Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 6, 1154–1170

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.6.1154

Reasoning Counterfactually: Making Inferences About Things That Didn't Happen

Valerie A. Thompson
University of Saskatchewan

Ruth M. J. Byrne
University of Dublin, Trinity College

The authors investigated the relationship between reasoners' understanding of subjunctive conditionals (e.g., if *p* had happened, then *q* would have happened) and the inferences they were prepared to endorse. Reasoners who made a counterfactual interpretation of subjunctive statements (i.e., they judged the statement to imply that *p* and *q* did not happen) endorsed different inferences than those who did not. Those who made a counterfactual interpretation were more likely to (a) judge the situation in which *p* and *q* occurred to be inconsistent with the conditional statement and (b) make negative inferences such as modus tollens (i.e., $\sim q$ therefore $\sim p$). These findings occurred with familiar and unfamiliar content, affirmative and negative conditionals, and conditional and biconditional relations.

Conditionals in the subjunctive mood, such as *if John had been wearing his seat belt, he would not have been injured*, are often interpreted as counterfactual. They seem to convey information about a presupposed factual situation in which John did not wear a seatbelt and was injured, and they hypothesize about the opposite situation in which he did wear a seatbelt and escaped injury. Counterfactual thinking is pervasive in everyday life (e.g., Byrne & McEleney, 2000; Kahneman & Tversky, 1982), perhaps because it may be linked with causal thoughts (e.g., McCloy & Byrne, 2000, 2002; Roese & Olson, 1995) and because it has consequences for the experience of a range of emotions, such as regret, and social ascriptions, such as blame (Byrne, Segura, Culhane, Tasso, & Berrocal, 2000). But very little is known about how people understand and reason from subjunctive conditionals. Although subjunctive conditionals have attracted considerable attention in philosophy and linguistics (e.g., Adams, 1970; Au, 1983; Ayers, 1965; Barwise, 1986; Dudman, 1988; Goodman, 1973; Kvart, 1986; Lewis, 1973; Pollack, 1986; Stalnaker, 1968) as well as in artificial intelligence (e.g., Ginsberg, 1986), they have received far less attention in psychology (with some notable exceptions; e.g., Carpenter, 1973; Fillenbaum, 1974; Miyamoto & Dibble, 1986). The purpose of this article is to investigate how people represent subjunctive conditionals and the implications that their representations have for the inferences they make.

Valerie A. Thompson, Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; Ruth M. J. Byrne, Department of Psychology, University of Dublin, Trinity College, Dublin, Ireland.

Some of the results of this research were presented at the International Colloquium on Cognitive Science at San Sebastien, Spain in 1997; at the International Workshop on Mental Models at Padua University, Italy in 1997; at the European Congress of Psychology at Dublin in 1997; and at the International Congress of Psychology at Stockholm, Sweden in 2000. Part of this research was carried out while Valerie A. Thompson was a visiting professor at the University of Dublin, Trinity College in 1996–1997. This research was funded by an operating grant from the Natural Sciences and Engineering Research Council of Canada. We acknowledge the support of Trinity College's Visiting Fellows and Professors Benefactions Fund.

We thank Alessandra Tasso, Cristina Quelhas, and Orlando Espino for discussions of subjunctive conditionals, and Philip Johnson-Laird for his very helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to Valerie A. Thompson, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, Saskatchewan S7N 5A5, Canada. E-mail: valerie.thompson@usask.ca

Subjunctive and Indicative Conditionals

Indicative Conditionals and Inference

Our theorizing about subjunctive conditionals is grounded in the extensive literature on indicative conditionals. Four sorts of inferences have been studied for indicative conditionals (see Evans, Newstead, & Byrne, 1993, for a review). Consider the following indicative conditional:

If Sarah went to Moose Jaw then Tom went to Medicine Hat.

For the modus ponens inference (MP), reasoners are given the true antecedent (TA), *Sarah went to Moose Jaw*, and they are asked to judge the validity of the true consequent (TC), *Tom went to Medicine Hat*. For the modus tollens inference (MT), reasoners are given the false consequent (FC), *Tom didn't go to Medicine Hat*, and are asked to judge the validity of the false antecedent (FA), *Sarah didn't go to Moose Jaw*. These two inferences are valid, regardless of whether one interprets the conditional as an *implication* relation (the antecedent is sufficient but not necessary for the consequent) or an equivalence *biconditional* relation (the antecedent is sufficient and necessary for the consequent). The findings indicate that reasoners tend to make the MP inference readily but that the MT inference is more difficult and many individuals conclude erroneously that nothing follows (see Evans et al., 1993).

Two further inferences are fallacies on the conditional implication interpretation, but they are valid on the biconditional equivalence interpretation. For the affirmation of the consequent inference (AC), reasoners are given the TC, *Tom went to Medicine Hat*, and they are asked to judge the validity of the TA, *Sarah went to*

Moose Jaw. For the denial of the antecedent inference (DA), reasoners are given the FA, *Sarah didn't go to Moose Jaw*, and they are asked to judge the validity of the FC, *Tom didn't go to Medicine Hat*. Reasoners often make these the AC and DA inferences from an indicative conditional (see Evans et al., 1993).

Mental Models and Indicative Conditionals

We propose that prior to making an inference, reasoners formulate a mental model of the premise information. Appendix A summarizes the information presented in this section. In the case of indicative conditionals, such as

If Sarah went to Moose Jaw then Tom went to Medicine Hat,

reasoners may construct a set of mental models such as the following:

Moose Jaw *Medicine Hat*
. . .

where the diagram uses *Moose Jaw* to represent *Sarah went to Moose Jaw* and *Medicine Hat* to represent *Tom went to Medicine Hat*.

Mental models are mental representations that correspond to the way the world would be if the assertion were true (Johnson-Laird, 1983), and the diagram represents separate models on separate lines (Johnson-Laird & Byrne, 1991). The models represent some information explicitly—the information mentioned in the assertion—and they represent other information implicitly—the three dots represent a model with no explicit content, corresponding to alternative situations that have not yet been mentally articulated (see, e.g., Johnson-Laird, Byrne, & Schaeken, 1992). The implicit model may be fleshed out to be explicit if necessary (see Appendix A):

Moose Jaw	*Medicine Hat*
not Moose Jaw	*not Medicine Hat*
not Moose Jaw	*Medicine Hat*

The three models in the fully fleshed out set correspond to a conditional interpretation of the assertion; on a biconditional interpretation only the first two models are consistent with the assertion (for discussion, see Evans, 1993; Johnson-Laird & Byrne, in press; Johnson-Laird, Byrne, & Schaeken, 1994; O'Brien, Braine, & Yang, 1994). The theory predicts that inferences based on the initial set of models are easier than inferences that require the models to be fleshed out. Hence, the MP inference is easier than the MT inference because MP can be made from the initial set of models but MT requires the models to be fleshed out (see Johnson-Laird & Byrne, 1991, for details).

Mental Models and Subjunctive Conditionals

The pattern of inferences described above for indicative conditionals is well established (see Evans et al., 1993, for a review). There is a much smaller literature available on the inferences reasoners make from subjunctive conditionals, but the available findings suggest that there are systematic similarities and differences in the inferences reasoners make from subjunctive and indicative conditionals (Byrne & Tasso, 1999). Recent evidence indicates that reasoners make more MT and DA inferences from a subjunctive conditional than an indicative conditional and about the same amount of MP and AC inferences (Byrne & Tasso, 1999).

The difference in inferences may reflect differences in the way people understand and represent subjunctive and indicative conditionals. It has long been known that when individuals try to remember a subjunctive conditional they have heard, they sometimes mistakenly identify paraphrases corresponding to the FA (Fillenbaum, 1974). For example, given a subjunctive conditional such as

If Sarah had gone to Moose Jaw then Tom would have gone to Medicine Hat,

reasoners may mistakenly identify *Sarah did not go to Moose Jaw* as a paraphrase. Indeed, subjunctive conditionals appear to convey information at two levels: The first concerns a factual situation in which Sarah did not go to Moose Jaw (FA) and Tom did not go to Medicine Hat (FC); in addition, reasoners hypothesize about the counterfactual state of affairs, namely what might have happened had Sarah indeed gone to Moose Jaw.

We propose that the information regarding FA and FC is incorporated into the reasoner's model of a conditional (see Appendix A), so that the initial representation of the conditional is more explicit than the initial representation of an indicative conditional (Byrne & Tasso, 1999; Johnson-Laird & Byrne, 1991):

factual:	*not Moose Jaw*	*not Medicine Hat*
counterfactual:	*Moose Jaw*	*Medicine Hat*

Moreover, the differences in representation between indicative and subjunctive conditionals should have implications for the inferences reasoners endorse. We propose that the MT inference is made more readily from the subjunctive than from the indicative because the initial set of models explicitly represents the FA–FC situation from the outset for the subjunctive (Byrne & Tasso, 1999). Thus, given the information that Sarah did not go to Moose Jaw, reasoners can infer readily on the basis of their initial models that Tom did not go to Medicine Hat. Likewise, the DA inference is made more readily from the subjunctive than from the indicative because of the explicit representation of the FA–FC situation. On this account, the MP and AC inferences are made at the same rate from subjunctive and indicative conditionals because these affirmative inferences require access to the TA–TC situation, which is explicitly represented in the initial models for both sorts of conditionals (Byrne & Tasso, 1999).

Implications, Consistent Situations, and Inferences

Our purpose in this article is to examine the nature of the mental representations that reasoners construct of subjunctive and indicative conditionals and to test the relationship between mental representations and reasoning. Our approach was to obtain converging evidence from a set of interrelated tasks. In two experiments, we gave participants three sorts of task: an implications task, a consistency-judgment task, and an inference task (i.e., the conditional-arguments task). The implications task allowed an independent assessment of reasoners' interpretations and representations of subjunctive and factual conditionals. The consistency-judgments task and the conditional-arguments task were then used to test our predictions of the consequences of factual and counterfactual representations.

Implications

The implications task allowed us a direct test of what subjunctive conditionals imply to reasoners. For this task, reasoners were provided with either an indicative conditional or a subjunctive conditional of the following sort:

John says that: *If Sarah had gone to Moose Jaw then Tom would have gone to Medicine Hat.*

and we asked them "What, if anything, do you think that John meant to imply?" We provided them with a selection of options, and they could circle as many of the options as they considered appropriate. These options corresponded to the TA, FA, TC, and FC cases:

(a) That Sarah went to Moose Jaw (TA),
(b) That Sarah did not go to Moose Jaw (FA),
(c) That Tom went to Medicine Hat (TC),
(d) That Tom did not go to Medicine Hat (FC), and
(e) Nothing was implied.

The logically accurate response option for both an indicative and a subjunctive conditional is the last: "nothing was implied." The speaker of a conditional hypothesizes a relation between the two propositions in the conditional but does not make a specific commitment to the truth or falsity of either proposition. However, the way in which people understand and reason from conditionals may lead them to believe that speakers do intend to imply something about the truth and falsity of the propositions. In particular, reasoners who have made a counterfactual interpretation of the subjunctive conditional should have represented not only the two affirmative propositions (the TA–TC situation) in their initial set of models but also the two negative propositions (the FA–FC situation). Consequently, we expected that they would judge that the subjunctive implies the FA situation, that Sarah did not go to Moose Jaw, and/or the FC situation, that Tom did not go to Medicine Hat, more often than reasoners who are given the indicative conditional.

Consistent Situations

A second aim was to test the situations with which reasoners judge a subjunctive to be consistent and inconsistent. For this purpose, we gave participants a consistency-judgment task (Byrne, Handley, & Johnson-Laird, 1995). As illustrated below, participants were asked to judge which of four possible event combinations were consistent or inconsistent with a conditional relation.

Sarah went to Moose Jaw and Tom went to Medicine Hat. (TA–TC)

Sarah did not go to Moose Jaw and Tom did not go to Medicine Hat. (FA–FC)

Sarah went to Moose Jaw and Tom did not go to Medicine Hat. (TA–FC)

Sarah did not go to Moose Jaw and Tom went to Medicine Hat. (FA–TC)

This is similar to a truth-table task (see Evans et al., 1993, for a review) but differs in several important respects. Some variants on the truth-table task ask participants to choose combinations of events that are true or false with respect to the conditional rule (see, e.g., Oaksford & Stenning, 1992). We chose to avoid asking for true–false judgments, because we propose that reasoners interpret counterfactuals by keeping in mind what is true as well as what is false but supposed hypothetically to be true. In this case, asking participants to judge what is true and false may confuse such distinctions for some of them, in ways that asking what situations are consistent and inconsistent would not. Another variant on the truth-table task asks reasoners to identify which of the four event combinations would be inconsistent with the conditional (e.g., Barrouillet & Lecas, 1998; Thompson, 1994, 2000) or to judge each of the four cases as true, false, or irrelevant with respect to the rule (e.g., Evans, 1983; Evans & Newstead, 1977). In these latter cases, a problem arises in interpreting the not selected, or irrelevant, combinations, because items may be deemed irrelevant for several reasons, including heuristic, attentional strategies, such as matching bias (e.g., Evans, 1998); alternatively, this option may be used as a substitute for "don't know." Because our goal was to build a picture of reasoners' models, rather than to measure heuristic and attentional processes, we asked for consistent–inconsistent judgments for each combination and did not allow a potentially ambiguous *irrelevant* option.

Findings with the truth-table task indicate that reasoners typically identify the TA–TC situation (in which Sarah goes to Moose Jaw and Tom goes to Medicine Hat) as consistent and the TA–FC situation (in which Sarah goes to Moose Jaw but Tom does not go to Medicine Hat) as inconsistent with indicative conditionals (see Evans et al., 1993). In contrast, we expected that reasoners who were given a subjunctive conditional would be more likely to consider the TA–TC situation to contradict the conditional. That is, if reasoners assume that a subjunctive conditional implies the falsity of its antecedent and consequent (FA and FC), then they should also consider that the occurrence of the antecedent and consequent (i.e., TA–TC) is inconsistent with it. For example, we believed that reasoners who interpret John's statement to imply that Sarah did not go to Moose Jaw and that Tom did not go to Medicine Hat should perceive a contradiction in the situation wherein Sarah goes to Moose Jaw and Tom goes to Medicine Hat. Conversely, we expected that reasoners who were given an indicative conditional should consider the FA–FC situation to contradict it more than would reasoners given a subjunctive conditional. Finally, we believed that reasoners given an indicative conditional will have represented the TA–TC situation in their initial set of models, and if they fail to flesh out their models to be more explicit, then they should judge the FA–FC situation to be inconsistent with their initial models.

Inferences

Our third aim was to examine the inferences that people make from subjunctive and indicative conditionals. We gave reasoners a conditional-arguments task, in which they were presented with the following sort of information:

John says that: *If Sarah had gone to Moose Jaw then Tom would have gone to Medicine Hat.*

Mary replies: *I know that Sarah went to Moose Jaw.*

We asked them "What, if anything, can they infer about Tom's whereabouts?" We provided them with three options and asked them to circle their answer:

(a) Tom went to Medicine Hat.
(b) Tom may or may not have gone to Medicine Hat.
(c) Tom did not go to Medicine Hat.

We expected to replicate recent evidence that reasoners are more likely to make the MT and DA inferences from the subjunctive than from the indicative conditional and are equally likely to make the MP and AC inferences (Byrne & Tasso, 1999). The reported inferential similarities and differences have been somewhat labile (e.g., previous studies have found increased rates of MT or DA but not always both; Byrne & Tasso, 1999). Our aim was to examine the similarities and differences in inferences between subjunctive and indicative conditionals over a range of different sorts of conditionals, including conditionals with affirmative and negative components, conditionals based on people-in-places content (as in the above example), causal content, definitional content, and conditionals with necessary antecedents (biconditionals) and nonnecessary antecedents (conditionals).

Individual Differences in Counterfactual Reasoning

One possibility is that the vagaries in the frequencies of inferences observed in previous studies arise because of individual differences. Perhaps surprisingly, there is very little research in the cognitive science of reasoning on individual differences (with some notable exceptions; e.g., Stanovich, 1999; Stanovich & West, 2000). Yet there are clear indications in the existing literature that individuals may differ in various ways (Torrens, Thompson, & Cramer, 1999), including in their interpretation of conditionals as biconditionals or conditionals proper (e.g., Markovits, 1984) and in their development and use of different strategies for complex deductions (e.g., Byrne & Handley, 1997).

Thus, it is possible that reasoners differ in the interpretations they make of subjunctive conditionals. Evidence to support this hypothesis comes from Byrne and Tasso (1999). In the Byrne and Tasso study, reasoners were asked to generate the two shapes that would best fit the subjunctive conditional (e.g., *if there had been a circle on the blackboard then there would have been a triangle*) and to generate the two shapes that would definitely go against it (Byrne & Tasso, 1999). Some reasoners generated as the best fit the TA–TC situation in which there is a circle and there is a triangle, and they generated as the situation that would definitely go against the subjunctive the FA–FC situation in which there is no circle and no triangle (or else the logically prudent TA–FC situation in which there is a circle but no triangle). Other reasoners did the opposite. They generated as the best fit the FA–FC situation, and as the situation that would definitely go against the subjunctive they generated the TA–TC situation (or else the logically prudent TA–FC situation).

These data hint at the existence of individual differences in the interpretational focus of subjunctive conditionals. That is, some reasoners may interpret subjunctive conditionals to be about the factual reality they presuppose (the FA–FC situation), whereas others may interpret them to be about the counterfactual possibility they suggest (the TA–TC situation). Our combination of tasks allowed us to examine both individual differences in how conditionals are interpreted and the relationship between their interpretations and the inferences they are prepared to endorse. For example, our approach permitted us to categorize reasoners into different groups on the basis of their judgments about what a subjunctive conditional implies and to test whether these different groups make different sorts of inferences.

Experiment 1

In our first experiment, we used materials that expressed unfamiliar relations (e.g., *If Sarah had gone to Moose Jaw then Tom would have gone to Medicine Hat*). By controlling for the familiarity of the stimuli in this manner, we were able to examine the interpretation given to subjunctive conditionals per se, as opposed to those induced by knowledge of the content. We were also interested in negative subjunctive conditionals (e.g., *If Sarah had not gone to Moose Jaw then Tom would not have gone to Medicine Hat*). Negative subjunctive conditionals have never been studied before. Subjunctive conditionals in everyday life are often negative (e.g., *if I hadn't switched investment companies I wouldn't have lost money*; Kahneman & Tversky, 1982), perhaps because counterfactual scenarios are frequently generated to mentally undo factual situations (Roese & Olson, 1995).

Method

Participants. The participants were 176 introductory psychology students from the University of Saskatchewan, Saskatoon, Saskatchewan, Canada, who took part in partial fulfillment of a course requirement. There were 106 women and 70 men; the mean age was 21 years (SD = 6.3, range = 16–50 years).

Materials. The conditional statements were all written in the past tense, and were presented (a) in one of two moods, either the indicative (e.g., *If Mike went to Calgary, then Barbara went to Edmonton*) or subjunctive (e.g., *If Mike had gone to Calgary, then Barbara would have gone to Edmonton*); (b) in one of two valences, either affirmative (i.e., as in the preceding examples) or negative (e.g., *If Mike had not gone to Calgary, then Barbara would not have gone to Edmonton*); and (c) with or without a context. A complete list of materials appears in Appendix B. As illustrated below, the goal of the context was to provide a meaningful rationale for the conditional statement. The no-context condition was identical, except that the following introductory paragraph was omitted:

> John and Mary manage a sales company that sends representatives to cities across Western Canada. Normally, they have a computer file that keeps track of where their representatives are, but today the computer is down. At the moment, they are trying to reconstruct their file and determine where their representatives are.

John says that: *If Mike had gone to Calgary, then Barbara would have gone to Edmonton.*

Tasks. Each participant completed three tasks: the conditional-arguments task, a consistency-judgment task, and an implications task, in that order (a complete example of each task is presented in Appendix B). For the conditional-arguments task, participants evaluated four logical arguments (i.e., MP, DA, AC, and MT). For each argument, participants were presented with a different conditional statement, each of which described the potential destinations of two of the sales representatives. The conditional statement was followed by a minor premise (TA, FA, TC, and FC for the MP, DA, AC, and MT forms, respectively). The order in which the minor premises were presented was randomized for every participant. The conditional statement and minor premise were presented as part of a conversation between John and Mary; following the minor premise, participants were asked to indicate what could be concluded on the basis of the information provided. They were given three options to choose from, always in the same order, as illustrated below.

Mary replies that: *I know that Barbara went to Edmonton.*

We asked them "What, if anything, can they infer about Mike's whereabouts?"

(a) Mike went to Calgary.

(b) Mike may or may not have gone to Calgary.

(c) Mike did not go to Calgary.

For the consistency-judgment task, participants were presented with a different conditional statement and were asked to determine which combination of events would be consistent with the conditional and which would be inconsistent. This was followed by four sentences, which corresponded to the TA–TC, TA–FC, FA–TC, and FA–TC combinations. The order in which these sentences were presented was randomized for every participant. Following each sentence were two options, "'consistent" and "inconsistent," and participants were instructed to circle the appropriate option for each combination of events.

For the implications task, participants were provided with another conditional sentence and were asked what, if anything, they thought that John meant to imply by his statement. They were presented with five options: TA, FA, TC, FC, and "Nothing was implied" and told to choose as many as seemed appropriate. The order in which the options appeared was randomized for every participant, except that the *nothing-implied* option was always presented last.

Procedure. The tasks were presented in the same order for every participant. The conditional-arguments task was presented first; all four questions (i.e., MP, AC, DA, and MT) were presented on a single page. This task was preceded by a paragraph informing participants that the purpose of the experiment was to examine conditional reasoning, and that they were to answer the questions that followed on the basis of the inferences that followed from the conditional statements. They were told to proceed carefully, to take as much time as they needed, and to ask if they had any questions. The consistency-judgment and implications tasks were presented on a separate page. The instructions for these tasks asked participants to answer the questions on the basis of how they interpreted the conditional statements by circling the appropriate answer for each question.

The assignment of conditional statements to tasks was as follows: For the context condition, we used the six conditionals describing Canadian places (see Appendix B). Each participant was presented with all six conditionals in the positive or negative valence; a different conditional statement was used for each of the four inferences on the conditional-arguments task (i.e., MP, MT, DA, and AC), another for the consistency-judgment task, and the final one was used for the implications task. Conditional statements were cycled through the tasks so that each statement was assigned equally often to each of the tasks and equally often to the four inferences on the argument task.

For the no-context condition, each participant completed each of the three tasks twice, once in the positive valence condition and once in the negative valence condition. To avoid repeating the content of the conditionals, we devised a second set using American places (see Appendix B); the American and Canadian contents were assigned equally often to the positive and negative valence conditions. Otherwise, the assignment of conditionals to tasks was as described above.

Design. Initially, this experiment consisted only of the context condition; valence (i.e., negative vs. affirmative) and mood (i.e., subjunctive vs. indicative) were tested as between-participants variables. However, we observed that a substantial proportion of reasoners did not give a counterfactual interpretation of the subjunctive conditionals. Accordingly, we reran the experiment omitting the context and with one minor difference, namely that valence was tested as a within-participants variable. The pattern of inferences was the same for both the context and no-context conditions,[1] therefore rather than present two essentially redundant experiments, we decided to combine them into a single experiment. Also, because one of our goals was to compare positive and negative statements, we conducted an extensive analysis of the valence variable; however, we observed very few differences between the positive and negative statements. As a result, we decided to average over valence, to simplify the reporting of the analyses.

Results and Discussion

Our predictions stemmed from the hypothesis that the initial representation of a subjunctive conditional is more complete than the initial representation of an indicative one. In particular, we hypothesized that the initial representation of an indicative conditional contains an explicit representation of only the TA–TC case, whereas the initial representation of a subjunctive conditional has been fleshed out to include an explicit representation of the FA–FC case as well, as illustrated in Appendix A. Moreover, these differences should have systematic implications for how subjunctive and indicative conditionals are treated on the three tasks. To test these hypotheses, we first analyzed the implications task, and on the basis of these findings, we classified reasoners as having made a factual or a counterfactual representation. We then examined the implications of these representations for performance on the consistency-judgment and conditional-arguments tasks.

Testing for counterfactual interpretations: Analysis of the implications data. The numbers in Table 1 indicate the proportion of reasoners in each condition who indicated that a given proposition was implied by the conditional. Two analyses of the data in Table 1 were conducted, one of the *implied* responses (i.e., the proportion of reasoners indicating that each of the four propositions, TA, TC, FA, and FC was implied) and another of the *nothing-implied* responses.

Note that for these, and for all subsequent analyses, when the assumption of sphericity was not met, the degrees of freedom were reduced using the Greenhouse–Geiser correction; thus, the degrees of freedom reported for the analyses below may be less than expected on the basis of the number of participants and may be reported as fractions. Only effects significant with less than .05 are reported.

The implications data were analyzed to confirm that the subjunctive conditionals were more likely to receive a counterfactual interpretation than the indicative conditionals. If this were the case, one would expect that subjunctive conditionals, but not indicative conditionals, should entail the falsity of the antecedent or consequent (i.e., FA or FC). Consistent with this hypothesis, a 4 (proposition) $\times$ 2 (mood) mixed analysis of variance (ANOVA) revealed a Mood $\times$ Proposition interaction, suggesting that different propositions were considered to be implied by subjunctive and indicative conditionals, $F(1.7, 296) = 51.97$, $MSE = 0.13$, $p < .01$; the main effects of both variables were likewise significant, $F(1, 174) = 16.23$, $MSE = 0.19$, $p < .01$, and $F(1.7, 296) = 3.96$, $MSE = 0.13$, $p = .026$, respectively. To verify that the pattern of the interaction corresponded to our predictions, we decomposed the interaction using t tests to compare the effects of mood (i.e., subjunctive vs. indicative) for each of the four propositions.

As expected, subjunctive conditionals were more likely than indicative conditionals to be interpreted counterfactually. Whereas approximately 50% of reasoners indicated that either FA or FC was implied by subjunctive statements, less than 5% indicated that

[1] The analyses for the conditional-arguments, consistency-judgment, and implications tasks were all computed using context as a factor. On the conditional-arguments task, reasoners accepted more inferences of all types (i.e., MP, MT, AC, DA) in the no-context than in the context condition, $F(1, 155) = 5.58$, $MSE = 0.43$, $p = .019$; otherwise, none of the interactions or main effects involving the context variable were significant ($p > .095$).

Table 1
Proportion of Propositions Implied in Experiment 1

	Mood	
Proposition	Indicative	Subjunctive
TA	.24	.11
TC	.44	.16
FA	.02	.48
FC	.01	.47
None	.54	.30

Note. $n = 85$ and 91 for indicative and subjunctive moods, respectively. TA = true antecedent; TC = true consequent; FA = false antecedent; FC = false consequent.

these were implied by indicative statements, $t(174) = 8.60$ and 8.66 for FA and FC, respectively ($ps < .001$). These data are consistent with the hypothesis that reasoners are more likely to represent the FA–FC case for subjunctive than indicative conditionals.

Somewhat surprisingly, reasoners given indicative conditionals also appeared to believe that these conditionals implied something about the truth or falsity of one of their propositions. Reasoners were more likely to perceive that indicative statements implied the truth of the antecedent or consequent than subjunctive statements, $t(174) = 2.28$ and 4.29 for TA and TC, respectively ($ps < .024$). In all, about 25% perceived that an indicative statement implied TA, and about 45% indicated that an indicative statement implied TC. One explanation for these findings is that reasoners represent indicative conditionals in terms of the TA–TC contingency without acknowledging the possibility that other contingencies may be possible (Barrouillet, Grosset, & Lecas, 2000; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). Thus, a conditional statement may be seen to imply one or both of its component propositions.[2]

Finally, the analysis of the *nothing-implied* responses indicated that, as expected, reasoners were more likely to indicate that there was nothing implied by indicative than by subjunctive statements $t(174) = 3.31$, $SE = .07$, $p = .001$.

Classifying reasoners' interpretations. Broadly speaking, these analyses are consistent with the experimental predictions, in that subjunctive statements were more likely to promote counterfactual interpretations than are indicative statements. However, the rates of these counterfactual interpretations were relatively low: Just over half of the reasoners indicated that either FA or FC was implied by the subjunctive statements. Thus, a large minority of the reasoners may not have made a counterfactual interpretation of the subjunctive statements. Consequently, rather than analyzing responses to the conditional-arguments and consistency-judgment tasks as a function of the mood of the conditional (indicative or subjunctive), we categorized participants on the basis of their responses to the implications task, and we used that categorization as the independent variable in subsequent analyses.

Participants were deemed to make a counterfactual interpretation if they indicated that the conditional implied the falsity of either antecedent or consequent (i.e., that FA or FC were implied and no other case). For the sake of simplicity, and to avoid creating further (and smaller) categories of interpretations, we assumed that the initial model formed by this group of reasoners represented both FA and FC (indeed, 77% of reasoners in the counterfactual group indicated that both FA and FC were implied by the conditional). All other participants, except those who gave an inconsistent response, were deemed to have made a factual interpretation of the premises.[3]

Participants who made inconsistent responses were eliminated from subsequent analyses. Of the 176 participants, 17 were eliminated on these grounds, leaving 159 participants. A participant was classified as giving an inconsistent response for two reasons. The first was that their response could be consistent with both a factual and counterfactual interpretation (e.g., when both TA and FA were seen to be implied). The second applied to the no-context condition only; in this condition, participants evaluated both a positive and a negative statement. As the data were collapsed over valence, only participants who gave the same response for positive and negative items could be retained in the analysis; 9 participants were eliminated for giving different interpretations to the positive and negative statements. The numbers of participants making factual and counterfactual interpretations are summarized in Table 2.

Consistencies and inconsistencies: Analysis of the consistency-judgment task. Table 3 presents the percentage of people indicating that each of four combination of events (i.e., TA–FC, FA–TC, TA–TC, and FA–FC) were inconsistent with the conditional. The data are categorized in terms of the interpretation (factual vs. counterfactual) that a reasoner made on the implications task. These data were analyzed using a 4 (event combination) $\times$ 2 (interpretation) mixed ANOVA with event combination as the within-subjects variable and interpretation as the between-subjects variable.

The analysis revealed a significant effect of event combination, $F(2.7, 427) = 253$, $MSE = 0.10$, $p < .01$, and more important, an Event Combination $\times$ Interpretation interaction, indicating the events judged to be inconsistent with a conditional varied as a function of the interpretation that was made, $F(2.7, 427) = 3.79$, $MSE = 0.10$, $p = .013$. To determine the exact form of the interaction, and to test the specific predictions we made regarding the effect of counterfactual interpretations and perceived inconsistencies, we computed planned comparisons for each event combination. Three predictions were tested.

1. It was predicted that reasoners who gave a counterfactual interpretation would consider the TA–TC combination to contradict the conditional. That is, if people assume that the conditional implies the falsity of the antecedent and consequent (FA and FC), they should consider the occurrence of the antecedent and consequent (i.e., TA–TC) to be inconsistent with the conditional. Consistent with this prediction, the TA–TC combination was perceived to contradict a counterfactual interpretation more often than a factual interpretation, $t(157) = 3.43$, $p < .01$.

2. Reasoners who gave a factual interpretation should be more likely to consider the FA–FC combination to contradict the conditional. Specifically, these reasoners were predicted to start with

[2] Thanks to Phil Johnson-Laird for pointing out this possibility.

[3] To simplify the analysis, we will not further subdivide the factual group. In fact, we observed very few differences (a) between reasoners who indicated that either TA or TC was implied or (b) between reasoners who made factual interpretations of indicative and subjunctive statements. This suggests that it is legitimate to combine these groups.

Table 2
Proportion of Reasoners in Experiment 1 Giving Counterfactual and Factual Interpretations

	Mood		
Interpretation	Indicative	Subjunctive	*n*
Counterfactual	0	.52	43
Factual	1	.48	116
n	77	82	159

a set of models in which the TA–TC situation was explicitly represented; if this set of models was not completely fleshed out, then the FA–FC combination should be perceived as inconsistent with the initial set of models. This prediction was not supported. The FA–FC combination was perceived to contradict factual and counterfactual interpretations equally, $t(157) = 1.29$, $p > .10$. The interpretation of this finding is qualified by a lack of power; there was 80% power to detect only a relatively large difference of .27 between the factual and counterfactual interpretations.

3. The degree to which the TA–FC and FA–TC combinations were perceived to contradict the conditional should not vary as a function of the interpretation made; this prediction was also borne out, $t(157) = 1.20$ and 0.72, respectively, ps $> .10$; this analysis had 80% power to detect differences of .17 and .23 between factual and counterfactual interpretations for the TA–FC and FA–TC combinations, respectively.

Analysis of the conditional-arguments task. Table 4 summarizes the proportion of people making each of the four inferences; an inference was deemed to have been made if a reasoner gave a determinant response (i.e., by choosing the *a* option when the minor premise was phrased in the affirmative or the *c* option when the minor premise was phrased in the negative). For example, when the valence of the conditional was positive (e.g., *if Mike went to Calgary, then Barbara went to Edmonton*), the minor premise for the MP inference was in the affirmative (e.g., *Mike went to Calgary*); in this instance, a reasoner was scored as making the MP inference when he or she chose the *a* option (i.e., *Barbara went to Edmonton*).

The data were analyzed using a 4 (inference) × 2 (interpretation) mixed ANOVA. There was an overall effect of inference type, $F(2.6, 406) = 12.25$, $MSE = 0.12$, $p < .01$, and interpretation, $F(1, 157) = 4.16$, $MSE = 0.44$, $p = .043$, and more important, an Inference Type × Interpretation interaction, $F(2.6, 406) = 9.28$, $MSE = 0.12$, $p < .01$; this finding supports the prediction that the inferences drawn from a conditional relation vary as a function of the interpretation that is made.

Table 3
Proportion of Scenarios in Experiment 1 Considered to Be Inconsistent With the Conditional Rule

	Interpretation	
Event combination	Factual	Counterfactual
TA–FC	.93	.98
FA–TC	.87	.91
TA–TC	.06	.23
FA–FC	.22	.13

Note. TA = true antecedent; TC = true consequent; FA = false antecedent; FC = false consequent.

Table 4
Proportion of Inferences Accepted in Experiment 1

	Interpretation	
Inference	Factual	Counterfactual
MP (TA ∴ TC)	.82	.77
AC (TC ∴ TA)	.61	.58
DA (FA ∴ FC)	.39	.67
MT (FC ∴ FA)	.46	.74

Note. TA = true antecedent; TC = true consequent; FA = false antecedent; FC = false consequent; MP = modus ponens; AC = affirmation of the consequent; DA = denial of the antecedent; MT = modus tollens.

To decompose the two-way interaction between interpretation and inference, and to test our predictions regarding the specific inferences that should be endorsed for each interpretation, we computed *t* tests comparing factual and counterfactual interpretations for each inference.

We reasoned that because the FA and FC situations were represented as part of the initial models of a counterfactual, but not a factual, interpretation, reasoners who made a counterfactual interpretation should accept inferences leading to the conclusion FA or FC (i.e., the DA and MT inferences). In contrast, reasoners who made a factual interpretation should be less likely to make either of these inferences; for these reasoners, the FA–FC situation was not predicted to be spontaneously represented in their initial models and should be available only during the fleshing out process. Consistent with this prediction, the effect of interpretation was robust for the DA and MT inferences, $t(157) = 3.38$ and 3.43, respectively, ps $< .001$, indicating that both inferences were made more often for counterfactual than factual interpretations.

In contrast, we reasoned that both types of interpretations should allow reasoners to conclude TC when given TA, and TA when given TC (corresponding to the MP and AC inferences), because the TA–TC situation was hypothesized to form part of the representation of both counterfactual and factual interpretations. Indeed, there were no differences between the factual and counterfactual interpretations for either the MP or the AC inferences ($t < 1$).

The relative predictive validity of mood and interpretation. We set out to investigate the interpretation of subjunctive conditionals; when we observed that a substantial proportion of reasoners did not make counterfactual interpretations of subjunctives, we decided to use reasoner's interpretations, rather than the mood of the conditional, to predict reasoning performance on the remaining tasks. What, then, can we say about subjunctive conditionals? To answer this question, we reanalyzed all of the responses in which we predicted differences between factual and counterfactual representations, namely the MT and DA inference from the conditional-arguments task and the TA–TC and FA–FC combinations from the consistency-judgment task. We used a multiple regression approach, entering two predictor variables, mood (subjunctive vs. indicative) and interpretation (factual vs. counterfactual), in steps.

For the TA–TC combination, as well as the MT and DA inferences, we obtained the following pattern of findings: Interpretation, when entered first, predicted a significant, and equivalent,

amount of variance in all three of the dependent measures ($R = .26$, $p = .001$). Mood predicted variance only in the MT inference ($R = .23$, $p = .004$), but not for either DA or TA–TC ($R \leq .15$, $p \geq .067$). Moreover, when interpretation was entered as the second variable in the equation, it predicted a significant portion of the variance over and above that accounted for by mood ($\Delta R \geq .16$, $p \leq .036$); on the other hand, mood did not predict any residual variance when entered second ($\Delta R \leq .01$, $p \geq .26$). For the FA–FC combination, neither mood nor interpretation predicted variance in responses on either the first or second step ($p > .1$). Overall, interpretations were a better predictor of inference patterns than was the mood of the conditional.

Conclusions. Our findings indicate that subjunctive conditionals are imperfect predictors of counterfactual interpretations. A slim majority of reasoners given a subjunctive conditional such as *If Sarah had gone to Moose Jaw then Tom would have gone to Medicine Hat* made a counterfactual interpretation in which it was assumed that Sarah did not go to Moose Jaw (FA) or Tom did not go to Medicine Hat (FC); a substantial minority made a factual interpretation instead. These reasoners interpreted the statement to imply that Sarah went to Moose Jaw (TA) or Tom went to Medicine Hat (TC) or as implying nothing.

In contrast, reasoners given an indicative conditional such as *If Sarah went to Moose Jaw then Tom went to Medicine Hat*, almost never interpreted it to imply either FA or FC. Instead, reasoners typically interpreted indicative conditionals to imply that TA or TC was the case (i.e., that Sarah went to Moose Jaw or Tom went to Medicine Hat) or as implying nothing. The finding that some reasoners perceive indicative conditionals to imply the truth of their component propositions was surprising; strictly speaking, conditionals express hypothetical relationships and do not presuppose that one or the other of the components is actually true. We hypothesized that some reasoners represent indicative conditionals only in terms of the TA–TC contingency, without acknowledging that other contingencies may be possible (Johnson-Laird et al., 1999) and thus interpret an indicative conditional as entailing the truth of TA, TC, or both.

Finally, we observed that reasoners who made a counterfactual interpretation drew different inferences than reasoners who made a factual interpretation. Consistent with the hypothesis that reasoners who make a counterfactual interpretation represent the FA–FC case as part of their initial model, we observed that reasoners (a) drew the inferences involving the FA and FC propositions (i.e., MT and DA) on the conditional-arguments task relatively frequently, and (b) were likely to perceive that the TA–TC case contradicted the conditional rule on the consistency-judgment task. We extended these findings to negative conditionals and found few differences between affirmative and negative conditionals in this regard.

These findings support our hypothesis that the FA–FC case is represented as part of a counterfactual interpretation. We also found evidence to support our hypothesis that the TA–TC case is also represented as part of a counterfactual interpretation; reasoners making a counterfactual interpretation accepted the MP and AC inferences as often as reasoners who made a factual interpretation. That is, reasoners made inferences about the TA and TC propositions equally often given a factual and counterfactual interpretation, suggesting that these propositions form part of both representations.

Moreover, we observed that the interpretation reasoners made (as inferred from the implications task) was a better predictor of their inferences than was the mood of the conditional, perhaps because of the fact that a sizable portion of reasoners given subjunctive conditionals did not make a counterfactual interpretation. In the analysis of Experiment 2 that follows, we explore further why only a small majority of our participants came to a counterfactual interpretation of the subjunctive conditional. We also examine the relation between counterfactual interpretations and necessary or biconditional interpretations of conditionals.

Experiment 2

Our aim in this experiment was to examine whether counterfactual interpretations are similar to biconditional interpretations of conditionals. A biconditional's antecedent is necessary and sufficient for its consequent to occur, whereas a conditional's antecedent is sufficient but not necessary for its consequent. The perceived necessity of the antecedent affects the inferences reasoners make; for example, reasoners make MP and MT from both sorts of conditionals, but they make DA and AC more often from conditionals with necessary antecedents (e.g., Byrne, 1989; Byrne, Espino, & Santamaria, 1999; Thompson, 1994, 1995, 2000). Our aim was to examine the relation between counterfactual interpretations and necessary or biconditional interpretations, in particular to establish whether reasoners tend to interpret subjunctive conditionals as biconditionals (i.e., as having necessary antecedents).

Biconditional and Counterfactual Interpretations

On the surface, the initial model of a biconditional interpretation may appear to resemble a counterfactual one. Consider the following conditional, which reasoners interpret as containing an antecedent that is both sufficient and necessary (Thompson, 1994):

If the butter was heated, then it melted.

Reasoners seem to represent conditionals with necessary antecedents using only two models:

heated	melted,
not heated	not melted.

This representation conveys the necessity of the antecedent: *Melting* cannot occur without *heating* also being represented (Johnson-Laird & Byrne, 1991, in press; Thompson, 2000).

On the surface, the model of a counterfactual conditional resembles the biconditional, in that the initial model of a counterfactual contains the same sort of information:

factual:	FA	FC,
counterfactual:	TA	TC.
	...	

Of course when the implicit model, represented by the three dots, is fleshed out, the fully explicit set of models captures the nonnecessity of the antecedent:

factual:	FA	FC,
counterfactual:	TA	TC, and
	FA	TC.

In our view, therefore, subjunctives are not interpreted as biconditionals (see also Byrne & Tasso, 1999), because subjunctives can be fleshed out to reveal the nonnecessity of the antecedent. Thus, even though there is a surface resemblance between the necessary and the counterfactual models, they differ in this fundamental respect.

Moreover, the differences in the initial models of subjunctive and necessary conditionals should have implications for the inferences reasoners draw.

1. The model of the necessary relation represents the FA–FC case only when it is fully fleshed out, whereas this case is represented spontaneously for the counterfactual from the outset. Consequently, reasoners should be more likely to draw inferences involving the FA and FC propositions (the DA and MT inferences) for counterfactual conditionals than for factual conditionals, even for necessary conditionals. This prediction represents a very strong test of our hypothesis, given that rates of DA are normally elevated for conditionals with necessary antecedents (e.g., Byrne, 1989; Byrne et al., 1999; Thompson, 1994, 1995, 2000).

2. The factors that prompt a complete fleshing out of an ordinary, factual conditional should also do so for a counterfactual conditional. In addition to the FA–FC case, a fully fleshed out factual model represents the FA–TC case. The probability that this FA–TC case is represented as part of a reasoner's model is known to increase when conditionals with nonnecessary antecedents are used (e.g., Byrne, 1989; Byrne et al., 1999; Thompson, 1994, 1995, 2000). This, in turn, reduces the probability that the AC and DA inferences are made.

Can such a decrease occur for counterfactuals? If counterfactual conditionals are equivalent to biconditionals, then the answer should be no, and a manipulation of the necessity of the antecedent should have no effect. That is, because the FA–TC case is proscribed for necessary antecedents, no fleshing out to include the FA–TC case can occur. In contrast, if counterfactual conditionals are not equivalent to biconditionals, then a manipulation of the necessity of the antecedent should have the effect of ensuring that the models are fleshed out to include the FA–TC case for nonnecessary antecedents. Because the inclusion of the FA–TC case is known to reduce the number of DA and AC inferences made, it follows that DA and AC inferences should be suppressed for nonnecessary antecedents, regardless of whether a factual or counterfactual interpretation has been made.

3. If counterfactual interpretations are equivalent to biconditional interpretations, then the tendency to make a counterfactual interpretation should be influenced by the necessity of the antecedent. That is, because biconditional interpretations occur more often for necessary than nonnecessary antecedents, if counterfactual interpretations were equivalent to biconditional ones, then reasoners should make more counterfactual interpretations of conditionals with necessary than nonnecessary antecedents.

Content and Counterfactual Inferences

We manipulated the necessity and nonnecessity of the antecedent by using conditionals for which it is well established that reasoners make a biconditional or conditional interpretation (Thompson, 1994). We used causal materials; for example:

If the butter had been heated, then it would have melted,

with antecedents that were either necessary or nonnecessary (see Appendix C). We also used definitional materials; for example:

If the animal had been warm blooded, then it would have been a mammal,

with antecedents that were either necessary or nonnecessary. The use of causal and definitional content allowed us to manipulate necessity readily.

A second aim was enabled by the use of causal and definitional content. We were able to explore further why only a small majority of participants in our first experiment came to a counterfactual interpretation of subjunctive conditionals. One possibility is that the people-in-places content of the first experiment did not assist reasoners in reaching a counterfactual interpretation. If so, the more realistic materials used in this experiment should remedy the problem. In particular, counterfactual and causal thinking have been closely linked by philosophers (e.g., Chisholm, 1946; Mackie, 1974) and psychologists (e.g., Mandel & Lehman, 1996; Roese, 1997). If the nature of the content also influences the interpretation reasoners provide for subjunctives, then we would expect that more counterfactual interpretations should occur for causal content than definitional.

The aim of identifying the determinants of counterfactual interpretations of subjunctive conditionals led us to pursue one additional reason for the relative paucity of counterfactual interpretations of subjunctives in our first experiment. In that experiment, reasoners made inferences before they judged what the conditionals implied, thus it is possible that the conditional-arguments task may have interfered with judgments on the implications task, leading to low rates of counterfactual interpretations on that task. Specifically, when reasoners are asked to make inferences about the MP and AC forms, they begin with the minor premise that either the antecedent or consequent is true (TA and TC respectively). Being asked to make this assumption may have overridden their natural interpretation of a subjunctive as implying FA or FC. That is, it may have struck reasoners as contradictory to reason about cases in which the antecedent and consequent are assumed to be true (TA and TC), given that they had interpreted the conditional to mean the opposite (i.e., that FA and FC were the case). As a result, some reasoners may have resolved this contradiction by forgoing their counterfactual interpretation and opting instead for a factual interpretation.

In this experiment, we took the precaution of presenting the affirmative inferences (MP and AC) for one conditional and the negative ones (MT and DA) for a different conditional. If the inferential task influences the interpretations reasoners provide for subjunctives, then more counterfactual interpretations should occur after the negative inferences than after the affirmative inferences.

Method

Participants. The participants were 151 introductory psychology students from the University of Saskatchewan who took part in partial fulfillment of a course requirement.

Materials. Eight conditional statements were chosen from Thompson's (1994) study. These statements expressed both necessary (e.g., *if the water was heated to 100 °C, then it boiled*) and nonnecessary relations (e.g., *if the car was out of gas, then it stalled*), as well as both causal (as the previous two examples) and definitional relations (e.g., *if the animal*

was warm blooded it was a mammal; *if the animal was a robin, then it was a bird*). There were two statements in each Necessity × Content cell. These statements had previously been rated for both perceived necessity as well as for content (see Thompson, 1994). The statements were all presented in the past tense and were either presented in the indicative (as the previous examples) or the subjunctive (e.g., *if the animal had been a robin, then it would have been a bird*) mood. A complete list of stimuli appears in Appendix C.

Design. In addition to the arguments, consistency-judgment and implications task, participants completed a rating task; this latter served primarily as a manipulation check of the content and necessity variables. Each participant reasoned about four different conditional statements: two causal and two definitional statements. As described below, for each statement, each participant completed four tasks: the arguments, consistency-judgments, implications, and ratings tasks. Perceived necessity and mood were manipulated between participants, and content (causal vs. definition) was manipulated within participants.

Tasks. The procedure for the conditional-arguments task varied slightly from Experiment 1. Instead of answering all four logical questions for each statement, participants answered questions about either affirmative minor premises (TA and TC in the case of MP and AC) or negative minor premises (FA and FC in the case of AC and MT) for a given conditional relation. That is, participants answered the MP and AC questions for one causal and one definitional statement and the DA and MT questions for the remaining causal and definitional statement. The assignment of questions to statements was counterbalanced across participants, so that each set of logical questions was asked for each conditional statement equally often.

Following each pair of logical questions, participants completed the consistency-judgment, implications, and ratings tasks. The conditional statement was underlined and was repeated for each of the arguments questions and again for each of the other tasks. Otherwise, the procedure for the consistency-judgment and implications task was as described in Experiment 1.

For the ratings task, participants were asked to make five judgments about each conditional relation. They were asked to rate, on a scale of 1 (*very well*) to 7 (*not at all*), how well each of five sentences coincided with their understanding of the conditional relation. Reasoners were asked to rate whether the conditional statement described a causal, contingency, or definitional relationship (e.g., *the water being heated to 100°C would cause it to boil*), and the degree to which the antecedent was perceived to be necessary and sufficient for the consequent (e.g., *it would be necessary for the water to have been heated to 100°C in order for it to have boiled*). These questions were asked in the same order for every participant (i.e., causal, contingency, sufficiency, necessity, definitional).

Results and Discussion

Analysis of the ratings data. Analysis of the ratings data affirmed that reasoners interpreted the situations as intended (note that some participants did not complete all of the questions, therefore the degrees of freedom for these analyses do not always add to 149). Necessary statements were rated as being more necessary than nonnecessary statements ($M = 1.80$ and 4.33, respectively), $t(146) = 11.60$, $p < .01$. Similarly, causal statements were rated to be more causal than definitional statements ($M = 1.25$ vs. 1.54), $t(147) = 3.21$, $p = .002$; definitional statements were judged to be more definitional than causal statements ($M = 1.70$ vs. 2.38), $t(146) = 4.90$, $p < .01$. Causal and definitional statements were judged equally poorly on the contingency rating scale ($M = 4.12$ vs. 4.01), $t(146) < 1$. Thus, reasoners appeared to interpret both content and necessity as intended.

Testing for counterfactual interpretations: Analysis of the implications task. As in Experiment 1, we classified reasoners as giving factual or counterfactual interpretations on the basis of their responses to the implications task.[4] As before, a response was scored as a counterfactual interpretation if the reasoner indicated that either FA or FC was implied by the conditional; all others (excluding four inconsistent responses) were scored as factual. An inconsistent response was scored when reasoners gave a response that would be consistent with both a factual and counterfactual interpretation (e.g., by indicating that both TA and FA were implied).

The proportion of counterfactual interpretations was analyzed as a function of content (definitional and causal), arguments question (MP–AC vs. DA–MT), necessity (nonnecessary and necessary), and mood (subjunctive and indicative) in a mixed ANOVA with content and arguments question as within-participants variables and the remaining factors as between-participants variables; these data are summarized in Table 5. The analyses addressed the following questions.

1. Were reasoners more likely to give counterfactual interpretations to subjunctive than indicative statements? The answer was affirmative; counterfactual interpretations were made more often for subjunctive than indicative conditionals ($M = .67$ vs. .10, respectively), $F(1, 143) = 49.37$, $MSE = 0.27$, $p < .01$. Because mood interacted with content and arguments question, $F(1, 143) = 5.46$, $MSE = 0.08$, $p = .02$, *t* tests were performed to confirm that the differences between subjunctive and indicative were observed for each level of content and question. As expected, more counterfactual interpretations were given to subjunctive than indicative statements, regardless of the other variables, $t(147) \geq 7.28$, $p < .01$.

2. Are counterfactual interpretations different from biconditional interpretations? If so, we would expect that the counterfactual interpretations would not be influenced by variables, such as perceived necessity, that influence biconditional interpretations. Indeed, the ANOVA indicated that the probability of making a counterfactual interpretation was the same for necessary and nonnecessary statements ($M = .37$ vs. .41, respectively), $F(1, 143) = 1.13$, $MSE = 0.27$, $p = .29$. Thus, it appears that counterfactual interpretations are made regardless of the perceived necessity of the conditional relation, supporting the hypothesis that counterfactual and biconditional interpretations are not equivalent. (For this analysis, there was 80% power to detect a difference of .24 between the two necessity conditions.)

3. Are counterfactual interpretations equally likely for causal and noncausal conditionals? Whereas the perceived necessity of the conditional relation did not affect counterfactual interpretations, the type of content represented by the statement did, $F(1, 143) = 30.04$, $MSE = 0.12$, $p < .01$. Counterfactual interpretations were more likely for causal than for definitional statements ($M = .45$ vs. .33, respectively).

4. Is the probability of a counterfactual interpretation influenced by the questions that are asked on the conditional-arguments task?

[4] We also replicated our analysis of the data in Table 1 in Experiment 1. Subjunctive conditionals were perceived to imply FA and FC ($M = .55$ and .47, respectively vs. $M = .10$ and .06, respectively, for indicative conditionals). Indicative conditionals were perceived to imply TA and TC ($M = .32$ and .50, respectively vs. $M = .10$ and .15, respectively, for subjunctives) or to imply nothing (.32 vs. .11, respectively). All differences were significant, $t(147) \geq 4.2$, $p < .01$.

Table 5
Proportion of Reasoners in Experiment 2 Giving Factual and Counterfactual Interpretations

Arguments question	Interpretation	Type of conditional relation: Indicative Nec.[a]	Indicative Nonnec.[a]	Indicative Total[b]	Subjunctive Nec.[c]	Subjunctive Nonnec.[c]	Subjunctive Total[d]
				Causal			
MP–AC	Factual	.95	.95	.95	.39	.21	.30
	Counterfactual	.03	.05	.04	.61	.79	.70
MT–DA	Factual	.76	.81	.78	.21	.13	.17
	Counterfactual	.24	.19	.22	.79	.84	.82
				Definition			
MP–AC	Factual	.95	1.00	.97	.58	.47	.53
	Counterfactual	.05	0.00	.03	.42	.53	.47
MT–DA	Factual	.89	.92	.91	.32	.24	.27
	Counterfactual	.11	.08	.09	.65	.76	.71

Note. Some cells do not sum to 1.00 because an observation was classified as *inconsistent.* Nec. = necessary; Nonnec. = nonnecessary; MP = modus ponens; AC = affirmation of the consequent; MT = modus tollens; DA = denial of the antecedent.
[a] $n = 37$. [b] $n = 74$. [c] $n = 38$. [d] $n = 76$.

Again, the answer is affirmative: Reasoners were more likely to give counterfactual interpretations after having made inferences about negated minor premises, the DA and MT inferences, than after making inferences involving affirmative minor premises, MP and AC ($M = .46$ vs. .31 respectively), $F(1, 143) = 19.01$, $MSE = 0.10$, $p < .01$. This finding lends credence to the concerns raised in Experiment 1. When reasoners had to make inferences in which they had to assume that either TA or TC was the case (i.e., after the MP and AC inferences), they were less likely to make a counterfactual interpretation than after making inferences based on the assumption that FA or FC was the case (i.e., the DA and MT inferences).

In summary, this experiment was somewhat more successful than Experiment 1 in eliciting counterfactual interpretations from reasoners (76% and 59% of causal and definitional subjunctive statements). Moreover, two factors were uncovered that may have suppressed the number of counterfactual interpretations made in Experiment 1. The materials in Experiment 1 expressed noncausal relations, and the results of Experiment 2 indicated that counterfactual interpretations are less likely when reasoners are given noncausal as opposed to causal relations. In addition, reasoners in Experiment 1 made inferences about affirmative and negated propositions at the same time, which may have suppressed counterfactual interpretations. However, although the proportion of counterfactual interpretations was influenced by the type of content expressed by a relation, as well as by the inferences made about that relation, it was unaffected by the perceived necessity of that relation. These findings are consistent with our claim that counterfactual interpretations are different from biconditional ones, given that factors (i.e., the necessity of the antecedent) that are known to promote a biconditional interpretation do not promote a counterfactual one.

Classifying reasoners' interpretations. As in Experiment 1, the effect of counterfactual interpretations on inference patterns was analyzed using individual interpretations rather than mood (subjunctive vs. indicative) as the independent variable. Even though we were more successful in invoking counterfactual interpretations than we were in Experiment 1, this distribution still fell short of the ideal (i.e., 32% of subjects given subjunctive statements made factual interpretations), so that an analysis on the basis of the mood of the conditional would mean that a sizable proportion of reasoners would be inappropriately classified. Consequently, as was the case in Experiment 1, analyses of the conditional-arguments and consistency-judgment tasks use reasoners' interpretations, rather than mood, as the independent variable.[5]

Unlike Experiment 1, however, reasoners in this Experiment interpreted four conditional relations. Although the interpretations given to the four conditionals tended to be correlated with each other ($.40 \leq r \leq .66$, $p < .01$, $M = .56$), not all reasoners gave the same interpretation to all four conditional statements. Thus, a reasoner's interpretation did not necessarily vary orthogonally across levels of the other within-subjects variables (i.e., arguments question and content), making a factorial analysis impossible.

To address this situation, we created a single category to classify each reasoner. This category was based on the number of counterfactual responses a reasoner made: A reasoner was classified as having given a counterfactual interpretation if at least half of their interpretations were classified as counterfactual (recall that all of the statements each reasoner received were either subjunctive or indicative). On the basis of this criteria, 80 participants (66 who were given indicative and 14 who were given subjunctive state-

[5] As was the case in Experiment 1, and to simplify the analysis, the factual group will not be further subdivided. As before, there were few systematic differences between those who indicated that either TA or TC was implied and those who indicated that nothing was implied or between those who gave factual interpretations to subjunctive versus indicative conditionals.

ments) were classified as factual; 67 (7 indicative and 60 subjunctive) were classified as counterfactual.[6]

Analysis of the consistency-judgment task. The proportion of reasoners who judged each of the four event combinations (TA–TC, TA–FC, FA–TC, FA–FC) to be inconsistent with the conditional is summarized in Table 6. Responses for each of the four combinations were analyzed using 2 (interpretation) × 2 (necessity) × 2 (content) mixed ANOVAs, with the first two factors between-participants factors and the last a within-participants factor.

The findings supported the hypothesis that the TA–TC case is represented in the initial model of a factual interpretation, whereas the model of a counterfactual interpretation represents the FA–FC case in the initial model. Consistent with Experiment 1, reasoners who gave a counterfactual interpretation were more likely to view the TA–TC situation to be inconsistent with the conditional than were reasoners who gave a factual interpretation ($M = .59$ vs. .11, respectively), $F(1, 142) = 114.21$, $MSE = 0.15$, $p < .01$. Also consistent with our hypothesis, reasoners who gave a factual interpretation viewed the FA–FC situation to be inconsistent with the conditional more often than those who made a counterfactual interpretation ($M = .46$ vs. .21, respectively), $F(1, 142) = 19.05$, $MSE = 0.22$, $p < .01$. Finally, neither the TA–FC nor the FA–TC judgments were affected by interpretation, $F(1, 141) = 1.25$ and .08, respectively, $MSE = 0.04$ and 0.11, respectively, ps $> .27$; this analysis had 80% power to detect a difference of .09 and .15 between interpretation conditions for the TA–FC and FA–TC situations, respectively.

Also as expected (cf. Thompson, 1994, 2000), reasoners more often gave responses consistent with a biconditional interpretation when the antecedent was necessary than when it was nonnecessary. A biconditional representation specifies that the relationship between the antecedent and consequent is one of exclusivity, such that the consequent always and only occurs when the antecedent does. Thus, situations in which the consequent occurs without the antecedent (i.e., FA–TC) should be viewed as inconsistent with the conditional when the antecedent is necessary. Consistent with this prediction, the FA–TC situation was viewed to be inconsistent with necessary conditional relations more often than nonnecessary relations ($M = .92$ vs. .73, respectively), $F(1, 141) = 24.66$, $MSE = 0.11$, $p < .01$.

Finally, it was observed that both the TA–TC and FA–FC judgments varied according to content. The TA–TC situation was considered to be inconsistent with causal statements more often than definitional statements ($M = .39$ vs. .28, respectively), $F(1, 142) = 14.40$, $MSE = 0.06$, $p < .01$; and the FA–FC situation was considered to be inconsistent with definitional statements more often than causal statements ($M = .38$ vs. .31, respectively), $F(1, 142) = 6.04$, $MSE = 0.06$, $p = .015$; an effect that interacted with perceived necessity, $F(1, 142) = 3.98$, $MSE = 0.06$, $p = .048$. No other main effects or interactions were significant ($p > .06$).

Analysis of the conditional-arguments task. The proportion of reasoners who accepted each of the four inferences is summarized in Table 7. Each of the four inferences was analyzed using 2 (interpretation) × 2 (necessity) × 2 (content) mixed ANOVAs, with the first two factors being between-participants factors and the last a within-participants factor.

Again, the findings replicated those of Experiment 1 regarding the differences between factual and counterfactual interpretations. Reasoners who made a counterfactual interpretation were more likely to accept both the DA ($M = .40$ vs. .57, respectively) and the MT inference ($M = .83$ vs. .65, respectively) than reasoners giving factual interpretations, $F(1, 102) = 8.89$, $MSE = 0.22$, $p = .004$, and $F(1, 139) = 11.14$, $MSE = 0.21$, $p = .001$.[7] More important, the increase in the DA and MT inferences was observed even for necessary relations. This finding represents a very strong affirmation of our hypothesis, given the normally elevated rates of DA for necessary conditionals. In contrast, the MP and AC inferences were accepted equally often for counterfactual and factual interpretations ($M = .81$ vs. .85, respectively, for MP, and $M = .54$ vs. .62, respectively, for AC), $F(1, 103) = 1.26$, $MSE = 0.17$, $p = .26$; and $F(1, 139) = 1.74$, $MSE = 0.18$, $p = .19$. This analysis had 80% power to detect a difference of .22 and .20 for the MP and AC inferences, respectively. These findings are consistent with our hypothesis that reasoners who make a counterfactual interpretation reason from a more fully fleshed out model (i.e., one that represents the FA–FC contingency) than do reasoners who make a factual interpretation.

In addition, these findings corroborated the analysis of the consistency-judgment task and confirmed that a biconditional interpretation was more likely for necessary than for nonnecessary statements (cf. Thompson, 1994, 2000). Reasoners accepted more AC ($M = .85$ vs. .32, respectively) and DA inferences ($M = .73$ vs. .35, respectively) for necessary than nonnecessary statements, $F(1, 139) = 106.97$, $MSE = 0.18$, $p < .01$, and $F(1, 102) = 30.99$, $MSE = 0.22$, $p < .01$. For the AC inference the main effect of necessity was qualified by an interaction with content; however, t tests verified that the difference between necessary and nonneces-

Table 6
Proportion of Scenarios in Experiment 2 Considered to Be Inconsistent With the Conditional Rule

Event combination	Interpretation					
	Factual			Counterfactual		
	Nec.	Nonnec.	*M*	Nec.	Nonnec.	*M*
	Causal					
TA–FC	.93	.93	.93	.87	.89	.88
FA–TC	.94	.74	.84	.95	.76	.85
TA–TC	.20	.13	.17	.71	.60	.65
FA–FC	.38	.47	.42	.18	.17	.17
	Definition					
TA–FC	.92	.97	.94	.90	.99	.95
FA–TC	.93	.67	.80	.87	.74	.80
TA–TC	.09	.04	.07	.60	.47	.53
FA–FC	.50	.47	.49	.32	.19	.25

Note. Nec. = necessary; Nonnec. = nonnecessary; TA = true antecedent; TC = true consequent; FA = false antecedent; FC = false consequent.

[6] A large majority (75%) of reasoners classified as counterfactual gave counterfactual interpretations to at least three of the four statements.

[7] The differences in the degrees of freedom for the analyses are due to the fact that some data were lost for both the DA and MP question because of a typographic error on the questionnaires.

Table 7
Proportion of Inferences Accepted in Experiment 2

	Interpretation					
	Factual			Counterfactual		
Inference	Nec.	Nonnec.	*M*	Nec.	Nonnec.	*M*
			Causal			
MP (TA ∴ TC)	.86	.71	.77	.67	.78	.74
AC (TC ∴ TA)	.83	.40	.62	.74	.41	.57
DA (FA ∴ FC)	.68	.30	.43	.88	.50	.62
MT (FC ∴ FA)	.63	.68	.65	.87	.85	.86
			Definition			
MP (TA ∴ TC)	.96	.90	.92	.80	.91	.87
AC (TC ∴ TA)	.98	.23	.62	.81	.22	.51
DA (FA ∴ FC)	.53	.27	.36	.88	.35	.52
MT (FC ∴ FA)	.54	.76	.64	.81	.80	.80

Note. Nec. = necessary; Nonnec. = nonnecessary; TA = true antecedent; TC = true consequent; FA = false antecedent; FC = false consequent; MP = modus ponens; AC = affirmation of the consequent; DA = denial of the antecedent; MT = modus tollens.

sary statement was observed for causal statements, $t(145) = 5.35$, $p < .01$, and definitional statements, $t(145) = 11.52$, $p < .01$. No other main effects or interactions were significant ($p > .07$).

The relative predictive validity of mood and interpretation. Finally, we replicated the analysis performed in Experiment 1, in which we compared the predictive validity of mood (subjunctive vs. indicative) and interpretation (factual vs. counterfactual). As before, we reanalyzed all of the responses in which we predicted differences between factual and counterfactual representations, namely the MT and DA inference from the conditional-arguments task and the TA–TC and FA–FC combinations from the consistency-judgment task. We used a multiple regression approach, entering two predictor variables, mood (subjunctive vs. indicative) and interpretation (factual vs. counterfactual), in steps. For this analysis, we collapsed over necessity and content to focus on the two variables of interest.

For the TA–TC combination, both mood and interpretation predicted variance in responses on both the first ($R = .59$ and .66, respectively, $p < .01$) and second step ($\Delta R = .02$ and .09, respectively, $p \leq .008$). For the FA–FC combination, as well as the MT and DA inferences, we obtained the following pattern of findings: Interpretation, when entered into the equation first, predicted a significant amount of variance in all three of the dependent measures ($R \geq .23$, $p \leq .018$). When entered first, mood predicted variance only for the FA–FC combination ($R = .30$, $p < .01$) but not in either DA or MT ($R \leq .13$, $p \geq .12$). Moreover, when interpretation was entered as the second variable in the equation, it predicted a significant portion of the variance over and above that accounted for by mood ($\Delta R \geq .06$, $p \leq .019$); on the other hand, mood did not predict any residual variance when entered second ($\Delta R \leq .02$, $p \geq .24$). These findings replicate those of Experiment 1 and support the conclusion that reasoners' interpretations are better predictors of their responses than is the mood of the conditional.

General Discussion

Our studies have uncovered a number of variables that influence the representation of conditional relations and revealed systematic relationships between reasoner's representations and the inferences that they make. We have focused primarily on two dimensions along which representations can be categorized: counterfactuality and biconditionality. We conclude that although biconditional and counterfactual representations are similar, they in fact arise from different interpretations of the conditional statement and are influenced by different variables. Counterfactual representations are influenced by the mood of the conditional and by the content of the conditional (causal vs. definition vs. unfamiliar). Biconditional representations are influenced largely by the perceived necessity of the relation and are normally insensitive to content (Thompson, 1994, 1995, 2000). Moreover, these dimensions are orthogonal, in that biconditional interpretations can be made of both counterfactual and factual relations, and vice versa.

Factual Versus Counterfactual Interpretations

Our findings also speak to the issue of the circumstances under which a counterfactual interpretation is likely to be made. Others have observed that the subjunctive mood is not necessary for a counterfactual interpretation (e.g., Dudman, 1988), and our experiments indicate that it is sometimes not sufficient either. Although counterfactual interpretations are more likely for subjunctive than indicative conditionals, the mood of the conditional is not a perfect predictor of counterfactual interpretations, and a substantial minority of reasoners given a subjunctive conditional make a factual, rather than a counterfactual, interpretation.

We did, however, find two variables that increased the probability of a counterfactual interpretation. First, we observed that more counterfactual interpretations were made of causal than of definitional materials, as the second experiment shows. This result is consistent with long-standing suggestions that thinking counterfactually may be related to thinking causally (e.g., Chisholm, 1946) and with recent suggestions that the primary functions of counterfactual thinking may be similar to those of causal thinking: the prediction and possible prevention of future outcomes. Second, we observed that counterfactual interpretations were more likely after reasoners made inferences about the FA and FC propositions (i.e., DA and MT inferences) than after they made inferences about the TA and TC propositions (i.e., MP and AC inferences). Thus, asking reasoners to make inferences in which they must assume the truth of the antecedent or consequent may override their natural interpretation of a subjunctive as implying FA or FC.

Interpretation, Mental Models, and Inference

We propose that the information that reasoners extract from a conditional relation is represented in a mental model of that relation. The default representation is one that represents only the TA and TC contingencies but allows for the possibility that other contingencies may be possible also. We have described this as a factual representation:

TA TC.

Reasoner's interpretation of the conditional will determine what information is added to or deleted from this initial representation.

Variables such as the mood of the conditional, the content of the conditional, and the perceived necessity of the antecedent may also be used to flesh out the initial representation. For example, both a counterfactual and a biconditional interpretation include the FA–FC contingency, and a nonnecessary interpretation includes the FA–TC contingency.

In addition, reasoners may represent a model that is less complete than the factual model. For example, the fact that many reasoners believe that a conditional implies either the TA or TC contingency, suggests that some reasoners form a representation that does not allow for further possibilities:

TA TC.

More important, we showed that the inferences people made varied systematically with the interpretations they had drawn. Reasoners who make a counterfactual interpretation are more likely to draw inferences involving the FA and FC cases than are reasoners who make a factual interpretation, and they are more likely to view the TA–TC case to be inconsistent with the rule. Our results indicate that this conclusion generalizes to both negative and affirmative conditionals, as the first experiment showed, as well as to causal and definitional contents and necessary and nonnecessary antecedents, as the second experiment showed.

Individual Differences

We observed large individual differences in how people represent and reason from both indicative and subjunctive conditionals. These findings highlight the importance of establishing the interpretations that reasoners have made of experimental materials before attempting to classify the nature of their inferences. Much previous work has attributed differences in inference patterns to differences in the content or context of the conditional relation (i.e., Byrne & Tasso, 1999; Cheng & Holyoak, 1985; Cosmides, 1989; Thompson, 1995; but see Markovits, 1984). Our findings extend these results by demonstrating that differences in inference patterns can be accounted for by differences in how this content is interpreted and represented.

References

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

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, 74,* 347–364.

Barrouillet, P., Grosset, N., & Lecas, J.-F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75,* 237–266.

Barrouillet, P., & Lecas, J.-F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67,* 209–253.

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, England: Cambridge University Press.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 1–83.

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

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: Human Experimental Psychology, 48*(A), 915–944.

Byrne, R. M. J., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1318–1331.

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 & Cognition, 28,* 264–281.

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

Carpenter, P. A. (1973). Extracting information from counterfactual clauses. *Journal of Verbal Learning and 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.

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. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 35*(A), 635–644.

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. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning, 4,* 45–82.

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

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.

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

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

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

Johnson-Laird, P. N., & Byrne, R. M. J. (in press). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review.*

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, Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88.

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.

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

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

Mackie, J. J. (1974). *The cement of the universe: A study of causation.* Oxford, England: Clarendon Press.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology, 71,* 450–463.

Markovits, H. (1984). Awareness of the "possible" as a mediator of formal

thinking in conditional reasoning problems. *British Journal of Psychology, 75*, 367–376.

McCloy, R., & Byrne, R. M. J. (2000). Counterfactual thinking and controllable actions. *Memory & Cognition, 28*, 1071–1078.

McCloy, R., & Byrne, R. M. J. (2002). Semifactual "even if" thinking. *Thinking and Reasoning, 8*, 41–67.

Miyamoto, J. M., & Dibble, E. (1986). Counterfactual conditionals and the conjunction fallacy. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (308–319). Hillsdale, NJ: Erlbaum.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 835–854.

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.

Pollack, J. L. (1986). *Subjunctive reasoning.* Dordrecht, the Netherlands: Reidel.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133–148.

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

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

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645–726.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22*, 742–758.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49*, 1–60.

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition, 76*, 209–268.

Torrens, D., Thompson, V. A., & Cramer, K. M. (1999). Individual differences and the belief-bias effect: Mental models, logical necessity and abstract reasoning. *Thinking and Reasoning, 5*, 1–28.

Appendix A

Models for Indicative and Subjunctive Conditionals

Indicative:
If Sarah went to Moose Jaw then Tom went to Medicine Hat

Subjunctive:
If Sarah had gone to Moose Jaw then Tom would have gone to Medicine Hat

Indicative		Subjunctive		
		Initial set of models:		
Jaw	Hat	factual:	not Jaw	not Hat
. . .		counterfactual:	Jaw	Hat
				. . .
		Abstract structure:		
TA	TC	factual:	FA	FC
. . .		counterfactual:	TA	TC
				. . .
		Fully explicit set of models:		
Jaw	Hat	factual:	not Jaw	not Hat
not Jaw	not Hat	counterfactual:	Jaw	Hat
not Jaw	Hat		not Jaw	Hat
		Abstract structure:		
TA	TC	factual:	FA	FC
FA	FC	counterfactual:	TA	TC
FA	TC		FA	TC

Appendix B

Materials Used in Experiment 1

A complete list of the conditional statements used in Experiment 1 appears below (only the indicative versions are listed). The Canadian people-and-places content, Items 1–6, was used for both the context and no-context conditions. The American people-and-places content, Items 7–12, was used for the no-context condition. Following the conditional statements are examples of each of the tasks used in Experiment 1. In the example, the same conditional statement is used throughout; in the Experiment, participants were provided with six statements, one for each of the conditional-arguments questions and another for the implications and consistent-situations tasks.

Conditional Statements

1. If Dave went to Vancouver, then Linda went to Victoria.
2. If Mark went to Regina, then Susan went to Winnipeg.
3. If Ellen went to Red Deer, then Tom went to Medicine Hat.
4. If Mike went to Calgary, then Barbara went to Edmonton.
5. If Helen went to Yorkton, then Steve went to Prince Albert.
6. If Sarah went to Flin Flon, then Rick went to Brandon.
7. If Jim went to Detroit, then Rachel went to Chicago.
8. If Bob went to Baltimore, then Dawn went to Washington.
9. If Kim went to Boston, then Ted went to New York.
10. If Kathy went to Los Angeles, then Ross went to San Diego.
11. If Doug went to Cleveland, then Kristen went to Pittsburgh.
12. If Wendy went to Seattle, then Curtis went to Spokane.

Conditional-Arguments Task

Inferences were scored as *accepted* when a reasoner chose the option marked with an asterisk.

John says that: *If Mike had gone to Calgary, then Barbara would have gone to Edmonton.*

(a) Mary replies that: *I know that Mike went to Calgary*. (MP)

What, if anything, can they infer about Barbara's whereabouts?
(a) Barbara went to Edmonton*
(b) Barbara may or may not have gone to Edmonton
(c) Barbara did not go to Edmonton

(b) Mary replies that: *I know that Mike did not go to Calgary*. (DA)

What, if anything, can they infer about Barbara's whereabouts?
(a) Barbara went to Edmonton
(b) Barbara may or may not have gone to Edmonton
(c) Barbara did not go to Edmonton*

(c) Mary replies that: *I know that Barbara went to Edmonton*. (AC)

What, if anything, can they infer about Mike's whereabouts?
(a) Mike went to Calgary.*
(b) Mike may or may not have gone to Calgary.
(c) Mike did not go to Calgary.

(d) Mary replies that: *I know that Barbara did not go to Edmonton*. (MT)

What, if anything, can they infer about Mike's whereabouts?
(a) Mike went to Calgary.
(b) Mike may or may not have gone to Calgary.
(c) Mike did not go to Calgary.*

Consistent Situations

John says that: *If Mike had gone to Calgary, then Barbara would have gone to Edmonton.*

Which combination of events would be consistent with John's statement? Which would be inconsistent?
Mike went to Calgary and Barbara went to Edmonton (TA–TC)
Mike did not go to Calgary and Barbara did not go to Edmonton (FA–FC)
Mike went to Calgary and Barbara did not go to Edmonton (TA–FC)
Mike did not go to Calgary and Barbara went to Edmonton (FA–TC)

Implications Task

John says that: *If Mike had gone to Calgary, then Barbara would have gone to Edmonton.*

What, if anything, do you think that John meant to imply?
(a) That Mike went to Calgary (TA)
(b) That Mike did not go to Calgary (FA)
(c) That Barbara went to Edmonton (TC)
(d) That Barbara did not go to Edmonton (FC)
(e) Nothing was implied.

(*Appendix follows*)

Appendix C

Materials Used in Experiment 2

The conditional statements used in Experiment 2 are presented below. Only the indicative version of each statement is presented.

Necessary Antecedent: Causal

1. If the butter was heated, then it melted.
2. If the water was heated to 100 degrees centigrade, then it boiled.

Necessary Antecedent: Definition

3. If the animal was warm-blooded, then it was a mammal.
4. If Beth was Sarah's mother's mother, then she was Sarah's maternal grandmother.

Nonnecessary Antecedent: Causal

5. If the car was out of gas, then it stalled.
6. If Fido tracked mud on the floor, then the floor was dirty.

Nonnecessary Antecedent: Definition

7. If the animal was a robin, then it was a bird.
8. If the card had a jack on it, then it was a face card.

Received April 21, 2000
Revision received May 15, 2002
Accepted May 15, 2002 ■