70.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98-112). Oxford: Blackwell.

NOTES

1. Our remarks apply equally to counterfactual and nonfactual conditionals, although for brevity we refer simply to counterfactuals.

2. Throughout, the statistical test we used is the 2 × 2 frequency table quick test (specific) from Meddis (1984, chap. 7), to which we will refer, for brevity, as the Meddis test. We chose it rather than, say, the Fisher exact test, because the latter assumes that the marginal totals in a 2 × 2 frequency table are constrained by the experimental situation, whereas for problems in the life sciences these totals are not necessarily fixed (Meddis, 1984, p. 109). The formula for the quick test rests on casting the data into a table of the following sort:

	MP	No MP
factual	A	B
nonfactual	C	D

and computing the following $z¢$ statistic:

$$\frac{AD - BC}{\sqrt{\frac{(A+B)(C+D)(A+C)(B+D)}{N-1}}}$$

(See Meddis, 1984, chap. 7 for further discussion of the relationships between the 2 × 2 frequency table quick test [specific], the chi-square test, and the Fisher exact test, and their relative merits.)

3. Each participant carried out a single inference of each sort, and in these analyses we treat each sort of inference separately.

4. The order in which participants were given the four inferences was random. A comparison of the frequency of the first inference with that of the fourth inference allowed us to compare making the specified inference *ab initio* with making it after each of the other three inferences had been made. A more fine-grained set of comparisons would require a design in which each of the four inferences is presented in every possible set of orders.

5. Participants' responses were either verbal or pictorial. A few responses could not be classified unambiguously (e.g., a square and a circle may be classified as not-p and not-q, or as not-p and p), and so we did not include them in any category.

6. The participants were told that the shapes were selected from the five shapes drawn. The probability of selecting the two shapes corresponding to p and q from the set of five shapes purely by chance was 2/5 or 40%. The probability of selecting two shapes that correspond to not-p and not-q from the set of five shapes purely by chance was 3/5 or 60%. These chance levels were the same for both factual and counterfactual conditionals.

(Manuscript received July 2, 1997;
revision accepted for publication August 3, 1998.)